Audio Engineering Society

Convention Paper

Simulating Hearing Loss in Virtual Training

Ramy Sadek, David Krum, and Mark Bolas

University of Southern California Institute for Creative Technologies, 12015 Waterfront Drive, Playa Vista, CA, 90066, USA

sadek@ict.usc.edu

ABSTRACT

Audio systems for virtual reality and augmented reality training environments commonly focus on high-quality audio reproduction. Yet many trainees may face real-world situations wherein hearing is compromised. In these cases, the hindrance caused by impaired or lost hearing is a significant stressor that may affect performance. Because this phenomenon is hard to simulate without actually causing hearing damage, trainees are largely unpracticed at operating with diminished hearing. To improve the match between training scenarios and real-world situations, this effort aims to add simulated hearing loss or impairment as a training variable. The goal is to affect everything users hear –including non-simulated sounds such as their own and each other's voices – without overt noticeability, risk to hearing, or requiring headphones.

1. INTRODUCTION

Often virtual reality (VR) and augmented reality (AR) training scenarios are used to expose trainees to situations that are too dangerous to simulate physically. In many such scenarios, interpersonal communication and situational awareness are crucial to success or survival (e.g. in emergency medical, police, or military operations). Training to function successfully with inhibited communication and diminished situational awareness may help trainees better cope with the effects of hearing impairment in the field. To accomplish this goal, the audio system must simulate the characteristics of hearing damage on real-world sounds such as trainee speech.

1.1. Relationship to prior work

Techniques for modeling hearing loss in a controlled environment are well known. Prior research in this area is extensive. Our work differs from the prior research in several ways. First, much existing work focuses on replicating the effects of cochlear hearing damage (via filtering an input signal) in a controlled environment (e.g. via headphones). Our aim is to mimic the effects of hearing damage *acoustically* rather than electrically, without headphones or other worn equipment. In other words we are trying to achieve the quantitative effects described in previous research (e.g. [2]), but in an uncontrolled setting by polluting the environment with

noise, rather than by manipulating controlled input signals.

2.   METHOD

Our method relies on the principle of auditory masking. By adding noise (sound that contains no information) into the training environment, a masking system can obfuscate or "block" content sounds (speech and environmental audio cues). An advantage of adding noise to the environment to simulate hearing loss is that it enables dynamic degradation of hearing without requiring trainees to wear additional equipment.

Furthermore, the donning of special devices that affect a trainee's hearing, such as earplugs, may tip off the trainee to the content of the training scenario and the training purpose, which may be undesirable. Finally, in scenario training for dangerous field assignments (e.g. military), trainees and trainers alike prefer to avoid any equipment that will not be used in the field to avoid the risk of negative training. Our system inhibits the real-world communication of participants by hiding noise signals within virtual world sound effects in order to disrupt speech intelligibility. For example, adding noise to a sound effect such as a helicopter affords significant disruption in the speech band without increasing the sound pressure to dangerous levels or changing the perceived nature of the sound effect.

2.1.   Background: Principles of Auditory Masking

Auditory masking refers to the psychoacoustic phenomenon wherein a sound signal is rendered inaudible by the presence of another sound (the masker). More specifically, a given sound has a threshold of audibility in a quiet environment below which it remains unheard; the presence of a masker raises the threshold of audibility so that the sound at its original volume cannot be heard (See Figure 1). The amount of masking is the difference between the audibility threshold in the presence of a masker and the threshold in quiet. The level of masking varies per combination of sound, masker, and listener [1] [3]. Broadband noise is the most common class of masking signal.

Masking systems are commonly used in workplaces to minimize worker distractions, in medical offices to ensure patient privacy, and in government facilities to ensure security. In each of these cases, the masker is

filtered broadband noise. The key to masking in these environments is that the masker calls no attention to itself. Broadband noise is appropriate because it contains no information, and is therefore easily ignored or "tuned out" [1].
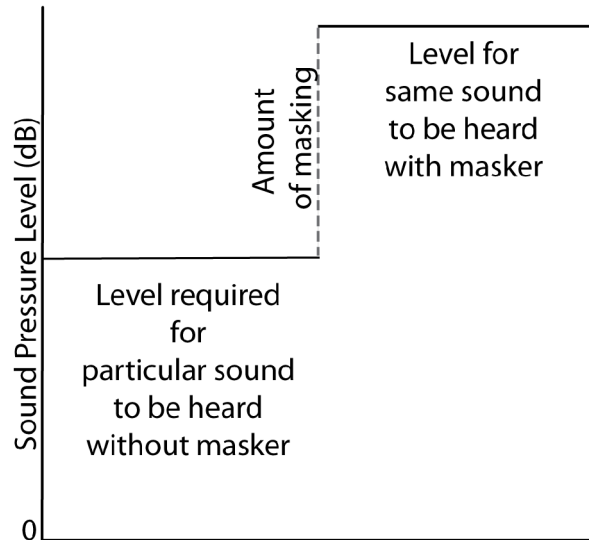

Figure 1 The presence of a masker decreases the audibility of a masked sound.

Inconspicuous maskers are also germane to VR and AR settings, since the presence of an obvious noise signal (e.g. audible rumble or hiss) could undermine the realism and believability of the scenario. The focus of this research effort is to *hide* masking signals so that they remain undetected by trainees.

2.2.   Approach

There are two places to hide a masking signal: in the sounds of the virtual environment and in the baseline noise of the physical training space. Commercial masking systems use the latter by filtering broadband noise to match the spectrum of baseline noise in the environment. When the masker is diffuse (i.e. not localizable) and spectrally similar to background noise, people are most likely to accept the masker as background and ignore it [1]. Although good spectral matching allows the masker to reach a loudness level significantly higher than the baseline noise before detection, beyond certain levels the noise calls attention to itself. To increase the amount of masking available in training, we hide additional maskers in the virtual world sound effects.

Audio from virtual world scenes is generally much louder than the baseline noise of the training space. As such, virtual audio offers additional room to hide maskers. There are two challenges with this approach. First, the time-varying spectra of most sound effects prevent the use of fixed-spectrum broadband noise since the masker is readily detectable when its spectral envelope does not match that of the environmental sounds. Second, quiet periods in the virtual world do not support masker hiding.

2.2.1. Time-Varying Spectra

Because virtual world sounds can be quite loud (e.g. a helicopter, truck motors, etc.) they offer plenty of energy with which to hide masking signals. In this sense they mask the masker, but for clarity we refer to this effect as *hiding*; that is, the sound effect *hides* the masker.

The time-varying spectral envelopes of such sound effects poorly match the fixed spectrum of filtered broadband noise. This mismatch prevents the sound effects from hiding the masking signal, calling attention to the masker itself. To remain hidden, the masker must share the time-varying spectral envelopes of the virtual world sound effects. We must shape broadband noise to match the spectra of virtual world audio. To do so, we leverage a technique common in computer music called *cross-synthesis* [5].

Cross-synthesis imposes the spectral envelope of one signal (the modulator) onto another (the carrier). This approach is the basis for several different synthesis schemes (e.g. different types of vocoders), which employ varying schemes for spectral envelope estimation, modulation, and synthesis. We experimented with several different schemes. Surprisingly, we found that sophisticated schemes such as phase vocoding, worked poorly: the modulated noise signal was easily detectable and its timbre was unacceptable, failing to blend with the virtual scene. Instead, the simple approach of replacing the frequency component of the carrier signal with that of the modulator, while preserving the carrier's phase, produced acceptable results.

We further avoid time correlation problems by choosing a relatively long FFT window of 8192 samples to capture adequate low-frequency information. For this window length, we found that a half-window overlap-add was sufficient to avoid audible boundary artifacts. In practice, we were able to bring the noise level within

-3dB of the main output signals before the noise was unacceptably noticeable. This rule of thumb held across many total SPL levels. The question of maximum undetectable noise level is an area for future investigation for us.
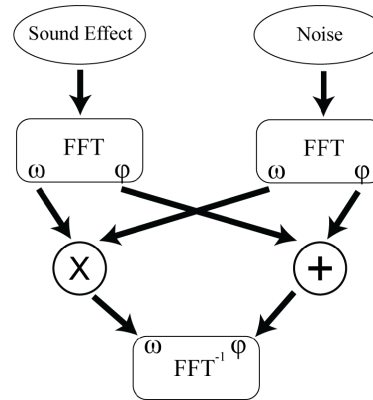

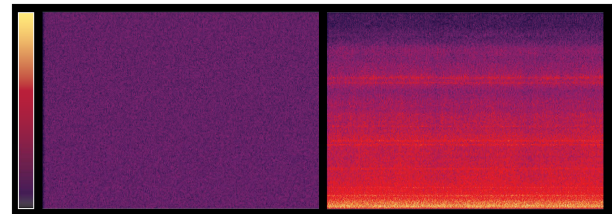Figure 2   Cross-synthesis signal flow

Figure 3 Two spectrograms. Left: pure white noise; Right, noise contoured by cross synthesis.

During quiet moments in the virtual world, there may be insufficient sound energy with which to hide maskers, leaving only the baseline environmental masker. We are currently investigating several approaches to overcome this problem.

3.   FUTURE WORK

One solution we are exploring is a semi-automated sound design algorithm, which takes as input an automatically annotated library of user-specified sound effects appropriate to a virtual scene. This solution has two parts, an annotation module and a sound design module. The annotation module parses the user-specified sound effects, and analyzes their time-frequency content to determine spectral envelope. The sound design algorithm will detect times when the level

of supported masking is too low, and select sound effects to play in order to allow for additional maskers. For example, during a lull, the algorithm may select several outdoor sound effects with broadband spectra such as wind, road noise, distant aircraft, crickets, or a babbling brook. The combined spectra of these sounds offer significant room for additional maskers.

Quiet moments may also be good times to add tinnitus simulation, imitating the high-pitched ringing sound commonly perceived after exposure to very loud sounds. We plan to introduce high-pitched tones into the environment to simulate this effect.

We intend to focus on quantifying and measuring the efficacy of our approach by comparing against known standards (e.g. Speech Transmission Index and Speech Intelligibility Index). Once we have quantitative measurement in place, we intend to automate level control of injected noise to minimize intelligibility of live speech. Finally, we hope to answer the question of how to compute the threshold level of inconspicuous noise.

4. CONCLUSION

We have made several preliminary findings that we believe are crucial to further work. First, matching the ambient noise spectrum in the manner similar to maskers used in office environments [1] is the first step to hiding noise. This approach allows for a modest baseline level of masking.

In order to allow additional masking energy by hiding maskers in virtual sound effects, the masker must share the same spectral envelope as the sound effect. This criterion requires time-varying filtering which we achieved via cross-synthesis. Matching time variance appears to be more important than smooth transitions in these hidden maskers.

Finally, the noise signal must be diffuse, otherwise it is detectable as an isolated sound. Using multichannel audio systems (e.g. 5.1 and 7.1), we implemented separate noise generators per channel to avoid correlation and phase effects, thus also preventing detection via localization.

By adhering to these criteria, we developed a system that successfully hides noise in virtual environment sounds, greatly suppressing audibility of live speech.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] R. Chanaud, *Sound Masking Done Right*, Magnum Publishing, 2008

[2] Baer, T. and Moore, B. C. J. Effects of spectral smearing on the intelligibility of sentences in noise. The Journal of the Acoustical Society of America. 94, 3, 1229-1241.

[3] E. Zwicker and H. Fastl, *Psychoacoustics*, Springer-Verlag, second edition, 1999.

[4] Gelfand, S. A. (2004). *Hearing : an introduction to psychological and physiological acoustics*. New York, Marcel Dekker.

[5] J.O. Smith, *Spectral Audio Signal Processing*, October 2008 Draft, http://ccrma.stanford.edu/~jos/sasp/, online book accessed 2009