Nadine E. Miner
Sandia National Laboratories
Albuquerque, NM

Timothy E. Goldsmith
Department of Psychology
University of New Mexico

Thomas P. Caudell
Department of EECE
University of New Mexico

Perceptual Validation Experiments for Evaluating the Quality of Wavelet-Synthesized Sounds

Abstract

This paper describes three psychoacoustic experiments that evaluated the perceptual quality of sounds generated from a new wavelet-based synthesis technique. The synthesis technique provides a method for modeling and synthesizing perceptually compelling sound. The experiments define a methodology for evaluating the effectiveness of any synthesized sound. An identification task and a context-based rating task evaluated the perceptual quality of individual sounds. These experiments confirmed that the wavelet technique synthesizes a wide variety of compelling sounds from a small model set. The third experiment obtained sound similarity ratings. Psychological scaling methods were applied to the similarity ratings to generate both spatial and network models of the perceptual relations among the synthesized sounds. These analysis techniques helped to refine and extend the sound models. Overall, the studies provided a framework to validate synthesized sounds for a variety of applications including virtual reality and data sonification systems.

1 Introduction

The perceptual evaluation of sounds, whether they be synthesized or digitized, is important for the successful implementation of sound in a virtual environment or data sonification system. Often, sounds are chosen in an ad hoc manner and are integrated into a project without conducting human perceptual studies to measure their effectiveness. Furthermore, when a sound synthesis technique is developed, formal validation experiments are seldom employed to measure the success of the technique. As a result, the sounds' contribution to the quality of the virtual environment or data sonification system is often unknown.

The purpose of this paper is to present a general method for validating the psychological reality of artificially generated sounds. Currently, standardized methods for evaluating sound stimuli do not exist. The proposed methods are presented as a starting point for systematically evaluating the perceptual properties of acoustic signals. The experiments and analysis techniques are described in enough detail to allow replication by those with little or no formal training in perceptual experimental methods. We illustrate the techniques by applying them to evaluate the success of a new sound synthesis method that uses wavelet analysis to create parameterized models capable of synthesizing a wide range of real-world sounds. Details of the synthesis method

Presence, *Vol. 11, No. 5, October 2002,* 508–524

are presented in a partner paper (Miner & Caudell, 2002). A more thorough battery of validation studies is described by Bonebright, Miner, Goldsmith, and Caudell (1998).

This paper describes the procedures, analysis techniques, and results from three psychoacoustic experiments: similarity rating, identification task, and context-based rating. Experimental results indicated that the wavelet-synthesized sounds created a variety of perceptually convincing aural images. Details and results for the three experiments follow.

1.1 Related Work

Gaver (1994), in developing a sound interface for human–computer interaction, proposed some physical-like models for synthesizing real-world sounds. A physical model captures an object's physical attributes (such as shape, material type, and weight) in the sound synthesis model. Using their synthesized sounds as stimuli, Gaver, Smith, and O'Shea (1991) performed psychoacoustic experiments to examine whether audio signals provided useful information about user-initiated events and processes in a virtual factory simulation. Subjects interacted with a factory simulation both with and without auditory feedback. Results showed that auditory feedback aided participants in monitoring the health of the factory and diagnosing problems. These experiments were the first examples of validating the use of synthesized sounds in a virtual environment with psychoacoustic methods.

Ballas (1993) investigated perceptual clustering of everyday sounds. Listeners rated sounds on various perceptual and cognitive scales, including aural properties (clarity, loudness, and timbre), identifiability, familiarity, existence of mental stereotypes, and number of similar sounds. Principal components analysis of the ratings identified three factors that were subsequently used in cluster analyses. Results suggested that a cognitive representation of sounds could be derived with perceptual, rather than physical, dimensions of the sounds. Using this result, the experiments presented in this paper attempt to identify the perceptual dimensions of synthesized event sounds.

An important issue for virtual environments is the role that context plays in identifying sounds. Ballas and Mullin (1991) showed that the identification of everyday sounds was influenced by contextual information. It is feasible that listeners can identify even imperfectly synthesized sounds given the right context. This could result in cost-effective sound synthesis and storage methods. The third experiment described in this paper attempts to quantify the success of sound synthesis given contextual information.

One general method that has proved useful in understanding people's perceptions of objects is psychological scaling. Scaling methods operate on a set of similarity relations to reduce noise in the data and uncover any inherit structure among the objects. Similarity relations can be collected in different ways, but they are commonly obtained by having people judge the similarity of pairs of objects on a rating scale. Scaling methods represent these data by formal models such as a multidimensional space or a network (graph). These representations can reveal perceptual or psychological dimensions of stimuli that may not otherwise be obvious.

One particular scaling method, multidimensional scaling (MDS), has become popular for investigating the underlying perceptual structure of stimuli (Schiffman, Reynolds, & Young, 1981). MDS represents each object as a point in multidimensional space. Objects are located such that their pairwise distances approximate their corresponding perceived similarities. Similar objects are close to one another, and dissimilar objects are farther apart. The goal is to represent the objects in a low dimensional space, but at the same time to maintain a good fit between the pairwise distances and perceived similarities. MDS helps to reveal psychologically meaningful dimensions for a set of objects. Generally, dimensions represent both qualitative and quantitative attributes of the stimuli. Because MDS does not require a priori knowledge of the perceptual characteristics of stimuli, it eliminates experimenter biases. Many perceptual areas have been explored using MDS including analysis of color (Shepard & Cooper, 1992), surface textures of objects (Hollins, Faldowski, Rao, & Young, 1993), taste

and smell (Schiffman et al., 1981), facial expressions (Green & Cliff, 1975), and vocal affect expression (Bonebright, 1996). Other investigators have used MDS to analyze various perceptual characteristics of auditory stimuli (Flowers & Hauer, 1993, 1995; Fox, 1985; Green & Cliff, 1975). Thus, MDS is an accepted psychological and psychoacoustic analysis procedure for exploring perceptual stimulus dimensions.

Another common scaling method is Pathfinder (Schvaneveldt, 1990), which represents each object as a node in a network with links interconnecting nodes. Related objects are nearer to each other in the network, as measured by the number of links in the shortest path distance. Pathfinder's goal is to represent the objects in a network with a small number of links while maintaining a good fit between the path distances and perceived similarities. The algorithm starts with a fully connected network with each link assigned a weight based on the similarity data matrix. Next, each link is examined to decide if it should be removed. If the link weight between two nodes is greater or equal to the shortest path distance between the nodes, the direct link is removed. Pathfinder differs from MDS both in the nature of the representation that is produced (network for Pathfinder versus space for MDS) and the method by which the ratings are analyzed. However, both operate on the same similarity data and have the common goal of minimizing the effects of noise inherent in these types of data. There is some evidence that Pathfinder might be better at representing higher-level conceptual relations, whereas MDS is better at representing perceptual relations (Goldsmith, Johnson, & Acton, 1991). In the present paper, we used both methods to analyze the similarity ratings of synthesized sounds.

The next section provides background information on the sound synthesis method used to generate the experimental sound stimuli.

1.2 Wavelet Synthesis Method

The sounds examined in this paper were generated by a new wavelet-based synthesis method. The first step in the approach is to develop sound models through a wavelet analysis process. The sound models are parameterized, meaning that the model can synthesize many different sounds depending on the parameter values. Wavelet analysis provides an effective method for extracting model parameters. The models are designed offline, prior to running a virtual reality simulation. Parameter modification and sound synthesis can be accomplished in real time during a virtual environment simulation. Thus, the sound synthesis is dynamic and provides variations in the sound environment according to changes in the virtual environment. Additionally, the sound models can yield perceptually convincing sounds for a variety of environments. Our results indicated that the synthesized sounds are perceptually convincing to human listeners.

The experimental sound stimuli were created by seven models developed with this wavelet-based synthesis method: rain, brook, 2000 RPM motor, electric motor, footstep, shuffling cards, and glass breaking. Each sound model started from a digitized sound sample to serve as the base sound. The base sounds were digitized at a 22,050 Hz sample rate with 16-bit resolution. The sounds were captured using a portable digital audio tape (DAT) recorder and a studio quality microphone. The base sound durations were 0.5 seconds (footstep), 1.2 seconds (shuffling), 2.0 seconds (glass breaking) and 1.0 seconds (rain, brook, and two motor sounds). The duration of the four "continuous" sound types (rain, brook, and the two motor sounds) was adjustable, whereas the other sounds had set durations. All stimuli were 16-bit, 22,050 Hz wavelet synthesized sounds. The sounds were saved in WAV files for later playback. (Thus, sound synthesis was not occurring in parallel with the perceptual experiment.)

Different sounds were created from the models by manipulating wavelet parameters. The parameters were wavelet coefficients obtained from a wavelet decomposition of the base sound using the discrete wavelet transform (DWT). Wavelet coefficients were grouped according to frequency characteristics. Wavelet decomposition involved two filters: a high-pass filter captured the high-frequency signal characteristics, or *detail (Dj) coefficients,* and a low-pass filter captured the low-frequency signal characteristics,

or *approximation (Aj)* coefficients. The shape and filter characteristics were determined by the choice of wavelet type. These models used the *Daubechies* wavelet type. Wavelet decomposition is a multilevel operation, with successive levels (j) creating finer resolution, or higher level of detail, in the coefficient groups. The sounds were generated with a level five ($j = 5$) decomposition for each model. The *Dj* and *Aj* coefficient groups served as the parameter vectors. These parameter vectors were modified such that, upon signal reconstruction (with the inverse DWT), the resulting sounds differed perceptually from the original base sound.

Sound stimuli were generated by two types of parameter manipulations: *magnitude scaling* and *scaling filter manipulation* (Miner & Caudell, 2002). The magnitude-scale manipulations operated on the level one detail (D_1) and level 5 approximation (A_5) coefficients obtained from a wavelet decomposition. Many different perceptually related sounds resulted from this type of scalar manipulation. For example, dividing the A_5 approximation coefficients by a scalar (such as 2, 4, or 8) reduces the low-frequency contribution and makes a car motor sound like a small toy engine. Multiplying a detail coefficient group, such as D_1, by a scalar (such as 2, 4, or 8) uniformly enhances the high-frequency information and transforms the sound of rain (from the rain model) to something like bacon frying, for example. Stimuli were generated by multiplying coefficient vectors by scalars with magnitudes of 2 to 10, because scale factors between 1 and 2 did not yield a significant perceptual change, and scale factors above 20 yielded perceptually unrecognizable sounds. Particular scalar values and decomposition levels for stimuli are included in the results section.

The scaling filter manipulations involved changing the length of the IDWT reconstruction filter. When the number of reconstruction filter points was increased, the filter was stretched resulting in a decrease in the sound frequency. When filter length was decreased, the filter was compressed and the sound was shifted up in frequency. Scaling filter manipulations can change the sound of a brook to the sound of a large, slow moving river (stretching scaling filter), or to the sound of a rapidly moving stream (compressing scaling filter). Scale filter manipulations ranged from decreasing the filter length in half (creating a six-point filter) to doubling the filter length (creating a 24-point filter). The results were perceptually convincing in some cases and others not, depending on the base sound characteristics. Specific values for filter length manipulations are included in the results section. More details of the wavelet-based synthesis method are contained in Miner & Caudell (2002) and Miner (1998). There are many references on wavelet theory; Miner (1998) and Misiti, Misiti, Oppenheim, and Poggi (1996) provide good introductions to the topic. More-detailed references include Ogden (1996), Meyer (1993), and Daubechies (1992).

2 Psychoacoustic Experiments

Empirical validation of the quality of sound synthesis techniques requires psychoacoustic experimentation with human subjects. Varieties of experiments are possible. Bonebright et al. (1998) presented a test battery for validating sound veracity, although these experiments have not been accepted as a standard. For this research, a set of three psychoacoustic experiments was conducted: similarity rating, free-form identification, and context-based rating. The goal of these studies was to validate the aural imagery produced from the synthesized sound models. The similarity rating experiment examined the relationships between sound models and suggested ways to extend the synthesis to a broader range of sounds. The free-form identification experiment evaluated the scope of the synthesized sounds from the subject's perceptual identification data. The context-based rating experiment measured sound synthesis quality by comparing the synthesized sounds against human expectations. These methods could be used to evaluate any type of sound stimuli. In addition, the experiments provided valuable cognitive and perceptual information for psychoacoustic researchers. In the following sections, we

describe the procedures, results, and analysis for the three experiments.

2.1 Similarity Rating Experiment

Similarity rating studies analyze proximity (or distance) data for a set of objects to help understand the interrelationships among the objects. Pairwise proximities can be collected in a number of ways, but most often they come from direct judgments of similarity between object pairs, category judgments, and pairwise confusion data. In the current study, we collected similarity ratings among synthesized sounds to examine the perceptual parameter space of the synthesis models. The parameter space was created by changing the model parameter values. Of particular interest was how the human perceptual space compared to the sound parameter space. Subjects listened to synthesized sound pairs and rated the similarity between the sounds. The similarity rating data were analyzed by both MDS (which resulted in a mapping of the synthesized sounds onto a multidimensional perceptual space (Shepard, 1980) and Pathfinder (which resulted in an undirected graph (Goldsmith et al., 1991)).

2.1.1 Subjects. Twenty-two subjects participated in the experiment (7 men and 15 women) ranging in age from seventeen to forty years old (Mean = 20.55, ± 4.72). The subjects were students in an introductory psychology class from the University of New Mexico and received class credit for their participation. Nine out of the 22 subjects had some musical experience. All subjects had normal hearing and normal (or corrected to normal) vision.

2.1.2 Stimuli. Four base sounds were used, with five parameter settings each, giving twenty unique sound stimuli. (The base sound itself counts as a null parameter setting.) The base sounds were rain, brook, car motor running at 2000 RPM, and electric motor. We decomposed all sounds using the same wavelet type (Daubechies, number 4) and level (five) to examine the perceptual effects of consistent wavelet manipulations across a variety of sounds. All sounds were played for 1 sec.

The sounds were played in pairs with an average inter-stimulus interval (time between each stimuli) of 1.46 sec. (±0.16 sec.). An inter-trial interval (time between successive trials) of at least twice the inter-stimulus interval was used; thus, subjects waited at least 3.0 sec. between consecutive pairs.

2.1.3 Apparatus. Experiments were run on computer workstations with a Network Computing Devices (NCD) model MCX smart terminal with a 17 in. display, an embedded soundboard, and a Sun Sparc Server 20 host computer. The experimental lab had five workstations (separated by at least 5 ft.). Sound stimuli were presented to subjects through AKG K240 stereo headphones. The experiment was computer driven using a graphical user interface (GUI) developed in Matlab. Instructions were presented using the GUI to ensure consistency. All performance data were collected automatically.

2.1.4 Procedure. The similarity rating and freeform identification (subsection 2.2) experiments were conducted sequentially using the same subject pool. Subjects completed both experiments in less than one hour including instructions, practice trials, and debriefing. The similarity rating experiment is discussed first.

Subjects completed an information screen (containing their initials, age, sex, academic major, musical experience, and hearing status), signed a consent form, and read experiment instructions. Next, subjects rated six practice sound pairs. Subjects then clicked on a Start button to begin the actual sound pairs. Sounds within a pair were randomized for each subject. After hearing a sound pair, subjects clicked on a number that indicated the sound similarity. A five-point rating scale was used, with 1 = least similar and 5 = most similar. Each sound pair was played once. Subjects were asked to rate the pairs quickly, giving their first impressions rather than pondering over a pair; however, response times were not limited. Once a sound pair was rated, the next pair played au-

Table 1. *Averaged Similarity Ratings Across 22 Subjects*

Sound description	#	Sound #																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Brook	1	0.0	1.6	2.0	1.6	2.0	3.7	3.5	3.6	3.5	3.6	4.3	4.2	4.0	4.0	4.0	4.0	4.7	4.6	4.5	4.5
Brook, A_5*4	2		0.0	1.1	2.1	2.3	3.6	3.6	3.1	4.0	3.7	3.4	3.5	4.0	4.0	4.2	3.9	4.1	4.0	4.4	4.6
Brook, A_5*8	3			0.0	1.9	2.8	3.3	3.4	2.9	3.8	4.2	2.9	3.3	3.2	4.2	4.3	4.4	4.4	3.9	4.6	4.3
Brook, D_1*4	4				0.0	1.4	2.4	3.3	3.4	3.6	3.9	4.1	4.0	4.1	4.2	4.0	4.3	4.0	4.6	4.0	4.5
Brook, D_1*8	5					0.0	2.6	3.2	3.3	3.7	3.5	3.8	4.4	4.3	3.9	3.7	3.7	4.2	4.2	3.4	4.2
Rain	6						0.0	1.7	2.0	1.9	2.5	4.5	4.5	4.5	3.8	3.8	4.6	4.6	4.6	4.5	4.4
Rain, A_5*4	7							0.0	1.4	2.2	2.7	4.3	4.1	4.0	4.0	4.0	4.6	4.4	4.7	4.7	4.5
Rain, A_5*8	8								0.0	2.5	3.4	4.5	4.0	4.0	4.1	4.1	4.5	4.7	4.7	4.5	4.6
Rain, D_1*4	9									0.0	1.1	4.6	4.5	4.5	4.4	3.4	4.5	4.7	4.9	4.4	4.5
Rain, D_1*8	10										0.0	4.6	4.9	4.6	4.2	4.2	4.5	4.7	4.6	4.4	4.5
Motor	11											0.0	1.5	2.1	1.6	1.6	3.0	3.8	3.5	3.7	3.7
Motor, A_5*4	12												0.0	1.5	2.2	2.3	3.8	3.5	3.1	4.0	4.1
Motor, A_5*8	13													0.0	2.0	2.3	4.3	3.9	3.4	4.2	4.0
Motor, D_1*4	14														0.0	1.1	3.6	3.3	3.9	3.5	3.8
Motor, D_1*8	15															0.0	3.5	3.7	3.1	3.7	4.0
Emotor	16																0.0	1.8	1.9	1.5	1.9
Emotor, A_5*4	17																	0.0	1.0	1.9	2.1
Emotor, A_5*8	18																		0.0	2.2	2.1
Emotor, D_1*4	19																			0.0	1.2
Emotor, D_1*8	20																				0.0

Converted to distances such that rating range of 1–5: 1 = Least Distant, 5 = Most Distant.

tomatically after the 3 sec. inter-trial interval. Rating responses could not be changed once entered.

Each of the twenty sounds was paired with every other sound, resulting in 190 sound pairs. The sequence of sound pairs was individually randomized for each subject. A set of twenty sound pairs selected randomly from the set of 190 was repeated to test for response reliability. The repeated pairs were unique to each subject. Subjects received two 30 sec. breaks, one-third and two-thirds through the experiment.

2.1.5 Results. All subjects completed the experiment. The average elapsed time to complete the experiment was 1748.4 seconds (±140.22 seconds). Table 1 contains the average similarity rating results across the 22 subjects. Prior to averaging, the similarity responses were converted to distances by subtracting each similarity value from 6. Each sound was assigned a number (for example, brook = 1). The contents of the cell at the intersection of a particular row and column indicate the average distance for that sound pair. For example, the average distance for the brook sound (1) and the rain sound (6) is 3.7.

2.1.6 Multidimensional Scaling Analysis and Discussion. The MDS alternating least-squares scaling (ALSCAL) algorithm (Young & Lewyckyj, 1979) contained in SPSS 7.5 for Windows was used to analyze the sound distance data. MDS moves stimuli around in an *n*-dimensional space to reduce the stress between points.

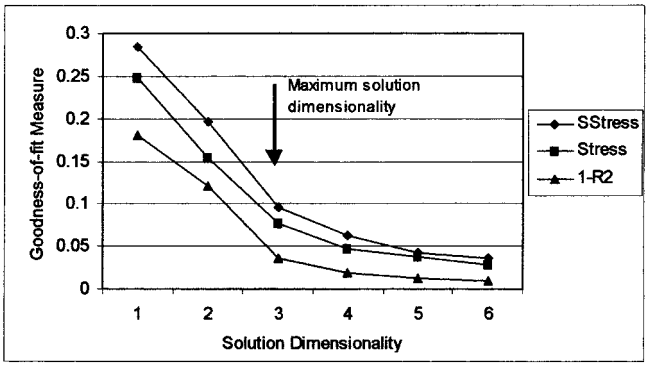

Figure 1. *Scree plot of three MDS ALSCAL goodness-of-fit measures.*

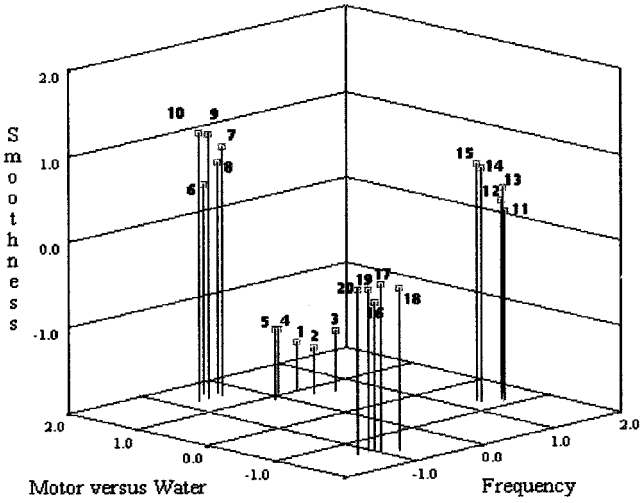

Figure 2. *Three-dimensional MDS solution from data in table 1. Axes are labeled with perceptual descriptions. Sound stimuli are numbered as in table 1.*

Typically, three goodness-of-fit measures influence the choice of *n*-dimensionality: Stress, SStress, and 1-R^2. Stress, as defined by Kruskal and Wish (1978) indicates how far the experimental distance data departs from the Euclidean distances for a particular space configuration. SStress is another distance measure (Takane, Young, & deLeeuw (1977). A minimum SStress value provides the best dimensionality fit. R^2 represented the proportion of variance in the ratings data accounted for by the MDS model. In general, solutions with an R^2 value greater than 0.8 are considered viable (Schiffman et al., 1981). 1-R^2 was calculated so that all of the goodness-of-fit values decreased with increasing dimensionality.

Scree plots serve as a heuristic guide in selecting the ideal dimensionality. Figure 1 shows the scree plots of the three goodness-of-fit measures versus the MDS dimensionality of $n = 6$. According to the scree criteria, the dimensionality at which a sharp "elbow" occurs indicates the maximum choice of solution dimensionality. For the sound similarity data, a distinctive elbow was apparent at a three-dimensional solution for each of the goodness-of-fit measures. Thus, the choice of a three-dimensional solution was clear.

Figure 2 contains a projection of the three-dimensional MDS solution with proposed perceptual axes labeled. The labels for the axes were obtained by subjective interpretation. For readability, the numbers in figure 2 refer to the numbered sounds in table 1.

For dimension 1, the motor sounds and water sounds were grouped together on either side of the axis midpoint. This was a logical grouping for these sound stimuli. Other labels for this axis could be naturalness, indicating artificial (mechanical) versus natural sound characteristics, or sound type, indicating motor sounds versus water sounds. It was evident by inspection of this axis that low-frequency manipulation of the brook sound began to convert it into a more mechanical sound. Interpreting this dimension offered a perspective on the amount of manipulation required to change the perception of the motor sound into rain or vice versa.

Dimension 2 showed the sounds progressing from high- to low-frequency as they moved from negative to positive values (respectively) along the axis. The highest-frequency sounds were the electric motor (Emotor), with rain, brook, and motor groups (in that order) having progressively lower frequencies. Inspection of dimension 2 revealed that subjects were able to perceive subtle variations in stimulus frequency. A possible label for dimension 3 is "smoothness." The smoothest sounds were found in the brook sound group, which contained variations of a flowing river. These were continuous sounds without clearly perceptible discrete com-

ponents. At the other extreme was the discrete rain sound group. Here the impacts of individual raindrops were clearly heard. High-frequency manipulation of the rain model emphasized its discrete nature, and this characteristic was reflected along dimension 3. The car motor sounds were less discrete than the rain sounds, although the perception of impacts was more evident. Again, high-frequency manipulation emphasized the discrete impacts in the motor sound. However, the impacts in the electric motor sound were so rapid and frequent that the sound might be perceived as continuous. This was reflected in the stimulus arrangement of dimension 3.

Overall, the MDS solution proved valuable in identifying sound dimensions salient to a human perceiver. Further, it was possible to show how manipulation of sound model parameters related to perceptually meaningful changes in the space. This type of understanding is potentially useful for reliably controlling the sound synthesis for a virtual environment to elicit a desired response from virtual environment participants.

2.1.7 Pathfinder Analysis and Discussion. The Pathfinder analysis technique provides an undirected network (graph) that indicates the conceptual relationships between sound stimuli (Schvaneveldt, 1990). In contrast to MDS, in which all stimuli equally influence the derived solution, the most-related stimulus pairs drive the Pathfinder solution. We used Pathfinder to answer two questions about the sound stimuli: (a) what, if any, clusters existed within the perceptual sound space (and how were these perceptual clusters related to the model parameter manipulations)?, and (b) how were the sounds related to one another? This information could be used to refine and extend the models to synthesize sounds that were perceptually related to the original sounds.

The average similarity data from all 22 subjects (table 1) served as input to the Pathfinder algorithm. It is often informative to examine several different Pathfinder networks that vary in number of links. Network connectedness ranges from a fully connected to a minimum spanning tree network (acyclic network with minimal total weight) based on Pathfinder criteria settings.

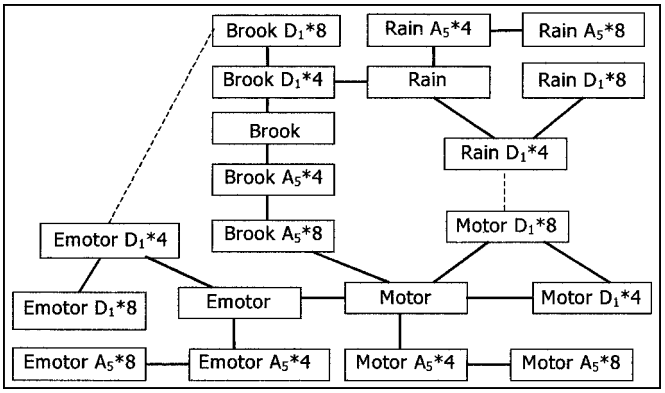

Figure 3. *Minimum spanning tree Pathfinder network solution for averaged subject data. Strongest links are solid lines; next most strongest links are dashed lines.*

The criteria to form a Pathfinder network (PFnet) is composed of two parameters: q and r. The q parameter constrains the search extent by restricting the number of links examined. If $q = 2$, Pathfinder searches alternate paths between any two objects with ≤ 2 links. If $q = n - 1$ (where n = number of objects = 20), all possible paths between any two nodes in the network are searched and this results in the sparsest graph. The r parameter is the path distance measure. Typically $r = \infty$, and the distance criteria becomes the max function. The max function says that the distance for any given path is equal to the maximum link along that path. The "weak links" are removed as the criteria on the parameter values are made more stringent.

To remove a link, Pathfinder compares the direct link weight between two nodes to the distance (calculated according to r) of all other paths of length q between the nodes. If the direct link weight is greater than or equal to any other path distance, the direct link is redundant and is removed. A network generated with particular values of q and r is referred to as a $PFnet(r, q)$. As the values of both the q and r parameters increase, the number of links in the network decreases. The acyclic network, with links connecting all vertices, having a minimal total weight is obtained with PFnet(∞, $n - 1$) and is known as the *minimum spanning tree.*

For the sound similarity data, Pathfinder analysis revealed a minimum spanning tree network, PFnet(∞, *n-1*) with twenty links, as shown in figure 3. The weak links (those links added when decreasing the *q* parameter by 1) are indicated by dotted lines. A network with 22 links was obtained with settings PFnet(∞,3). Thus, the only difference between the minimum spanning tree and the next most complex tree (next most related objects) were two links represented by dashed lines in figure 3.

The links in the minimum spanning tree represented the strongest conceptual relationships between the sound stimuli.

Whereas the MDS solution helped define the perceptual parameter dimensions, the Pathfinder networks provided a means for analyzing the conceptual relatedness. In this context, a "sound concept" refers to the generalized idea of a sound class. Clustering in the networks identified sound concepts. The Pathfinder network revealed distinct sound clusters for all four base sounds. A small number of links separated these sound clusters.

By examining the network links between sound concepts, it was possible to determine which model parameters were responsible for changing from one sound concept to another. The motor sound group was the most tightly coupled cluster with a maximum link distance of three, whereas the other groups had a maximum link distance of four. Thus, the Pathfinder network confirmed that the base sounds and their parameterized derivatives were highly related to each other in a conceptual sense. Furthermore, the network showed that the conceptual relatedness was consistent with the model parameter manipulations. Increasing the coefficient scale factors had the effect of increasing the conceptual distance from the base sound. This key finding establishes an important connection between the model parameter manipulations and the conceptual parameter space. By increasing the coefficient scale factors, perceptually related but increasingly distinct sounds are created.

Furthermore, examining how the sound clusters were related suggested ways to extend the scope of the sound model. The links between sound cluster groups in the network indicated the strongest conceptually related sounds. If one wanted to expand a sound model, the path links between sound clusters could suggest possible parameters to manipulate. For example, the rain sound had a link distance of 1 to the "brook, D_1*4" sound. This was the strongest link between these two sound groups. Further manipulation of the "brook, D_1*4" model parameters may yield sounds similar to those obtained from the rain sound model. Thus, the idea of extending the brook model to encompass rain sounds (or vice versa) is brought to light by examining the Pathfinder network.

Both the MDS and Pathfinder analyses of the similarity data helped to elucidate the relation between the sound model's parameters and human perceptions of the generated sounds. The MDS results showed how modifying the sound parameters moved a sound along three perceptual dimensions. The Pathfinder analysis showed how parameter manipulations related to changes in meaningful sound clusters. Both analyses provided guidance for refining and extending the sound model to a broader class of sounds.

The similarity rating experiment provided a tool for examining the psychological relations among a set of sounds. In the following studies, we examine the perceptual extent of aural images that could be synthesized with the wavelet models, and offer a metric for assessing the sound synthesis quality.

2.2 Freeform Identification Experiment

The freeform identification experiment examined the perceptual labels for synthesized sounds without providing a context (that is, verbal phrase or picture). The experiment answered the question "what aural image comes to mind when you listen to this sound?" Subjects listened to synthesized sounds and entered a description. This was a freeform type of identification experiment similar to that conducted by Ballas (1993) and Mynatt (1994). The purpose of the experiment was twofold. First, the experiment tested whether the parametrically manipulated synthesized sound resembled the base sound (or target sound) strongly enough to elicit a

freeform identification without any verbal or visual context. Second, the experiment identified perceptually related sound labels that were different from the base sound when parameters were modified. These perceptually related labels demonstrated the variety and scope of sounds that can be synthesized from an individual sound model.

This experiment was conducted directly after the similarity rating experiment with the same subjects. These experiments were considered complementary because exposure to the synthesized sounds during the similarity rating task served to clarify the aural imagery for the freeform identification. Thus, any perceptual ideas that were formed during the similarity rating experiment served to enhance the freeform label quality.

2.2.1 Subjects. The same 22 subjects from the similarity rating study participated in the freeform identification experiment. One subject left the experiment prior to completion.

2.2.2 Stimuli. Seven wavelet-based synthesized sounds served as the base sound stimuli. Four sounds were the same as used in the similarity rating experiment, except that here they were 3 seconds in duration rather than 1 second: rain, brook, car motor running at 2000 RPM, and electric motor. Additional shorter duration sounds used were footstep on gravel (0.5 second), glass breaking (2.0 seconds), and shuffling a deck of cards (1.2 seconds). In addition to these base sounds, four different parameter settings for each sound were used, giving a total of 35 stimuli.

2.2.3 Apparatus. The hardware apparatus was identical to that used in the similarity rating experiment, although different GUI screens guided subjects.

2.2.4 Procedure. Immediately following the completion of the similarity rating experiment, subjects began the freeform identification experiment. Subjects read the GUI instructions and could ask questions. Next, subjects clicked the mouse on the Begin button to bring up the main experiment screen. Subjects clicked on the Play Sound button to hear a sound. Each sound could be played as many times as desired. No time limit was imposed, although response times were recorded. After hearing the sound, subjects typed in a sound identification description in the Description Box and pressed the Enter key to finalize their entry. At this point, the description could not be changed. Identification descriptions included a noun and any relevant descriptive adjectives. Subjects were asked to think of the identification phrases in terms of the object(s) that could be creating the sound. Subjects could not replay previous sounds or modify their descriptions of these sounds, and there were no breaks or practice trials for this experiment. The order of the 35 synthesized sounds was individually randomized for subjects, and no sounds were repeated. The procedure continued until a sound description for each stimulus was obtained.

2.2.5 Results. Subjects took an average of 15.23 seconds (±6.64 seconds) to listen to and identify the sounds. The overall average time to complete the experiment was 768.93 seconds (±255.63 seconds).

A total of 367 unique freeform responses were obtained for the 35 different sounds with an average of 10.49 different descriptions per sound (±4.13). Responses were combined when they were identical in content or meaning. For example, *water from a sink running* and *faucet* were combined and given a response frequency of 2. In addition, *running river water, water running in a river, river noise,* and *flowing river* were combined into a single term with a count of 4. The terms *rain falling against a window* and *rain* were not combined because the first term provides additional information that would be lost if it were combined with the simpler term *rain*. In addition, the terms *bath water* and *running water* were not combined because the first term implies a specific perceptual quality to the sound of *running water*. The response frequency indicated how many subjects identified a sound with the same or similar phrase.

Table A in appendix 1 contains the average response times and standard deviation for each term (across subjects), the freeform label with the highest response fre-

quency across all subjects and all responses, the frequency of that response, and the total number of unique phrases.

2.2.6 Discussion. The first question addressed was whether the synthesized sounds with no parameter manipulations were identified correctly. Table A in appendix 1 shows the most-frequent responses for each of the 35 stimulus. A complete list of all responses can be found in Miner (1998). In six out of seven cases, the synthesized sound with no parameter manipulations was most frequently identified correctly (that is, consistently with the base sound being synthesized). In the case of the footstep, the most frequent response was "Eat/crunch/biting/chewing." However, when all responses are considered, nine subjects out of 22 identified the sound as some type of footstep. The footstep descriptions were varied however: stepping on fallen leaves (three responses), footstep in the snow (three), walking on gravel (two), and stepping into mud (one). To preserve the additional aural imagery information, these sounds were not combined into a single entry. Thus, this experiment showed that the synthesized sounds, with no parameter manipulations, resembled the base sounds strongly enough to elicit a free-form identification without any verbal or visual context.

A second question addressed was whether perceptually related sounds would be obtained by manipulating sound model parameters. Results showed that 367 unique sound labels were obtained for the 35 different synthesized sounds, or an average of 10.49 ($\pm$3.67) unique terms per sound. There were many different labels given for each parameterized sound. These labels were generally more descriptive of the sound (that is, specific contexts) heard and were thus not combined into a single entry. The sound labels were consistent with the base sounds but also identified perceived sounds that were quite different from the base sound. Labels for mechanically oriented sounds often emerged as the frequency of the synthesized sounds increased. Sound labels indicating larger objects were more common with low-frequency synthesized sounds. Thus, manipulating the model parameters resulted in a predictable change in aural imagery across a variety of different sound types.

Changing the base sound parameter values resulted in the most frequent description changing to a different sound for the rain, brook, and shuffling card sounds. For the rain base sound, the most consistent response for the parameterized sounds was "water running in the shower." Parameter manipulations of the brook model changed the most frequent response from "water running in a river" to "rain." Parameter manipulations of the card-shuffling model changed the most frequent response from "shuffling cards" to "breaking twigs" and "starting a motorcycle." This result supported the hypothesis that manipulating the parameters creates a variety of aural images beyond that of the base sound.

The motor sound model produced the greatest variety of labels with an average of 13.8 unique labels. Many of the labels for the motor sound described some type of engine sound, including car motor, truck, diesel engine, helicopter, lawn mower, airplane, tractor, and vacuum cleaner. Another group of responses referred to different types of machinery or construction equipment, including jackhammer, generator, drill, and bulldozer. This finding was consistent with the Pathfinder network showing the motor sounds with the highest connectivity to other stimuli.

The glass-breaking sound model produced the lowest number of unique responses with an average of 3.4 labels per sound. Correspondingly, the glass-breaking model had the lowest average response time and the highest average number of common responses per label. Thus, the glass-breaking sound was perceptually salient and easily identifiable. There were some variations in the adjective descriptions of the sounds such as "crystal" or "delicate glass breaking" versus "heavy vase" or "window breaking," but much less variation in the responses than for any other sound. In fact, only one response given for this sound group was outside the category of "glass breaking," the identification phrase was "keys falling to the ground" given by one subject for the D_1*8 parameter

settings. The low response variation implies that the parameter manipulations succeeded in creating subtle sound variations, but they were not successful in transforming the base sound into perceptually different sounds. Results may also indicate that the breaking-glass sound had characteristic components not shared by other sounds. A low unique response count may indicate the synthesis of a globally convincing aural image.

One point to consider when examining the average response times was that some sounds were 3 seconds in duration, while others ranged from 0.5 second to 2 seconds. Response times were measured from the end of the sound stimulus until a subject completed entering an identification label. If subjects chose to repeatedly hear the sound stimulus before entering a label, this time was included in the response time average. Thus, the average response time for the continuous sounds would be higher than for the shorter sounds. As mentioned previously, the footstep sounds had the longest average response time, and the glass-breaking sound had the shortest average response time. Response times may indicate the compellingness of the aural image generated, with faster times indicating a more compelling image and slower times a less compelling image.

Examination of the individual responses revealed that there was a cross-pollination of labels between groups. For example, the "brook, D_1*8" sound was identified by a few subjects as a "printing press" or "machine noise" sound. Also, the "rain with seventeen-point reconstruction filter" sound was identified as "machinery" and "truck driving." These results may have been influenced by other sound stimuli (for example, hearing a machine noise previously may have influenced a subject to identify successive sounds as machine sounds) rather than creating a convincing synthesis of that sound. In this sense, the group of experimental sound stimuli provided a sound context. It was not possible to discern the influence of one sound on another because the cross-pollination of labels did not occur consistently enough across subjects.

Overall, the results from the freeform identification experiment helped to validate the correct identification of the base sounds. In addition, the results provided a variety of perceptually related sound labels. Furthermore, results showed that manipulation of model parameters resulted in predictable changes in the aural imagery, providing guidelines for developing new sound models. However, the freeform identification task did not provide a measure of goodness for the freeform labels. Thus, it was not clear how well subjects' responses matched the sounds they heard. Such information would be important for generating realistic sounds. The next experiment used a rating scale to evaluate the goodness, or veracity, of the sound labels obtained in the freeform identification experiment.

2.3 Context-based Rating Experiment

The third experiment was designed to provide a sound quality metric for each synthesized sound within a verbal context. This experiment also showed how changes in the sound parameters succeeded in creating corresponding perceptual changes. Phrases obtained from the freeform identification experiment were paired with synthesized sounds to provide a context. Subjects were asked to rate how well the phrases matched the sounds they heard. These sound ratings measured how well the synthesized sounds corresponded to the perceptual labels.

2.3.1 Subjects. Twenty-seven subjects from the same subject pool as the first two studies (five men and 22 women), ranging in age from 18 to 26 years (mean = 18.93, ± 2.15), completed the experiment. None of the subjects participated in the previous two experiments. Thirteen out of the 27 subjects had some musical experience. All subjects had normal hearing and normal (or corrected to normal) vision.

2.3.2 Stimuli. A subset of sound stimuli and labels from the freeform identification experiment were used. The number of labels used for each sound model was based on the most-frequent responses

Table 2. *Highest Rated Sound-Phrase Pairs for Rain Stimuli*

	Sound descriptions (parameter settings)		
Rating	Rain base sound	Rain base w/D_1*8	Rain base w/A_5*4
Very good	Hard rain	Light drizzle of rain	Hard rain
Good	Shower water running	Shower water running	Large waterfall
Match	Lots of people typing	Lots of people typing	Large fire
Match	Small waterfall	Bacon frying	Small waterfall

Rating label scores: Very good = 4.25–5.0, Good = 3.5–4.25, Match = 2.75–3.5. Sound descriptions indicate sound-phrase pairs receiving a rating (For example, For Rain Base Sound, The Phrase "Hard Rain" Received a Match Rating of "Very Good").

from the freeform identification experiment. In addition, cross-matching labels between sound types provided for a full range of possible responses. The number of labels was as follows: rain (ten labels), brook (eleven), 2000 RPM motor (ten), electric motor (ten), footstep (ten), glass breaking (six), and shuffling cards (twelve), for a total of 69 labels. Example labels are contained in table A in appendix 1 and the rating results are in table 2 through 5. (See Miner (1998) for a complete listing.) Parameter manipulations consisted of the original base sound and two different parameter settings for each model (details coefficients on level 1 (D1) scaled by 8 and approximation coefficients on level 5 (A5) scaled by 4). Thus, a total of 207 sound and text label (or phrase) pairs were individually randomized and presented to each subject. In addition, fifteen pairs were repeated to measure response reliability. The repeated pairs were selected from the randomized list of 207 terms; thus, the set of repeated pairs was unique for each subject. At the end of the experiment, subjects evaluated the repeated stimuli. Overall, subjects evaluated 222 sound-phrase stimulus pairs.

2.3.3 Apparatus. The hardware apparatus for this experiment was identical to that used in the previous two experiments. Different GUI screens guided subjects through the task.

2.3.4 Procedure. First, subjects completed the information screen, signed the consent form, and read the experiment instructions. There were no practice trials or breaks for this experiment. Subjects clicked on the start button on the main screen to begin the experiment. The first context-based phrase was displayed for subjects to read. After a 2-second pause, the corresponding synthesized sound was played. Subjects had time to read and understand the context-phrase prior to hearing the sound. Each sound could be heard only once. Subjects were asked to click on the number that indicated how well the phrase and the sound matched. Subjects used a five-point rating scale to indicate their judgments: 1 = no match, 2 = below average match, 3 = moderate match, 4 = good match, 5 = excellent match.

Subjects were asked to rate each sound-phrase pair quickly, using their first impressions rather than pondering each for a long time. Response times were not limited. After the inter-trial interval elapsed, the next sound was played. Rating responses could not be changed. The stimuli order was randomized for each subject.

2.3.5 Results. The average time to complete the experiment was 1665.5 seconds (±159.03 seconds). All of the 222 sound-phrase pair average rating results are contained in Miner (1998). Of particu-

Table 3. *Highest Rated Sound-Phrase Pairs for Footstep Stimuli*

	Sound descriptions (parameter settings)		
Match rating	Footstep base sound	Footstep base w/D_1*8	Footstep base w/A_5*4
Good	Footstep in the snow	Footstep in the snow	Footstep in the snow
Good	None	Stepping on gravel	Shaking out a rug
Match	Stepping on gravel	Shaking out a rug	None
Match	Shovel digging dirt	None	None

"None" indicates no additional phrases with that rating (that is, sound, for footstep base, only 1 phrase obtained a Good rating).

Table 4. *Highest Rated Sound-Phrase Pairs for Breaking, Glass Stimuli*

	Sound descriptions (parameter settings)		
Match rating	Glass base sound	Glass base w/D_1*8	Glass base w/A_5*4
Very good	Breaking fragile crystal	Breaking fragile crystal	A window breaking
Good	A window breaking	None	Breaking fragile crystal
Good	Breaking a plate	None	Breaking a plate
Good	None	None	Breaking a heavy vase
Match	Breaking a heavy vase	Breaking a heavy vase	None
Match	None	Breaking a plate	None
Match	None	A window breaking	None

lar interest are the sound-phrase pairs that received the highest ratings (that is, the best sound-phrase match). We used the following labels from the five-point scale to describe matching ranges: match = 2.75 to 3.5; good match = 3.5 to 4.25; very good match = 4.25 to 5.0. Due to space limitations, we focus on the results (and discussion) from four of the seven models. Tables 2 through 4 contain the highest rated (≥2.75) sound-phrase pairs for rain, footstep and glass breaking.

Examining the best matches for each sound in different groups indicated how successful the parameter manipulations were at changing the base sound perception. For example, table 2 shows the best matches for the rain sound group. The label that best matched the original sound was "hard rain." By increasing the high-frequency sound content through parameterization, the best perceptual sound label match changed to a "light drizzle of rain." The "large waterfall" label was a good match for the low-frequency parameterization. Thus, these results indicated that parameter manipulations were successful in changing the perceptual auditory imagery.

Table 3 shows that the footstep sound obtained a match rating for the "stepping on gravel" label rather than a very good rating as might have been expected. (Note that the base sound for this model was the sound of a footstep on gravel.) This result suggests that digitizing a sound does not guarantee a convincing aural image. In fact, the "stepping on gravel" syn-

thesized sounds (that is, manipulating sound model parameters to enhance the sound) received a higher rating than did the original digitized sound of a footstep on gravel. This result is an example of what we call the *Foley effect*. The Foley effect is when human listeners find a digitally embellished (or synthetically enhanced) sound more compelling than an actual sound recording. This experiment provided a method for examining which sounds created the most compelling aural images. A Foley, or virtual world builder, could use these results to select sounds for a given scenario.

Table 4 (breaking-glass sound group) shows that identical labels can result in different degrees of matching depending on the parameter settings. The parameterization that emphasized the high-frequency components (D_1*8), resulted in a very good match for "breaking fragile crystal" and a match for "a window breaking" and "breaking a heavy vase." The parameterization emphasizing the low-frequency sound components (A_5*4) resulted in a very good match for "a window breaking" and a good match for "breaking a heavy vase" and "breaking fragile crystal." Again, manipulation of sound model parameters influenced the perceptual interpretation.

Overall, these results validated the quality of the sounds resulting from the wavelet-based sound model. The results suggest that parameter changes and other general manipulations can be applied to wavelet-based sound models to obtain perceptually compelling synthesized sounds. This experiment showed that the freeform identification experiment did not adequately quantify the sound synthesis quality. The freeform responses provided a label for the sound but lacked any indication of label quality. Furthermore, this experiment provided numeric information about how the aural imagery changed as the model parameter settings changed. Thus, this experiment numerically validated the perceptual success of the model parameter manipulations.

Additional context-based experiments could provide metrics for quantifying the sound synthesis veracity in other settings. For example, still images or video could provide the context for evaluating the quality of corresponding sounds. Synthesized sounds should be perceived as more compelling when coupled with realistic and compelling visual stimuli. Another study could examine sound quality in the context provided by an immersive virtual environment. The compellingness of the synthesized sounds could be compared to the results from both the verbal and the video context for evaluating the veracity of the virtual experience as a whole.

6 Conclusions

We have described three psychoacoustic experiments that investigated the underlying perceptual characteristics of synthesized sounds and then validated their realism as perceived by human observers. We believe that experiments of this type are important for confirming the success of a sound synthesis technique, and for ensuring the effective use of sound in applications. Separate comparative experiments could be performed between differing VR sound techniques to evaluate relative merit. The experiments presented in this paper validated that the wavelet-based sound synthesis technique created perceptually convincing sounds. Furthermore, results showed that model parameter manipulation was effective for creating a variety of convincing sounds. This information both validated the sound synthesis technique and provided avenues for extending the sound model scope. Application designers can use these types of experiments to ensure that sounds included in a virtual simulation enhance rather than detract from the simulated experience.

Acknowledgment

Sandia National Laboratories supported this work under its Doctoral Study Program. Thanks to the reviewers who provided helpful comments, and thanks also to experiment volunteers for their participation.

Appendix 1. *Freeform Identification Summary Statistics. Highest-Frequency Responses Listed*

Sound stimuli	Most frequent freeform label	Resp freq	Avg resp time (sec.)	Std dev	# Unique IDs	Avg # IDs per term
Rain	Rain	8	14.25	7.80	7	
Rain, D_1*8	Water running in the shower	9	14.65	7.15	8	
Rain, A_5*4	Water running in the shower	7	12.68	4.48	9	
Rain, 17-pt filter	Tie: 1) a large fire 2) shower	3	14.33	7.04	13	
Rain, 7-pt filter	Water running in the shower	9	15.15	8.38	5	8.4
Motor	Construction machine/tractor	5	15.11	9.15	13	
Motor, D_1*8	Machine engine/noise	4	15.78	6.80	12	
Motor, A_5*4	A tractor/heavy equipment	4	15.02	5.62	12	
Motor, 17-pt filter	Truck/large truck engine	3	17.15	5.39	17	
Motor, 9-pt filter	Drill/drilling/high pitched drill	3	15.49	4.36	15	13.8
Brook	Water running in a river	6	14.51	10.18	10	
Brook, D_1*8	Rain	5	15.98	8.18	11	
Brook, A_5*4	Rain	5	22.74	40.96	12	
Brook, 17-pt filter	Rain	4	15.25	7.81	16	
Brook, 9-pt filter	Tie: 1) Heavy rain 2) TV static	3	15.94	6.71	16	13
Electric motor (Emotor)	Tie: 1) an electric razor 2) drill/drilling/dentist drill	3	11.37	4.17	13	
Emotor, D_1*8	An electric razor	8	13.54	4.92	7	
Emotor, A_5*4	An electric razor	4	12.04	4.15	12	
Emotor, 17-pt filter	Machine noise/factory machines	3	24.53	17.89	13	
Emotor, 7-pt filter	An electric razor	10	12.70	6.19	11	11.2
Footstep	Eat/crunch/biting/chewing	5	16.92	9.87	13	
Footstep, D_1*8	Stepping on dried leaves/twigs	6	18.70	8.34	11	
Footstep, A_5*4	Biting/chewing/eating	3	16.28	6.31	16	
Footstep, 17-pt filter	Eat/crunch/biting/chewing	3	18.07	10.18	16	
Footstep, 9-pt filter	Bite apple/crunch cereal/eating	9	18.15	10.14	11	13.4
Glass	Glass breaking	19	11.97	4.90	3	
Glass, D_1*8	Glass breaking	16	11.30	6.25	3	
Glass, A_5*4	Glass breaking	15	15.58	11.78	3	
Glass, 20-pt filter	Glass breaking	10	10.76	5.68	6	
Glass, 7-pt filter	Glass breaking	14	10.71	4.06	2	3.4
Cards	Shuffling (large) deck of cards	10	13.49	5.38	10	
Cards, D_1*8	Breaking twigs/noodles/celery	9	15.28	6.62	6	
Cards, A_5*4	Shuffling (large) deck of cards	6	19.86	11.79	10	
Cards, 17-pt filter	Starting a motorcycle (Harley)	5	15.54	6.69	12	
Cards, 8-pt filter	Shuffling cards	4	12.41	4.57	13	10.2
Overall statistics			15.23	3.08	367	10.49

"Resp freq" is number of times that response was given. Average response time (and standard deviation) is average time to respond after sound is played first time. "Num unique ids" is the total number of unique stimulus phrases obtained for that stimulus. Data are across all subjects.

References

Ballas, J. (1993). Common factors in the identification of an assortment of brief everyday sounds. Journal *of Experimental Psychology: Human Perception and Performance, 19*(2), 250–267.

Ballas, J., & Mullin, T. (1991). Effects of context on the identification of everyday sounds. *Human Performance, 4*, 199–219.

Bonebright, T. L. (1996). *Vocal affect expression: A comparison of multidimensional scaling solutions for paired comparisons and computer sorting tasks using perceptual and acoustic measures.* Doctoral dissolution, University of Nebraska.

Bonebright, T., Miner, N., Goldsmith, T., & Caudell, T. (1998). Data collection and analysis techniques for evaluating the perceptual qualities of auditory stimuli. *Proceedings of the International Conference on Auditory Displays.* Available online at http://www.icad.org/websiteV2.0/Conferences/ICAD98/icad98progamme.html

Daubechies, I. (1992). *Ten lectures on wavelets.* Philadelphia: SIAM.

Flowers, J. H, & Hauer, T. A. (1993). "Sound" alternatives to visual graphics for exploratory data analysis. *Behavior Research Methods, Instruments & Computers, 25* (2), 242–249.

———. (1995). Musical versus visual graphs: Cross-modal equivalence in perception of time series data. *Human Factors,* 37, 553–569.

Fox, R. A. (1985). Multidimensional scaling and perceptual features: Evidence of stimulus processing or memory prototypes? *Journal of Phonetics*, *13, 205–217.*

Gaver, W. (1994). Using and creating auditory icons. In *Auditory display: Sonification, audification, and auditory interfaces, G. Kramer(Ed.), Proceedings vol. XVIII* (pp. 417–446). Reading, MA: Addison-Wesley.

Gaver, W., Smith, R., & O'Shea, T. (1991). Effective sounds in complex systems: The ARKola simulation. *Proceedings of CHI 1991,* 85–90.

Green, R. S., & Cliff, N. (1975). Multidimensional comparisons of structures of vocally and facially expressed emotion. *Perception and Psychophysics, 17*(5), 429–438.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge *Journal of Educational Psychology, 83*(1), 88–96.

Hollins, M., Faldowski, R., Rao, S., & Young, F. (1993). Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis. *Perception & Psychophysics, 54*(6), 697–705.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling.* Beverly Hills: Sage Publications.

Meyer, Y. (1993). *Wavelets: Algorithms and applications.* Philadelphia: SIAM.

Miner, N. E. (1998). *Creating wavelet-based models for real-time synthesis of perceptually convincing environmental sounds.* Doctoral dissertation, University of New Mexico.

Miner, N. E. & Caudell, T. P. (2002). A wavelet synthesis technique for creating realistic virtual environment sounds. *Presence: Teleoperators and Virtual Environments, 11*(5), 493–507.

Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J. (1996). *Wavelet toolbox for use with matlab.* MathWorks, Inc.

Mynatt, E. D. (1994). Designing with auditory icons. *Proceedings of the Second International Conference on Auditory Display (ICAD),* 109–119.

Ogden, R. (1996). *Essential wavelets for statistical applications & data analysis.* Boston: Birkhauser.

Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organization.* New Jersey: Ablex Publishing.

Schiffman, S., Reynolds, M., & Young, F. (1981). *Introduction to multidimensional scaling: Theory, methods and applications.* New York: Academic Press.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210*(October 24), 390–398.

Shepard, R. N., & Cooper, L. A. (1992). Representations of colors in the blind, color-blind, and normally sighted. *Psychological Science, 3*(2), 97–104.

Takane, Y., Young, F. W., & deLeeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least-squares method with optimal scaling features. *Psychometrika*, *42*, 7–67.

Young, F. W., & Lewyckyj, R. (1979). *ALSCAL-4: Users guide.* Chapel Hill: Data Analysis and Theory Associates.