Phonic Frequencies

shaping networked realities

Tamas Szakal, Christoph Groenegress, Wolfgang Strauss, Predrag Peranovic

MARS – Exploratory Media Lab

Fraunhofer Institute for Media Communication, Sankt Augustin, Germany

E-mail: {szakal, groenegress, strauss, peranovic}@imk.fraunhofer.de

ABSTRACT: The paper introduces a networked multi-user installation and interactive mixed reality environment that combines the fields of interactive art, telecommunication and streaming technologies. Phonic Frequencies is an audiovisual data space whose appearance can be altered via networked communication devices. Visitors are active agents and participate in shaping spaces as they control audiovisual data with their telephones.

Linking the physical space to the digital network space, Phonic Frequencies opens a hybrid reality for distributed sound exchange and visual communication, blurring the boundaries between verbal communication and digital information. The work attempts to break the conventional communication cycle of bidirectional dialogue and to consider the meaning of in- and output from a different perspective.

KEYWORDS: mixed reality, interactive art, multi-user installation, collaborative environment, remote collaboration, networked performance, telephone network, mobile phone, telecommunication network, streaming video

INTRODUCTION

Over the past years, a lot of research in mixed reality (MR) has been devoted to the concept of space and interaction techniques within hybrid environments. In contrast to purely virtual environments, for instance, the physical body retains its significance and role as an interface to the surrounding environment. One concept that has arisen in this context is that of body-centred interaction [1], which promotes the idea of *physical* interaction as a key factor to presence in any kind of environment. A complete shift away from the immersive concept as pointed out by many VR installations [2] occurred with the concept of a multi-user environment based on bodily interaction [3], which are aimed at the development of a body language by playful interaction in order to establish a sense of co-awareness to other participants. Extending the display-related taxonomy of mixed reality [4] towards social interaction and the inclusion of architectural space therefore requires multi-modal interfaces.

Aims of Research

A number of artistic works have coupled the idea of participation of multiple users in observing and manipulating remote sites [5] with the use of telecommunication networks [6; 7; 8]. Other projects focused on arranging sound and music via digital networks [9], or visualizing sound information [10; 11].

Phonic Frequencies is aimed at integrating some of these ideas in order to expand communication networks for multi-directional creative exchange of information and to create an interactive environment offering collaborative creation of sound and image presented in a shared physical space via audio/video display. The installation space can be shaped, extended and viewed from remote locations, simultaneously by multiple agents. The work is thus aimed at integrating telecommunication devices and ambient environments in order to explore the interconnection of information networks and physical spaces.

By investigating the use of telephones as a user interface for controlling remote sites and for collaboration in an artistic context, we wish to extend communication in order to establish a database of space shaping material. Also, how do users react to using a device that is accepted as a communication tool rather than an instrument for influencing space? In this context it is also important to define suitable but entertaining and effective modes of interaction: what degree of complexity is acceptable in such environment to ensure usability and understanding of influence?

Finally, we analyse user acceptance and the (perceived) quality of the interaction by examining the users' choices among presets and their inputs.

BACKGROUND

The tele-interactive net audio installation *Dialtone* [12] was a first experiment to test familiar telecommunication devices for distributed collaboration and mixing audio using telephone lines.

i2TV [13] is an experiment aimed at integrating remote and on-site participants into an event taking place in a real physical location by addressing presence and presentation issues. This human-centred approach for connected spaces is demonstrated in a distributed poetry play allowing users to alter the word sequence of Ernst Jandl's famous poem 'Ottos Mops'.

Another work [14] focuses on user interaction in virtual environments, addressing the question of whether humans are able to remap their understanding and perception of their own body functions in order to gain control of the virtual components of the environment. Users can alter the shape, colour and position of abstract virtual 3D objects using inherent parameters of their own voice.

Notion of space

All MR installations are inevitably related to one or more

types of space. Most commonly, they are grounded in real space. Other kinds of spaces are then attached or linked to it in some way. In our work, we attempt to connect people through various kinds of spaces that are connected to a real space. We use various notions of space such as architectural space, physical space or installation space. Clearly, these three expressions refer to the same thing, which is the real space. By using different terms for this or other kinds of spaces we do not wish to confuse the reader. We rather wish to convey a certain property of this space and to emphasize one aspect that we consider important in relation to other aspects of this work.

PROJECT DESCRIPTION

Concept

Phonic Frequencies is an interactive audiovisual data space, whose appearance can be altered via telephones and communication of multiple users. Access to the installation space is given by telephone numbers. The installation consists of *public* displays acting as audio-visual output interfaces [Figure 1], and a *personal* interface: telephones. Conditions of the space presented by the audiovisual displays in turn depend on the activity of active participants.

Figure 1: Phonic Frequencies prototype

Here telephone and computer networks build the framework for the responsive space that displays social dynamics, and thus constantly changes.

Phonic Frequencies extends technological concepts of telematic spaces in a poetic manner: the transformation and extension of the space through verbal input. Messages are understood as signals offering the parameters to modify the space.

System Description

The installation is centred on a physical room comprised of three or more video projections and an audio system for audio-visual display.

Incoming calls of participants are received by voice modems (each connected to a single computer). The messages are transformed into sound files and subsequently saved on the computer's hard disk. Modified communication software allows participants to send dual tone multiple frequency (DTMF) touch tones and, of course, voice or sound messages themselves to the line output of the computer's soundcard. Each sound stream is then analysed using Max/Msp [15] and coupled with a colour (red, green and blue of the animation controlled with Nato [16]. So, from the perspective of participants there is a red, a green, and a blue telephone line. (See Diagram 1 in the Appendix for more clarification of the system)

Interface

The fact, that telephones are firmly tied in our everyday life, makes them a compelling, though limited, tool for collaborative tasks for a number of reasons: firstly, a greater number of people can possibly partake than in technically more advanced environments (e.g. as discussed above [17; 18; 19]). It is accessible to local and remote audience and thus forms an open system. Secondly, the installation changes and enhances the meaning of the device itself, allowing users to experience and understand the shift of its functionality. The device contains a simple 'audio transmitter' (i.e. the microphone), and a keyboard which can be used for navigation in a complex system. So, it also challenges the user to become an active part of the work. Finally, visitors need not undergo any training in order to use them.

User Interaction

People connect to the system by dialling one of the numbers and a pre-recorded message explains the details about the modes of interaction. This has multiple layers. At its most basic level users can use their telephone's keypad to change the background colour of the projection and therefore the ambient lighting conditions of the installation space as well. In addition, they can use the buttons to browse an archive of previously recorded messages in order to replay them (play, stop, forward, rewind, skip).

The most significant mode of interaction, however, is the use of the phone to transmit sounds (words, noise, music), which are analysed and transformed into graphics in real time. Once a user disconnects, a further trace is left in form of a correspondingly coloured line, the length of which indicates the duration of the call [Figure 2].

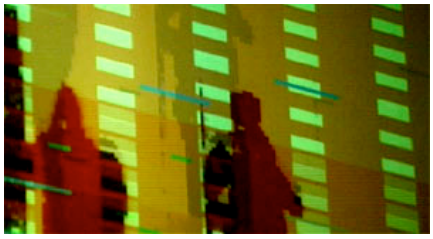


Figure 2: Sign of presence: the space is captured by a camera. The red line is in use – green and blue lines represent the duration of the previously recorded messages.

People in the installation space become part of the imagery, as they are filmed by a camera. Essentially, it is a tele-mirror that shows an abstracted view of reality.

Sound

The audio feedback is presented on a multi-channel speaker system, and each telephone input is assigned to a unique speaker, so that the extension of the real space by the 'virtual' ones represented by each caller becomes more apparent. The speakers are coloured according to a single phone line and located in different parts of the room to stress this fact. Therefore, calling the system via a different number results in the sounds being played from a different speaker in a different location.

Visuals – Image Processing

We analyse amplitude, peak and envelope of the incoming sound signal, which parameters are the basis for simple image processing operations. The processed inputs are coupled with a small number of random parameters to ensure that the visual output is not monotonous yet easily comprehensible. We use lines as a simple geometric form to keep the visuals clear and easy to follow. The audio parameters affect the width, height, position of the forms and the motion of images.

The installation space is filmed by a camera. Each frame is thresholded, subsampled and the background subtracted so that only people present in the room can be seen in the imagery. This is then mixed with the graphics visualizing the activity of users so that remote users receive visual feedback about visitors in the installation space.

TELEMATIC AND MIXED REALITY SPACE

Essentially, the work consists of two layers determining the mixed reality space. Firstly, the installation space acts as the main shared physical space; it is accessible to both viewers and actors. Furthermore, it consists of a number of public and private interfaces: The projection screens and the speakers are available to everyone, whereas access via private interfaces (i.e. telephones) is limited and therefore only a few participants can use them at any time.

The physical space is extended by other physical locations (location of callers) and a shared data space, attempting to involve remote participants and viewers. The data space can be split into an active and a passive channel. Both channels are one-way only. The active channel deals with user input – by only using the telephone to send sounds to the system and to use keys in order to control, users do not receive any feedback – while the passive one can be regarded as an presentation channel, where live audio and video from the main installation space is streamed on the web. The active channel is used to send and leave traces in the room and to navigate through the system, while the passive channel is used to give feedback. Each channel can be used independently from the other.

Local visitors will experience the installation directly [shown in Figure 3], but they cannot see remote participants; they can rather view an abstract visual representation of the participants' actions. On the other hand, remote users cannot experience the installation on the same level as local visitors since they can only interact with the others through the environment; using the given modes of interaction however they can establish communication to other participants and achieve a sense of co- or telepresence.

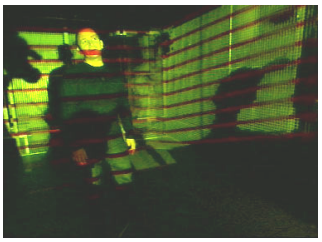


Figure 3: Visitor in the installation space.

Since we treat each loudspeaker in the installation space as an extension to this space, we regard them as dynamic boundaries. These boundaries connect the physical installation space and other virtual spaces representing a phone line, which in turn can be occupied or visited by callers. Since callers are always present in a real physical space, however, the installation can also be said to *virtually* link a real physical space (i.e. the installation space) with other real physical spaces (representing the location of a caller). Note that the latter set of locations may coincide with the former, meaning that the space could be (and often is) self-referential.

Streaming

When users send data to the system, the desired parameters are extracted and linked to the audiovisual output. Information from the telephone network is therefore projected into another medium.

The audio and video stream containing the mixed audiovisual data enables remote participants to see and hear the results of their inputs. The stream is presented on the internet to attract a wider audience from any location [Figure 4]. This ensures, that remote users can view (and hear) the results of their actions. During the first installation, the stream was also transmitted into another room near the installation venue in order to test the remote scenario.

Delay is an essential characteristic of digital networks regarding the transmission of audio-visual data [20]. An important aspect was to minimize delays rather than to eliminate them completely to ensure that the cycle between interaction and system feedback is easily understood. We tested different streaming formats using our recently developed Mobile Streaming Lab [21] and finally used Real Media, in order to reach best compatibility and a latency starting from five seconds. This results in sound and image aesthetics, which clarify collaboration over distance.

Figure 4: Remote participant viewing the Real Media stream.

RESULTS, DISCUSSION AND EVALUATION

Phonic Frequencies is an environment that connects different people in different locations and on different levels, enabling them to shape a physical space, to collaborate with other participants in form of sound and vision. During the first public presentation at the workshop *'Networked Archives'* [22] in October 2002 a number of people experimented with the installation from various locations. Although we did not conduct a formal user study, we gathered a number of points from observation and discussion with visitors.

Most users noted that, after a short period of interaction, they stopped speaking into the phone but rather started constructing sounds using objects and sound sources present in the room, in order to test how this would affect the graphical output. Indeed, there was a big difference between voice and other forms of sound. In this sense, the acceptance of the telephone as a user interface matched the scope of our expectations, because not only did users understand the new functionality of the device, they also realised that, by using different sound inputs, they could affect the environment in completely different ways. Some users also stated that they felt uneasy and more inhibited using their voice. This, however seemed to be a problem for local visitors rather than to remote participants. Callers from remote locations did not report such feelings; they rather felt they were invisible to others and therefore 'safe' to speak freely. The performance – which also relates to the acceptance of the interface – could be related to the emotional state of the user: the more self-conscious and uneasy he or she feels the lower the performance and acceptance.

The experiment failed however in one aspect: though multiple users tried the installation simultaneously, it did not create any real sense of co-presence or collaboration. Though the users were *aware* of each other's actions in form of the audiovisual output they created, there was no sense of mutual interaction. Even though multiple inputs were perceived as more interesting by the majority of the audience and participants, it merely helped users to understand themselves as part of an interactive installation instead of a collaborative one.

Most remote actors, however, felt that the requirement of having a phone *and* a connection to the internet at the same time in order to send input and to receive feedback was frustrating to handle.

CONCLUSIONS

We have presented an interactive environment whose ambient character can be controlled using telephones. The system transforms sound input made by the user into an abstract graphical representation that is displayed in the space together with the sounds provoking the visual response. It also provides viewing facilities for remote by streaming the output and a live view of the space on the web. This apparatus can be regarded as a dynamic mixed reality space: a physical installation 'stage' forms the centre of the space, which is dynamically extended by virtual spaces represented by user input.

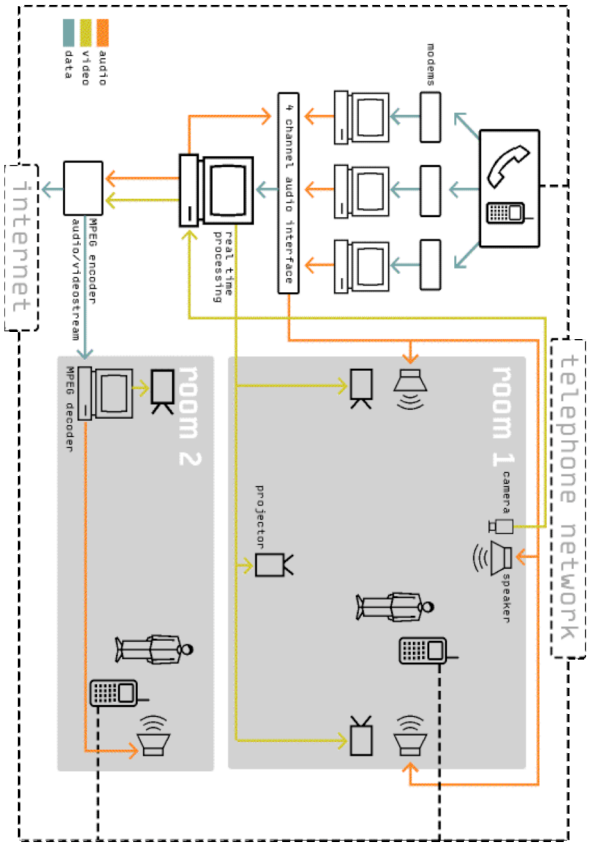
The installation demonstrates that users understand the interaction cycle for the given application scenario, providing an entertaining and stimulating atmosphere for remote and local users. Secondly, we were able to link arbitrary virtual spaces to the physical installation space via audiovisual output, creating a dynamic mixed reality environment.

Since we did not conduct a formal user study, we can not directly confirm any of our assumptions. Thus, it would be interesting to study some aspects of the installation in greater detail. For instance, can users directly establish a link between *voice* input and visual output? Can one directly determine which sound parameter (e.g. volume, pitch etc.) is linked to a certain output parameter? Since a sound wave is always characterised by the parameters we analysed, there is no way for users to experiment with just one parameter, simply because the others are never absent. This is a serious flaw in the concept of using inherent sound parameters for control, because, at any given time, all parameters are present in the sound and will therefore trigger something. We could go even further, and claim, that because of this flaw, direct or absolute control over the environment is never fully given to the user, even though most users did report a feeling of control.

ACKNOWLEDGEMENTS

Within the framework of the 'digital sparks 01' [22] production award, Phonic Frequencies was designed and developed by Tamás Szakál at the MARS- Media Arts & Research Studies Lab at the Fraunhofer Institute for Media Communication in 2002 with help and advice from within the interdisciplinary research group under the direction of Monika Fleischmann.

APPENDIX: SYSTEM OVERVIEW (DIAGRAM 1)



REFERENCES

1. Rokeby, D., The Construction of Experience: Interface as Content, in: "Digital Illusion: Entertaining the Future with High Technology," Clark Dodsworth, Jr., ACM Press, 1998.

2. Fleischmann, M., Strauss, W., Images of the body in the house of illusion, in Art@Science, Sommerer, C. (editor), Springer, March 1998.

3. Strauss, W., Fleischmann, M., Thomsen, MR, Novak J., Zlender, U., Kulessa, T., Pragasky, F., Staging the space of mixed reality – reconsidering the conceptof a multi-user environment. VRML99, Paderborn, Germany, February 23-26 1999.

4. Milgram, P., and Kishino, F., A taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems, Special issue on Networked Reality, Decemver 1994.

5. Shanken Edward A. Tele-Agency: Telematics, Telerobotics, and the Art of Meaning, Art Journal, 2000: pp. 64-77

6. Kac, Eduardo, Ornitorrinco and Rara Avis: Telepresence Art on the Internet. Leonardo, Vol. 29, N. 5, pp. 389-400, 1996

7. Ken Goldberg: http://telegarden.aec.at/

8. Blinkenlights: http://www.blinkenlights.de/

9. Sensorband: http://sensorband.com

10. GolanLevin: http://www.flong.com

11. Granular Synthesis :http://www.granularsynthesis.info

12. Szakal, T, Dialtone, http://dial.tone.hu,

13. Novak J., Fleischmann, M., Strauss, W., Peranovic, P., Seibert, C., The i2tv System: A mixed reality communication interface, ConTEL 2001, 13-15 June 2001a

14. Groenegress, C., Thomsen M.R., Slater, M., Designing Intuitive Interfaces for Virtual Environments. Masters Dissertation, UCL, London, 2001.

15. Max/Msp: http://www.cycling74.com

16. Nato: http://www.eusocial.com

17. Koleva, B., Taylor, I., Benford, S., Fraser, M., Greenhalgh, C., Schnädelbach, H., vom Lehn, D., Heath, C., Row-Farr, J., Adams, M., Orchestrating a mixed reality Performance. CHI 2001 ACM Press, 2001.

18. Greenhalgh, C., Benford, S., Taylor, I., Bowers, J., Walker, G., Wyver, J., Creating a live broadcast from a Virtual Environment. SIGGRAPH'99, pp375-384, Los Angeles, 1999.

19. Benford, S., Greenhalgh, C., Reynard, G., Brown, C., Koleva, B. Understanding and constructing shared spaces with mixed reality boundaries. ACM Transactions on Computer-Human Interaction, Vol. 5 (3), pp182-223, 1998

20. Tanaka, Atau, Netmusic – a perspective, Tokyo, Japan, 1999.Originally published in the catalogue from the Festival du Web, 1999, Webart, Paris

21. Pfuhl, D. and Peranovic, P. (2002): Mobile Streaming Lab - leading to a modular learning environment. In Proceedings of the eight international conference on virtual systems and multimedia - Creative digital culture, VSMM 2002, Gyeongju, Korea, 25th-27th September 2002 (pp. 578-584). Seoul, Korea: Kiwissoft

22. NetworkedArchives: http://netzspannung.org/workshops/online-archives/

23.DigitalSparksURL: http://netzspannung.org/digital-sparks/