Sounding Objects

Davide Rocchesso
University of Verona, Italy

Roberto Bresin
Royal Institute of Technology, Sweden

Mikael Fernström
University of Limerick, Ireland

Interactive systems, virtual environments, and information display applications need dynamic sound models rather than faithful audio reproductions. This implies three levels of research: auditory perception, physics-based sound modeling, and expressive parametric control. Parallel progress along these three lines leads to effective auditory displays that can complement or substitute visual displays.

When designing a visual widget we attempt to understand the users' needs. Based on initial requirements, we can sketch the idea on paper and start to test it. We can ask potential users for advice and they might tell us that, for example, "it should be like a rectangular button, with a red border, in the upper left hand corner, with a frame around it." As testing and new sketches evolve, we can then code the widget and do further tests.

Designing auditory-enhanced interfaces immediately poses a number of problems that aren't present in the visual case. First, the vocabulary for sound is vague—for example, "when I do this, it goes whoosh," or "it should sound like opening a door." What does whoosh or an opening door really sound like? Is one person's whoosh the same as another person's whoosh? What can I do, as a designer, if I want to continuously control the sound of a door opening? A sample from a sound effect collection is useless in this case. This article aims to shed some light on how psychologists, computer scientists, acousticians, and engineers can work together and address these and other questions arising in sound design for interactive multimedia systems.

Sounding Object

It's difficult to generate sounds from scratch with a given perceived identity using signal-based models, such as frequency modulation or additive synthesis.[1] However, physically based models offer a viable way to get naturally behaving sounds from computational structures that can easily interact with the environment. The main problem with physical models is their reduced degree of generality—that is, models are usually developed to provide a faithful simulation of a given physical system, typically a musical instrument. Yet, we can build models that retain direct physical meaning and are at the same time configurable in terms of physical properties such as shape, texture, kind of excitation, and so on. We used this approach in the Sounding Object[2] (http://www.soundobject.org) project that the European Commission funded to study new auditory interfaces for the Disappearing Computer initiative (http://www.disappearing-computer.net).

Based on our experience with the Sounding Object project, and other recent achievements in sound science and technology, we'll explain how

- the sound model design should rely on perception to find the ecologically relevant auditory phenomena and how psychophysics can help in organizing access to their parameters;
- physics-based models can be "cartoonified" to increase both computational efficiency and perceptual sharpness;
- the physics-related variables have to be varied in time and organized in patterns to convey expressive information (this is the problem of control); and
- we can construct interactive systems around sound and control models.

We'll illustrate each of these points with one example. You can download the details, software implementations, and sound excerpts of each example from http://www.soundobject.org.

Perception for sound modeling

Humans sense the physical world and form mental images of its objects, events, and processes. We perceive physical quantities according to nonlinear, interrelated scales. For instance, humans perceive both frequency and intensity of sine waves according to nonlinear scales (called, respectively, *mel* and *son*), and intensity perception is frequency dependent. The discipline of psychoacoustics has tried, for more than a century, to understand how sound perception

1070-986X/03/$17.00 © 2003 IEEE

Published by the IEEE Computer Society

Previous Research

The introduction of sound models in interfaces for effective human–computer interaction dates back to the late 1980s, when William Gaver developed the SonicFinder for Apple's MacIntosh, thus giving an effective demonstration of auditory icons as a tool for information and system status display.[1,2] The SonicFinder extended Apple's file management application finder using auditory icons with some parametric control. The strength of the SonicFinder was that it reinforced the desktop metaphor, creating an illusion that the system's components were tangible objects that you could directly manipulate. Apple's later versions of MacOS (8.5 onward) implemented appearance settings—which control the general look and feel of the desktop metaphor, including sound—based on the SonicFinder.

During the last 10 years, researchers have actively debated issues such as auditory icons, sonification, and sound design while the International Community for Auditory Display (http://www.icad.org) has emerged as an international forum, with its own annual conference. In addition, several other conferences in multimedia and computer–human interaction have hosted an increasing number of contributions about auditory interfaces. Researchers have developed several important auditory-display-based applications that range from computer-assisted surgery to continuous monitoring of complex systems to analysis of massive scientific data.[3] However, the relevant mass of experiments in sound and multimedia have revealed a substantial lack of methods for designing meaningful, engaging, and controllable sounds.

Researchers have attempted to fill this gap from two opposite directions. Some researchers have tried to understand and exploit specific sound phenomena. For example, Gaver[4] studied and modeled the sound of pouring liquids. Others have constructed general and widely applicable sound information spaces. For instance, Barrass proposed an auditory information space analogous to the color information spaces based on perceptual scales and features.[5]

The need for sounds that can convey information about the environment yet be expressive and aesthetically interesting, led to our proposal of sound spaces constructed on dynamic sound models. These sound spaces build on synthesis and processing models that respond continuously to user or system control signals. We proposed that the architecture for such a sound space would be based on perceptual findings and would use the sound modeling technique that's most appropriate for a given task. Sound events have both identity and quality aspects,[6] so physical models are appropriate for representing a sound's identity and signal-based models are more appropriate for adjusting a sound's quality.[7] In fact, the nature of a sound's physical structure and mechanisms (such as a struck bar, a rubbed membrane, and so on) determines a sound source's identity while the specific instances of physical quantities (for example, metal bars are brighter than wood bars) determine its qualitative attributes.

References

1. W.W. Gaver, "The Sonic Finder: An Interface that Uses Auditory Icons," *Human–Computer Interaction*, vol. 4, no. 1, Jan. 1989, pp. 67-94.
2. W.W. Gaver, "Synthesizing Auditory Icons," *Proc. INTERCHI*, ACM Press, 1993, pp. 228-235.
3. G. Kramer et al., *Sonification Report: Status of the Field and Research Agenda*, 1997, http://www.icad.org/websiteV2.0/References/nsf.html.
4. W.W. Gaver, "How Do We Hear in the World? Explorations in Ecological Acoustics," *Ecological Psychology*, vol. 5, no. 4, Sept. 1993, pp. 285-313.
5. S. Barrass, "Sculpting a Sound Space with Information Properties," *Organised Sound*, vol. 1, no. 2, Aug. 1996, pp. 125-136.
6. S. McAdams, "Recognition of Sound Sources and Events," *Thinking in Sound: The Cognitive Psychology of Human Audition*, S. McAdams and E. Bigand, eds., Oxford Univ. Press, 1993, pp. 146-198.
7. M. Bezzi, G. De Poli, and D. Rocchesso, "Sound Authoring Tools for Future Multimedia Systems," *Proc. IEEE Int'l Conf. Multimedia Computing and Systems*, vol. 2, IEEE CS Press, 1999, pp. 512-517.

works.[3] Although the amount of literature on this subject is enormous, and it can help us guide the design of auditory displays, most of it covers perception of sounds that rarely occur in real life, such as pure sine waves or white noise. A significantly smaller number of studies have focused on sounds of musical instruments, and an even smaller body of literature is available for everyday sounds. The latter are particularly important for designing human–computer interfaces, information sonification, and auditory displays.

Even fundamental phenomena, such as the auditory perception of size and shape of objects or the material they're made of, have received attention only recently (see Vicario et al.[4] for an annotated bibliography). More complex phenomena, such as bouncing and breaking events or liquids pouring into vessels, have drawn the attention of ecological psychologists and interface designers,[5] mainly because of their expressiveness in conveying dynamic information about the environment.

As an example, we report the results of a test that we conducted using a large subject set (197 undergraduate computer science students) to evaluate the effectiveness of liquid sounds as an

auditory substitute of a visual progress bar. The advantages for this application are evident, as the user may continuously monitor background activities without being distracted from the foreground work. We used 11 recordings of 0.5- and 1-liter plastic bottles being filled or emptied with water. The sound files' lengths ranged from 5.4 to 21.1 seconds. The participants used headphones to listen to the sounds. They were instructed to respond using the "0" and "1" keys on their keyboards. We divided the experiment into three sections to detect

- the action of filling or emptying,
- whether the bottle was half full or empty, and
- whether the bottle was almost full.

We randomized the order between sections and between sound files. When we asked the participants to respond if the sound was filling or emptying, 91.8 percent responded correctly for emptying sounds and 76.4 percent for filling sounds. In the section where they responded to when the bottle was half full or empty, responses during filling had a mean of 0.40 (range normalized to 1) with a standard deviation of 0.13. During emptying responses had a mean of 0.59 with a standard deviation of 0.18. In the section where users responded to whether the bottle was almost full (just about to overflow) or almost empty, the mean value for filling sounds was 0.68 with a standard deviation of 0.18, and for emptying sounds the mean was 0.78 with a standard deviation of 0.18. Based on these results, we envisage successfully using this kind of sound, for example, as an auditory progress bar, because users can distinguish between full or empty and close to completion. From informal studies, we've noted that users also can differentiate between the bottle sizes, pouring rate, and viscosity.

Perception is also important at a higher level, when sounds are concatenated and arranged into patterns that can have expressive content. Humans have excellent capabilities of deducing expressive features from simple moving patterns. For instance, the appropriate kinematics applied to dots can provide a sense of effort, or give the impression that one dot is pushing another.[6] Similarly, temporal sequences of sounds can convey expressive information if properly controlled.

Cartoon sound models

In information visualization and human–computer visual interfaces, photorealistic rendering is often less effective than nicely designed cartoons. Stylized pictures and animations are key components of complex visual displays, especially when communication relies on metaphors.[6] Similarly, auditory displays may benefit from sonic cartoons[5]—that is, simplified descriptions of sounding phenomena with exaggerated features. We often prefer visual and auditory cartoons over realistic images and sounds[7] because they ease our understanding of key phenomena by simplifying the representation and exaggerating the most salient traits. This can lead to a quicker understanding of the intended message and the possibility of detailed control over the expressive content of the picture or sound.

Sound design practices based on the understanding of auditory perception[8] might try to use physics-based models for sound generation and signal-based models for sound transformation. In this way it's easier to impose a given identity to objects and interactions (such as the sound of impact between two ceramic plates). Some quality aspects, such as the apparent size and distance of objects, may be adjusted through time-frequency manipulations[9] when they're not accessible directly from a physical model.

Physics-based models

The sound synthesis and computer graphics communities have increasingly used physical models whenever the goal is a natural dynamic behavior. A side benefit is the possibility of connecting the model control variables directly to sensors and actuators since the physical quantities are directly accessible in the model. We can use several degrees of accuracy in developing a physics-based sound model. Using a finite-element model of interacting objects would make sense if tight synchronization with realistic graphic rendering is required, especially in the cases of fractures of solids or fluid-dynamic phenomena.[10] Most often, good sounds can be obtained by simulating the rigid-body dynamics of objects described by piecewise parametric surfaces.[11] However, even in this case each impact results in a matrix of ordinary differential equations that must be solved numerically. If the display is mainly auditory, or if the visual object dynamics can be rendered only poorly (for example, in portable low-power devices), we can rely on perception to simplify the models signifi-

cantly, without losing either their physical interpretability or dynamic sound behavior.

As an example, consider the model of a bouncing object such as a ball. On one extreme, we can try to develop a finite-element model of the air cavity, its enclosure, and the dynamics of impact subject to gravity. On the other hand, if we have a reliable model of nonlinear impact between a striking object and a simple mechanical resonator, we can model the system as a point-like source displaying a series of resonant modes. The source can be subject to gravity, thus reproducing a natural bouncing pattern, and we can tune the modes to those of a sphere, thus increasing the impression of having a bouncing ball.

Of course, this simplification based on lumping distributed systems into point-like objects introduces inevitable losses in quality. For instance, even though we can turn the ball into a cube by moving the modal resonances to the appropriate places, the effect wouldn't resemble that of a bouncing cube just because the temporal pattern followed by a point-like bouncing object doesn't match that of a bouncing cube. However, we can introduce some controlled randomness in the bouncing pattern in such a way that the effect becomes perceptually consistent. Again, if the visual display can afford the same simplifications, the overall simulated phenomenon will be perceived as dynamic.[12]

The immediate benefit of simplified physics-based models is that they can run in real time in low-cost computers, and gestural or graphical interfaces can interactively control the models. Figure 1 shows the graphical Pure Data patch of a bouncing object, where sliders can control the size, elasticity, mass, and shape. (For more information about Pure Data, see http://www.puredata.org.) In particular, the shape slider allows continuous morphing between sphere and cube via superellipsoids and it controls the position of resonances and the statistical deviation from the regular temporal pattern of a bouncing ball.

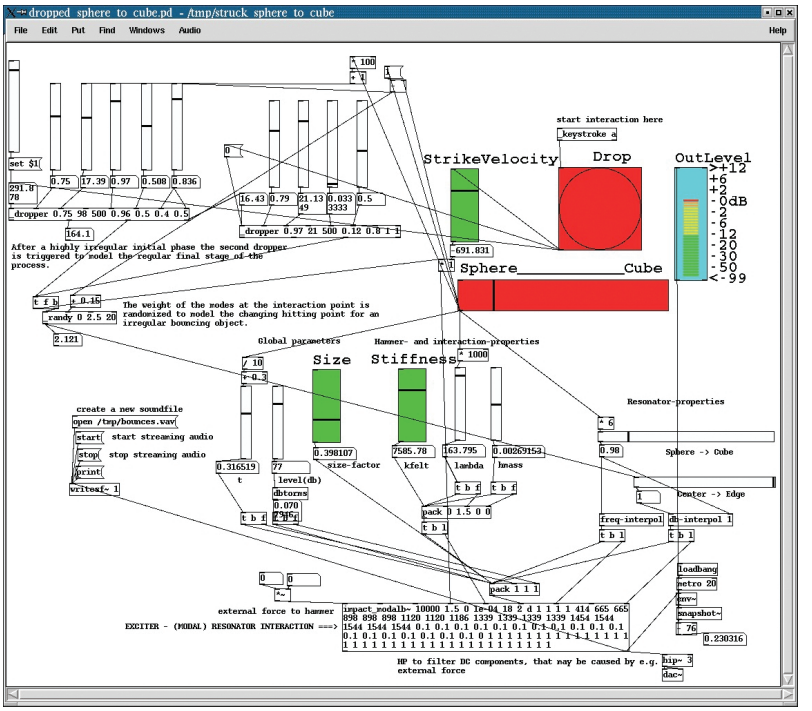

Figure 1. Pure Data patch for a bouncing object.

Figure 2. Maisy drinking with a straw. (Illustration from "Happy Birthday Maisy," ©1998 by Lucy Cousins. Reproduced by permission of the publisher Candlewick Press, Inc., Cambridge, Mass., on behalf of Walker Books Ltd., London.)

Drinking lemonade with a straw

In physical sciences it's customary to simplify physical phenomena for better understanding of first principles and to eliminate second-order effects. In computer-based communication, when the goal is to improve the clarity and effectiveness of information, the same approach turns out to be useful, especially when it's accompanied by exaggeration of selected features. This process is called *cartoonification*.[5]

Consider the cartoon in Figure 2, displaying a mouse-like character drinking lemonade with a straw. The illustration comes from a children's book, where children can move a carton flap

Sound control can be more straightforward if we generate sounds with physics-based techniques that give access to control parameters directly connected to sound source characteristics.

that changes the level of liquid in the glass. With the technology that's now available, it wouldn't be too difficult to augment the book with a small digital signal processor connected to sensors and to a small actuator, so that the action on the flap would also trigger a sound. Although many toys feature this, the sounds are almost inevitably prerecorded samples that, when played from low-cost actuators, sound irritating. Even in experimental electronically augmented books with continuous sensors, due to the intrinsic limitations of sampled sounds, control is usually limited to volume or playback speed.[13] The interaction would be much more effective if the sound of a glass being emptied is a side effect of a real-time simulation that responds continuously to the actions. Instead of trying to solve complex fluid-dynamic problems, we use a high-level analysis and synthesis of the physical phenomena as follows:

1. take one micro event that represents a small collapsing bubble;
2. arrange a statistical distribution of micro events;
3. filter the sound of step 2 through a mild resonant filter, representing the first resonance of the air cavity separating the liquid surface from the glass edge;
4. filter the sound of step 3 with a sharp comb filter (harmonic series of resonances obtained by a feedback delay line), representing the filtering effect of the straw; and
5. add abrupt noise bursts to mark the initial and final transients when the straw rapidly passes from being filled with air to being filled with liquid, and vice versa.

The example of the straw is extreme in the context of cartoon sound models, as in reality this physical phenomenon is usually almost silent. However, the sound model serves the purpose of signaling an activity and monitoring its progress. Moreover, because we built the model based on the actual physical process, the model integrates perfectly with the visual image and follows the user gestures seamlessly, thus resulting in far less irritating sounds.

On the importance of control

Sound synthesis techniques have achieved remarkable results in reproducing musical and everyday sounds. Unfortunately, most of these techniques focus only on the perfect synthesis of isolated sounds, thus neglecting the fact that most of the expressive content of sound messages comes from the appropriate articulation of sound event sequences. Depending on the sound synthesis technique, we must design a system to generate control functions for the synthesis engine in such a way that we can arrange single events in naturally sounding sequences.[14]

Sound control can be more straightforward if we generate sounds with physics-based techniques that give access to control parameters directly connected to sound source characteristics. In this way, the design of the control layer can focus on the physical gestures rather than relying on complex mapping strategies. In particular, music performance studies have analyzed musical gestures, and we can use this knowledge to control cartoon models of everyday sounds.

Control models

Recently, researchers have studied the relationship between music performance and body motion. Musicians use their body in a variety of ways to produce sound. Pianists use shoulders, arms, hands, and feet; trumpet players use their lungs and lips; and singers use their vocal chords, breathing system, phonatory system, and expressive body postures to render their interpretation. When playing an interleaved accent in drumming, percussionists prepare for the accented

stroke by raising the drumstick up higher, thus arriving at the striking point with larger velocity.

The expressiveness of sound rendering is largely conveyed by subtle but appropriate variations of the control parameters that result from a complex mixture of intentional acts and constraints of the human body's dynamics. Thirty years of research on music performance at the Royal Institute of Technology (KTH) in Stockholm has resulted in about 30 so-called performance rules. These rules allow reproduction and simulation of different aspects of the expressive rendering of a music score. Researchers have demonstrated that they can combine rules and set them up in such a way that generates emotionally different renderings of the same piece of music. The results from experiments with expressive rendering showed that in music performance, emotional coloring corresponds to an enhanced musical structure. We can say the same thing about hyper- and hypo-articulation in speech—the quality and quantity of vowels and consonants vary with the speaker's emotional state or the intended emotional communication. Yet, the phrase structure and meaning of the speech remain unchanged. In particular, we can render emotions in music and speech by controlling only a few acoustic cues.[15] For example, email users do this visually through emoticons such as ;-). Therefore, we can produce cartoon sounds by simplifying physics-based models and controlling their parameters.

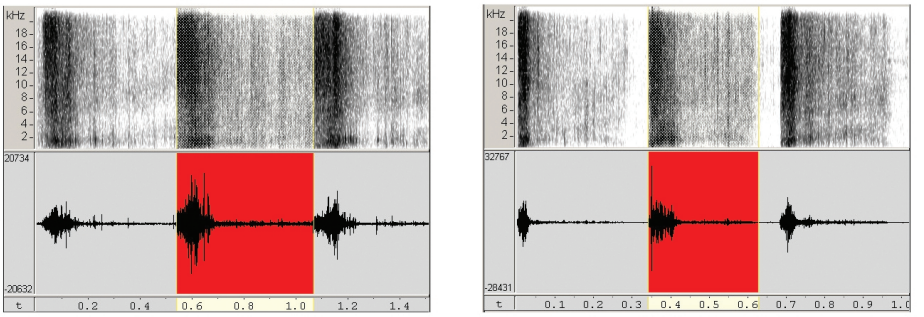

Figure 3. Waveform and spectrogram of walking and running sounds on gravel. (a) Three adjacent walking gaits overlap in time. The red rectangle highlights the time interval between the onset time of two adjacent steps. (b) Vertical white strips (micropauses) correspond to flight time between adjacent running gaits. The red rectangle highlights the time interval between the onset and offset times of one step.

Walking and running

Music is essentially a temporal organization of sound events along short and long time scales. Accordingly, microlevel and macrolevel rules span different time lengths. Examples of the first class of rules include the Score Legato Articulation rule, which realizes the acoustical overlap between adjacent notes marked legato in the score, and the Score Staccato Articulation rule, which renders notes marked staccato in the score. Final Retard is a macrolevel rule that realizes the final ritardando typical in Baroque music.[16] Friberg and Sundberg demonstrated how their model of final ritardando was derived from measurements of stopping runners. Recently, researchers discovered analogies in timing between walking and legato and running and staccato.[17] These findings show the interrelationship between human locomotion and music performance in terms of tempo control and timing.

Friberg et al.[18] recently studied the association of music with motion. They transferred measurements of the ground reaction force by the foot during different gaits to sound by using the vertical force curve as sound-level envelopes for tones played at different tempos. Results from listening tests were consistent and indicated that each tone (corresponding to a particular gait) could clearly be categorized in terms of motion. These analogies between locomotion and music performance open new possibilities for designing control models for artificial walking sound patterns and for sound control models based on locomotion.

We used the control model for humanized walking to control the timing of the sound of one step of a person walking on gravel. We used the Score Legato Articulation and Phrase Arch rules to control the timing of sound samples. We've observed that in walking there's an overlap time between any two adjacent footsteps (see Figure 3) and that the tendency in overlap time is the same as observed between adjacent notes in piano playing: the overlap time increases with the time interval between the two events. This justifies using the Score Legato Articulation rule for walking. The Phrase Arch rule used in music performance renders *accelerandi* and *rallentandi*, which are a temporary increase or decrease of the beat rate. This rule is modeled according to velocity changes in hand movements between two fixed points on a plane. We thought that the

Web Extras

Visit *IEEE MultiMedia*'s Web site at http://www.computer.org/multimedia/mu2003/u2toc.htm to view the following examples that we designed to express the urgency of dynamic sound models in interactive systems:

- Physical animation (AVI video) of simple geometric forms. The addition of consistent friction sounds (generated from the same animation) largely contributes to construct a mental representation of the scene and to elicit a sense of effort as experienced when rubbing one object onto another.
- Sound example (Ogg Vorbis file) of gradual morphing between a bouncing sphere and a bouncing cube. It uses the cartoon model described in the "Cartoon sound models" section.
- Sound example (Ogg Vorbis file) of Maisy drinking several kinds of beverages at different rates and from different glasses. We based this on the cartoon model described in the "Cartoon sound models" section.
- Sound example (Ogg Vorbis file) of an animated walking sequence constructed from two recorded step samples (gravel floor) and one that uses a dynamic cartoon model of crumpling (resembles steps on a crispy snow floor). The rules described in the "On the importance of control" section govern this animation.

Players for the Ogg Vorbis files are available at http://www.vorbis.com for Unix, Windows, Macintosh, BeOS, and PS2.

Phrase Arch rule would help us control walking tempo changes on a larger time scale.

Human running resembles staccato articulation in piano performance, as the flight time in running (visible as a vertical white strip in the spectrograms in Figure 3) can play the same role as the key-detach time in piano playing. We implemented the control model for stopping runners by applying the Final Retard rule to the tempo changes of sequences of running steps on gravel.

We implemented a model for humanized walking and one for stopping runners as Pure Data patches. Both patches allow controlling the tempo and timing of sequences of simple sound events. We conducted a listening test comparing step sound sequences without control to sequences rendered by the control models presented here. The results showed that subjects preferred the rendered sequences and labeled them as more natural, and they correctly classified different types of motion produced by the models.[4] Recently, we applied these control models to physics-based sound models (see the "Web Extras" sidebar for a description of these sound examples available at http://computer.org/multimedia/mu2003/u2toc.htm).

The proposed rule-based approach for sound control is only a step toward the design of more general control models that respond to physical gestures. In our research consortium, we're applying the results of studies on the expressive gestures of percussionists to impact-sound models[19] and the observations on the expressive character of friction phenomena to friction sound models. Impacts and frictions deserve special attention because they're ubiquitous in everyday soundscapes and likely to play a central role in sound-based human–computer interfaces.

New interfaces with sound

Because sound can enhance interaction, we need to explore ways of creating and testing suitable auditory metaphors. One of the problems with sounds used in auditory interfaces today is that they always sound the same, so we need to have parametric control. Then, small events can make small sounds; big events make big sounds; the effort of an action can be heard; and so on. As Truax[20] pointed out, before the development of sound equipment in the 20th century, nobody had ever heard exactly the same sound twice. He also noted that fixed waveforms—as used in simple synthesis algorithms—sound unnatural, lifeless, and annoying. Hence, with parametric control and real-time synthesis of sounds, we can get rich continuous auditory representation in direct manipulation interfaces. For instance, a simple audiovisual animation in the spirit of Michotte's famous experiments[6] (see the geometric forms animation in the "Web Extras" sidebar) shows how sound can elicit a sense of effort and enhance the perceived causality of two moving objects.

Design

Barrass[21] proposed a rigorous approach to sound design for multimedia applications through his Timbre-Brightness-Pitch Information Sound Space (TBP ISS). Similar to the Hue-Saturation-Lightness model used for color selection, sounds are arranged in a cylinder, where the radial and longitudinal dimensions are bound to brightness and pitch, respectively. The timbral (or identity) dimension of the model is restricted to a small collection of musical sound samples uniformly distributed along a circle derived from experiments in timbre categorization. A pitfall of this model is that it's not clear how to enrich the palette of timbres, especially for everyday sounds. Recent investigations found that perceived sounds tend to cluster based on shared physical properties.[22] This proves beneficial for structuring sound information spaces because there are few sound-producing fundamental mechanisms as compared to all possible sounding objects. For instance, a large variety of sounds (bouncing, breaking, scraping, and so on) can be based on the same basic impact model.[12] Bezzi, DePoli, and Rocchesso,[8] proposed the following three-layer sound design architecture:

- *Identity*: A set of physics-based blocks connected in prescribed ways to give rise to a large variety of sound generators.
- *Quality*: A set of signal-processing devices (digital audio effects[19]) specifically designed to modify the sound quality.
- *Spatial organization*: Algorithms for reverberation, spatialization, and changes in sound-source position and size.[9]

Such a layered architecture could help design a future sound authoring tool, with the possible addition of tools for spatio-temporal texturing and expressive parametric control.

Organization and access

The complexity of the sonic palette calls for new methods for browsing, clustering, and visualizing dynamic sound models. The Sonic Browser (see Figure 4) uses the human hearing system's streaming capabilities to speed up the browsing in large databases. Brazil et al.[23] initially developed it as a tool that allowed interactive visualization and direct sonification. In the starfield display in Figure 4, each visual object represents a sound. Users can arbitrarily map the location, color, size, and geometry of each visual object to properties of the represented objects. They can hear all sounds covered by an aura simultaneously, spatialized in a stereo space around the cursor (the aura's center). With the Sonic Browser, users can find target sounds up to 28 percent faster, using multiple stream audio, than with single stream audio.

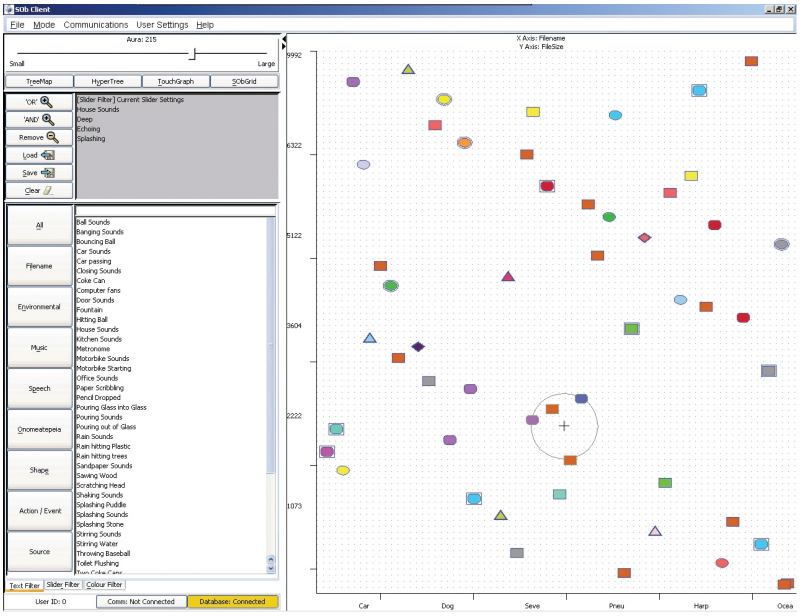

Figure 4. Sonic browser starfield display of a sound file directory. Four sound files are concurrently played and spatialized because they're located under the circular aura (cursor).

In the context of our current research, we use the Sonic Browser for two different purposes. First, it lets us validate our sound models, as we can ask users to sort sounds by moving visual representations according to auditory perceptual dimensions on screen. If the models align with real sounds, we consider the models valid. Second, the Sonic Browser helps users compose auditory interfaces by letting them interactively access large sets of sounds—that is, to pick, group, and choose what sounds might work together in an auditory interface.

Manipulation and control

Real-time dynamic sound models with parametric control, while powerful tools in the hands of the interface designer, create important problems if direct manipulation is the principal control strategy. Namely, traditional interface devices (mouse, keyboards, joysticks, and so on) can only exploit a fraction of the richness that

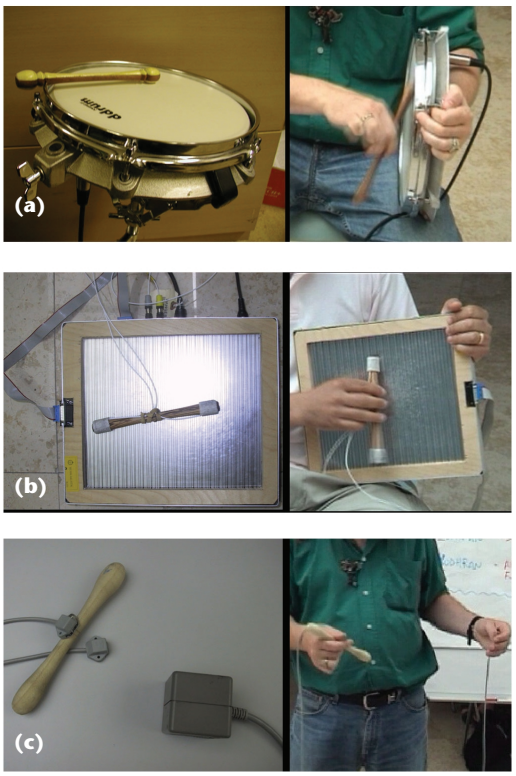

Figure 5. Three controllers for the Vodhran: (a) Clavia ddrum4, (b) Max Mathew's Radio Baton, and (c) Polhemus Fastrack.

emerges from directly manipulating sounding objects. Therefore, we need to explore alternative devices to increase the naturalness and effectiveness of sound manipulation.

The Vodhran

One of the sound manipulation interfaces that we designed in the Sounding Object project is based on a traditional Irish percussion instrument called the bodhran. It's a frame drum played with a double-sided wood drumstick in hand A while hand B, on the other side of the drum, damps the drumhead to emphasize different modes. This simple instrument allows a wide range of expression because of the richness in human gesture. Either the finger or palm of hand B can perform the damping action, or increase the tension of the drumhead, to change the instrument's pitch. Users can beat the drumstick on different locations of the drumhead, thus generating different resonance modes. The drumstick's macrotemporal behavior is normally expressed in varying combinations of duplets or triplets, with different accentuation. We tuned our impact model to reproduce the timbral characteristics of the bodhran and give access to all its controlling parameters such as resonance frequency, damping, mass of the drumstick, and impact velocity.

To implement a virtual bodhran, the Vodhran, we used three devices that differ in control and interaction possibilities—a drumpad, a radio-based controller, and a magnetic tracker system (see Figure 5). Each device connects to a computer running Pure Data with the impact model used in Figure 1.

Drumpad controller. For the first controlling device, we tested a Clavia ddrum4 drumpad (http://www.clavia.se/ddrum/), which reproduces sampled sounds. We used it as a controller to feed striking velocity and damp values into the physical model. The ddrum4 is a nice interface to play the model because of its tactile feedback and the lack of cables for the drumsticks.

Radio-based controller. We used Max Mathew's Radio Baton[24] to enhance the Vodhran's control capabilities. The Radio Baton is a control device comprised of a receiving unit and an antenna that detects the 3D position of one or two sticks in the space over it. Each of the sticks sends a radio signal. For the Vodhran, we converted the two sticks into two radio transmitters at each end of a bodhran drumstick, and played the antenna with the drumstick as a normal bodhran. The drumstick's position relative to the antenna controlled the physical model's impact position, thus allowing a real-time control of the instrument's timbral characteristics. Professional player Sandra Joyce played this version of the bodhran in a live concert. She found that it was easy to use and that it had new expressive possibilities compared to the traditional acoustical instrument.

Tracker-based controller. With the Polhemus Fastrack, we can attach the sensors to users' hands or objects handled by them, so that they can directly manipulate tangible objects of any kind with the bodhran gestural metaphor. During several design sessions, leading up to the public performance in June 2002, we evaluated numerous sensor configurations. With a virtual instrument, we could make the instrument bodycentric or geocentric. A real bodhran is bodycentric because of the way it's held. The configurations were quite different. We felt a bodycentric reference was easier to play,

although the geocentric configuration yielded a wider range of expression. In the Vodhran's final public-performance configuration, we attached one Polhemus sensor to a bodhran drumstick (held in the right hand), and the player held the second sensor in her left hand. We chose the geocentric configuration and placed a visual reference point on the floor in front of the player. We used normal bodhran playing gestures with the drumstick in the player's right hand to excite the sound model. The distance between the hands controlled the virtual drumhead's tension, and the angle of the left hand controlled the damping. Virtual impacts from the player's hand gestures were detected in a vertical plane extending in front of the player. We calibrated this vertical plane so that if the player made a virtual impact gesture with a fully extended arm or with the hand close to the player's body, the sound model's membrane was excited near its rim. If the player made a virtual impact gesture with the arm half-extended, the sound model was excited at the center of the membrane.

Conclusion

An aesthetic mismatch exists between the rich, complex, and informative soundscapes in which mammals have evolved and the poor and annoying sounds of contemporary life in today's information society. As computers with multimedia capabilities are becoming ubiquitous and embedded in everyday objects, it's time to consider how we should design auditory displays to improve our quality of life.

Physics-based sound models might well provide the basic blocks for sound generation, as they exhibit natural and realistically varying dynamics. Much work remains to devise efficient and effective models for generating and controlling relevant sonic phenomena. Computer scientists and acousticians have to collaborate with experimental psychologists to understand what phenomena are relevant and to evaluate the models' effectiveness. Our work on Sounding Objects initiated such a long-term research agenda and provided an initial core of usable models and techniques.

While continuing our basic research efforts, we are working to exploit the Sounding Objects in several applications where expressive sound communication may play a key role, especially in those contexts where visual displays are problematic. **MM**

Acknowledgments

The European Commission's Future and Emergent Technologies collaborative R&D program under contract IST-2000-25287 supported this work. We're grateful to all researchers involved in the Sounding Object project. For the scope of this article, we'd like to acknowledge the work of Matthias Rath, Federico Fontana, and Federico Avanzini for sound modeling; Eoin Brazil for the Sonic Browser; Mark Marshall and Sofia Dahl for the Vodhran. We thank Clavia AB for lending us a ddrum4 system and Max Mathews for providing the Radio Baton.

References

1. C. Roads, *The Computer Music Tutorial*, MIT Press, 1996.
2. D. Rocchesso et al., "The Sounding Object," *IEEE Computer Graphics and Applications*, CD-ROM supplement, vol. 22, no. 4, July/Aug. 2002.
3. W.M. Hartmann, *Signals, Sound, and Sensation*, Springer-Verlag, 1998.
4. G.B. Vicario et al., *Phenomenology of Sounding Objects*, Sounding Object project consortium report, 2001, http://www.soundobject.org/papers/deliv4.pdf.
5. W.W. Gaver, "How Do We Hear in the World? Explorations in Ecological Acoustics," *Ecological Psychology*, vol. 5, no. 4, Sept. 1993, pp. 285-313.
6. C. Ware, *Information Visualization : Perception for Design*, Morgan-Kaufmann, 2000.
7. P.J. Stappers, W. Gaver, and K. Overbeeke, "Beyond the Limits of Real-Time Realism—Moving from Stimulation Correspondence to Information Correspondence," *Psychological Issues in the Design and Use of Virtual and Adaptive Environments*, L. Hettinger and M. Haas, eds., Lawrence Erlbaum, 2003.
8. M. Bezzi, G. De Poli, and D. Rocchesso, "Sound Authoring Tools for Future Multimedia Systems," *Proc. IEEE Int'l Conf. Multimedia Computing and Systems*, vol. 2, IEEE CS Press, 1999, pp. 512-517.
9. U. Zölzer, ed., *Digital Audio Effects*, John Wiley and Sons, 2002.
10. J.F. O'Brien, P.R. Cook, and G. Essl, "Synthesizing Sounds from Physically Based Motion," *Proc. Siggraph 2001*, ACM Press, 2001, pp. 529-536.
11. K. van den Doel, P.G. Kry, and D.K. Pai, "Foley Automatic: Physically-Based Sound Effects for Interactive Simulation and Animation," *Proc. Siggraph 2001*, ACM Press, 2001, pp. 537–544.
12. D. Rocchesso et al., *Models and Algorithms for Sounding Objects*, Sounding Object project consortium, 2002, http://www.soundobject.org/papers/deliv6.pdf.

13. M. Back et al., "Listen Reader: An Electronically Augmented Paper-Based Book," *Proc. CHI 2001*, ACM Press, 2001, pp. 23-29.
14. R. Dannenberg and I. Derenyi, "Combining Instrument and Performance Models for High-Quality Music Synthesis," *J. New Music Research*, vol. 27, no. 3, Sept. 1998, pp. 211-238.
15. R. Bresin and A. Friberg, "Emotional Coloring of Computer Controlled Music Performance," *Computer Music J.*, vol. 24, no. 4, Winter 2000, pp. 44-62.
16. A. Friberg and J. Sundberg, "Does Music Performance Allude to Locomotion? A Model of Final Ritardandi Derived from Measurements of Stopping Runners," *J. Acoustical Soc. Amer.*, vol. 105, no. 3, March 1999, pp. 1469-1484.
17. R. Bresin, A. Friberg, and S. Dahl, "Toward a New Model for Sound Control," *Proc. COST-G6 Conf. Digital Audio Effects* (DAFx01), Univ. of Limerick, 2001, pp. 45-49, http://www.csis.ul.ie/dafx01.
18. A. Friberg, J. Sundberg, and L. Frydén, "Music from Motion: Sound Level Envelopes of Tones Expressing Human Locomotion," *J. New Music Research*, vol. 29, no. 3, Sept. 2000, pp. 199-210.
19. R. Bresin et al., *Models and Algorithms for Control of Sounding Objects*, Sounding Object project consortium report, 2002, http://www.soundobject.org/papers/deliv8.pdf.
20. B. Truax, *Acoustic Communication*, Ablex Publishing, 1984.
21. S. Barrass, "Sculpting a Sound Space with Information Properties," *Organised Sound*, vol. 1, no. 2, Aug. 1996, pp. 125-136.
22. J.M. Hajda et al., "Methodological Issues in Timbre Research," *Perception and Cognition of Music*, J. Deliege and J. Sloboda, eds., Psychology Press, 1997, pp. 253-306.
23. E. Brazil et al., "Enhancing Sonic Browsing Using Audio Information Retrieval," *Proc. Int'l Conf. Auditory Display*, ICAD, 2002, pp. 132-135.
24. J. Paradiso, "Electronic Music Interfaces," *IEEE Spectrum*, vol. 34, no. 12, 1997, pp. 18-30.

Davide Rocchesso is an associate professor in the Computer Science Department at the University of Verona, Italy. His research interests include sound processing, physical modeling, and human–computer interaction. Rocchesso received a PhD in computer engineering from the University of Padova, Italy.

Roberto Bresin is a research assistant at the Royal Institute of Technology (KTH), Stockholm. His main research interests are in automatic music performance, affective computing, sound control, visual feedback, music perception, and music education. Bresin received an MSc in electrical engineering from the University of Padova, Italy, and a PhD in music acoustics from KTH.

Mikael Fernström is director of the master's program in interactive media at the University of Limerick, Ireland. His main research interests include interaction design, sound, music, and electronics. Fernström studied electronic engineering and telecommunications at the Kattegatt technical college in Halmstad, Sweden, and has an MSc in human–computer interaction from the University of Limerick.

Readers may contact Davide Rocchesso at the Computer Science Dept., Univ. of Verona, Strada Le Grazie 15, 37134 Verona, Italy; Davide.Rocchesso@univr.it.