Augmented Audio Reality: compositing mobile telerobotic and virtual spatial audio

Yasuhiro Yamazaki, Michael Cohen, Jie Huang, and Tomohide Yanagi

University of Aizu
Aizu-Wakamatsu 965-8580; Japan
{*yam@spica., mcohen@, j-huang@, m5041133@*}*u-aizu.ac.jp*

Abstract

Keywords: R^3 (**r**ealtime **r**emote **r**obotics), spatial audio, telerobotics, CVE (**c**ollaborative **v**irtual **e**nvironments), groupware, CSCW (**c**omputer-**s**upported **c**ollaborative **w**ork).

Topic Areas: Augmented and Mixed Reality, Visual and Auditory Displays, Communication with Realistic Sensations, VR Interaction and Navigation Techniques, Teleëxistence/Telepresence.

1. Introduction

In the visual domain, techniques associated with approaches variously known as "augmented," "enhanced", "hybrid," "mediated," or "mixed" reality/virtuality overlay computer generated imagery (CG) on top of a real (photographic) scene, or composite sampled data into virtual scenery. Idealized notions of "reality" and "virtuality" can be thought of as endpoints on a continuum [MK94], an instance of the former approach corresponding to a "see-through" display, an instance of the later to texture-mapped image-based rendering. Analogously, in the audio domain, computer-synthesized or -manipulated sounds can be mixed "on top of" the natural ambient soundscape or into directly acquired channels [CAK93]. Such combinations are illustrated by Fig. 1.

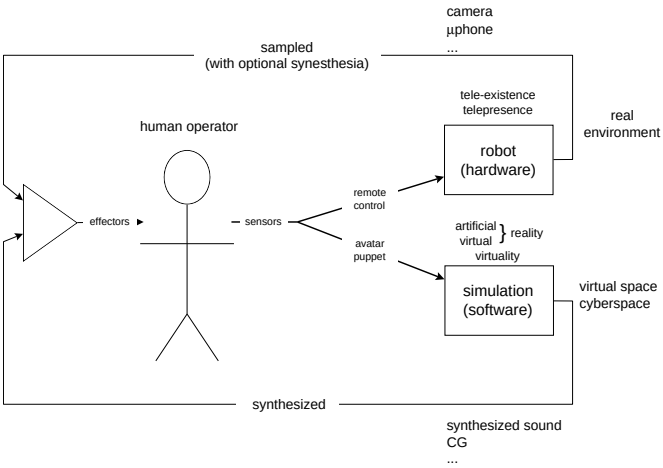

Fig. 1: Augmented/enhanced/hybrid/mediated/mixed Reality/Virtuality: mixing sampled and synthesized data

Sample-based functionality represents teleëxistence (a.k.a. tele-existence or telepresence): instead of accessing a computer-modeled or -synthesized universe, teleëxistence gives the user a sense of being somewhere else in the familiar universe we call 'reality.' Teleëxistence delegates a robot slave to act on behalf of a human master, like *bunraku* (puppet theatre) with feedback. Such a delegate can be monitored and controlled from afar, the human's various sensors corresponding to the puppet's effectors, the human seeing and hearing what the drone senses, including collision alarms, through stereo sight and sound. Such technology can be deployed in hazardous environments like nuclear power plants, fires, and toxic waste dumps.

2. Hybrid audio: robotic telepresence and synthetic spatialization

We have incorporated a mobile telerobot [HST+99], shown in Fig. 2, into a multimodal groupware suite, integrating graphical, audio, and eventually haptic I/O modalities into a mixed reality system [YCHY01]. "Hero," for **he**aring **ro**bot,[1] is based upon the Applied AI Systems LABO-3 platform[2] (not to be confused with eponymous kits made by Heath/Zenith). It is a wheel-based mobile indoor robot, driven by two-wheel differential steering. The robot uses a bidirectional UHF radio modem for host↔robot communication: four FM transmitters and a video transmitter are used for quadraphonic audio and video signal transmission.

The robot is designed to support both autonomous and interactive (human-piloted) modes. For instance, the robot might automatically patrol an area, checking for suspicious activity. Upon encountering some unusual phenomenon, the robot could relinquish control to a human pilot, who would directly steer the robot to investigate.

The robot visually captures scenes with its cyclopean (monocular) fixed-direction frontally-mounted CCD camera. This visual data is radioed to a base station, where it is compressed and streamed over the internet. Graphical output can be displayed to a pilot through a workstation or laptop computer with a wireless internet connection (like Apple's Airport [a.k.a. "Wi-Fi"] technology,[3] which supports data rates up to 11 Mbps).

We have extended the robot control program to sup-

[1] www.u-aizu.ac.jp/~j-huang/Robotics/robotics.html
[2] www.aai.ca/products/robots/labo3.html
[3] www.apple.com/airport/

Fig. 2: "Hero" Labo-3 **he**aring tele**ro**bot: operable in piloted and autonomous (automatic obstacle avoidance) modes, with cyclopean (one-eyed) camera and quadraphonic tetrahedral microphone array, shown along with a Java3D model used in simulations.

port a mixed audio reality/virtuality display by compositing artificially directionalized audio sources with the soundscape directly captured by the robot, as shown in Fig. 3.

The robot server can catch control messages from client using TCP socket. A control message is processed immediately in robot control class. The robot control class has three instances— virtual robot (emulator), network robot, and actual robot— selected as required. The first instance is an emulator to predict and calculate a behavior of the robot. The second instance works as proxy to send control messages to another server via the network using a TCP socket. The third instance is connected to a transceiver to send messages translated into the robot opcode set to remotely control an actual robot. Independent of the controller, video/audio signals can be caught by the receiver, and then captured by hardware (graphics card and sound card). We use the Java Media Framework (JMF[4]) [GT99] for capturing and sending these streams.

Our system *samples* environmental sources, *spatializes* virtual sources, and *synchronizes* displays that track and control the location and orientation of the telerobot and virtually collocated sources. These functions are described in the following sections.

2.1 Sampling

The Hero robot has four ears (microphones), arrayed on a tetrahedron, but currently we use streams from only two of them— corresponding to the left and right, to simplify

[4] `java.sun.com/products/java-media/jmf/`

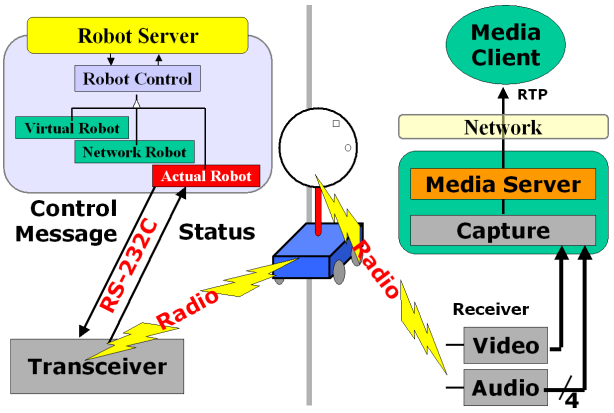

Fig. 3: Actual robot system with media streaming

the mixdown with synthesized signals to a stereo pair— compressed and streamed via the same wireless interface as the video. Since the sphere upon which the robot's microphones are mounted is both larger than a human head (≈30 cm vs. ≈21 cm) and acoustically transparent, the binaural cues are not natural (exagerated time delays and understated level differences), but we hope to eventually resynthesize, from the detectable time differences, the missing corresponding level differences, as proposed by [Mar00].

2.2 Spatializing

An audio source can be spatialized by a supporting application [CS00] [MC01], like that shown in Fig. 4 that pans by intensity stereo. A virtual source can be used as a compass or homing beacon, or for collision avoidance warnings (like a sonic lighthouse), as they are aligned in the "real world" through which the robot navigates. The directionalized sources can be files (PCM [`aiff`, `au`, `snd`, `wav`], MIDI [type 1 and type 2] and RMF) or (near) realtime audio streams, captured by a workstation microphone (about 1 s delay). The composite sampled/synthesized soundscape is presented to a human pilot via normal stereo headphones or via nearphones (for "near earphones") mounted near a chair's headrest, presenting unencumbered binaural sound with soundscape stabilization for multichannel sound image localization.

2.3 Synchronizing

Over the last several years, our group has developed heterogeneous interfaces implemented in Java (with Java3D, JMF, and Swing) to support various multimedia presentation modalities— including the virtual spiral spring GUI, shown in Fig. 4, and a 2.5D dynamic map [MSC01], shown in Fig. 6— allowing not only planar translation but also rotation. Various controls, displays, and widgets are synchronized by a client/server architecture and groupware framework [KCNH01] which uses a `get()`/`set()`-like protocol to distribute state events. The groupware architecture is shown in Fig. 5 A groupware server multicasts shared data— for example, orientation and location— in a session. Our proxy client, which connects to this group-

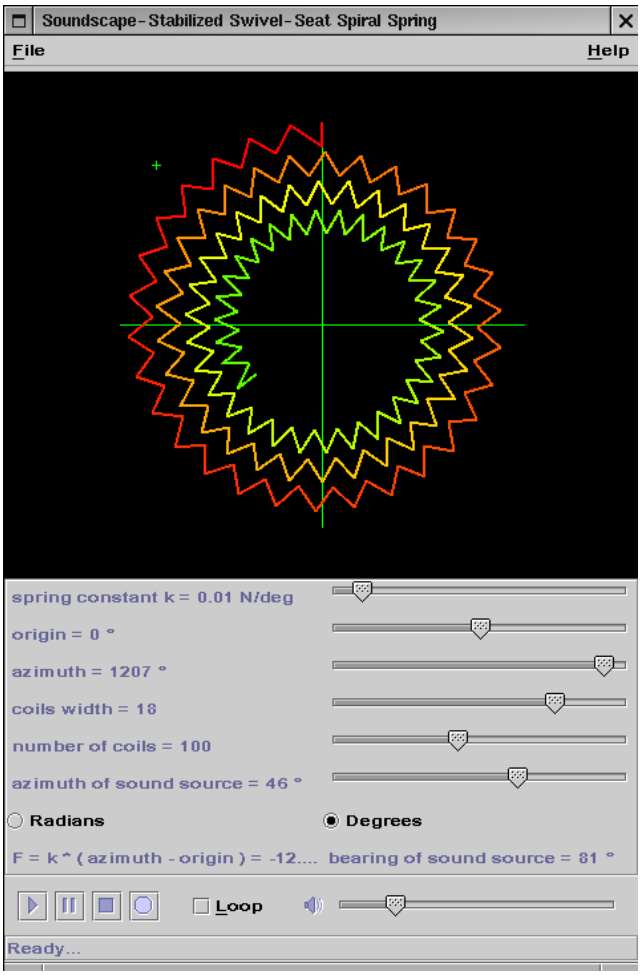

Fig. 4: Soundscape-stabilized spiral-spring screen shot. The "+" mark in the perimeter (NW) indicates the direction of a virtual sound source.

ware server, exchanges robot control data with the robot controller. A virtual sound spatializer generates a stereo pair using the shared data to configure the soundscape. The virtual sound is composited with captured "real" stereo sound from robot in a mixer, so the user can hear an integrated display of both real and virtual sources.

3. Conclusion

In groupware situations like teleconferencing or chat spaces, robot position can be coupled with iconic representations, avatars in a virtual world, enabling social situation awareness via coupled visual displays, soundscape-stabilized virtual source locations, and direction-dependent projection of non-omnidirectional sources [Coh00].

By wrapping a Java wrapper around our robot's control/display software, we have integrated it with a common management server. We hope to composite into its displays various synesthetic (sensory substitution, or capability amplification) cues. For example, parallel research [HOS95] [HOS97] [Hua00] [Hua01] is exploring computational scene analysis, processing signals from the four-microphone array to determine source location. Such

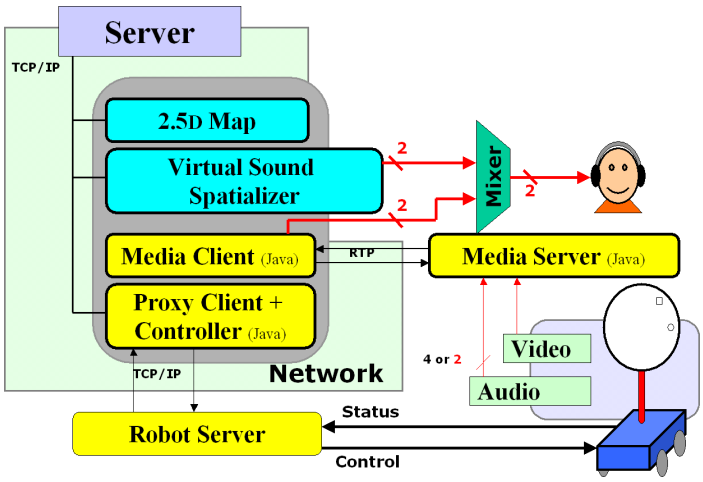

Fig. 5: Our robot in architecture of groupware suite

inferred positions could be graphically rendered, on the same map as the other objects, giving human pilots a non-fleeting visual hint about the source of such acoustic activity.

We expect the compositing of naturally captured and artificially projected information to magnify the synergy between the interface and telerobot, as auditory and visual I/O modalities complement each other through a powerful human-machine interface that is literally sensational.

4. Acknowledgments

The soundscape-stabilized spiral-spring interface was developed by Kenta Sasa (佐々 健太) and Shiyougo Ihara (伊原 正悟) with Takashi Wada (和田 貴志), the 2.5D dynamic map by Masataka Shimizu (清水 雅高), and the VR_4U_2C browser by Noor Alamshah Bolhassan. The client/server architecture and framework were designed by Toshifumi Kanno (菅野 才文). This research has been supported by a grant from the Fukushima Prefectural Foundation for the Advancement of Science and Education.

5. References

[CAK93] Michael Cohen, Shigeaki Aoki, and Nobuo Koizumi. Augmented audio reality: Telepresence/VR hybrid acoustic environments. In *Proc. Ro-Man: 2nd* IEEE *Int. Workshop on* **Ro***bot and Hu***man** *Communication*, pages 361–364, Tokyo, November 1993. ISBN 0-7803-1407-7.

[Coh00] Michael Cohen. A Design for Integrating the Internet Chair and a Telerobot. In *Proc. IS2000: Int. Conf. on Information Society in the 21st Century*, pages 276–280, Aizu-Wakamatsu, Japan, November 2000. IPSJ, IEICE, IEEE.

[CS00] Michael Cohen and Kenta Sasa. An interface for a soundscape-stabilized spiral-spring

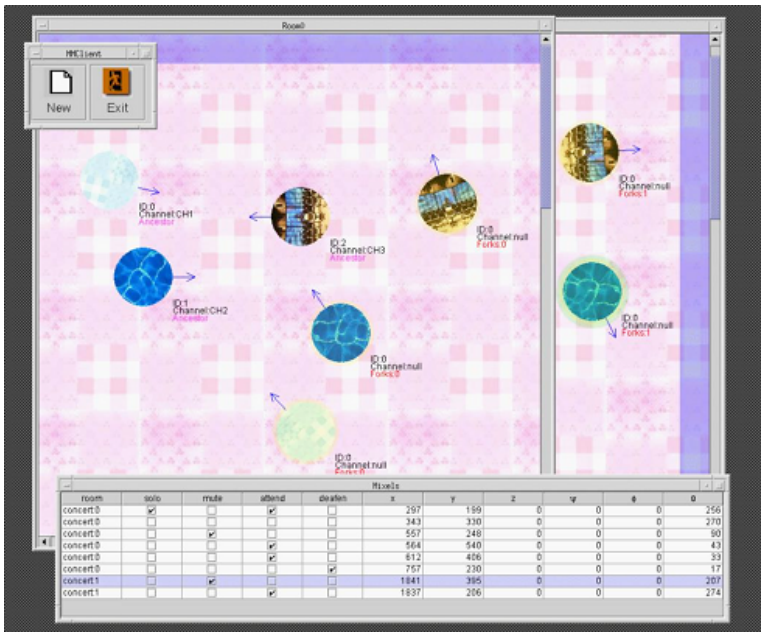

Fig. 6: 2.5D Dynamic Map interface allows icons to circulate, translating and rotating.

swivel-seat. In *Proc. WESTPRAC VII:* 7th **West***ern* **P***acific* **R***egional* **A***coustics* **C***onf.*, pages 321–324, Kumamoto, Japan, October 2000. ISBN 4-9980886-1-0 and 4-9980886-3-7.

[GT99] Rob Gordon and Stephen Talley. *Essential JMF — Java Media Framework*. Prentice Hall, 1999. ISBN 0130801046.

[HOS95] Jie Huang, Noboru Ohnishi, and Noboru Sugie. A biomimetic system for localization and separation of multiple sound sources. *IEEE Trans. Instrumentation and Measurement*, 44(3):733–738, June 1995.

[HOS97] Jie Huang, Noboru Ohnishi, and Noboru Sugie. Building ears for robots: Sound localization and separation. *Artificial Life and Robotics (Springer-Verlag)*, 1(4):157–163, 1997.

[HST+99] J. Huang, T. Supaongprapa, I. Terakura, F. Wang, N. Ohnishi, and N. Sugie. A model based sound localization system and its application to robot navigation. *Robotics and Autonomous Systems (Elsevier Science)*, 27(4):199–209, 1999.

[Hua00] Jie Huang. Spatial sound processing for a hearing robot. In *Proc. IS2000: Int. Conf. on Information Society in the 21st Century*, pages 281–287, Aizu-Wakamatsu, Japan, November 2000. IPSJ, IEICE, IEEE.

[Hua01] Jie Huang. Developing a multimodal telerobot — spatial auditory processing. In *Proc. CIT:* 2nd *Int. Conf. on* **C***omputer and* **I***nformation* **T***echnology*, pages 147–151, Shanghai, September 2001. ISSN 1007-6417.

[KCNH01] Toshifumi Kanno, Michael Cohen, Yutaka Nagashima, and Tomohisa Hoshino. Mobile control of multimodal groupware in a distributed virtual environment. In *Proc. ICAT:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence*, Tokyo, December 2001.

[Mar00] William L. Martens. Pseudophonic listening in reverberant environments: Implications for optimizing auditory display for the human user of a telerobotic listening system. In *Proc. IS2000: Int. Conf. on Information Society in the 21st Century*, Aizu-Wakamatsu, Japan, November 2000. IPSJ, IEICE, IEEE.

[MC01] William L. Martens and Michael Cohen. Virtual Acoustic Research at the University of Aizu. *JVRSJ:* **J**. **V***irtual* **R***eality* **S***ociety of* **J***apan*, 6(3), December 2001. ISSN 1342 6680.

[MK94] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual dispays. IEICE *Trans. Inf. Sys.*, E77-D(12):1321–1329, December 1994.

[MSC01] Takashi Mikuriya, Masataka Shimizu, and Michael Cohen. A collaborative virtual environment featuring multimodal information controlled by a dynamic map. *3D Forum: J. of Three Dimensional Images*, 15(1):133–136, 3 2001. ISSN 1342-2189.

[YCHY01] Yasuhiro Yamazaki, Michael Cohen, Jie Huang, and Tomohide Yanagi. Augmented audio reality: compositing mobile telerobotic and virtual spatial audio. In *Proc. ICAT:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence*, Tokyo, December 2001.