Using Wavelets to Synthesize Stochastic-based Sounds for Immersive Virtual Environments

Nadine E. Miner
Sandia National Laboratories
P.O. Box 5800, MS 1008
Albuquerque, NM 87185-1008
+1 505 845-9717
neminer@isrc.sandia.gov

Thomas P. Caudell
Dept. of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131-1356
+1 505 277-5637
tpc@eece.unm.edu

ABSTRACT

Stochastic, or non-pitched, sounds fill our real world environment. Humans almost continuously hear stochastic sounds such as wind, rain, motor sounds, and different types of impact sounds. Because of their prevalence in real-world environments, it is important to include these types of sounds for realistic virtual environment simulations. This paper describes a synthesis approach that uses wavelets for modeling stochastic-based sounds. Parameterizations of the wavelet models yield a variety of related sounds from a small set of models. The result is dynamic sound models that can change according to changes in the virtual environment. This paper contains a description of the sound synthesis process, several developed models, and the on-going perceptual experiments for validating the sound synthesis veracity. The developed models and results demonstrate proof of the concept and illustrate the potential of this approach.

Keywords

Sound synthesis, wavelets, virtual reality, immersive environments, audio perception.

INTRODUCTION

This paper describes a sound synthesis approach to modeling stochastic-based environmental sounds for immersive environments. Stochastic sounds can be divided into two basic classes: continuous sounds, such as motors, fans, wind, rain, scraping and sliding sounds, and impulsive sounds, such as a door knock, gun firing, glass breaking and cigarette lighter. All of these sounds are primarily non-pitched and consist mainly of stochastic noise. Because of their prevalence in the real world, including stochastic sounds in a virtual experience is important for obtaining realistic virtual environments.

Most current virtual reality (VR) systems utilize pre-digitized sounds rather than synthesized sounds. Pre-digitized sounds are static and do not change in response to user actions or to changes within a virtual environment.

Often times, obtaining an application specific sound sequence is difficult and requires sophisticated sound editing hardware and software [6]. Creating an acoustically rich virtual environment requires thousands of sounds and variations of those sounds. Obtaining this very large digitized sound library is impractical. The alternative to using pre-digitized sound is to use sound synthesis. Although sound synthesis may be preferred, there are essentially no virtual sound systems available today which provide flexible, real-time sound synthesis tools for the virtual world builder. The approach described in this paper is a first step towards filling this void.

The use of pre-digitized sound provides high fidelity, static sounds. In contrast, the sound synthesis approach can yield perceptually convincing sounds and provide flexibility through model parameterization. By manipulating the model parameters, a variety of related, but perceptually different sounds are generated. These sounds group into what are called "clusters" of perceptually related sounds. Ballas investigated perceptual clustering of everyday sounds in [1]. He defined perceptual clusters as sounds that are often confused with each other. In our research, we investigate the potential for obtaining clusters of perceptually convincing sounds by adjusting the sound model parameters.

The synthesis method described in this paper uses wavelets for modeling stochastic-based sounds. Parameterization of the wavelet models yields a variety of related sounds from a small set of models. The result is dynamic sound models that allow the sound synthesis to change according to changes in the virtual environment. We describe the on-going perceptual experiments used to validate the veracity (perceptual accuracy or precision) of the sound synthesis. Preliminary experimental results indicate that the synthesized sounds are perceptually convincing to human listeners. Finally, we describe several

different parameterized stochastic sound models developed to demonstrate the functionality and potential of this sound synthesis approach.

RELATED WORK

Some related work in synthesizing real-world sounds using dynamic, parameterized models exists in the literature. Gaver, in developing a sound interface for human-computer interaction, proposed some physical-like models for real-world sounds [4]. Gaver implemented parameterized models for impact, scrapping, breaking and bouncing sounds. The synthesis algorithms succeeded in creating parameterized sounds in real-time, however, the results were somewhat "cartoon-like" and required training to interpret. Doel and Pai proposed a general framework for producing impact sounds in [13]. Their approach uses physical modeling of the vibration dynamics of physical bodies. The models were parameterized based on the material and object shape, and the collision force and location. Prototype sound simulations also produced somewhat cartoon-like impact sounds. Smith used a “digital waveguide” method for developing physical models of string, wind and brass instruments [12]. This method yields excellent quality music synthesis and some high-end synthesizer keyboards are based on this technology.

One main difference between these methods and the one presented here is the emphasis on the importance of modeling the stochastic sound components. Serra shows that incorporating stochastic components in a sound model result in sound simulations with more realism [10]. The method presented here is likely to be successful in synthesizing pitched sounds as well as stochastic-based sounds.

OBJECTIVES

The overall goal of this research is to develop methods for synthesizing perceptually compelling sounds for use in immersive environments. The aim is to provide an approach for creating flexible sound models that yield a variety of sounds and increase the overall richness and realism of an immersive experience.

One specific goal of this research is to create perceptually convincing sounds rather than physically precise ones. The first motivation for this approach is to obtain models that produce many perceptually different sounds rather than producing only one sound very accurately. This idea stems from the potential for cross-identification between perceptually close sounds. For example, the sound of a drawer closing may be identified as a hammer striking a block of wood, and the sound of water dripping may be identified as a clock ticking. With the visual context provided by a virtual environment, cross-identification of sounds may be more prevalent. The second motivation is to reduce the computational complexity required for the synthesis. Creating physically accurate simulations of complex sounds is compute intensive. It is anticipated that synthesizing perceptually convincing sounds will be less compute intensive because evaluation of complex physics equations are not required.

Parameterization of the sound model is a major strength of the approach presented here. There are two reasons for choosing this approach. First, parameterization provides the possibility of obtaining a variety of sounds from a single model. For example, one parameterized rain model might generate the sound of light rain, medium rain, heavy rain and even a waterfall sound for example. The second reason is to create dynamic sound models. Manipulating the sound model parameters in real-time can yield a dynamically changing sound. With the rain model example, changing the parameters as the virtual simulation evolves allows the rain sound to progressively increase in intensity as the graphics simulation shows increasing and darkening clouds. Overall, model parameterization provides flexibility such that a variety of dynamic sounds result from a small model set.

RESEARCH APPROACH

The sound synthesis method described here uses wavelets for modeling stochastic-based sounds [7]. Wavelets were chosen over the more standard Fourier transform because Fourier methods do not adequately model the time varying nature of real-world signals. Windowed Fourier transforms capture the frequency information for different sections of time, but the resolution is limited and fixed by the choice of window size. Wavelet analysis provides a time-based windowing technique with variable-sized windows. Wavelets examine the high-frequency content of a signal with a narrow time-window and the low-frequency content with a wide time-window. Fast wavelet algorithms provide the potential for synthesizing wavelet-modeled sounds in real-time. The fast wavelet algorithms are comparable in terms of compute time to the Fast Fourier Transform algorithms according to Ogden [9].

Development of a wavelet sound model is accomplished through a four-phase process: analysis phase, parameterization phase, synthesis phase and validation phase as shown in Figure 1.

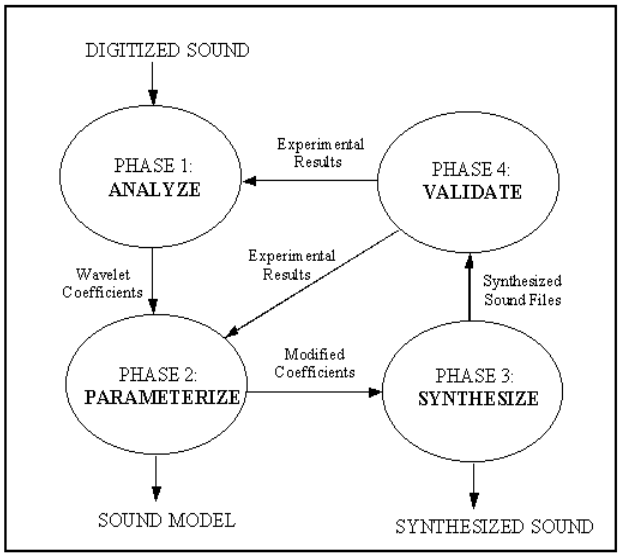

Figure 1: Four-phase Synthesis Model Development Process

Phase 1: Analysis Phase

The analysis phase begins with a digitized sound sample. Detailed examination of the digitized sound determines the best wavelet shape for the signal decomposition. For the parameterized models presented here, wavelet function (ψ) and corresponding scaling function (ϕ) selection was from the standard Daubechies family of wavelets [3]. The original digitized sound is decomposed using the discrete wavelet transform (DWT) which employs a set of filtering and decimation (or down sampling) operations to obtain two sets of coefficients which completely describe the original sound. Refer to [2], [3] or [9] for a description of the wavelet decomposition process. Using the wavelet decomposition coefficients, a model of the original digitized sound is constructed. The original sound is exactly reconstructable from the model. Thus, as with the FFT, no information is lost through the wavelet decomposition process.

Phase 2: Parameterization Phase

The second phase of the process is parameterization. The wavelet decomposition coefficients are the source of the parameters for the sound synthesis model. Manipulating the model parameters yields a variety of sounds related to the original digitized sound. Essentially unlimited control in amplitude, time and frequency are available; however, the parameters are not directly related to the physical characteristics of the sound source, as is the case with other parametric approaches ([13] for example). Determining the sound model parameterization is largely an iterative process. For example, increasing the low-frequency content of a model results in the perception of a larger sound source having generated the sound. By manipulating the low frequency and high frequency parameters of an engine model turns the sound of a standard sized car engine into the sound of a large truck or a small toy car respectively. Scaling function parameter manipulations can shift the sound in frequency. Manipulations of this type can change the sound of a brook to the sound of a large, slow moving river, or to the sound of a rapidly moving stream. More sophisticated parameter manipulations create perceptually convincing sounds that are beyond the scope of the original sound. For example, manipulating the parameters of a rainstorm can result in the sound of applause or a machine room. Model parameter manipulation translates into a new set of wavelet coefficients.

Manipulation of the sound model parameters can be represented in a perceptual sound space. Figure 2 depicts an idealized example of a synthesized sound space. The center of each fuzzy circle represents the original digitized sound from which the model was developed. Parameter manipulation extends the sound perception into many dimensions. It is feasible to move from one type of sound source to another by changing the parameter settings as indicated in Figure 2 by the overlapping sound spheres. For example, manipulating the rain model parameters creates a sound that includes the sound of light rain, medium rain, a heavy, rapid rainfall, a small waterfall, and some motor sounds.

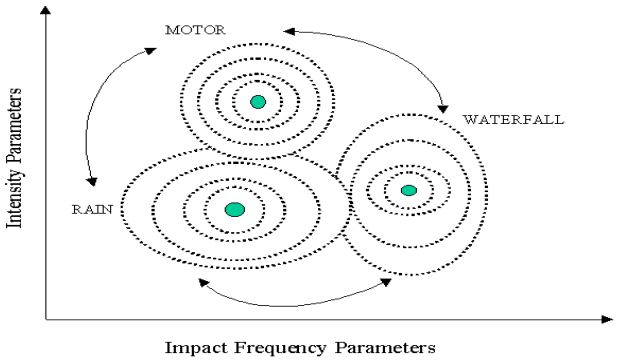

Figure 2: Example Perceptual Sound Space with Parameter Manipulations

Phase 3: Synthesis Phase

The synthesis phase employs the Inverse Discrete Wavelet Transform (IDWT). The modified parameter coefficients are the inputs to the IDWT. The IDWT consists of up-sampling the modified coefficients (inserting zeroes) and filtering using appropriate reconstruction filters. The scale function and wavelet type used during decomposition determines the choice of reconstruction filters. The set of decomposition filters together with the associated reconstruction filters form a system of quadrature mirror filters. Refer to [2], [3] or [9] for a detailed description of the wavelet reconstruction process. The output of this phase is a synthesized sound for use in VR applications and validation experiments.

Phase 4: Validation Phase

Validation is the final phase of the sound synthesis process. Because the goal is to create perceptually convincing sounds, performing a rigorous mathematical proof is not feasible to validate the success of the sound synthesis. Instead, three psychoacoustic experiments are planned to validate the sound synthesis veracity.

The first experiment illuminates the sound space and possible sound clustering. The self-similarity technique from psychophysics is used to accomplish this. Self-similarity experiments are used to understand the interrelationships among the important concepts in a knowledge area [5]. In this experiment, subjects rate the similarity between two synthesized sounds on a 5-point rating scale. Every possible combination of sound pairs is presented in random order. The similarity rating data is analyzed with two different methods. The first method derives a graph representing the rating data using the Pathfinder scaling algorithm [5]. The second method uses multidimensional scaling (MDS) which results in a mapping of the synthesized sounds onto a multidimensional perceptual space [11]. Examination of these analysis results provides a better understanding of the perceptual sound clustering occurring through parameter manipulation.

The second experiment examines the perceptual identification of the synthesized sounds. Subjects listen to synthesized sounds and enter an identification description. Identification phrases include a noun and descriptive adjectives. Subjects are asked to think of the sound source when formulating the descriptions. There is no time limit and subjects are permitted to replay the sounds. Response times are measured so that uncertainty values can be calculated. This is a free form identification experiment similar to that run by Ballas [1] and Mynatt [8].

The third experiment measures the perceptual sound veracity. Phrases obtained from the second experiment are paired with synthesized sounds. The phrases provide a perceptual context for the sounds. Subjects are asked to rate how well the phrases match the sounds they hear. Ratings are on a 5-point scale, with 1 = no match and 5 = perfect match. Both digitized and synthesized sounds are included in the experiment. Examining the digitized sound ratings provides a standard to which the synthesized sound ratings can be compared. In this way, evaluation of sound veracity within a verbal context is obtained.

These experiments are on going. Examination of perceptual experiment results indicates whether design iteration is necessary. Iteration of the process refines the synthesis model to obtain the desired perceptual characteristics. Reanalysis of the model involves iterating through the process starting either with phase 1 (and a new wavelet analysis) or phase 2. The result is a parameterized sound model capable of producing a variety of perceptually convincing sounds.

SOUND SYNTHESIS MODELS

Several parameterized sound models have been developed using this four-phase process to demonstrate the validity of the approach. Many more models and parameter manipulations are possible for both stochastic and pitched sounds. Below is a description of the continuous and impulse-based stochastic sound models developed.

Continuous stochastic models

Continuous stochastic sound models consist primarily of non-pitched sound and do not have a finite duration. To demonstrate the potential of the wavelet synthesis approach, four continuous stochastic models have been developed and are described below.

Rain

This model is parameterized to simulate light rain, medium rain and progressing to heavy rain. The perception of increasing wind accompanies the sound of increasing rain to convey the sense of a large rainstorm. Other perceptually grouped sounds that might emerge from this model are bacon-frying, machine room sounds, a waterfall, a large fire, and applause.

Car Engine

This model simulates the sound of a car engine idling with parameter adjustments for different sized cars, different type of engines and different RPMS. Adjusting one set of parameters results in synthesis of a large diesel truck, a standard truck, a mid-sized car, a small car and a toy car. Work is in progress to parameterize the model for increasing the RPMS for the same sized car engine. Further parameterizations may include engine load, or different engine types (e.g. lawn mower or blender).

Electric Motor

This model simulates the sound of different sizes and RPMs for electric motors used in small, hand held equipment such as drills, electric screwdrivers, vacuum cleaners, electric yard equipment. Different perceptually grouped sounds include the natural sounds of a bee buzzing and a small bird's wings flapping. Other sounds identified during perceptual experiments for this sound model are electric razor, static on a television, and a welding machine.

Brook

This model simulates the sound of a babbling brook with parameter adjustments for stream activity level (calm to raging). Additional parameter adjustments vary the stream size from a very wide stream to a narrow stream. With these controls, the sound of a babbling brook is converted into the sound of a wide, calm, deep river and further converted into the sound of a waterfall. Other parameter adjustments yield the perception of a heavy rainstorm, water from faucet, water running into a bathtub, and a printing press as evidenced by preliminary perceptual experiments.

Other models under development include wind sounds, fan sounds, and ocean sounds.

Impulse-based stochastic models

Impulse-based stochastic sound models are non-pitched and have a finite duration. Often distinct impact sounds are evident. Three example models of this type which have been developed are described below.

Footsteps

This model simulates the sound of footsteps on gravel. Parameter manipulations result in the perception that the footsteps are on different material types such as dirt, leaves, a hard concrete floor or a wood floor. Further parameter adjustments yield the perception of different weights of the person walking. Perceptually grouped sounds identified during preliminary perceptual experiments are chewing, crumbling paper, crushing a can, stomping of horse hooves, lighting a gas grill, and gunfire.

Glass Breaking

This model simulates the sound of breaking glass with parameter adjustments for the glass thickness (or density), the surface hardness on which the glass is breaking, and the force with which the glass is thrown. Exercising this sound synthesis model can result in the sound of dropping a thick glass on a wood floor or a throwing a fine piece of crystal against a concrete floor.

Shuffling a Deck of Cards

This model simulates the sound of a deck of cards being shuffled. Perceptually grouped sounds identified during preliminary perceptual experiments include: wind hitting a loose object, breaking of spaghetti noodles, wings flapping, paper burning, and a motorcycle starting up.

Other models under development include various explosions and impact sounds.

FUTURE EXTENSIONS

One future extension will be to merge several different models into one generalized model. For example, merging the electric and car motor models may yield a general *motor model*. This is desirable because end users would have a variety of engine sounds, engine loads, RPMs, etc. all from one model. Another example would be a general *running water model* that could provide synthesis of rain, brook, rivers, waterfalls, water from faucets, and more.

Real-time sound synthesis is possible for the approach described within this paper. Completing the analysis and parameterization phases in non-real time produces the parameterized model. The parameter manipulation and synthesis phases can be computed in real-time in parallel with graphical and environmental VR simulations. Real-time implementations of wavelet transforms are becoming available on today's digital signal processing platforms and on Intel's MMX platform. A dedicated sound server may be necessary to perform the inverse wavelet transform and parameter manipulations in real-time. This server could also perform the 3-D sound localization. To further enhance the real-time performance, it may be possible to compress wavelet coefficients thereby reducing the number of coefficients and synthesis time without compromising the perceptual sound synthesis quality.

Implementing a real-time sound synthesis system makes integration into a VR system possible. A cost-effective approach for integrating into a VR system is to choose a combination of pre-digitized sound (for static sound sources) and sound synthesis (for dynamic sound sources). Changes to the dynamic sound environment within the VR simulation would result in parameter setting updates on the sound server. The sound server would update the sound synthesis and localize the sound in the environment. The interface to the sound server must be useable so that developers need not become sound experts in order to utilize sound in their virtual worlds.

CONCLUSIONS

We have described a four-phase development process for a new stochastic sound synthesis approach. The iterative nature of the process allows continuous model refinement according to perceptual sound quality results. The analysis and synthesis phases utilize the discrete wavelet transform and the inverse discrete wavelet transform respectively. The parameterization phase creates dynamic, flexible sound models which, when exercised, are capable of producing sounds with a variety of perceptual qualities. We describe the on-going perceptual validation experiments designed to elucidate the sound clustering and rate the sound synthesis veracity. Several different continuous and non-continuous stochastic-based sound models have been developed using this method including models for: rain, car engine, electric motor, brook, glass breaking, shuffling cards and footstep sounds. These models provide evidence of the validity and potential of this approach. Several steps are required before these sound synthesis models are available to end users, including further model development, real-time implementation, development of an intuitive user interface, and integration with virtual reality simulation systems.

ACKNOWLEDGMENTS

This work was supported by Sandia National Laboratories under their Doctoral Study Program Fellowship. We thank the reviewers who provided helpful comments. We also thank the experiment volunteers whose feedback has improved the quality of the synthesis models.

REFERENCES

1. Ballas, J. Common Factors in the Identification of an Assortment of Brief Everyday Sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 1993, Vol. 19, No. 2, 250-267.
2. Cohen, A. and Ryan, R. D. *Wavelets and Multiscale Signal Processing*. Chapman & Hall, London, 1995.
3. Daubechies, I. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.
4. Gaver, W. Using and Creating Auditory Icons. In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, edited by G. Kramer, Santa Fe Institute Studies in the Science of Complexity, Proc. Vol. XVIII. Addison-Wesley, 1994, 417-446.
5. Goldsmith T. E., Johnson, P. J. and Acton, W. H., Assessing Structural Knowledge. *Journal of Educational Psychology*, 1991, Vol. 83, No. 1, 88-96.
6. Miner, N. Using Voice Input and Audio Feedback to Enhance the Reality of a Virtual Experience. *Proceedings of the 1994 IMAGE Conference*, Tucson, AZ, June 1994.
7. Miner, N. A Wavelet Approach to Synthesizing Perceptually Convincing Sounds for Virtual Environments and Multi-Media. PhD dissertation. University of New Mexico. To be completed Spring 98.
8. Mynatt, E. D. Designing with Auditory Icons. *Proceedings of the Second International Conference on Auditory Display (ICAD)*. Nov. 1994, 109-119.
9. Ogden, R. *Essential Wavelets for Statistical Applications & Data Analysis.* Birkhauser, 1997.
10. Serra, X. A System for Sound Analysis/ Transformation/Synthesis based on a Deterministic Plus Stochastic Decomposition. Ph.D. Dissertation, Report No. STAN-M-58. CCRMA and Department of Music, Stanford, University, October 1989.
11. Shepard, R. N. Multidimensional Scaling, Tree-Fitting, and Clustering. *Science*, Vol. 210, 10/24/80, 390-398.
12. Smith, J. O. Physical Modeling Using Digital Waveguides. *Computer Music Journal*, MIT Press, Vol. 16, No. 4, Winter 1992.

13. Van Den Doel, K. and Pai, D. K. Synthesis of Shape Dependent Sounds with Physical Modeling. *Proceedings of the International Conference on Auditory Displays*, Santa Clara, CA, Nov. 1997