Beaming: An Asymmetric Telepresence System

Anthony Steed, William Steptoe, Wole Oyekoya, Fabrizio Pece, Tim Weyrich, and Jan Kautz
University College London

Doron Friedman
Interdisciplinary Center Herzliya

Angelika Peer
Technische Universität München

Massimiliano Solazzi, Franco Tecchia, and Massimo Bergamasco
Scuola Superiore Sant’Anna

Mel Slater
ICREA University of Barcelona

Most collaboration tools such as videoconferencing and collaborative virtual environments (CVEs) provide symmetric access to the shared medium. For example, in videoconferencing, each person usually sees a view of the other participants and their surroundings. Although these systems can be configured similarly to face-to-face meetings, they lack some of those meetings' immediacy. Researchers (including some of us) have argued that this is due partly to the systems' technical limitations, including the lack of representation of the full 3D space of the conversation.

One alternative that supports full 3D space is collaborative VR in which each user is surrounded by and immersed within the display systems. However, technologies tend to be laboratory based and relatively uncommon. So, participants normally can’t access these systems without leaving their usual work or living spaces.

The Beaming project (Being in Augmented Multimodal Naturally Networked Gatherings; www.beaming-eu.org) has tackled technological and access issues head on. We've abandoned the symmetry of access to the collaboration technology and the notion of a novel virtual environment in which collaboration happens. Instead, we focus on recreating, virtually, a real environment and having remote participants visit that virtual model. The display systems can be in any reasonable space such as an office or meeting room, domestic environment, or social space.

An Overview

Figure 1 gives an overview of the Beaming system. The destination is a real space populated with people we call locals. The transporter is a high-end VR system equipped with 3D surround visuals, 3D surround audio, tactile and haptic systems, and biosensing. The visitor is the transporter’s user. The system aims to capture the destination and display it to the visitor and simultaneously capture the visitor and display him or her to the locals.

One goal of Beaming is that the destination shouldn’t be a laboratory space with carefully calibrated equipment. Any technical intervention must be portable or mobile, self-calibrating, and dynamically configurable. It should also be as unobtrusive as possible so that it doesn’t interfere with the locals’ behavior. Figure 2 represents potential interventions, which include mobile robots, situated displays, wall or environment projections, augmented reality, camera capture, and audio capture.

User Experience Goals

Amy Voida and her colleagues noted that throughout more than 20 years of media space research, a recurrent theme has been the pursuit of symmetry, which has both social and technological components. Social symmetry is desirable so that
no participant is disadvantaged in his or her ability to interact with or influence other participants. Technologies such as videoconferencing and CVEs generally feature technological symmetry to ensure that the same sensory cues are available to all parties. Both these technologies have limitations. Videoconferencing poorly supports eye gaze between participants and spatial references between users and objects in their environment. Videoconferencing poorly supports eye gaze between participants and spatial references between users and objects in their environment. CVEs typically involve a restricted interface so that users see the virtual world as if through a window. 

Steve Benford and his colleagues suggested that the interest in spatial approaches to computer-supported cooperative work might be viewed as a shift of focus toward supporting the context in which distributed work takes place, rather than the process of the work itself. The destination-visitor paradigm in Beaming is fundamentally asymmetric but aims to support symmetric social interaction between the visitor and locals. That social interaction should be able to exploit the objects at the destination because this is a key component of the context in which the interaction takes place.

The long-term goal for Beaming-like systems is to provide collaborative mixed-reality environments that grant symmetrical social affordances and sensory cues to all connected users whether they're locals or visitors. Put another way, although the mediating technologies are highly asymmetric between the destination and transporter sites, visit user behavior shouldn't be hindered because of their remote location. Also, they should be represented to the locals with a virtual or physical embodiment. Using terminology from the virtual-environments field, we might say that we strive to give a sense of spatial presence within the destination for visitors and a sense of copresence among both locals and visitors.

Consequently, we claim that the visual display at the visitor site must be an immersive display such as a head-mounted display (HMD) or a display similar to a CAVE (cave automatic virtual environment). This is because, to strive for social symmetry, the system must provide similar sensory experiences, particularly the dominant visual mode, to all parties. Locals need no visual mediation to perceive the destination as being realistic and spatial because they perceive the actual physical location. However, stimuli representing the destination must be transmitted in real time to
Moreover, the technological display properties used by the visitor must foster the impression of being physically at the destination. To this end, Ralph Schroeder considered immersive technology as an end state in that synthetic and multiuser environments that are displayed to the senses can’t be developed further than fully surrounding and spatial immersive systems. We expect that immersive presentation will increase the social symmetry in an asymmetric, heterogeneous system architecture.

A Beaming Example

One test case for the Beaming concept has been remote rehearsal between two actors and a director. Figure 3 shows the capture and display devices used to connect the actors. In this example, the visitor was an actor using an HMD and wearing a motion capture suit so that her movements were accurately tracked (see Figure 3a). The single local was also an actor in an office environment. In that office, a SphereAvatar display showed the visitor’s head and viewing direction, and a large projection screen showed her body movements (see Figure 3b). The local could choose between the two representations.

At this time, we hadn’t used any augmented-reality or robotic representations. We captured the local with Microsoft Kinect and a surround camera, a Point Grey Research Ladybug 3. We had constructed the office offline via photogrammetry. So, the visitor could see one of three visualizations: a surround image (see Figure 3c), a 3D model with an inserted depth video of the local (see Figure 3d), and the 3D model with an animated avatar of the local (see Figure 3e). The two actors wore wireless microphones and talked to each other using Skype.

A theatrical rehearsal aims to practice and refine the performance of a scene. We chose a scene consisting of rich conversation, spatial action and interaction, direction of attention, and object handling, which are all common practice in general social interaction and collaborative work. So, through repeated and evolving run-throughs of the scene with professional actors and directors, this structured activity formed an excellent basis for analytic knowledge regarding the system’s successes and failures.

The rehearsals explored the central aspects of spatiality and embodiment. Overall, the common spatial frame of reference (the destination) was
highly conducive to acting and directing, allowing blocking, gross gestures, and unambiguous instructions to be issued. The main lesson learned was thus that we achieved spatial interaction but that multiple representations were confusing. Visitors tended to prefer the 3D model with the avatar representation of the local because it was visually consistent and presented a spatial reference relative to their own location. Locals tended to prefer the wall display because it showed the visitor’s body language.

The communication’s central limitation was the relative lack of expressivity of the actors’ embodiments. This meant that moments relying on performing and reacting to facial expressions and subtle gestures were less successful. We expect SphereAvatar to be more useful when it can dynamically present the visitor’s facial expressions.

The Beaming Technical View
The user experience goals set many technical challenges. In the Beaming project, we’re investigating many of these in a range of modalities including visual, audio, and haptic scanning, representation, and display, and emotion recognition and display. We eventually want to reconstruct the whole destination in real time. However, in early demonstrations we’re using a variety of prebuilt models of spaces and objects (visual, audio, and haptic), either directly for rendering or to act as a source for trackers. Here, we give examples of technical demonstrators we built and how we’re integrating them to achieve the project’s long-term goal.

Robot Representation
Beaming uses two forms of robots.

Encountered-type haptic devices. Interaction with encountered-type haptic devices (see Figure 4a) lets the visitor perceive interaction with the objects at the destination or with locals. Such haptic devices approach the visitor only when contact needs to be rendered, thus guaranteeing perfect rendering of free space.

Different end effectors, including a prosthetic hand and a universal object that can morph its shape, display a series of human-human and human-object interactions. For example, in Figure 4a, the robot is presenting a hand to mimic a local’s actions. In this situation, the visitor is seeing a representation of the local and the destination through the HMD. The visitor can see the local’s movement, which the robot matches.

In this situation, the challenge is to haptically reconstruct the destination in real time and remotely, so that the encountered-type haptic device can present all physical interactions. Currently, we do this using tracking of locals employing a Kinect array and reconstructed models of the surfaces at the destination. The encountered-type haptic device can present both data types. We’ve tested the overall system’s functionality; a detailed evaluation is part of future research.

Mobile robot avatars. This avatar represents the visitor at the destination (see Figure 4b). Unlike the encountered-type haptic device, which the visitor and locals don’t see, the robot avatar is broadly anthropomorphic, with two robotic arms and hands and an emotion-expressing head. The visitor’s arm and hand movements are tracked by the motion capture suit and mapped to the robot’s movements. A camera looks underneath the HMD to capture facial expressions; the system analyzes them to recognize emotional states, which the emotion-expressing head then conveys.

Key aspects of the robot avatar’s development are the remote streaming of the visitor’s movements and real-time remapping of the visitor’s movements to the robot’s degrees of freedom. This shares some similarities with previous telepresence research, such as the telexistence systems. However, Beaming emphasizes reconstruction of a 3D model of the destination for the visitor to interact with, so as to allay some issues of latency and dynamics of movement. Unlike previous research on remote haptic interaction, our two haptic devices aren’t directly slaved together. They’re better described as loosely coupled; the encountered-type haptic device mimics the destination, and the robot avatar mimics the visitor’s movements.

A Situated Display to Represent the Visitor
An emerging form of telecollaboration is situated displays at a physical destination that virtually represent remote visitors. SphereAvatar (see Figure 4c) is a situated display that’s visible from all sides so that locals can determine the visitor’s identity from any viewpoint. Flat displays are visible from only the front and lack spherical displays’ 360-degree and multiview capabilities. Exploiting spherical displays’ unique characteristics might
bestow greater social presence on the participant represented on the sphere. SphereAvatar is small enough to situate almost anywhere in a room; for example, it could be on a seat next to a table or on a mobile platform. It can show either video captured from camera arrays around the visitor or a photorealistic computer-generated head.

**Depth Camera Streaming from the Destination**

Cameras that can acquire a continuous stream of depth images are commonly available—for instance, Microsoft Kinect. They’re particularly useful for Beaming because they let us easily capture the destination’s geometric and texture information (for example, see Figure 3d). Streaming the raw data consumes significant bandwidth, so compressed streaming is desirable. Although much research exists on compressed streaming of (8-bit color) videos, little research exists on streaming (16-bit) depth videos. In particular, no codecs for depth compression are publicly available.

We adapted existing video encoders for depth streaming for two reasons. First, they’re widely available. Second, using the same video codec to transfer both color and depth frames enhances consistency and simplifies the streaming architecture. We use a robust encoding of 16-bit depth values into $3 \times 8$ bits. This is essentially a variant of differential phase shift encoding, such that the depth maps suffer from few compression artifacts. Using this encoding, we can stream depth maps over standard VP8 and H.264 codecs. This fur-
ther allows us flexibility in where analysis of the depth maps occurs. Different machines can now perform different types of analysis on the same data (for example, extracting both point clouds and skeleton data).

Other Research Areas
The following R&D areas add to Beaming’s uniqueness as a collaborative system.

Body Representation
Beaming explicitly uses recent cognitive-neuroscience research on body ownership illusions. Multisensory data streams that imply a changed body representation are readily interpreted as changes to the body. The most famous example of this is Matthew Botvinik and Jonathan Cohen’s rubber-hand illusion.11 Tapping on a rubber hand on a table in front of a participant while tapping on the participant’s corresponding obscured real hand led to the illusion that the rubber hand felt as if it were the participant’s real hand. Researchers have applied this technique at the whole-body level using video streaming through HMDs to generate out-of-the-body-type illusions and whole-body transformation illusions.

A body transformation illusion can also work strongly in VR. In one example, researchers gave men the illusion that their body was that of a small girl.12 This research also showed that seeing the virtual body from the first-person perspective contributed critically to the experience of ownership.

An important goal of Beaming is to give visitors the strong sensation of ownership regarding their body representation at the destination. Because visitors see their body representation naturally from the first-person perspective, a degree of such ownership will likely occur. In addition, the haptic and visual feedback together provides multisensory feedback that enhances the ownership illusion. In applications in which it’s desirable to transform the visitor’s representation (for example, in acting rehearsals), such transformations have also been shown to preserve body ownership.

A Visitor Proxy
We can envisage a future in which some sort of automated proxy partly fulfills people’s communication needs. In Beaming, we’ve implemented a communication proxy that can not only represent visitors for a short period of time (for example, a phone interruption) but also replace them in a whole meeting or presentation. (For instance, see the video at youtu.be/43i739kKffk.)

The communication proxy thus isn’t an autonomous virtual agent or a standard virtual avatar. It must be based on models of the owner’s behavior and be aware of its owner’s goals and preferences. The proxy can learn some of its owner’s behaviors by observing the owner’s motion and speech. For example, the proxy can learn certain aspects of its owner’s body language and gestures. Other behaviors and higher-level goals would need to be explicitly encoded or modeled.

An important goal of Beaming is to give visitors the strong sensation of ownership regarding their body representation at the destination.

The proxy’s implementation presents several challenges. The software must observe both the visitor and the real environment (that is, sensors at the destination). Beaming’s software architecture supports this through a loosely coupled heterogeneous platform in which sensors, actuators, and displays collaborate through a shared data service. The proxy then needs traditional AI components such as gesture recognition and speech interpretation.

The proxy can also operate in a mixed mode. For example, while representing its owner in a meeting, if it’s asked a question it doesn’t understand, it can open a voice connection to the owner and relay the question. The owner can answer the question, taking over the representation for a short while before returning control to the proxy.

Extending Haptic Feedback for the Visitor
We’re investigating finger-mounted portable devices that can display the transition between contact and noncontact of the fingers (see Figure 4d). Such devices have three potential advantages: unlike with desk-mounted devices, their range of operation has no intrinsic limits; they’re not likely to interfere with the user; and constructing systems with multiple contacts for the fingers is easier.

By acknowledging the varied ways in which much collaborative work and socializing occur, Beaming’s destination-visitor paradigm could provide an integrated approach to high-quality telecommunication. The paradigm presents interesting technical challenges: high-fidelity real-time reconstruction of the destination and locals in
multiple modalities and presentation of the visitor in multiple modalities to the locals. Our experience from early example applications and integrations is that participants can understand the situation and exploit the 3D spatial nature of the communication in both directions. This has been true for both the acting application we discussed and a medical application that naturally exploits Beaming’s asymmetry.13

Clearly, the Beaming project and concept raise serious ethical and legal issues. For example,

- someone could misuse transformation or the proxy to mislead locals,
- someone in one country could use a robotic embodiment to commit a crime in another country, or
- a body transformation experience could cause psychological problems.

Widespread implementation of this concept would entail profound changes in law and societal relationships. The Beaming project takes these issues seriously, dedicating an entire strand of research to them.

Recent advances in camera technology, particularly depth cameras, have made some aspects of the system more easily deployable. We expect significant advances in the ability to visually reconstruct populated spaces in real time in the near future, and we’ll exploit these advances to build multimodal representations of the destination. Our next-generation systems will also combine robots with novel types of display to afford more effective representations of the visitor.

Acknowledgments
The European Commission supports Beaming under the FP7 Information and Communication Technologies Work Programme. Investigators in the project include Martin Buss (Technische Universität München), Antonio Frisoli (Scuola Superiore Sant’Anna), Maria Sanchez-Vives (August Pi i Sunyer Biomedical Research Institute), Miriam Reiner (the Technion), Giulio Ruffini (Starlab Barcelona), Matthias Harders and Gábor Székely (Swiss Federal Institute of Technology Zurich), Benjamin Cohen (IBM Haifa Research Lab), and Dorte Hammershøi (Aalborg Universitet). We also thank the associated research fellows and students.

References

Anthony Steed is the head of the Virtual Environments and Computer Graphics group in the Department of Computer Science at University College London. Contact him at a.steed@ucl.ac.uk.

William Steptoe is a postdoctoral research fellow in the Virtual Environments and Computer Graphics group in the Department of Computer Science at University College London. Contact him at w.steptoe@ucl.ac.uk.
Wole Oyekoya is a postdoctoral research fellow in the Department of Computer Science at University College London. Contact him at w.oyekoya@cs.ucl.ac.uk.

Fabrizio Pece is a PhD student in the Virtual Environments and Computer Graphics group in the Department of Computer Science at University College London. Contact him at f.pece@cs.ucl.ac.uk.

Tim Weyrich is a senior lecturer in the Virtual Environments and Computer Graphics group in the Department of Computer Science at University College London. Contact him at t.weyrich@cs.ucl.ac.uk.

Jan Kautz is a professor of visual computing in the Department of Computer Science at University College London. Contact him at j.kautz@ucl.ac.uk.

Doron Friedman heads the Advanced Virtuality Lab at Interdisciplinary Center Herzliya. Contact him at doronf@idc.ac.il.

Angelika Peer is a senior researcher and lecturer at the Institute of Automatic Control Engineering and Carl von Linde Junior Fellow at the Institute for Advanced Study of Technische Universität München. Contact her at angelika.peer@tum.de.

Massimiliano Solazzi is an assistant professor in the Portable Haptic Interfaces group at Scuola Superiore Sant’Anna. Contact him at m.solazzi@sssup.it.

Franco Tecchia leads the Computer Graphics and Virtual Environments Group at Scuola Superiore Sant’Anna. Contact him at f.tecchia@sssup.it.

Massimo Bergamasco is a full professor of applied mechanics at Scuola Superiore Sant’Anna. Contact him at massimo.bergamasco@sssup.it.

Mel Slater is a research professor at the Catalan Institution for Research and Advanced Studies at Universitat de Barcelona and a professor of virtual environments at University College London. He’s the Beaming technical manager. Contact him at melslater@ub.edu.

Contact department editors Bernd Froehlich at bernd.froehlich@uni-weimar.de and Mark Livingston at mark.livingston@nrl.navy.mil.

Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.