Marvin: Supporting Awareness through Audio in Collaborative Virtual Environments

Martin Kaltenbrunner
FH Hagenberg
Hauptstrasse 117
A-4232 Hagenberg
Austria
modin@yuri.at

Avon Huxor
Centre for Electronic Arts
Middlesex University
Cat Hill, Barnet
UK
a.huxor@mdx.ac.uk

0. Abstract

This paper describes Marvin, an awareness support agent written in Java. The system provides audio cues, text-to-voice and voice recognition, and currently operates as a bot in ActiveWorlds, an Internet-based, shared 3D virtual environment. The ActiveWorlds space was designed to facilitate chance encounters between members of distributed work groups, and Marvin was written to overcome the problems that arose in use. Audio allows us to free the user from both attending to the screen, and also from being present at only one virtual location within the world, drastically enhancing the chances of encounter. It also blurs the boundaries between the virtual and physical workplace.

1. Introduction

This paper builds on recent work that explored the use of AlphaWorld (AW), a simple multi-user Internet-based virtual environment to support chance encounters (Huxor 1999). That is, rather than being a virtual space in which scheduled meetings for distributed teams might occur, it supports the unplanned meetings that take place in the conventional workplace. These meetings, often occurring in corridors, or by the coffee machine, have been shown to be very important for the functioning of an organisation team (Backhouse & Drew 1992) and risk being lost in a distributed organisation. The informal flow of information, of contacts to maintain trust, are central, and are often not fulfilled by existing software tools such as email, which require an intent to communicate.

As part of a longer research program to develop a 'virtual Centre for Electronic Arts', a shared virtual space was build in AlphaWorld, the oldest and largest world that is available from the ActiveWorlds[1] client. This client supports easy navigation of the space and real-time text-chat between users. The virtual office 'space' within AlphaWorld was created with the aim of supporting new working forms in which people are working not only in the traditional office, but also at home, on the road, at customer premises or other venues. The authors use the space to access collaborative documents that are stored in an Internet-based server called BSCW[2], which allows for collaboration across institutions, and to informally meet both colleagues, and 'weak ties' (those people who pass through a space but are not immediate colleagues). These

[1] http://www.activeworlds.com
[2] http://bscw.gmd.de

encounters were spatially managed, in that task-related content were placed in stable rooms in the virtual office, content which drew users relevant to these tasks. Encounter occurred visually, in that each user has an avatar that can be observed passing through the space. This ActiveWorlds virtual office[3] has been active for some years now, and has proved itself useful in supporting both chance encounters and weak ties, as intended. However a number of problems arose, but the most important of these were those that prevented chance encounter taking place as often as possible.

1.1 Problems of the CVE

The space aimed to support encounter and awareness of other users, but awareness failed in many cases. These failures can be grouped into three types:

1. The user is at the machine, but the task they are undertaking employs the full screen, so that the virtual space window is hidden below that currently in use.
2. The user is nearby the machine, but not attending to it as they are reading a paper document, on the telephone or talking to colleagues, for example.
3. The user is away from the machine visiting another office, en route elsewhere, at lunch etc.

These problem of missed encounters are crucial, due the critical mass which is required to make such social media function. As I missed various visitors, they appear to visit less often, the chances for encounter drops further, and a vicious circle develops. This must be broken to allow for the communication process to be enabled.

It was recognised that the problem arose from a contradiction between the goals of the space, in terms of supporting mobility in the workplace, and its actual effects. Although the shared virtual space was employed to support a more flexible form of working, one in which a user is not tied a single office desk, it actually bound the user in two important ways. Firstly, they are tied to their computer monitors so as to see any passing avatars in the shared space, and secondly, they are further tied within the virtual space to a single location. The first problem calls for a means of indicating presence in the world that is non-visual, the second a reconsideration of the nature of presence in virtual spaces.

It retrospect, it seemed unnecessary that we adopt all the constraints of structuring action that are derived from the spatial metaphor. In the physical world users are adopting many other techniques, such as mobile phones, to overcome these, so why re-introduce them in virtual spaces? That is, can we separate the positive from the negative aspects of the spatial metaphor, to give users more benefit, a common problem in metaphorical interface design? It was concluded that it is the spatial management of tasks that remains central to collaborative spaces, and that we can be more flexible in our use of the term ‘avatar’. If each user has many avatar proxies, they can have a ‘foot in many camps’, depending on the range of tasks they are working on at the time.

[3] The office can be visited in AlphaWorld at co-ordinates 188S 34E

It was decided to address these concerns through two means:

1. Use of audio as an awareness mechanism, one that allows the user to not attend to either the screen or, if working with another application onscreen, to ActiveWorlds.
2. Allow users to have multiple presence in the virtual world, each with an associated set of audio cues. Just as we can listen to the physical office next to our own for cues, so audio facilitates attention to diverse spaces.

2. Audio Cues and Audio Spaces

Sound has been a neglected part of interface and systems design, but has recently seen a growth in interest. The requirements of an audio interface to a shared virtual environment can draw upon a wide range of work on audio use in computing. In one major area of research, sound is being used to support the standard GUI interface (Gaver 1989), especially for small screen PDAs, and mobile telephone access to information services (e.g. Brewster 1998), as such devices are seen as a major market segment in the future. Other work has investigated how real-time audio links between users and physical spaces can support distributed workgroups (Mynatt et al. 1998, Hindus, et al. 1996), and Sawhney & Schmandt (1999) have looked at how mobile access to services can be supported by a wearable headset/microphone device.

Our work on extending ActiveWorlds looks to all these. Many of the issues that apply to audio interfaces, namely the design of audio icons, sound effects and text-to-voice use, apply to supplementing the current AlphaWorld visual interface. Equally, as our concern is with facilitating the maintenance of a social sense throughout a distributed workgroup, lessons from audio spaces are relevant. For Hindus et al. (1996) found that audio only spaces can lead to 'social spaces', and although ActiveWorlds currently supports text-chat only, we are looking to augment this with voice over IP (see, for example, Onlive Traveler[4], a shared 3D world that employs voice rather than text-chat).

The audio supplement to ActiveWorlds (AW), the browser technology for AlphaWorld, is a system called Marvin, described in more detail below. Marvin creates a presence in various points in AW worlds, and gives audio cues representing various events in the world. In addition it uses text-to-speech to allow users to listen in on chat in the worlds, and we are investigating voice recognition technology to support hands-off navigation within them.

The detailed aspects of the sound design it employs are, where possible, drawn from published empirical results. For example, door knock and door opening worked well to indicate comings and goings (Cohen 1994), and such door sound cues are used to represent other users entering or leaving the area of interest. James (1996) notes that users preferred natural sounds to artificial sounds, even when poorly chosen, and typical sounds (e.g. the high-low tones of a standard doorbell) were identified quicker than atypical sounds (Ballas 1994), so these have also been used. However, we have also been aware that it was also discovered that, if too similar, it could confuse users between activities in the virtual and the physical spaces. With regard to text-to-voice, text-to-speech need 'prefacing' so users could be prepared to hear the main message properly (Cohen 1994). Also, James (1997) found that multiple voices were valuable

[4] http://www.onlive.com

when used to speak online documents, to mark macro-structures, such as headings, in a document. Although our application differs from James', we have used different voices to represent different users in the space, and we can easily add prefacing comments before the quoted text-chat.

This audio complement to the standard AW interface provides additional functionality over the 3D space, in that it allows users to be in more than one place at a time. Cohen (1994) found that priority attribution for notification seems task dependent: certain other users and/or files were important at different times. Exploiting our ability to attend to multiple audio channels, users can have agent bots present in various places in the virtual environment, depending on which tasks are relevant at the time. The *locales* (Fitzpatrick et al. 1996) that maintain task-related content and encounters can thus act as a simple mechanism for managing the setting of priorities within the audio awareness component. That is, we can avoid the situation where the user must modify notification priorities manually, by letting these be managed by a change of place, just as different physical spaces create new affordances. This approach follows Erickson's (1993) idea that increasingly, as we move to shared applications, the Interface can best be understood as an Interplace, a place or places that structure activity.

3. Marvin: An Auditory Awareness Bot

3.1 Implementation

The Marvin bot application is a simple programmable agent, which can enter multiple information servers (such as AlphaWorld) as a proxy and report noteworthy events via speech and audio to the user. Since with the help of the robot one is constantly aware what is going on, he can if desired then directly turn his attention to the appropriate application and enter the locale instead of the robot. The proxy itself isn't an independent intelligent agent; it is an extension to the user's senses for extended awareness of events in info-space.

Marvin is implemented in the Java programming language, version 1.1 or above. This decision was made mainly because of Java's platform independence and the availability of various multimedia features like Speech, Sound, and Media APIs. We use the IBM ViaVoice Runtime as the speech recognition and synthesis engine. The application runs and is developed on Windows NT, but it should also run on a Linux or Solaris machine since there is also Java compatible speech recognition software available for this platform, though we never tested it. If IBM also releases a Speech API package for ViaVoice on Apple Macintosh, Marvin should also run on this platform without any further modification.

The application consists of two layers:

1. The Marvin kernel, which provides the basic features like speech recognition and synthesis, sound output, logging and so on. Upon start-up this core loads all present plug-ins and starts them as independent threads. The plug-ins then can use Marvin's event processing interface.

2. The plug-in interface, which allows the easy addition of robots for any information service. At the moment we have only implemented an Active Worlds Robot. But due to this architecture and the fact that Marvin is released under the GPL (Gnu General Public License) it is easy for third party developers to add their own plug-ins for various other network information services or multi-user environments, such as BSCW/Nessie, IRC or ICQ.

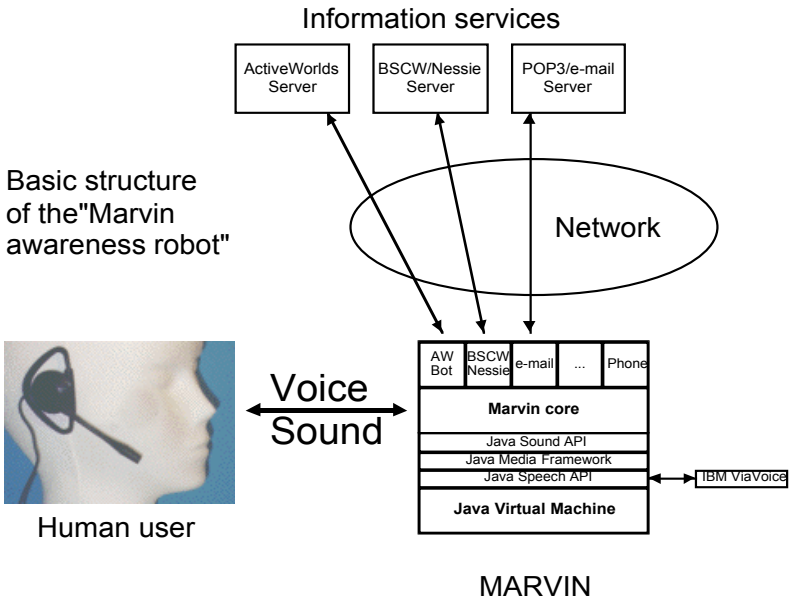

Fig. 1 Structure of Marvin system

3.2 Event processing functionality

As mentioned above, the Marvin kernel provides basic audio notification features for all available plug-ins. This means the simple playback of audio file or the output of synthesised speech, also provides an interface for the voice control of all components. The plug-in applications cannot directly access the sound and speech methods. These methods are combined in a central event-processing interface. Depending on their priority, events are either only logged to a file (no priority) or the user is informed with speech and audio (high priority). The table below shows the exact priority scheme we currently use. Higher priority level events implicitly include the processing of their lower priority levels.

no priority	logged to file
low priority	logged to screen
medium priority	audio clip
high priority	speech output
urgent priority	mobile notification (not yet implemented)

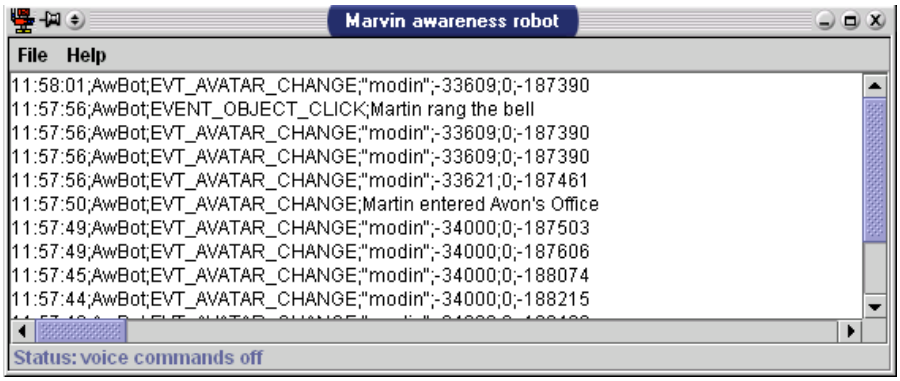

Fig. 2 Example Log of Events

3.3 The Active Worlds (AW) Plug-In[5]

The AWBot plug-in uses ThierryNabeth's Java native interface for the Active Worlds Software Development Kit. Applications implementing this interface can place a remote-controlled avatar into the Active Worlds space, which can interact with other persons or events in this virtual environment. Once such a robot is placed in a certain predefined area of a world within the AW server, it will notice any event in its surroundings. These events are then caught be the robot application, analysed and then processed by the Marvin core according to their priority.

Since, as discussed above, a robot only can notice events in its nearest surroundings, multiple instances of the robot – called probes – were created and placed in various areas of the AW space. To ease the modification of the robots, all the variable parameters such as position, user names or sound files are stored in separate configuration files that are processed upon start-up.

3.4 Marvin in Use

Marvin is a separate application that can be started to supplement the standard Active Worlds browser. The user can then assign a number of proxy avatars of Marvin in places of interest. For example, in addition to one's own virtual office, one might place proxies in parts of the space that link to content relevant to ongoing projects. When another user enters the space, an audio cue of a door is heard, and Marvin greets them using both text-chat in the window, but also with text-to-voice. The former lets the visitor know they have been acknowledged, the text-to-voice allows any user with Marvin to hear the greeting (which includes the name of the guest), letting them know who has arrived. Similarly, users departing from the zone specified are giving a farewell in the text-chat box, and this is also spoken aloud. Clearly, I could be working away from my machine but still be aware of activity in the space, and respond accordingly. It is also significant that it allows persons sharing a physical office to be mutually aware of activity in the others spaces, adding to a sense of working community. Both authors share a physical office and this has been a noticeable result in adding to the overall sense of awareness within the workplace.

[5] Because the Active Worlds SDK is only available for the Windows operating systems, this plug-in is not platform independent.

Marvin also repeats (with text-to-voice) the words that have been 'spoken' (using AW text-chat) by other users, so that one can listen into a conversation while doing other tasks, and jump in when appropriate. This feature already proved interesting in the limited time that the authors were using Marvin. It allowed a conversation between two colleagues: one in a different part of the same building, and another across the city, to be followed (even though it was occurring on a machine across the room). The avatar, representing Marvin, assured that these other participants were aware of my proxy presence. For certain designated individuals, those that are particularly relevant to our collaboration, specific voice types (age and gender) have been assigned so that they can be recognised.

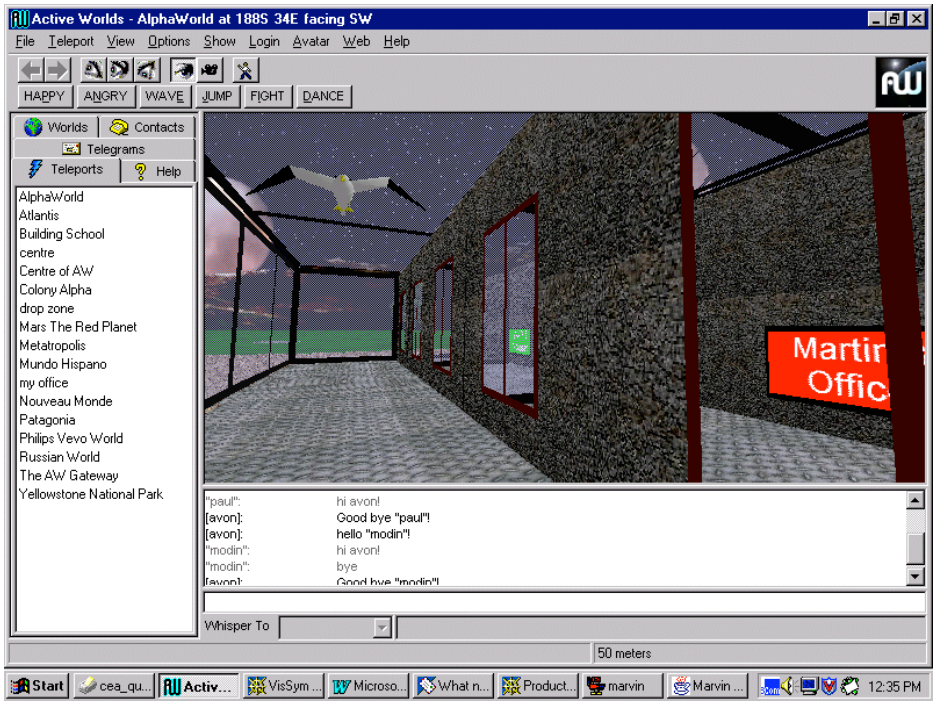

Fig. 3 Marvin in AlphaWorld

Other sound cues indicate if other users interact with objects that belong to me in the virtual space, such as those that hyperlink to web content. Thus, I can also be aware of task related work activity by other users in addition to supporting personal encounters.

The figure above shows Marvin, whose avatar is a bird, hovering in the virtual space. The text-chat box below the space captures the interaction between Marvin, who as my proxy uses the name [avon][6], and visitors.

[6] The square brackets indicate, in AW, that an avatar is a 'bot'

4. Issues Arising during Use

In addition to providing an auditory interface to overcome the problems identified in the original AW space, the implementation and use of Marvin has lead to a number of issues.

Firstly, it is easier to create 'mixed' spaces with an audio enhanced interface, as sound emanate from the both physical and virtual and are perceived in a similar manner. This differs from the visual aspect of the virtual space, which presently has a very distinctive appearance from physical space. It was already noticeable that during limited use, this similarity creates a different relationship to the virtual environment and events within it. The door sounds and phone ringing sounds feel similar to those of doors and phones from adjacent rooms. And it is noteworthy that, increasingly, the events in the physical world to which we respond have a virtual component. For example, an informal study of the soundscape of CEA was undertaken to assist in the design of Marvin's cues. It was found that the most important cues included phone rings (whether answered or not), and telephone answering machines. This hints at a view in which the conventional distinction between virtual spaces and physical ones break down, for sound has a 'physical' sense (Gaver 1993).

Secondly, audio also has problems that arise from its being so pervasive, namely annoyance of the user and others, and it particularly presents a problem for privacy. However, it was found that features from the virtual space could help in the management of these. The fact that the interactions are spatially set assists in managing the sense of what is public and what is private due to the ownership of spaces. Also, having avatar assists in preserving symmetry in an online interaction: if one has a proxy presence in another space, it is visible through the personification of the avatar. Not knowing who is listening in on an audio space is a frequent complaint. It would be possible for a user to have multiple presence within ActiveWorlds, but use an invisible avatar, but this would give an impression of surveillance. Therefore, we have personalised the various proxies of the user as one of the standard AW avatars. In the current version, it is a bird, as this conveys the sense of being aloof, of partial presence.

Finally, Gaver (1993) reports work in which blindfolded students could orient themselves by 'acoustic landmarks', resonance, echoes and near/far sounds. This suggests that it may be valuable for the sounds emanating from a virtual space to represent its visual spatial form to assist in identification. Thus we may have a large or a small virtual space within the world. One possibility that this suggests is that we can use different acoustic qualities to 'tie' sounds, which may be from a multiplicity of locations within the virtual space, to one in particular. For example, a large room in AlphaWorld would have reverberation, which would affect the sounds from it, and could be recognised as such by users directing their response to the appropriate place.

4. Conclusions and Future Work

This paper has argued that:

1. Audio can improve awareness in collaborative 3D virtual environments by freeing the user from being bound to the monitor.
2. Audio cues can also support multiple proxies for each user in the virtual space, allowing users to have access to many work contexts at any moment. Each place in the environment, however, maintains its spatial management, its sense of *locale* (Fitzpatrick et al. 1996).
3. This idea of using locales to manage different notification priorities illustrates the notion of Interplace replacing the conventional interface. Users do not modify a menu, they just move from one place to another.

In the introduction it was pointed out that the original AW space failed in supporting awareness in various ways. The work described above sought to address the problem for the user who is nearby to their computer, but future work seeks to extend the principle to support users who are away from any machine. It is inspired by work that has similar aims, such as the Nomadic Radio project ofSawhney & Schmandt (1999), but they use specialised hardware devices. We aim next to explore how awareness cues from the virtual space can be made available to users away from their physical office, on route to a colleague, in the coffee area, etc. As mobile phones have become so ubiquitous, it seems likely that they will become a natural portal to online resources. This can only contribute further to blurring of the distinction between the physical and the virtual, creating a unified 'space of work'.

Acknowledgements

The authors would like to acknowledge the efforts of Thierry Nabeth of the Department of Technology Management at the Centre for Advanced Learning Technologies, INSEAD, France. His Java port of the AlphaWorld SDK made the Marvin system possible, and his speedy updates from our comments were invaluable.

Bibliography

Backhouse, A. & P. Drew (1992) The design implications of social interaction in a workplace setting. *Environment and Planning B: Planning and Design*, 19: 573-584.

Ballas, J. A. (1994) Delivery of Information through Sound. In: Gregory Kramer (ed.) *Auditory Display*, SFI Studies in the Sciences of Complexity, Proc. Vol. XVIII, Addison-Wesley, pp. 79-94.

Brewster, S.A. (1998). Using non-speech sounds to provide navigation cues. *ACM Transactions on Computer-Human Interaction*, 5(2), pp 224-259.

Cohen, J. (1994) Monitoring Background Activities. In: Gregory Kramer (ed.) *Auditory Display*, SFI Studies in the Sciences of Complexity, Proc. Vol. XVIII, Addison-Wesley, pp. 499-531.

Erickson, T. (1993) From Interface to Interplace: The Spatial Environment as a Medium for Interaction. *Proc. of Conf. on Spatial Information Theory*. Heidelberg: Springer-Verlag.

Fitzpatrick, G., Mansfield, T. & S. M. Kaplan (1996) Locales Framework: Exploring foundations for collaboration support. *IEEE Proc of the 6th Australian Conf on Computer-Human Interaction OZCHI'96*, Hamilton, NZ, pp. 34-41.

Gaver, W. W., (1989). The SonicFinder, a prototype interface that uses auditory icons. *Human Computer Interaction* (4): 67 - 94.

Gaver, W. W. (1993) What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5(1): 1-29.

Hindus, D. et al. (1996) Thunderwire: A Field Study of an Audio-Only Media Space. *Proc. ACM Conf. on Computer Supported Cooperative Work (CSCW'96)*, pp. 238-247.

Huxor, A (1999) The Role of 3D Shared Worlds in Support of Chance Encounters in CSCW. In: Vince, J. & Earnshaw, R. (eds.) *Digital Convergence: The Information Revolution.* Springer-Verlag

James, F. (1996) Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext. *ICAD '96 Proceedings*. Xerox PARC, 4-6 November 1996, pp. 97-10.

James, F. (1997) AHA: Audio HTML Access. *Proc. of The Sixth International World Wide Web Conference.* Santa Clara: CA, pp. 129-139.

Macaulay, C. and Crerar, A. (1998) Observing the Workplace Soundscape: Ethnography and Interface Design. *In Proceedings of the International Conference on Auditory Display (ICAD '98)*, Glasgow, November 2-5.

Mynatt, E. D., Back, M. & R. Want (1998) Design Audio Aura. *Proceedings of the Computer-Human Interaction (CHI) Conference*, Los Angeles, April 1998, pp. 566-573.

Sawhney, N. and C. Schmandt. (1999) Nomadic Radio: Scaleable and Contextual Notification for Wearable Audio Messaging. *ACM SIGCHI Conference on Human Factors in Computing Systems*, Pittsburgh, May 15-20.