ACOUSTIC CUES FOR 3-D SHAPE INFORMATION

Davide Rocchesso

Dipartimento Scientifico e Tecnologico
Università degli Studi di Verona
Strada Le Grazie, 37134 - Verona, Italy
rocchesso@sci.univr.it

ABSTRACT

Three-dimensional resonators, such as cavities or rooms, affect the perceived timbral character of any sound source there enclosed. It is well understood how the resonator size, or the material of the enclosure, are conveyed to the listener by means of specific features of acoustic signals. On the other hand, the perception of the shape of a resonator is a much subtler issue that we investigate in this paper, taking the sphere and the cube as reference cases. The perceptual study is motivated by the availability of a compact resonator model whose parameters can be tuned to represent different shapes.

1. INTRODUCTION

"Can one hear the shape of a drum?" asked Mark Kac in 1966 [1]. His question about the existence of isospectral domains, indeed of profound mathematical nature, could be turned into the physical dilemma of knowing whether there are membranes of different geometries exhibiting the same eigenfrequencies. While it has been proved that such different membranes exist and can be constructively derived [2], turning the question into perceptual terms (as the original question seems to suggest) is still difficult. The reason of such difficulty is that there is no clear understanding of the acoustic features that convey a sense of shape to the listener. Furthermore, such sense of shape is not even firmly assessed.

Even though most of psychoacoustic research has focused on the perception of signal properties rather than source properties, the field of ecological psychology has recently been enriched by several studies looking at the physical and geometric properties of simple objects. In this paper we are interested in how spatial features of objects are conveyed to the listener by means of sound, so we restrict our attention to those prior works that focused on the same kind of features. One of the most important spatial features of an object is its size, and Carello et al [3] showed that listeners are able to scale the lengths of rods properly without any standard of comparison. This is not too surprising, since the pitch of a long and thin resonator is in simple relationship with its length, and our perception of pitch is very accurate. After assessing the accuracy in perception of 1-D lengths by sounds, the next natural question is wether we can give an estimate of 2-D properties of objects, such as length and height. An answer to this question was given by Lakatos et al. [4] for the rectangular cross section of struck bars, and by Kunkler-Peck and Turvey [5] for suspended rectangular plates. In the latter work two experiments of shape identification for non-rectangular plates were also reported, and the analyses showed that the distributions of modal frequencies are the structural information that are most likely to be used for shape identification.

Provided that shape information can be encoded into sounds, our interest is mainly focused on how to do such encoding. In other words, we would like to be able to control shape parameters in sound models. In this paper we focus on 3-D objects, as we already developed compact 3-D resonator models whose form factor can be controlled by simple parametric changes [6].

The paper is structured as follows. In section 2 we briefly recall the acoustics of spherical and cubic resonators, and explain the additive-synthesis models that have been implemented and used in subjective experimentation. In section 3 we describe the experiments conducted to test the relationship between pitch and volume, and the ability to identify the shape of a resonator as it impresses an acoustic signature onto sound sources. In section 4, a sound analysis based on auditory models is used to reveal interesting patterns that might be used by the hearing system to estimate the degree of "roundness" of a resonator. In section 5 we discuss how the results of subjective experimentation can be used to control the parameters of a versatile and compact resonator model, and how we can produce "cartoon" acoustic shapes, i.e. over-simplified models that still retain their shape features.

2. MODELS OF SPHERES AND CUBES FOR EXPERIMENTATION

An acoustic 3-D resonator is, with excellent approximation, a linear system and, therefore, it is thoroughly described by its impulse response or by its frequency domain counterpart, the frequency response. A rectangular resonator has a frequency response that is the superposition of harmonic combs, each having a fundamental frequency

$$f_0, lmn = \frac{c}{2}\sqrt{(l/X)^2 + (m/Y)^2 + (n/Z)^2}\ , \qquad (1)$$

where c is the speed of sound, l, m, n is a triple of positive integers with no common divisor, and X, Y, Z are the edge lengths of the box [8].

A spherical resonator has a frequency response that is the superposition of inharmonic combs, each having peaks at the extremal points of spherical Bessel functions. Namely, said z_{ns} the s^{th} root of the derivative of the n^{th} Bessel function, the resonance frequencies are found at

$$f_{ns} = \frac{c}{2\pi a} z_{ns}, \qquad (2)$$

where a is the radius of the sphere [7].

In [6] we extended the ball-within-a-box (BaBo) model [9] to provide a unified 3-D resonator model, based on a feedback delay network, that allows independent control of wall absorption, diffusion, size, and shape. Namely, the shape control is exerted by changing the parameters of allpass filters that are cascaded with the delay lines. In this way we can have a single computational structure that behaves like a rectangular box, or a sphere, or like an intermediate shape. The availability of such a model raised new questions about the perceptual significance of this shape control.

To investigate the perception of resonator shapes in an experimental framework, we prefer to construct impulse responses by additive synthesis rather then using the BaBo model, as the latter relies on some approximations whose significance has not been thoroughly assessed. The impulse response of a sphere or a rectangular box can be modeled by summing the contributions of exponentially damped sinusoids, each tuned at the position of a theoretical resonance frequency.

The additive-synthesis models for the sphere and the rectangular box have been implemented as MATLAB functions[1]. The functions have, as parameters, the size of the resonator, the material of its enclosure, the sample rate, and a flag that allows to introduce randomization of amplitude and/or phases in the damped sinusoids. This makes sense because, by changing the position of the source or the listener within the cavity, we apply different complex weighting at the eigenmodes. The material of the enclosure can be chosen from a small set of options, specified by frequency-dependent absorption curves. So far, we have only coded specific varieties of marble, wood, and drape.

3. SUBJECTIVE EXPERIMENTS

The main problem, in assessing the capabilities of humans in distinguishing shapes from acoustic signatures, is the dominance of pitch as a perceived feature. Pitch is associated with size, as smaller cavities tend to resonate at higher pitches, but it is not clear how to equalize the pitches of two different shapes. Indeed, 3-D shapes such as cubes and spheres have spectra that are far from harmonic, so that it is difficult to define mathematically what the perceived pitch is. The first step in our tests has the objective to understand how people compare the pitches of two impulse responses, one of the sphere and one of the cube. A first result from this preliminary stage is that we can correlate the perceived pitch with the volume of the object, a fact that was already noticed in [6]. In a second stage, we tried to measure human performance in a 3D-shape matching task. In all these experiments, subjects were volunteer computer science students and they listened to stimuli played through closed headphones (Beyerdynamic DT-770) at a comfortable level. In the first stage (pitch comparison) they listened to the impulse responses of the cavities. In the second stage (shape matching), they listened to the convolution of a complex sound with the impulse response of the cavities.

3.1. Pitch equalization of spheres and cubes

Both cubic and spherical resonators do not have harmonic or quasi-harmonic spectra. Therefore, it is not clear whether we can have a sensation of pitch and, if this is the case, what frequency value corresponds to such pitch. According to standard definitions in psychoacoustics, pitch is that acoustic attribute that allows to rank a sound as higher or lower than a reference. Therefore, the simplest way to know the pitch properties of 3-D cavities is to compare the impulse responses by means of the method of constant stimuli, i.e. by comparison of a fixed stimulus (namely, the spherical impulse response) with another stimulus (namely, the cubic impulse response) chosen from a finite set. For our purpose, this method is advantageous over up-down methods [10] because it is not clear a priori if pitch is well defined as a monotonic function of size. With the method of constant stimuli we can use a small set of precomputed responses and collect data about the pitch proximity of each couple of responses of different shapes.

To generate the impulse responses we used the additive-synthesis models described in section 2. In order to give the maximum richness to the stimuli, we set maximum excitation for all modes, and we randomized phases to avoid artifacts due to modal phase alignment. The decay time of each modal frequency was computed using the Sabine reverberation formula

$$T = 0.163 * V./(\alpha * A) \ , \tag{3}$$

where V is volume, A is surface area, and α is the absorption coefficient. The absorption curve was computed by interpolation between the following values, which can be considered as representative of a smooth wood-like enclosure:

$$f = [0, 125, 250, 500, 1000, 2000, 4000, F_s/2] \text{ Hz} \ ; \tag{4}$$
$$\alpha = [0.19, 0.15, 0.11, 0.10, 0.07, 0.06, 0.07, 1.00] \ , \tag{5}$$

and the sample rate was set to $F_s = 22050$Hz. We investigated sizes ranging from 30cm to 100cm in diameter. The use of Sabine formula might be criticized, especially for the range of sizes that we investigated. Indeed, using the Eyring formula or even exact computation of decay time does not make much difference for these values of surface absorption [11]. Moreover, it has been assessed that we are not very sensitive to variations in decay time [13], so we decided to use the simplest formula. This choice, together with the absorption coefficients that we chose, give quite a rich and long impulse response, even too much for a realistic wooden enclosure. However, for the purpose of this experiment it is definitely better to have rich responses so the ear has more chances to discover shape-related information.

Another question is what step size should be used to differentiate contiguous stimuli. We converted frequency JNDs, as found in psychoacoustic textbooks and measured for pure tones, into length differences:

$$\Delta l = c/\Delta f \ , \tag{6}$$

where c is the speed of sound in the cavity.

In a pre-experiment with 9 subjects, we chose a spatial step equal to twice Δl and we extended our range along 9 steps. The central size for the cubic box was chosen so that it has the same volume as the fixed comparison sphere, the latter having diameter $d = 36$cm. Since we found that most subjects could easily answer to the higher vs. lower question for most of the box sizes, we refined the spatial step to match a single frequency JND, and we used 13 steps, with 7 subjects.

Keeping the sphere size fixed, each subject was asked to listen to all the sphere-cube pairs, each repeated ten times. The whole set of 130 couples was played in random order. The question was "is the second sound higher or lower in pitch than the first sound?".

In figure 1 (left) we report the mean and standard deviation of the pitch judgement for different box sizes in comparison with the ball ($d = 36$cm), for 7 subjects. In practice, such statistics are

[1] The models are available from `http://www.soundobject.org`

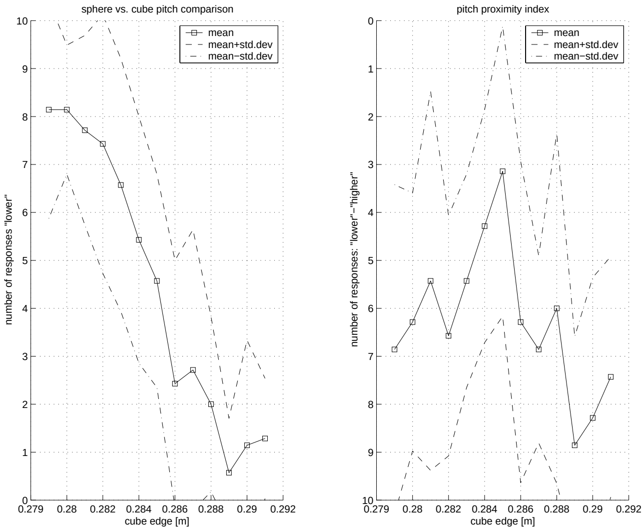

Figure 1: *Mean and standard deviation of pitch comparison between a sphere ($d = 36$cm) and cubes differing by an integral number of length JNDs.*

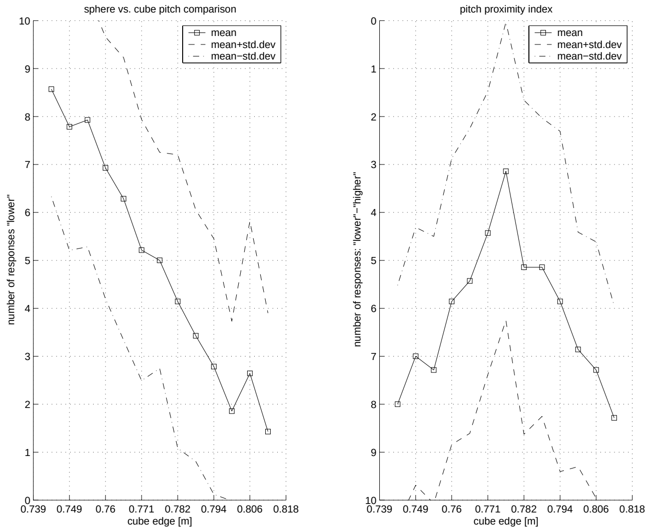

Figure 2: *Mean and standard deviation of pitch comparison between a sphere ($d = 100$cm) and cubes differing by an integral number of length JNDs.*

computed on one of the two rows of the confusion matrix of the responses of each subject. On the other hand, if we compute the difference of the two rows of the confusion matrix, we can come up with a sort of pitch proximity index, where maximum proximity corresponds to maximum confusion, i.e., it is found where the difference between the two rows is minimal. Figure 1 (right) depicts the mean and standard deviation of the pitch proximity index. The peak corresponds to the point of maximal confusion, which can be identified with the point of equal pitches. We decided to keep all of the 7 subjects, even though a pair of them gave answers very different from the average curve of fig. 1. All of the subjects of this test were volunteer computer science students, and no one classified himself as a musician. Even by including the outliers in the analysis, we have a very clear pitch effect (ANOVA: $F(12, 78) = 12.23$, $F_{\text{crit}} = 1.88$, $p < 0.05$). Looking at the peak of the pitch proximity index, we can say with reasonable accuracy that pitch equalization occurs for equal volumes.

We repeated the experiment with more subjects (14) and with larger resonators ($d = 100$cm for the sphere) and we got the results reported in fig. 2. The pitch effect is confirmed (ANOVA: $F(12, 169) = 11.87$, $F_{\text{crit}} = 1.81$, $p < 0.05$). Again, we did not remove the outliers from the statistics even though a few subjects performed very differently from the average. For instance, one subject almost systematically ranked the cube as lower in pitch than the sphere, and another one did the opposite. Most of the subjects were computer science students, and can be classified as "naive" listeners. However, one of the outliers is a trained musician, so his response may be due to analytic rather than casual listening. In fact, he said that he could hear two pitches in the cube, and he was confused by this fact. Another expert listener reported that, while he could give a definite pitch to the sphere, he had a definite problem to do that with the cube.

At room temperature, the lowest resonance for the ball of diameter $d = 36$cm is found at about 640Hz, corresponding to the fundamental frequency of the modal series associated with the order-1 Bessel function. For the same volume, the lowest resonance of the cube is below 600Hz. The JND at that frequency is lower than 4Hz. So, the pitch of the two shapes can not be trivially associated with the fundamental frequency. Looking at the frequency responses of the two cavities (see fig. 3) does not help much the task of deducing a "reasonable" pitch. However, with other representations (see sec. 4) the task may be easier.

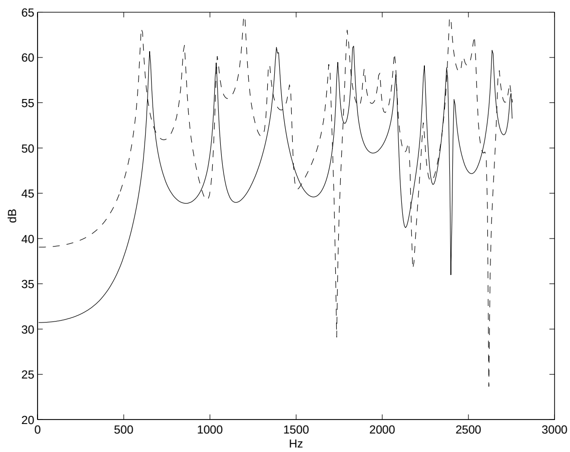

Figure 3: *Low-frequency spectra of the responses of a ball ($d = 0.36$ - solid line) and a box having the same volume (dashed line).*

3.2. Shape matching

When there is a pitch sensation, this tends to dominate over other acoustic cues, so it is important to know how to equalize the pitch of sounding objects in order to highlight other subtler timbral features. From an ecological viewpoint, one such feature is certainly the resonator shape, and we are now interested in understanding whether a cubic and a spherical resonator can be correctly classified.

In shape matching and classification experiments, we did not use the impulse responses, as we found them very unnatural. In fact, most people never experienced the impulse response of small cavities and we verified in informal experiments that the task was very difficult using the bare impulse response. It is much more natural to experience the shape of a resonator by means of the filtering effect that it impresses onto a known sound source. The choice of the sound source was driven by the need of exciting a large part of the frequency response without destroying the identity of the source. We tried with an anechoic voice source, but results were poor, probably because the voice has a strong harmonic content and only a few resonances of the frequency response can be excited in a short time segment. A source such as an applauding audience turned out to be unsuitable because its identity changes dramatically when it is filtered by a one-meter box. This is an interesting phenomenon that should be investigated, but for the scope of this work we sticked with a sound source that keeps its identity and that is rich enough to reveal the resonances of the cavities. We chose a snare drum pattern as a source.

We prepared three couples of stimuli, each couple corresponding to the snare drum convolved with the impulse responses of a sphere and a cube of the same volume. We chose volumes of spheres having diameters 106cm, 60cm, and 36cm. These three couples of stimuli where used to train the subjects, who could listen to the sounds as many times as they liked, and they could read the shape that each sound came from. For the real test, five different volumes were used, corresponding to spheres having diameters 100cm, 90cm, 70cm, 50cm, and 30cm. We did not choose the same sizes used in the training phase because we would like to avoid short-term memory effects and we want to assess the generalization ability of the subjects. In principle, a good listener should be able to decouple shape from pitch during training and to apply the cues of shape perception to other pitches.

One might argue that shape recognition should be assessed without any training. However, as it was pointed out in [6], the auditory shape recognition task is difficult for most subjects just because in real life we can use other senses to get shape information more reliably. This might not be the case for blind subjects [16], but with our (sighted) subjects, we found that training was necessary. Therefore, the task may be described as classification by matching [15].

We used 19 subjects. Each of them listened to 100 sounds and had to say, for each of them, if it came from a sphere or from a cube. The 100 sounds were composed by random shuffling of ten responses for each size for each shape. The results of the experiment are summarized in fig. 4, where we can see two interesting phenomena:

- The classification is significantly better than random choice for both shapes. For the large cubes, classification is around the threshold chosen in 2AFC experiments (e.g., 75%);
- The task is easier for larger volumes, converging to random choice for volumes of diameter smaller than 50cm. This observation is supported by the ANOVA run on the cumulative data represented in the third plot of fig 4 ($F(4,90) = 3.07$, $F_{\text{crit}} = 2.47$, $p < 0.05$).

Looking at fig. 4 one might say that the classification task is easier for the cube than for the sphere. Indeed a column-by-column t-test run on these data does not allow us to draw such

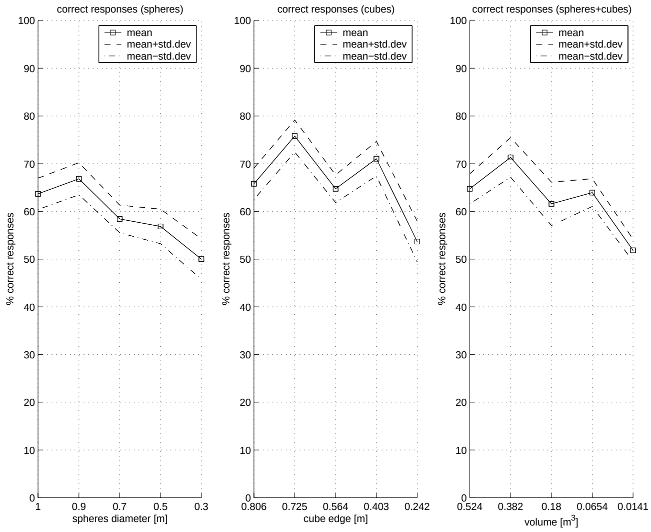

Figure 4: *Results of the shape classification task. Left: correct classification for spheres. Center: correct classification for cubes. Right: correct classifications.*

conclusion, since the variability is too high. However, looking at the direction of the difference in subjective performance is also important [14]. In this case, shape classification seems to be more reliably done for boxes, regardless of size. We found that, by removing the four subjects with the poorest perfomances, we got curves similar to those of fig. 4, with an incrase by about $5--10\%$ in the number of correct responses.

Indeed, in the analysis we didn't throw away any outliers even though there were some. In particular, some subjects classified resonators of certain sizes (especially the smaller ones) consistently with the same label. This may be due to a non-accurate training or to a mental association between pitch and shape. On the other hand, there were subjects who performed very well, such as the one whose responses are depicted in fig. 5. In that chart, color black in the lower row (and white in the top row) indicates that spheres (or cubes) have been consistently classified throughout the experiment. Grey-levels are used to indicate mixed responses. It is clear that for this subject the task was easy for larger volumes and more difficult for smaller volumes.

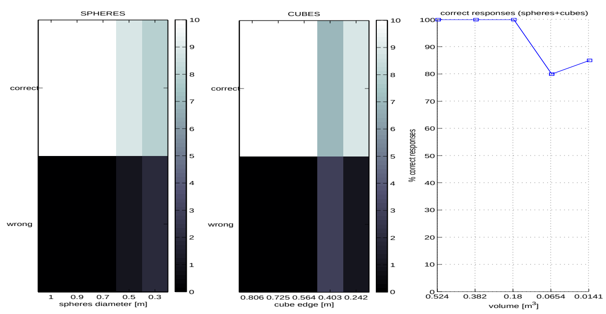

Figure 5: *Results of shape classification for the best subject.*

3.2.1. On the role of brightness

Brightness, often measured by the spectral centroid, plays a key role in sound source recognition [15]. When subjects were asked to give a qualitative description of the sounds played in the two different cavity shapes, they often said that the sphere sounds brighter. This is confirmed by measuring the brightness of the impulse responses. We did that using the routine used to compute the spectral centroid of sounds contained in the repository of the COST-G6 Action on Digital Audio Effects [2]. For instance, for cavities with wooden walls and volume of 1m^3, the centroid is 5570Hz and 5760Hz for the cube and the sphere, respectively. This change in brightness could be expected since, for the same volume, the sphere has a smaller surface area than the cube. Therefore, absorption acts more effectively in the cube than in the sphere.

The effect of brightness is mitigated by the fact that the listeners heard to different combinations of pitch (size) and shape in random order, and brightness is also pitch dependent. If the analysis of sec. 4 has any perceptual relevance, the role of brightness does not seem to be central in shape perception. However, this has to be carefully verified with further experiments.

From a theoretical standpoint, our approach to shape classification is purely ecological, as it uses models that are completely determined by physical features of objects (size, shape, material). In this framework, the role of brightness might be neglected. However, signal attributes such as brightness or harmonic structure are important to understand how shape information are conveyed to the listener. Such understanding is a prerequisite for the development of simplified yet effective sound models.

4. ANALYSIS BASED ON AUDITORY MODELS

The correlogram is a representation of sound as a function of time, frequency, and periodicity [12]. Each sound frame is passed through a cochlear model and split into a number of cochlear channels, each representing a certain frequency band. Cochlear channels are nonlinearly spaced according to the critical bands. The signal in each band is autocorrelated to highlight its periodicities. The autocorrelation magnitude is expressed in grey levels in a 2-D plot.

We use the correlogram to analyze the impulse responses of the cube and the sphere. Since the cavities are linear and time invariant, a single frame is enough to characterize the behavior of the whole impulse response. We use the impulse responses of resonators having marble walls, just because the images are more contrasted. However, using wooden resonators does not change the visual appearance of correlogram patterns appreciably. Figure 6 depicts the correlogram of the impulse response of a cube (edge length equal to $0.5m$) and that of the impulse response of a sphere having the same volume. Superimposed on the figure, we notice some patterns that emerge from the analysis and that, we conjecture, are the signature of the particular shape.

To support our conjecture, we make the following remarks:

- The cube has more than one vertical alignment of peaks, which corresponds to the fact that expert listeners could indeed hear the superposition of more than one pitch;
- The curved pattern of the sphere gets more clear as the size is increased (see figures 7 and 8). For small spheres it is barely noticeable, and this is confirmed by the poor results in shape discrimination for small cavities;

[2] `http://echo.gaps.ssr.upm.es/COSTG6/`

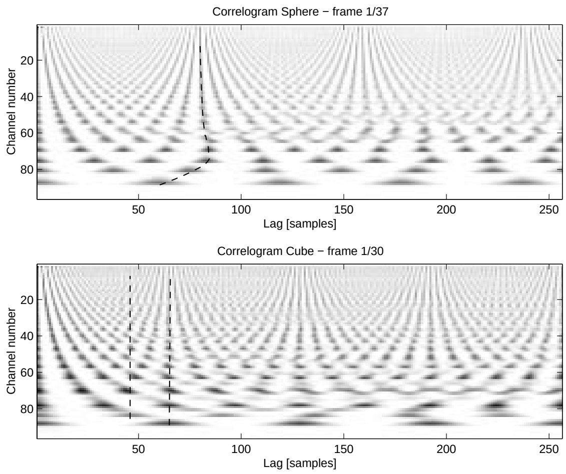

Figure 6: *Correlograms for the cube (edge $0.5m$) and for the sphere having the same volume.*

- The pitch of the spherical impulse responses, as found in the subjective experiments, may lay close to the asymptotic value of the pattern depicted in figure 6.

Even though we should be cautios in inferring that this is the kind of preprocessing that is used by the hearing system to make the shape identification, we can certainly say that the correlogram can be a useful tool to do the shape-from-sound detection.

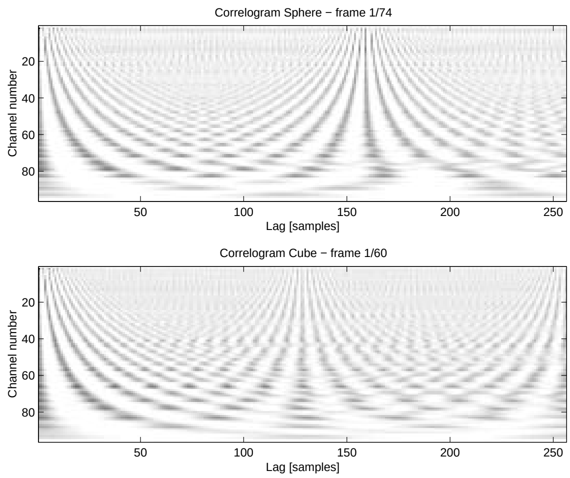

Figure 7: *Correlograms for the cube (edge $1.0m$) and for the sphere having the same volume.*

5. THE SHAPE-ENHANCED BABO MODEL

The BaBo model was proposed in [9] as a rectangular-resonator physical metaphor that can be used to control a feedback delay network. The model was extended in [6] to spherical cavities, and it was proposed that a shape control handle could control the degree of “roundness” by proper selection from a set of allpass filter coefficients.

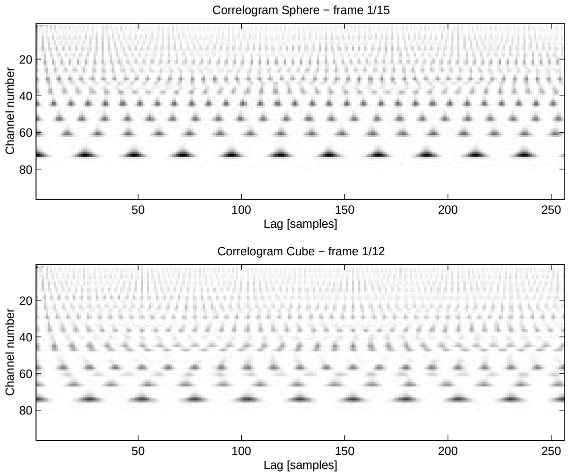

Figure 8: *Correlograms for the cube (edge $0.2m$) and for the sphere having the same volume.*

The analysis of section 4 suggests that only a few resonances are responsible for the patterns of figures 6 and 7 and, maybe, only these resonances have to be properly located to convey the correct sense of shape. For instance, fig. 9 shows a correlogram frame obtained by keeping only the modal frequencies f_{11}, f_{22}, f_{02}, and f_{42} of the spherical frequency response. The curve of the pattern is preserved. Similarly, only a few resonances are needed to preserve most of the vertical alignments visible in fig. 6.

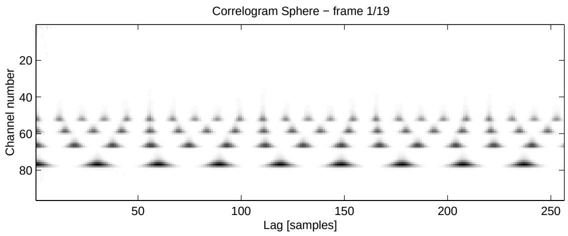

Figure 9: *Correlogram for a simplified model of spherical resonator that retains only four resonances.*

In our future investigations, we will use these simplified cavity models (i.e., parallel connections of a few second-order resonators) to see if (i) the same volume-based pitch matching is preserved; (ii) the shape classification task can be made easier and, possibly, turned into a real shape recognition.

6. ACKNOWLEDGMENT

Laura Ottaviani conducted the experiments and helped with the analysis of data. Andrea Tevoi and Giulio Verzini developed the Matlab functions for generating the impulse responses of the sphere and the cube. Nikolaus Mattern and Andrea Margoni did a preliminary analysis of the impulse responses by means of the auditory correlogram. Federico Avanzini and Federico Fontana provided useful comments on a preliminary paper draft.

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

7. REFERENCES

[1] M. Kac, "Can one hear the shape of a drum?," *Amer. Math. Monthly*, vol. 73, part II, pp. 1–23, 1966.

[2] S.J. Chapman, "Drums that sound the same," *Amer. Math. Monthly*, vol. 102, pp. 124–138, 1995.

[3] C. Carello, K. L. Anderson, and A. J. Kunkler-Peck, "Perception of object length by sound," *Psychological Science*, vol. 9, n. 3, pp. 211–214, 1998.

[4] S. Lakatos, S. McAdams and R. Caussé, "The representation of auditory source characteristics: Simple geometric form," *Perception and Psychophysics*, vol. 59, n. 8, pp. 1180–1190, 1997.

[5] A. J. Kunkler-Peck and M. T. Turvey, "Hearing shape," *Journal of Experimental Psychology*, vol. 26, n. 1, pp. 279–294, 2000.

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

[7] M. R. Moldover, J. B. Mehl and M. Greenspan, Gas-filled spherical resonators: Theory and experiment, *Journal of the Ac. Soc. of America*, vol. 79, no. 2, pp. 253–272, 1986.

[8] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, McGraw-Hill, New York, 1968.

[9] D. Rocchesso, The Ball within the Box: a sound-processing metaphor, *Computer Music Journal*, vol. 19, no. 4, pp. 47–57, 1995.

[10] H. Levitt, Transformed up-down methods in psychoacoustics, *Journal of the Ac. Soc. of America*, vol. 49, no. 2, pp. 467–477, 1970.

[11] W. B. Joyce, Exact effect of surface roughness on the reverberation time of a uniformly absorbing spherical enclosure, *Journal of the Ac. Soc. of America*, vol. 64, no. 5, pp. 1429–1436, 1978.

[12] M. Slaney and R. F. Lyon, On the importance of time – A temporal representation of sound, in *Visual Representations of Speech Signals*, M. Cooke, S. Beete, and M. Crawford, eds., J. Wiley and Sons., Sussex, England, pp. 409–429, 1993. Available from `http://www.slaney.org/malcolm/pubs.html`.

[13] T. Tolonen and H. Järveläinen, Perceptual study of decay parameters in plucked string synthesis, in Proc. *109-th AES Convention*, Los Angeles, Sept. 2000. Available from `http://www.acoustics.hut.fi/publications/`.

[14] D. Levitin, Experimental design in psychoacoustic research, in *Music, Cognition, and Computerized Sound*, P. R. Cook, ed., MIT Press, Cambridge, MA, 1999.

[15] S. McAdams, "Recognition of sound sources and events", in *Thinking in Sound: the cognitive psychology of human audition*, S. McAdams and E. Bigand, eds., Oxford University Press, 1993.

[16] R. McGrath, T. Waldmann, and M. Fernström, "Listening to rooms and objects", in *Proc. 16-th AES Conference on Spatial Sound Reproduction*, Rovaniemi, Finland, April 1999, pp. 512–522.