RECOGNITION OF ELLIPSOIDS FROM ACOUSTIC CUES

Federico Fontana, Laura Ottaviani, Matthias Rath and Davide Rocchesso

University of Verona
Dipartimento di Informatica
strada Le Grazie, 14
37134 Verona - Italy
{fontana,ottaviani,rath,rocchesso}@sci.univr.it

ABSTRACT

Ideal three-dimensional resonators are "labeled" (identified) by infinite sequences of resonance modes, whose distribution depends on the resonator shape. We are investigating the ability of human beings to recognize these shapes by auditory spectral cues. Rather than focusing on a precise simulation of the resonator, we want to understand if the recognition takes place using simplified "cartoon" models, just providing the first resonances that identify a shape. In fact, such models can be easily translated into efficient algorithms for real-time sound synthesis in contexts of human-machine interaction, where the resonator shape and other rendering parameters can be interactively manipulated. This paper describes the method we have followed to come up with an application that, executed in real-time, can be used in listening tests of shape recognition and together with human-computer interfaces.

1. INTRODUCTION

Recently, research topics in the field of psychophysics have been concerned with the faculty of human beings to *hear the shape*, both in the two-dimensional (2-D) and three-dimensional (3-D) case [1]. This means, for example, that sounds coming from square rather than circular membranes after an excitation, or resonances that are produced by cubic rather than spherical empty cavities, containing a sound source in their interior, may convey to the listener cues accounting for the shape of the resonator.

Several results [1, 2] seem to testify that, in the case of 3-D shapes, a fundamental role in this type of recognition is played by the spectral content of the sounds. Since, in the case of ideal resonators, the sequence of resonance modes depends only on the resonator shape, it makes sense hyphotesizing that 3-D resonators convey perceptually relevant cues that are in strong correlation with their shapes.

The way these cues are perceived by the listener is a matter of investigation for ecological psychologists [3]. We decided to focus our attention in the spectral *mode series* such as a *label* of the resonator, meanwhile taking care of preserving as far as possible the constancy of all other physical and geometrical parameters. In particular, a variation of the resonator size leads to a proportional shift in frequency of the mode series: a solution must be found to govern those shifts during changes in shape of the resonator or, in other words, a rule to infer the size must be found once a shape is given.

The simplest idea would be constraining the fundamental mode to a unique value during changes in shape. This approach is quite non-ecological. Rather, a psychophysically more well-founded rule suggests to *preserve the constancy of the resonator volume* during changes in shape [4]: following this approach, the fundamental mode shift is minimal for intermediate shapes between the cube and the sphere.

A rule is needed to map one or more *morphing* parameters into correspondent shapes. Superquadrics [5] have been adopted here to realize direct and versatile maps: using just one parameter, geometries that are consistent with the problem can be selected via a simple set of equations. These geometries can be used to initialize models of resonators, if these models can be directly "shaped" exactly like the resonators should be. Waveguide meshes [6, 7] comply with this requirement, and represent a good modeling solution, with several pros and cons that have been explained in more detail in previous literature [8].

In particular, Waveguide meshes allow to select the excitation and acquisition points in the resonators. In this way, the mode series can be detected in regions where the modal density is particularly rich, and, conversely, in regions that are *nodal* with respect to many resonances [9]. Since our investigation needs only to deal with the first part of the mode series (i.e., few tens of modes), the waveguide models allow to assess with enough precision all the resonances that are present in the mode series: this is done detecting the signal on acquisition points that are located near the corners of the resonator. Alternatively, for reasons of symmetry and on a practical and ecologically-consistent basis, the center has been chosen such as the region where only some resonances are audible.

Smooth changes in shape using the morphing parameter translate into progressive changes in the resonances positions. In the meantime, smooth changes in the acquisition point, moving between the center and the corner of the resonator, translate into correspondent variations in amplitude of the resonance peaks. Hence, the controls of shape and position map into intuitive features of the frequency responses.

All these features can be easily reproduced using a filter bank of second-order filters, where each filter is tuned to one particular resonance frequency [10]. Moreover, this filter bank has a precise and physically meaningful interpretation [11]. Simple methods like linear interpolation can be used to interpolate between responses that have not been simulated.

We have developed a pd-module [12] that implements such a filter bank. It is controlled in the parameters of shape (between sphere and cube), and listening points (between center and corner). Using this module, proper sounds can be convolved as if they were listened from a point located in a 3-D cavity having a given shape. This module realizes a so-called *cartoon* model [13], that can be

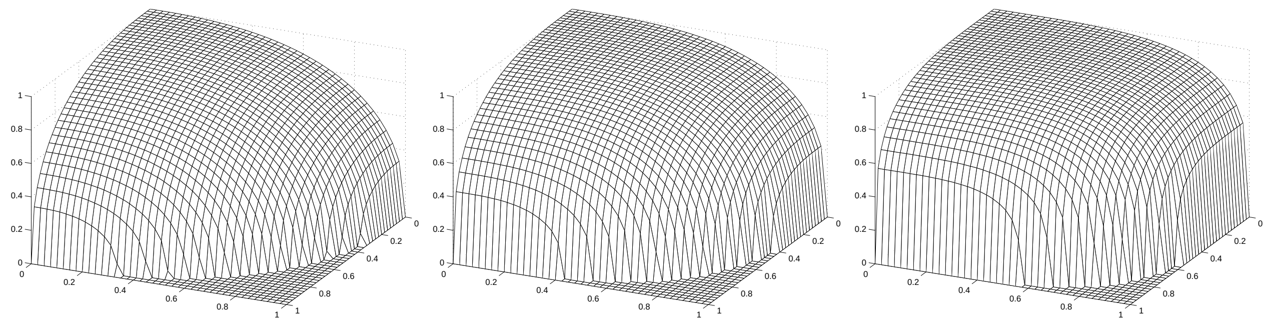

Figure 1: Positive sections of geometries obtained using (1). γ has been set to $\{2.5\ ,\ 3\ ,\ 4\}$, starting from left.

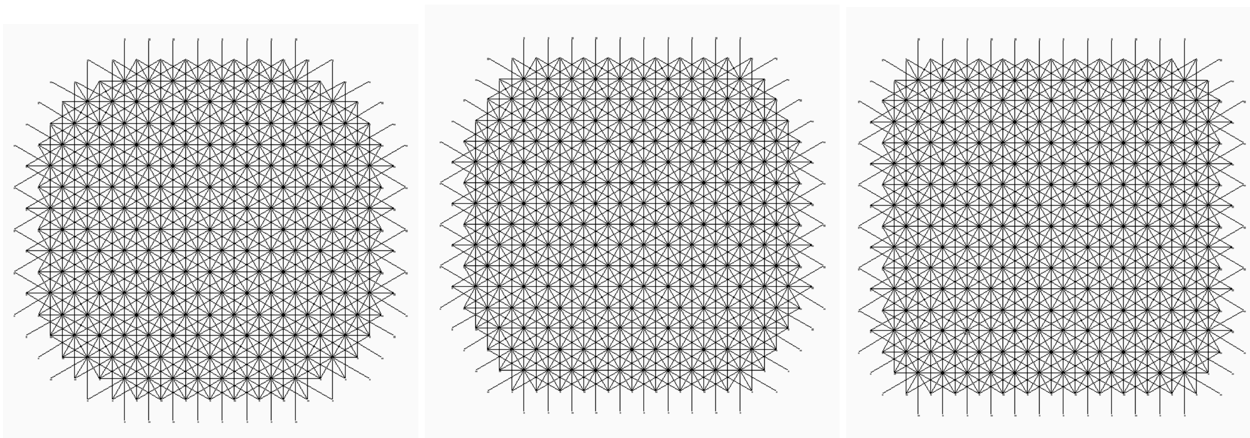

Figure 2: Orthogonal projections of Waveguide meshes modeling the geometries seen in figure 1. γ has been set to $\{2.5\ ,\ 3\ ,\ 4\}$, starting from left.

used in interactive real-time environments. In particular, we are going to use it in listening tests of shape recognition.

2. GEOMETRIES

One possible morphing from spheres to cubes can be realized if we constrain the geometries to be ellipsoids. Superquadrics are, in this sense, a versatile family of ellipsoids. For our purpose, we restrict their use to a set of geometries that is defined by the following equation in the 3-D space [5]:

$$|x|^{\gamma} + |y|^{\gamma} + |z|^{\gamma} = 1 \tag{1}$$

Changes in shape are simply determined by varying γ, that acts like a morphing parameter:

- sphere: $\gamma = 2$
- ellipsoid between sphere and cube: $2 < \gamma < \infty$
- cube: $\gamma \rightarrow \infty$.

We have analyzed eight shapes, including the sphere and the cube, which have been built according to (1). The correspondent values of the morphing parameters are the following ones:

$$\begin{array}{llll} \gamma_1 = 2 & \gamma_2 = 2.2 & \gamma_3 = 2.5 & \gamma_4 = 3 \\ \gamma_5 = 4 & \gamma_6 = 6 & \gamma_7 = 20 & \gamma_8 = 100 \end{array}$$

γ_8 results in a geometry that approximates the cube with good precision. Positive sections (i.e., volumes limited to $\{(x, y, z) : x > 0\ ,\ y > 0\ ,\ z > 0\}$) of some of these geometries (γ_3, γ_4 and γ_5) are depicted in figure 1, starting from left.

Then, 3-D Waveguide mesh models have been designed, in such a way that they reproduce ideal resonators having a shape that matches, as close as possible, the geometry given by the correspondent superquadric. Figure 2 depicts, for the same geometries and with the same ordering seen in figure 1, orthogonal projections of the mesh models that have been used in this context. The reader can find further details of the modeling strategies that have been adopted in this research in a previous paper [8].

3. SIMULATIONS

For each chosen geometry two impulse responses have been computed. The resonator model was fed with energy in a way that all the modes in the scope of our analysis were excited[1]. The responses were acquired from junctions located near the corner and at the center. In this way two mode series were collected from each shape, one accounting for all the modes that a resonator can

[1] in practice this required to feed several junctions of the mesh with an impulse.

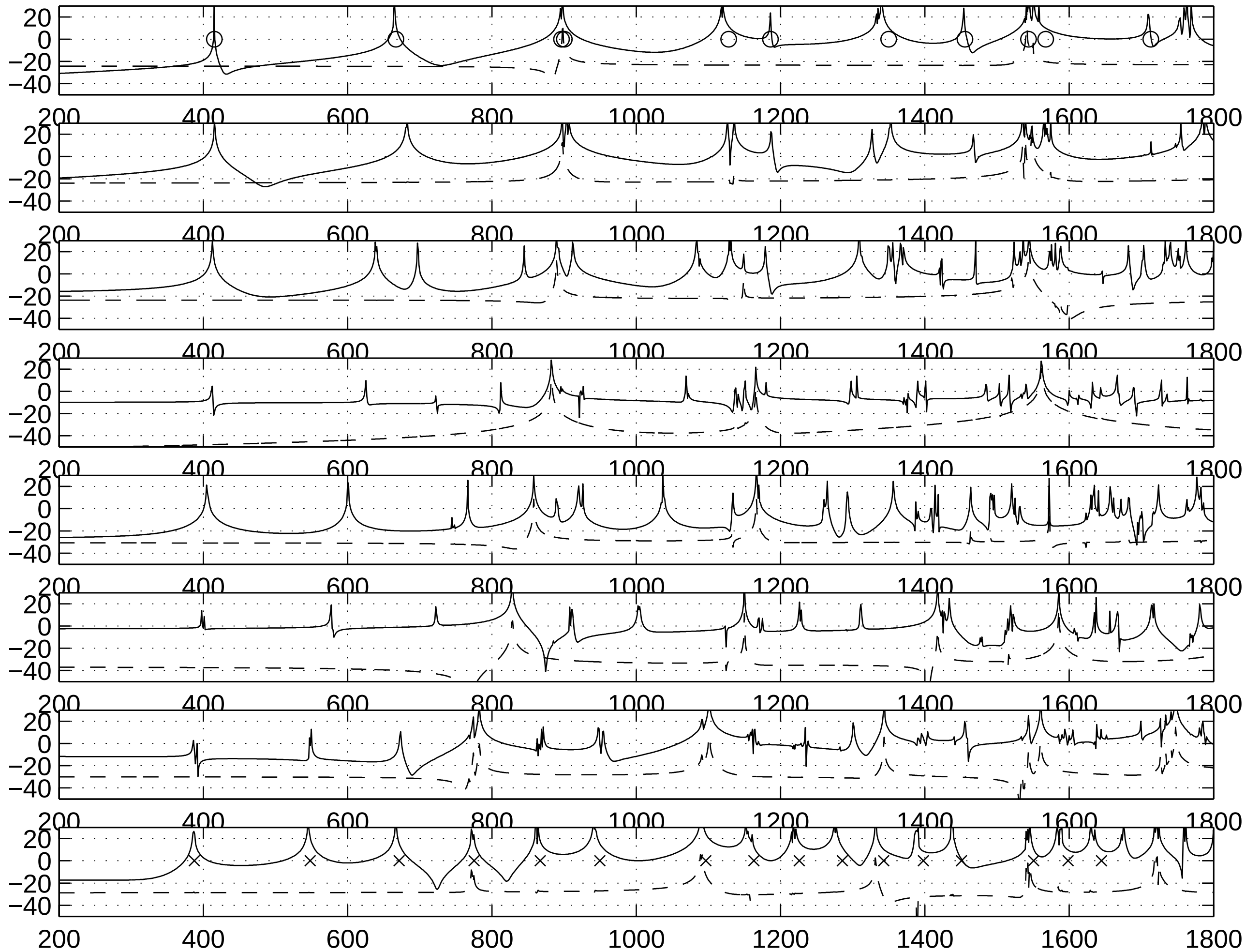

Figure 3: *Plots of frequency responses from resonators defined by morphing parameters $\gamma_1, \ldots, \gamma_8$, starting from above. Solid line: acquisition near the boundary; dashed line: acquisition at the center. All frequency domains are in Hz, gains are in dB. Theoretical positions of the resonances in the sphere are depicted with '$\circ$'; theoretical positions of the resonances in the cube are depicted with '$\times$'. All resonators resized to maintain the volume constant with shape. Nominal frequency of the fundamental mode in the sphere has been set to 415 Hz.*

produce in the lower part of the frequency domain, the other presenting only the modes resonating at the center of a 3-D ideal resonator, respectively.

Since changes, both in shape and in the acquisition position, do not introduce discontinuities, the correspondent responses exhibit smooth and continuous variations as well. Changes in the acquisition point result in mode cancellations that depend on the nodal regions falling on that point. Such cancellations are present in the spectra in the form of missing resonances.

The effect of changes in shape is more complicate: they result not only in shifts of the modes, but also in resonance splits and merging. By this phenomenon, the higher mode density in the case of the cube turns to be possible.

Figure 3 depicts all the responses that have been calculated in our analysis, both acquiring the signal at the center (dashed line) and at the boundary (solid line). Parameters $\gamma_1, \ldots, \gamma_8$ are ordered starting from above. All frequency domains are in Hz, and gains are in dB. Each resonator has been resized to maintain the volume constant with shape.

The fundamental frequency value is uninfluential in our experience. It has been set in the sphere to a nominal value of 415 Hz. Other fundamental frequencies are constrained by volume constancy: in practice, they slightly move with shape toward a lower frequency [14].

Mode canceling due to changes in the acquisition point are evident for all the geometries. It can be interesting to notice that the canceled modes do not change with shape, so that the responses taken from the center exhibit an overall homogeneity of behavior. A definite homogeneity is quite evident also for the responses that are acquired near the boundary. Mode splitting and merging is figured out in particular focusing on the very first modes.

Finally, the theoretical mode positions (depicted in figure 3 with '$\circ$' for the sphere, and with '$\times$' for the cube), as they are calculated using analytical tools, match well with the resonances computed by the simulations. This suggests a correct use of the Waveguide mesh models in our research.

4. CARTOON MODELS

Transfer functions having magnitude plots such as the ones shown in figure 3 can be realized straightforwardly. The most versatile and probably best-known solution makes use of second-order tunable equalization filters [15]. In spite of this we have adopted, like in the case of the volume constancy rule seen in Section 3, a solution that has a more consistent physical background, although respecting the requirement of efficiency and versatility.

Given the first N resonance modes generated by an ideal resonator, we can reproduce them using a one-dimensional (1-D) physical system consisting of N elementary blocks in series, each one being made of one mass and one spring. A damper is added to each block to provide a lossy component, giving physical consistency and realism to the model. In this way each elementary block independently governs the correspondent mode. More precisely, the parameters of frequency position and decay time of a mode are computed by simple functions of the mass, the spring constant and the damping factor [11]. After discretization, we obtain a physical model of a 1-D resonator in the form of a parallel second-order filter bank, where each filter in the bank accounts for a single mode of the resonator.

The 1-D resonator is, so, a cartoon model of the cavity [13]. Although justified from a physical modeling viewpoint, this model also allows a quite straightforward control of the position and the amplitude of each mode. As seen in Section 3, these two parameters can be seen as resulting from a particular choice of the resonator shape and sound acquisition point. Hence, we can think to set up a rule that maps couples of (shape,position) into couples of (frequencies, amplitudes):

$$(\gamma, r) \longrightarrow (\boldsymbol{\omega}, \boldsymbol{G}) = ([\omega_1, \ldots, \omega_N], [G_1, G_N]) \qquad (2)$$

where r is the distance of the acquisition point from the center, measured along a pre-determined direction common for all shapes ($0 \leq r \leq R$, R being the distance between the boundary and the center), $\omega_i, i = 1 \ldots N$ are positions in frequency and $G_i, i = 1 \ldots N$ are gains of the N modes at the acquisition point; γ, of course, selects the shape ($\gamma_1 \leq \gamma \leq \gamma_8$).

A careful design of such a map would require the knowledge of several responses from each geometry, since the nodal regions combine in a wide variety over different acquisition points. Moreover, a precise reproduction of the modes may complicate the map expressed by (2) up to a point where a control in real time of the 1-D model could in principle become difficult. For this reason, we have “cartoonified” also the control layer, linearly interpolating between couples (γ, r) where the image $(\boldsymbol{\omega}, \boldsymbol{G})$ is not known.

Suppose to set the input parameters to (γ_s, r_s), such that the N mode positions and amplitudes, $(\boldsymbol{\omega}_s, \boldsymbol{G}_s)$, require interpolation. *Bi-linear* (Lagrange) interpolation requires to calculate $\boldsymbol{\omega}_s$ and $\boldsymbol{G}_s$ using relations that involve four interpolated points, where the mode positions and amplitude are known, and the distance between these points and the interpolation point. If (see figure 4) the interpolated points are respectively labeled with $(n, 0)$, (n, R), $(n+1, 0)$, $(n+1, R)$ (n is a number between 1 and 7), such relations become:

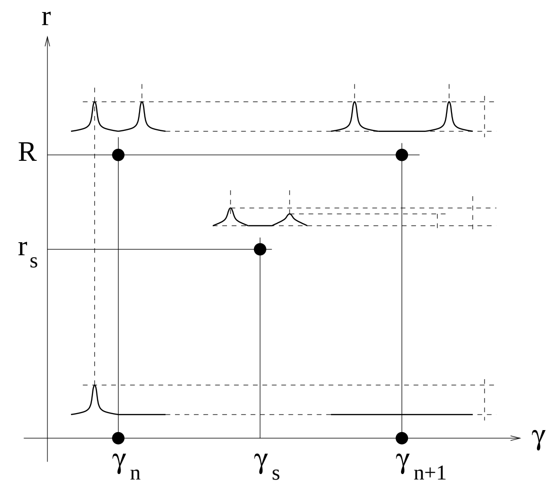

Figure 4: *Lagrange interpolation of gains and frequency positions.*

$$\begin{aligned}
\boldsymbol{\omega}_s &= \boldsymbol{\omega}_n + \frac{\gamma_s - \gamma_n}{\gamma_{n+1} - \gamma_n}(\boldsymbol{\omega}_{n+1} - \boldsymbol{\omega}_n) \\
\boldsymbol{G}_s &= \frac{R - r_s}{R}\frac{\gamma_{n+1} - \gamma_s}{\gamma_{n+1} - \gamma_n}\boldsymbol{G}_{n,0} + \frac{r_s}{R}\frac{\gamma_{n+1} - \gamma_s}{\gamma_{n+1} - \gamma_n}\boldsymbol{G}_{n,R} \\
&+ \frac{R - r_s}{R}\frac{\gamma_s - \gamma_n}{\gamma_{n+1} - \gamma_n}\boldsymbol{G}_{n+1,0} + \frac{r_s}{R}\frac{\gamma_s - \gamma_n}{\gamma_{n+1} - \gamma_n}\boldsymbol{G}_{n+1,R}
\end{aligned}$$

Note that bi-linear interpolation reduces to linear interpolation between two points in the first equation. In fact, mode positions are independent from the acquisition point.

5. MODEL IMPLEMENTATION

The described model has been implemented as a *pd module* under a PC running Linux and pd [12]. After acquiring a sound, it performs a convolution introducing the selected resonances in the audio signal. The simplicity of the module results in a low computational load for the hardware, so that enough resources are left to the graphic interface.

Figure 5 shows a snapshot of the module interface, that provides an interactive, real-time environment where the user can modify the parameters of shape and the acquisition point, receiving an immediate auditory feedback from the system.

This implementation is going to be used in listening tests of shape recognition. During these tests, particular care must be taken in the choice of the sound that will be processed by the filter bank, since poor or inadequate sounds can result in inconsistent tests. Scope of these tests is investigating if the auditory cues that are conveyed by our model can evoke sensations of shape, or improve the effectiveness of a multi-modal display devoted to convey information about 3-D shapes. For this reason, the model is going to be implemented also in a system for human-machine interaction involving a more sophisticated interface.

6. ACKNOWLEDGEMENTS

This research has been supported by the European Commission under contract IST-2000-25287 (“SOb - the Sounding Object”). `http://www.soundobject.org`.

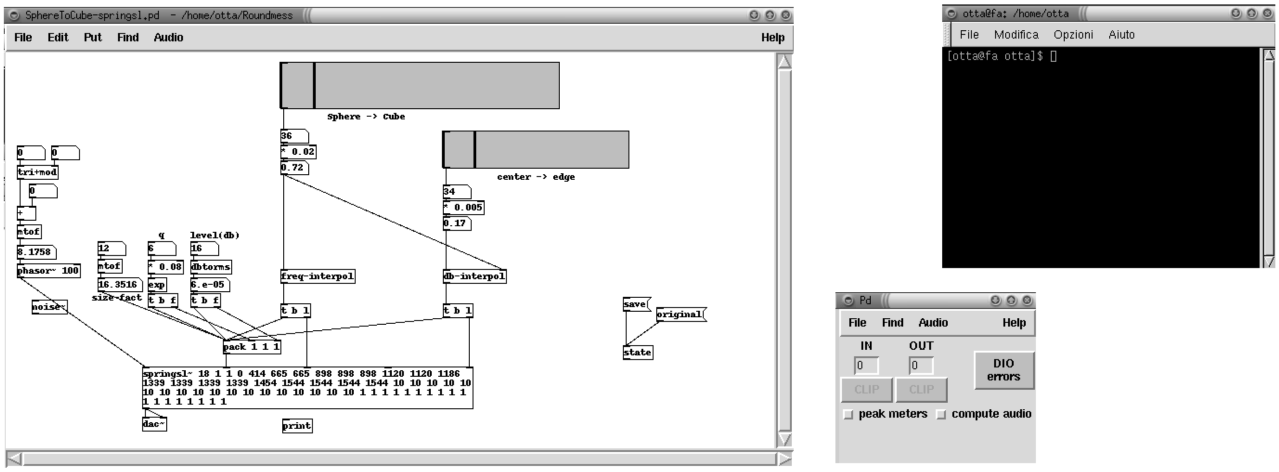

Figure 5: *Implementation of the model as a* pd module*: a snapshot of the graphic user interface.*

The Authors are grateful to Mr. Federico Avanzini for his helpful suggestions.

7. REFERENCES

[1] A. J. Kunkler-Peck and M. T. Turvey, "Hearing shape," *Journal of Experimental Psychology*, vol. 26, no. 1, pp. 279–294, 2000.

[2] D. Rocchesso, "Acoustic cues for 3-d shape information," in *Proc. of the 2001 Int. Conf. on Auditory Display*, July 2001, pp. 175–180.

[3] W. Gaver, "How do we hear in the world? explorations in ecological acoustics," *Ecological Psychology*, vol. 5, no. 4, pp. 285–313, Apr. 1993.

[4] D. Rocchesso and L. Ottaviani, "Can one hear the volume of a shape?," in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, In press.

[5] S. K. Kumar, S. H. Han, and K. Bowyer, "On recovering hyperquadrics from range data," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17, no. 11, pp. 1079–1083, Nov. 1995.

[6] S. A. Van Duyne and J. O. Smith, "Physical modeling with the 2-D digital waveguide mesh," in *Proc. Int. Computer Music Conf.*, Tokyo, Japan, 1993, ICMA, pp. 40–47.

[7] F. Fontana, D. Rocchesso, and E. Apollonio, "Using the waveguide mesh in modelling 3d resonators," in *Proc. Conf. on Digital Audio Effects (DAFX-00)*, Verona - Italy, Dec. 2000, COST-G6, pp. 229–232.

[8] F. Fontana, D. Rocchesso, and E. Apollonio, "Acoustic cues from shapes between spheres and cubes," in *Proc. Int. Computer Music Conf.*, La Habana, Cuba, Sept. 2001, In press.

[9] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer-Verlag, New York, 1991.

[10] S. K. Mitra, *Digital Signal Processing. A computer-Based Approach*, McGraw-Hill, New York, 1998.

[11] F. Avanzini and D. Rocchesso, "Modeling collision sounds: Non-linear contact force," in *Proc. Conference on Digital Audio Effects (DAFX-01)*, Limerick, Ireland, Dec. 2001, Elsewhere in these Proceedings.

[12] M. Puckette, "New public-domain realizations of standard pieces for instruments and live electronics," in *Proc. Int. Computer Music Conf.*, La Habana, Cuba, Sept. 2001, In press.

[13] W. Gaver, "Using and creating auditory icons," in *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, Ed., pp. 417–446. Addison-Wesley, 1994.

[14] D. Rocchesso and P. Dutilleux, "Generalization of a 3d resonator model for the simulation of spherical enclosures," *Applied Signal Processing*, vol. 2001, no. 1, pp. 15–26, 2001.

[15] P. A. Regalia and S. K. Mitra, "Tunable digital frequency response equalization filters," *IEEE Trans. on Speech and Audio Processing*, vol. ASSP-35, no. 1, pp. 118–120, Jan. 1987.