Faculty and Researcher Publications

2002

Entertainment Industry Sound Design Techniques to Improve Presence and Training Performance in VE

Shilling, Russell

Monterey, California: Naval Postgraduate School.

Entertainment Industry Sound Design Techniques to Improve Presence and Training Performance in VE

LCDR Russell Shilling, Ph.D.
MOVES Institute
833 Dyer Road, Room 254
Naval Postgraduate School
Monterey, CA 93943-5118
(831) 656-2543
shilling85@alumni.wfu.edu

ABSTRACT: *The design and development of auditory interfaces in virtual environments has lagged behind visual interfaces. However, the auditory interface should be considered an essential component to VE that adds ambience, emotion, and a sense of presence to the simulation. The entertainment industry has long recognized the importance of sound to add these perceptual elements to film and videogames. As a result, the film industry has devoted many of its resources to developing techniques for producing sound effects and ambiences that evoke emotional responses and immerse the viewer in the film. Although interactive media present unique challenges to these techniques, many of them can be readily applied. Additionally, audio engineering techniques widely used by the film industry can be applied to improving the quality of virtual environments. If the same care and artistry is applied to virtual environments as in film, the sense of presence should be improved and it is possible that training effectiveness will also be enhanced. The current paper will summarize efforts of the MOVES Institute Immersive Audio Research Laboratory to work with the entertainment industry (film, videogame, and engineering) to adapt these techniques to virtual environments. Research examining the use of physiological response measures(temperature, electrodermal activity, heart rate, etc) to assess emotion and presence in VE will also be examined. In particular, we have shown that speaker systems contribute more to emotion (and we argue presence) than headphone systems.*

1. Introduction

Imagine entering a meeting with a room full of people. When you enter the room, you realize that the speaker's voice is emanating from all points in the room, yet the room is totally anechoic. In addition, you see other attendees moving in the room, yet there are no additional noises in the room except the speaker's voice. Despite walking into a "real" environment, your sense of reality would be challenged. The entertainment industry has long recognized the importance of properly designing sound effects and sound systems to add realism, emotion, and a sense of immersion to film and to video games. As a result, the film industry has allocated significant resources to developing techniques for the design of sound effects and ambient sounds that evoke a sense of realism and manipulate the emotional response of the viewer. It is difficult to imagine that all sound heard in the battle scenes of "Saving Private Ryan" were added in layers after the film was shot, yet they were. In the opening scenes depicting the Normandy invasion, the audio effects, including the actor's voices, are completely synthetic; added to the film after it was shot. The audio effects were spatialized using a surround sound system to immerse the audience in the sound field. Even though the medium is non-interactive, a movie with a properly designed audio track represents a better example of an auditory virtual environment than most high-end simulations developed by industry and the military. Although film is a non-interactive medium (for most people), videogames are another matter. These same audio design techniques have been adapted by the electronic gaming industry to add both realism and excitement to their titles. Although, not reported here, current research in our laboratory is aimed at deriving objective rather than subjective measures for the contributions these techniques add to virtual environments using a variety of techniques including psychophysical measurement, electrophysiology and even basic memory experiments.

There are many lessons to be learned from the entertainment industry. First is in the field of sound design. One theme has emerged from discussions with professional sound designers, producers, and audio engineers. Sound is emotion! Anyone in the audio industry will tell you this. A visual display lacking a properly designed audio component will be emotionally flat. Sound design for interactive environments is not a trivial matter in either its importance or implementation. As will be discussed, sound design needs to be a carefully planned process.

The second lesson to be learned is in the design of the listening space. Both speech intelligibility and emotional impact are dependent upon the quality of audio reproduction and the acoustical properties of the listening space. The best example of the importance placed on reproduction and physical acoustics by the entertainment industry is THX certification. George Lucas formed THX because of his belief that "sound is fifty percent of the movie experience" [1, 2]. He was distressed that the impact of his movies was dampened by poor sound reproduction and acoustics in the movie theaters. Thus, THX was formed to standardize and certify audio reproduction in theaters to assure that they produced intelligible speech and that audio reproduction was as close to the product created in the original mixing room as possible.

2. Sound Design

The first rule of sound design is, "see a sound, hear a sound." [3] Unless we supply the appropriate background sounds (machinery, artillery, animals, footsteps, etc) the participant will probably feel detached from the action. Footsteps sound differently when you're in the grass as opposed to the pavement or in a hallway. Likewise, the sound of the action of an M-16 changes depending on whether you're located inside a building or outside. Explosions and other concussive events are not just auditory; they are tactile, full-body experiences. Vehicle noise is very specific to a particular vehicle and becomes part of the expectation for using that device. These are the types of things that create immersion, ambience and emotion in movies, and the same should hold true in virtual environments. Four of the audio elements we've been focusing on in our development of virtual environments and videogame based simulations are footsteps, vehicular sounds, weapons, and general ambience.

2.1 Footsteps

As we are discovering in our military videogame development activities, these techniques should be applicable to the types of virtual environments and simulations currently under development by military establishments. The entertainment industry employs foley actors to create footsteps for films. Footsteps are a rather small detail that sound designers believe (an unanswered empirical question) adds to the sense of presence in a film, even in conditions where footsteps may not actually be heard in the real world. Current game editors allow us to create the sound of footsteps on different terrains and textures. This should be a *de facto* standard for all VE development for dismounted infantry. In fact, it has been our experience that the implementation of audio in the editors of gaming engines are far more advanced than in so-called, "high-end simulations".

2.2 Vehicular Sounds

Likewise, vehicle noises should be as accurate and as detailed as possible. Recordings and sound levels should be available for sounds inside, behind, and in front of the vehicles [3]. For instance, for various military amphibious vehicles, the sounds of wheels on different terrain surfaces may be very important for creating a sense of presence. Details such as engine idle, engine revs, and gear changes may need to be included, both in the water and on land. Our laboratory has recently applied professional recording techniques for creating audio for an LCAC using professional sound designers and recording engineers to excellent results. Reproduced over a high quality audio system, the result is very accurate and very compelling. Other platforms will soon be following suit.

2.3 Weapons Sounds

Weapon sounds provide a particularly difficult challenge for simulations designers. One of the first things you will hear from customers of military simulations is that they want the weapons to sound accurate. Practically speaking, this is impossible. Although the mechanisms and actions of weapons (foley) can be accurately recorded and replicated, the actual sound of a weapon firing is very problematic. First, microphones cannot capture the full dynamic range of the majority of concussive weapons. Likewise, commercial speaker and headphone systems cannot produce this dynamic range. Therefore, an M16 will never sound exactly like an M16. From discussions with various videogame companies and sound designers for films, attempts to use accurate recordings of weapons result in an uninteresting and emotionless perception of the auditory event. Listeners complain that they don't sound realistic. There are several solutions to this problem. One solution is to record weapons with multiple tape recorders using different types of microphones that accentuate different parts of the frequency spectrum, then mix the recordings together [3]. This tends to be expensive and tedious since you need two sets of expensive recording gear. For smaller budgets, the solution is to dive into your sound effects libraries and look for weapons sounds that combine the audio attributes you desire, then mix them together [3]. These techniques produce effects that are "hyper-real" [1]. The sound of an actual M-16 firing may not sound as exciting as that of an M-16 mixed with other weapons to produce something more compelling. Remember, the best you can accomplish is producing a gunshot that sounds SIMILAR to the actual weapon. The worst thing you can do is simply use your editing software to equalize the audio and boost the bass response. A large part of the emotional response is due to the high frequency crack of the weapon and not, as many people believe, the low frequency boom.

2.4 General Ambience

The audio element which is probably the most important aspect of evoking the sense of immersion in a VE is ambient sound. It's also the least likely to be implemented. The sound of a breeze, a computer fan, an air conditioner, or an electrical generator may not be obvious to the main training focus of a simulation, but they will contribute to the sense of presence or immersion. It may seem like a trivial point at first, but we do not live in a silent world. We create surrealistic unreality in VE by failing to include these basic cues. We don't notice them when they are there. But, we do notice their absence. One of the things we are currently attempting to demonstrate is that these ambient sounds do, in fact, impact performance and training in VE.

3. Sound Libraries ~~V~~vs Custom Recording

There are some very fine prepackaged professional audio libraries. Good libraries are not cheap, but worth the expenditure. We've found that sound libraries provide excellent weapons foley, footsteps, and a wide variety of vehicles. They do not typically do a fantastic job for explosions and weapon firing. The sound designer will have to do significant amounts of work to get these sounds to have a significant impact on the listener. Not surprisingly they also lack the more esoteric weapons platforms and vehicles. Hence our recent foray to record an LCAC. If you do not have your needed sound in a library, you will either have to synthesize your sounds from recordings of similar equipment, or go out and record it yourself.

Detailed instructions for obtaining these recordings are readily available from experts and publications on the topic of "sound in film" [1,2]. However, be warned that recording audio is as much of an art as a science. Digital recording is not always better, especially when recording explosions. There is a dizzying array of microphones, each with different frequency response and directionality. Just because you've gone out and bought a $1500 digital tape recorder, does not mean you're ready to start recording effects. So, be advised that recording equipment can be very expensive, and you're going to have to learn to use it properly. Our laboratory is currently using a Tascam DA/P1, which is an excellent quality digital recorder commonly used as a field recorder and is capable of handling professional quality microphones. We will be purchasing (or leasing) an analog recorder in the near future to handle the weapons sounds mentioned earlier.

If you don't want to spend large amounts of time learning the art of recording and audio postproduction and you have a limited budget, it might be better and cheaper to hire a professional to do the work for you. The bottom line is that sound design is a very complex issue and deserves far more attention in the development of virtual environments and training systems than it is currently awarded.

4. Design of Listening Spaces

One of the first decisions that need to be made when creating a VE or training system is whether audio needs to be presented over headphones or speakers. Each presentation method has strengths and weaknesses. Headphone spatialization is typically far more expensive and needs to be customized to the individual listener for optimum effect, especially if you want to simulate elevation cues. Bass response on headphones is limited to the ear canal. Speaker systems have a much larger footprint and are sensitive to room acoustics. Elevation cues are more difficult (though not impossible) to simulate.

4.1 Headphone Systems

A detailed explanation of Spatial Auditory displays using headphones and some rudimentary information on speaker system design can be found in Reference [4]. Using signal-processing techniques, it is possible to generate stereo (binaural) signals over headphones that contain most of the normal spatial cues available in the real world (azimuth, distance & elevation). Spatial audio uses pinna cues in the form of head-related transfer functions to simulate the way the ear naturally processes the location of sound, combining pinna cues and interaural time and intensity cues. In fact, if properly rendered, headphone presented spatialized audio is indistinguishable from free-field presentation [5]. When coupled with a headtracking device, spatialized audio provides a true virtual auditory interface. Using a spatialized auditory display, a variety of sound sources can be presented simultaneously at different directions and distances. Most head-mounted displays are currently outfitted with headphones of sufficient quality to reproduce spatialized audio, making it relatively easy to incorporate spatialized audio in an immersive VR system. A complete lexicon for understanding and developing auditory displays can be found in Reference [6].

4.2 Speaker Systems

However, for imparting realism and emotionality to a VE, speaker systems hold many advantages. It has long been recognized in film making that low frequency sound represents “threat” to the listener [2]. In 1974, a short-lived and kitschy sound format was developed known as “sensuround”. Sensuround gave an emotional boost to such movies as “Midway”, “Battlestar Galactica” and “Earthquake” by adding low frequency tones at 120 dB between 5- 40 Hz. These intense low frequencies add to emotionality, in part, by vibrating the whole body. In 1977, “Star Wars” used a surround format with a separate low frequency channel to achieve something of the same effect. Current theater sound typically incorporates a 5.1 surround sound system, although 10.2 surround sound systems are now being demonstrated. In 5.1, three speakers are located in the front of the theater at left, center, and right positions. Two “surround” speakers are in the rear. The “.1” speaker is a large subwoofer (or separate low-frequency channel) located in the front of the theater. IMAX theaters incorporate a massive 3000 lb, 3200 watt subwoofer into the system.

Lucasfilm is at the forefront of professional sound design. George Lucas was not satisfied with the quality of theater sound, so his engineers developed a theater standard for faithfully reproducing sound in movie theaters. This certification is called “THX” after the George Lucas film “THX 1138” or alternately after his chief sound engineer, “Tom Holman’s X-over” or “Tom Holman’s eXperiment”. THX is not surround sound, but a quality control process. THX has now been expanded to include certification of home theater systems. If properly installed, THX certified sound systems can faithfully reproduce the sound of a movie almost exactly as it was heard during the mixing phase of film production. The acoustical properties of small rooms are vastly different from a large theater, so THX has also developed systems specifically tailored to small listening environments such as home theaters [7]. The lessons learned from these endeavors are directly applicable to the problems we have developing sound in VE.

However, if speaker systems are not properly placed and installed in a room, even the best sound systems will sound inferior. Because of boundary effects and standing waves, improperly placed or installed speakers can reduce speech comprehension, blur localization, reduce the perception of spaciousness, destroy the sense of immersion, and dramatically reduce bass response in a system [7]. This is especially true when dealing with small theaters. The IAC standard home theater room size is 9’ x 16’ x 21’. Since many virtual environments are smaller than home theaters, special care needs to be taken when designing these systems. If the simulation is being presented in a larger room, or demonstrated at an exhibition, care must be taken to design an audio system with enough headroom to handle a larger room space. If the system is installed properly, there will be a uniform (flat) frequency response at the listening area [2].

One example of a four speaker system (two front and two surround) is described in an International Telecommunications Union (ITU) recommendation and places speakers at ± 30º in front of the listener and at ± 110º behind the listener [8]. It is further recommended that the signals emanating from the two surround channels be decorrelated to increase the sense of spaciousness. Correlated mono signals may give a sense of lateralization rather than localization. Lateralization is a perceptual phenomenon where the sound is perceptually inside the head, rather than externalized. If a subwoofer is used, it is usually placed in front of the room. Placing the subwoofer too close to a corner may increase bass response, but may result in a muddier sound. The subwoofer should be moved to achieve the best response in the listening area. Unfortunately, speaker placement will vary depending on the dimensions and shape of the room, as well as the number of speakers employed. If the system is mobile, the sound system will have to be readjusted for every new location, unless the simulation incorporates it~~’~~s own enclosure. If the simulation will be housed in different sized rooms, the audio system (amplifiers and speakers) must have enough headroom (power) to accommodate both large enclosures as well as small. When possible, acoustical tile and diffusers should be employed where appropriate to reduce reverberation and echoes.

In addition, virtual environment systems tend to be relatively noisy because of electronics, HVAC systems, and computer fan noise. Acceptable levels of background noise should be below 30 dB(A). Also, a heavy bass response can cause objects in a room to rattle loudly. If immersion is a concern, care must be taken to reduce these sources of background noise and distraction.

5. Objective Measures for Audio Contributions to Emotion and Presence

Recently, thesis research in our laboratory has focused on devising objective measures for assessing the contribution of professionally produced sound on emotion and presence in VE. Using the argument that emotion is an important, but overlooked, component of immersion/presence, research was recently completed which measured differential physiological responses from sound delivered by headphone systems, speaker systems, and headphone/subwoofer hybrid systems. The results indicate a significant decrease in temperature when sound was presented over speakers as opposed to headphones and headphone/subwoofer hybrids while subjects played a commercial quality videogame [9]. Both these measures are indicative of emotional arousal. All sound presentation methods demonstrated a significant increase in arousal over conditions which were silent.

The thesis also discovered that these measures of arousal do not correlate with traditional questionnaire results examining "presence" or "immersion" in VE. Presence questionnaires do not assess emotionality evoked by simulation. The absence of emotional content in current measures of presence is a bit perplexing. It would be difficult to argue that a simulation aimed at teaching Marines or SWAT teams how to perform room clearance evokes a sense of presence, if it did not also evoke some of the emotion of actually storming a room containing hostile combatants. This strongly suggests to us, that the definition of "presence" needs to be revised to encompass emotional response. This is especially important when we examine our data showing that sound plays a major role in evoking an emotional response in simulation. It is quite possible that current measures of presence underestimate the contribution of audio in evoking a sense of presence.

6. Conclusions

In conclusion, it is apparent that well-designed sound in virtual environments is more than just cosmetic. Properly designed sound is crucial to the sense of presence in VE and essential to injecting emotional context into the simulation. Furthermore, we are making the case that emotionality is an overlooked component to the traditional measures of "presence". The entertainment industry offers the best guidelines to properly implement audio (and emotionality) into simulations, though most of their work has focused on motion pictures, rather than interactive entertainment. The one exception is the use of sound in videogames. The sound capabilities of cutting edge videogame engines are far superior to that of most VE development platforms. To that end, research at NPS is focusing on the development of videogames for use in military applications. The NPS Moves Institute has assembled a team of professional videogame designers, including artists, programmers and level designers and has licensed a commercial game engine and even developed new engines when needed. The results coming from this effort are products which are at the cutting edge of the entertainment industry and far superior in quality to the majority of military projects in both the quality of graphics and sound. To take advantage of this capability, we are applying the sound design techniques discussed in this paper (and many more). Audio research in the next several years will aim at improving audio techniques used in videogames and adapting them to high-end military simulations. Finally, the videogame platform provides an excellent test platform for testing new techniques for simulating accurate room acoustics, auditory cueing, and implementing headphone based spatial audio.

7~~65~~. References

[1] Holman, T. (1997). Sound for Film and Television. Boston, MA: Focal Press.

[2] Holman, T. (2000). 5.1 Surround Sound Up and Running. Boston, MA.: Focal Press.

[3] Yewdall, D. (1999). Practical Art of Motion Picture Sound. Boston, MA: Focal Press.

[4] Letowski, T., Karsh, R., Vause, N., Shilling, R., Ballas, J., Brungert, D. & McKinley, R. (2001). Human Factors Military Lexicon: Auditory Displays. ARL Technical Report, ARL-TR-2526; APG (MD).

[5] Shilling, R.D., Shinn-Cunningham, B. (2002). Virtual Auditory Displays. Virtual Environments Handbook, Kaye Stanney,. New York, Erlbaum.

[6] Langendijk, E. H., and Bronkhorst, A.W. (2000). Fidelity of three-dimensional-sound reproduction using a virtual auditory display. Journal of the Acoustical Society of America 107(1): 528-537.

[7] THX Certified Training Program (2000). Presentation Materials. June 20-23. San Rafael, CA.

[8] International Telecommunications Union. (1994). Multi-channel Stereophonic Sound System With and Without Accompanying Picture. Recommendation ITU-R BS.775-1. Hendrix, C. and Barfield, W. (1996). The sense of presence in auditory virtual environments. Presence 5(3): 290-301.

[9] Sanders, R., and Scorgie, R. (2002). The Effect of Sound Delivery Methods on the User's Sense of Presence in a Virtual Environment. Thesis Research. Naval Postgraduate School, Monterey, CA.

Author Biography

LCDR RUSSELL SHILLING, **Ph.D.,** is a Navy Aerospace Experimental Psychologist and an Associate Professor at the Naval Postgraduate School (NPS) in Monterey, CA. He is the Technical Director for Immersive Technologies for the MOVES Institute and the sound designer for the NPS videogame lab. He is an active researcher in the ONR VIRTE program and a variety of C4I projects. He has been instrumental in certifying simulations in the MOVES Institute as "Dolby Digital" compatible and working with other industry leaders in the audio and entertainment world. He received his B.A. from Wake Forest University in 1985 and his Ph.D. in Experimental Psychology at the University of North Carolina at Greensboro in 1992. He has been studying auditory psychophysics since 1986. Further information can be found at: http://shillingresearch.net .