

Chapter 6

Size, shape, and material properties of sound models

Davide Rocchesso, Laura Ottaviani and Federico Fontana
Università di Verona – Department of Computer Science
Verona, Italy

davide.rocchesso@univr.it, ottaviani@sci.univr.it,
fontana@sci.univr.it

Federico Avanzini
Università di Padova – Department of Information Engineering
Padova, Italy

avanzini@dei.unipd.it

Recent psychoacoustic studies dealt with the perception of some physical object features, such as shape, size or material. What is in many cases clear from everyday experience, that humans are sensitive to these features, has been proved experimentally in some controlled conditions, using either real-world or synthetic sounds.

A 3D object can be described by its shape, size, material, position and orientation in space, color, surface texture, etc.. Leaving on a side those features that are purely visual (e.g., color), and those features that are only relevant for the “where” auditory subsystem, we are left with shape, size, and material as relevant “ecological” dimensions of sounding objects. This chapter addresses each of these three dimensions as they are representable in signal- or physics-

based models. Section 6.1 uses geometry-driven signal models to investigate how size and shape attributes are conveyed to the listener. We bear in mind that ecological acoustic signals tell us a lot about how the objects interact with each other, i.e., about the excitation mechanisms. Indeed, it is the case that different kinds of excitation highlight different features of resonating objects. Even simple impact sounds can elicit the perception of two objects simultaneously, a phenomenon called “phenomenal scission” by some perceptionists (see chapter 2). Section 6.2 stresses the importance of accurate simulation of interaction mechanisms, in particular for calibrating the physical parameters in a sound model in order to render an impression of material by listening to sound sources.

6.1 Spatial features

Of the three object attributes, size, shape, and material, at least the first two are perceptually related in a complex way. The relationship between the ecological properties size and shape reflects a relationship between the perceived sound properties pitch and timbre.

When looking at prior studies in the perception of shape of resonating objects, we find works in:

- 1D - Carello et al. [49] showed that listeners are able to reliably evaluate, without any particular training, the lengths of rods dropped on a hard surface. Rods are essentially one-dimensional physical systems, so it is a degenerate case of shape perception. Indeed, subjects are perceiving the pitch of complex, non-harmonic tones;
- 2D - Lakatos et al. [145] showed that the hearing system can estimate the rectangular cross section of struck bars, and Kunkler-Peck and Turvey [142] did similar experiments using suspended rectangular plates. Even though these objects are 3D, attention is focused on their 2D aspect, and it is argued that the distribution of modal frequencies gives cues for shape perception.

Apparently, there is lack of research results in auditory perception of 3D shape. Some direct observations (see chapters 2 and 5) indicate that impact sounds of small solid objects give no clue on the shape of the objects themselves. On the other hand, some prior research [207] indicated that a rough sensation of shape can be elicited by the filtering effect of 3D cavities excited

by a source. In that work, the ball-within-a-box (BaBo) model [204] was extended to provide a unified 3D 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, it is possible to 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 3D resonator shapes in an experimental framework, we used models of spheres and cubes that are controllable in size and material of the enclosure. We chose to construct impulse responses by additive synthesis, since closed-form expressions of the resonance distributions of cubic and spherical enclosures are available.

A rectangular resonator has a frequency response that is the superposition of harmonic combs, each having a fundamental frequency

$$f_{0,l,m,n} = \frac{c}{2} \sqrt{(l/X)^2 + (m/Y)^2 + (n/Z)^2} \quad , \quad (6.1)$$

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 [181].

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

$$f_{ns} = \frac{c}{2\pi a} z_{ns} \quad , \quad (6.2)$$

where a is the radius of the sphere [178].

The impulse response of a sphere or a cube can be modeled by damping the modes according to the absorption properties of the cavity, introducing some randomization in mode amplitudes in order to simulate different excitation and pickup points, and stopping the quadratic frequency-dependent increase in modal density at the point where the single modes are no longer discriminable. In particular, the decay time of each modal frequency was computed using the Sabine reverberation formula

$$T = 0.163 \frac{V}{\alpha A} \quad , \quad (6.3)$$

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:

$$f = [0, 125, 250, 500, 1000, 2000, 4000, F_s/2] \text{ Hz} \quad ; \quad (6.4)$$

$$\alpha = [0.19, 0.15, 0.11, 0.10, 0.07, 0.06, 0.07, 1.00] \quad , \quad (6.5)$$

and the sample rate was set to $F_s = 22050$ Hz.

This sound model has been used to produce stimuli for tests on perception of cubic and spherical shape and size. In particular, we investigated sizes ranging from 30 cm to 100 cm in diameter. The use of the 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 [130]. Moreover, it has been assessed that we are not very sensitive to variations in decay time [234], 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.

6.1.1 Size

When comparing the impulse response of a spherical cavity with that of a cubic cavity, we may notice that one sounds higher than the other. Therefore, there is a pitch relationship between the two shapes. In order to use shape as a control parameter for resonators, for instance in auditory display design, it is important to decouple it from pitch control.

In an experiment [225], we submitted couples of impulse responses to a number of subjects, in random order and random sequence: one of a sphere with fixed diameter and the other of a cube. We used a fixed sphere and thirteen cubes, of which the central one had the same volume as the comparison sphere, while the others had edge length that varied in small steps, equal to Δl , calculated by converting frequency Just Noticeable Differences (JND), as found in psychoacoustic textbooks and measured for pure tones, into length differences:

$$\Delta l = \frac{c}{2} \left(\frac{1}{f_0 - \Delta f} \right) - l_0 \quad , \quad (6.6)$$

where c is the speed of sound in the cavity, and $l_0 = c/(2f_0)$ is the reference size length. For instance, for a reference size length $l_0 = 1$ m, the length JND is about 18 mm.

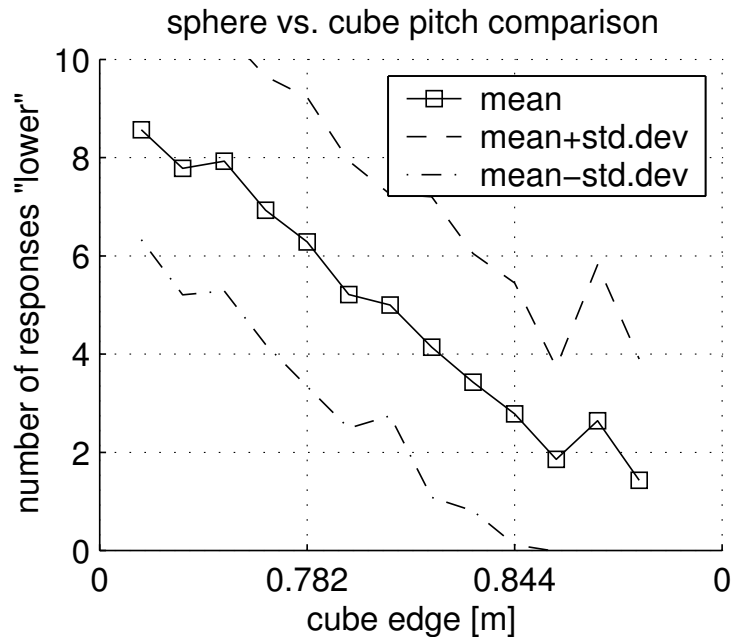


Figure 6.1: Mean and standard deviation of pitch comparison between a sphere ($d = 100$ cm) and cubes differing by an integral number of length JNDs.

Therefore, we had a central cube with the same volume as the fixed comparison sphere, 6 cubes smaller than the central one, and 6 bigger.

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?”.

The experiment was repeated for two values of sphere diameter, $d = 36$ cm and $d = 100$ cm. Results for the 100 cm sphere and for 14 subjects are plotted in Figure 6.1. A large part of the deviation from the mean curve is due to two outliers who could segregate more than one pitch in the cube impulse response. These two subjects, indeed skilled musicians, seemed to be using a analytic listening mode rather than the holistic listening mode used by the other subjects [30]. This is symptomatic of how the same sonic signal can carry different information for different kinds of listeners.

In both experiments, the Point of Subjective Equality (i.e., where the mean curve crosses the 50% horizontal line) is found where the two shapes are roughly equalized in volume, thus meaning that a good pitch (or size) equalization is obtained with equal volumes. This is an important finding in an ecological perspective. However, the next question is how to give a number (in Hz) to this perceived pitch. Conventional spectral analysis do not give any hint to find a “reasonable pitch”.

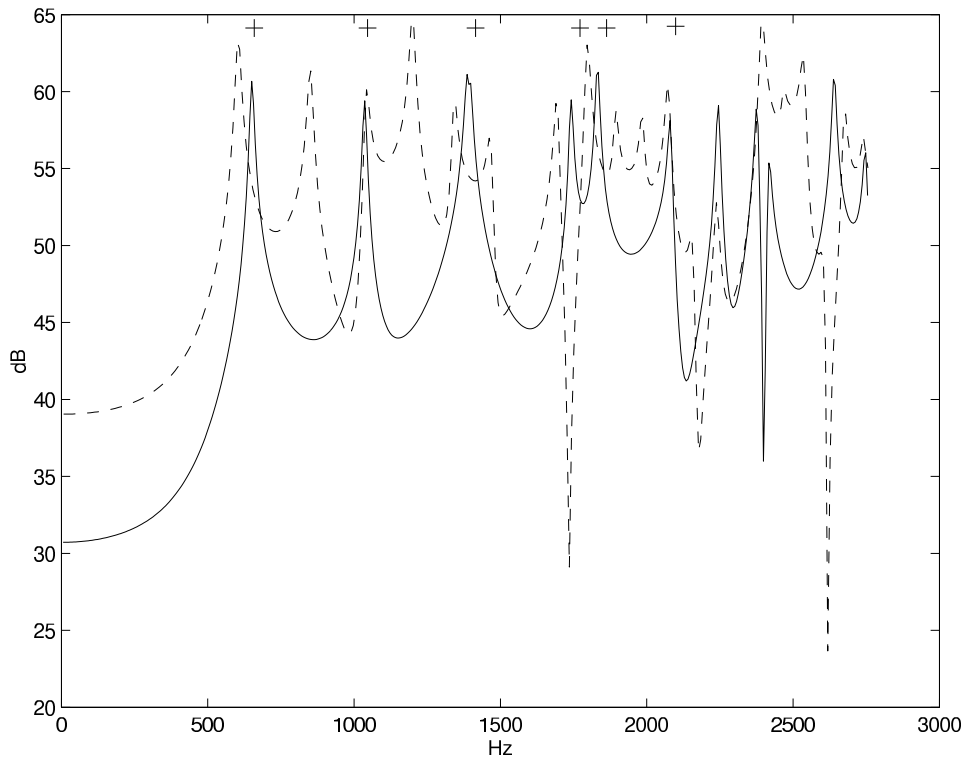


Figure 6.2: Low-frequency spectra of the responses of a ball ($d = 0.36$ - solid line) and a box having the same volume (dashed line). Crosses indicate resonance frequencies obtained after a waveguide mesh simulation.

Figure 6.2 plots the frequency responses of a sphere of diameter $d = 36$ cm and a cube with the same volume. Clearly, the pitch of the two cavities in this shape comparison task can not be trivially associated with the fundamental frequency.

In a second experiment [225], we tried to “measure” the perceived pitch comparing the spherical impulse response with an exponentially-damped sinusoid. We chose the spherical response because it was perceived by the subjects to have a higher “pitch strength” than that of the cube. In this pitch comparison test, we used a simple staircase method [152], keeping the ball impulse response fixed (diameter $d = 0.5$ m), and varying the frequency of the test sine wave. The participants to the experiment were 28, that are 25 computer science students, two of the authors, and a research collaborator. They listened to a couple of sounds two times before answering to the question: “Is the first sound higher in pitch than the second sound?”. With the staircase method each subject could converge to a well-defined value (in Hz) of the pitch of the sphere.

Figure 6.3 reports the histogram that shows the distribution of frequency

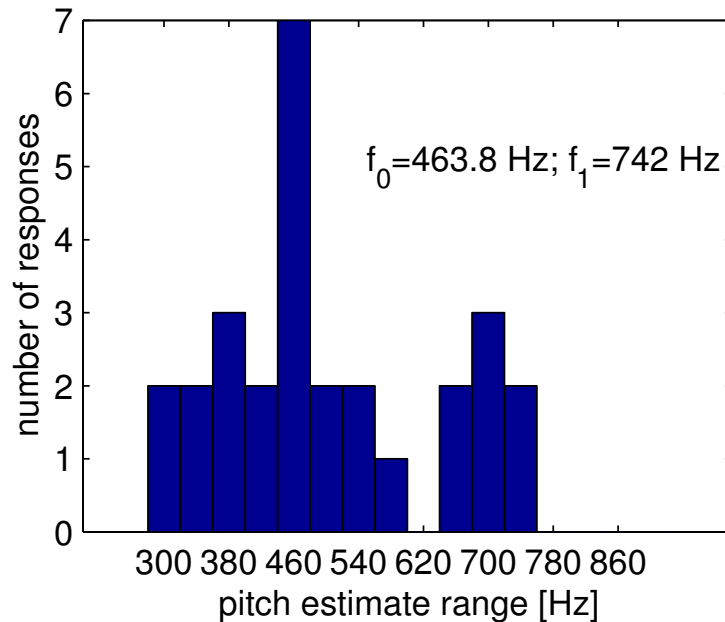


Figure 6.3: Distribution of the pitch estimation values of a ball ($d = 0.5$ m).

values. It can be noticed that one fourth of the subjects estimated the perceived pitch in the range $[460, 480]$ Hz. Within this range, the estimated pitches were 444.8, 455.8, 461.9, 468.1, 468.9, 473.3, 474.2 Hz. The lowest partial of the spherical frequency response is found at 463.8 Hz. Thus, one fourth of the subjects found a pitch close to the lowest partial. Moreover, Figure 6.3 shows that seven subjects estimated the pitch in the range $[640, 760]$ Hz, that is a neighborhood of the second partial (found at 742 Hz).

In this latter experiment, the difference in the nature of the two stimuli is such that a different, more analytic, listening mode is triggered. Indeed, many subjects tried to match one of the partials of the reference sound. Interestingly, most subjects found this pitch comparison task difficult, because of the different identity of the two stimuli.

On the other hand, from the first experiment it seems that, when a subject tries to match the pitch of two different shapes, he or she is able to extract some information about the physical volume of the two shapes and to match those volumes. That experiment provides evidence of direct perception of volume of cavities, even though the question asked to the listeners did not mention volumes at all.

As far as the use of sound of shapes in information sonification is concerned, one should be aware of the problems related to the relationship size/shape, or pitch/timbre, and to the listening attitude of different listeners, espe-

cially if pitch is used as a sonification parameter.

6.1.2 Shape

Besides the size perception of different shapes, it is interesting to understand whether a cubic or a spherical resonator can be correctly classified. If this is the case, a scatter plot where data are represented by simple geometric objects may be sonified by direct shape sonification.

Since most people never experienced the impulse response of small cavities, we made some informal tests by convolving “natural” sounds with the cavity impulse response, to see what could be a suitable excitation signal. The constraint to be remembered is 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. Therefore we were looking for 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 to generate the stimuli set consisting of 10 sounds, i.e. five couples of stimuli, one from a sphere and one from a cube of the same volume, each couple different from the others for their volumes. The diameters of the spheres were 100 cm, 90 cm, 70 cm, 50 cm, and 30 cm.

The 19 volunteers, who participated to the experiment, listened to isolated sounds, belonging to the stimuli set and each one repeated 10 times in random order. Therefore, the whole experiment consisted in listening to 100 sounds. For each of them, the participants had to say whether it was produced in a spherical or in a cubic enclosure.

The experiment was preceded by a training phase, when the subjects could listen to training sounds as many times as they liked, and they could read the shape that each sound came from. For generating the training sounds, we chose spheres having different sizes than the ones used in the main experiment (diameters 106 cm, 60 cm, and 36 cm), because we wanted to avoid memory effects and 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 [207], 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 [172] but with our (sighted) subjects, we found that training was necessary. Therefore, the task may be described as classification by matching.

The experiment [225] showed that the average listener is able to classify sounds played in different shapes with an accuracy larger than 60% when the cavity diameter is, at least, equal to 50 cm. This is not a very strong result. Again, a major reason is that there were some notable outliers. 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.

One of them is visually impaired and her responses are depicted in Figure 6.4. It is clear that for this subject the task was easy for larger volumes and more difficult for smaller volumes. This specific case indicates that there are good possibilities for direct shape sonification, at least for certain subjects.

An auditory feature related to shape is brightness, which is often measured by the spectral centroid and which plays a key role in sound source recognition [171]. We noticed that, 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 sounded brighter. Therefore, we decided to investigate this aspect. Moreover, in order to study the sound characteristics involved in shape recognition, we analyzed the spectral patterns, by means of auditory models. We report the results of these investigations, that are interesting to examine further.

For measuring the brightness of the impulse responses we used a routine that computes the spectral centroid of sounds [205]. For cavities with wooden walls and volume of 1 m^3 the centroid is located at 5570 Hz and 5760 Hz for the cube and the sphere, respectively. This change in brightness must be expected since, for equal volume, the sphere has a smaller surface area than the cube. Therefore, absorption is greater for the cube than the sphere.

The effect of brightness was smoothed, as listeners had to listen to random combinations of pitch (size) and shape; furthermore, brightness is pitch-dependent as well.

We examined the spectral patterns of the spherical and cubic enclosures by analyzing the correlogram calculated from their impulse responses. The correlogram is a representation of sound as a function of time, frequency, and periodicity [222]: Each sound frame is passed through a cochlear model, then split

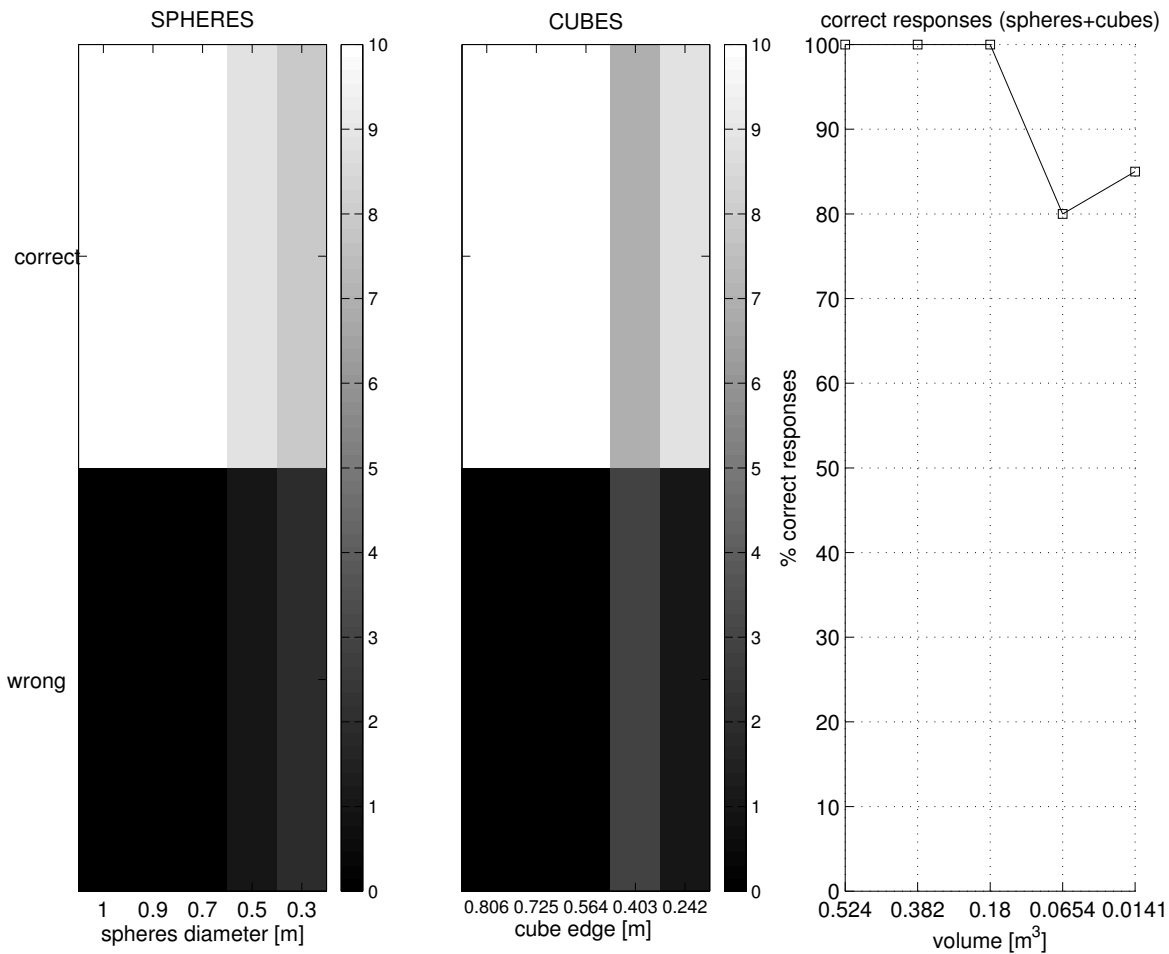


Figure 6.4: Results of the shape classification for the best subject. Color black in the lower row (and white in the top row) indicates that spheres (or cubes) have been consistently classified throughout the experiment. Grey-levels are used to indicate mixed responses.

into a number of cochlear channels, each one representing a certain frequency band. Cochlear channels are nonlinearly spaced, according to the critical band model. The signal in each critical band is self-correlated to highlight its periodicities. The autocorrelation magnitude can be expressed in gray levels in a 2D plot.

Since the cavities are linear and time invariant, a single frame gives enough information to characterize the behavior of the whole impulse response. We used the impulse responses of resonators having marble walls, just because the images are more contrasted. However, the use of wooden resonators would not have changed the visual appearance of the correlogram patterns appreciably. Figure 6.5 depicts the correlogram of a cube (sized 0.5 m) and a sphere having

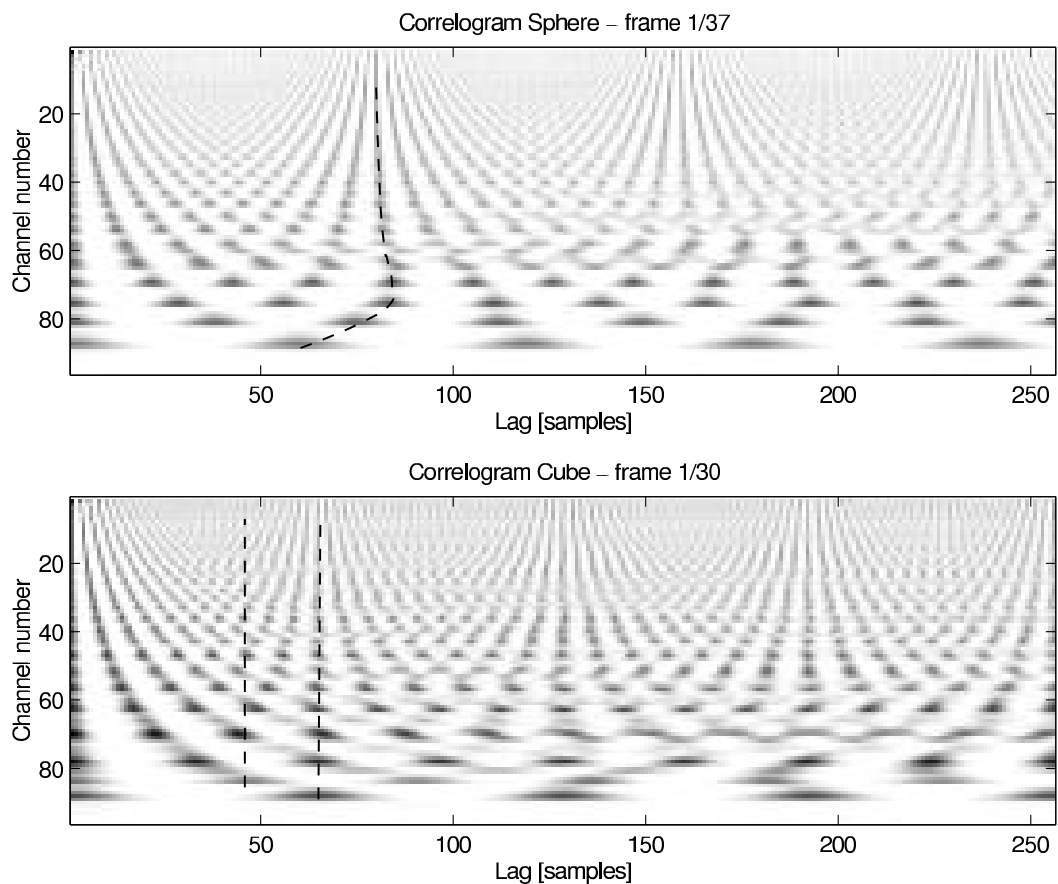


Figure 6.5: Correlograms for the cube (edge 0.5 m) and for the sphere having the same volume.

the same volume. If we superimpose them, we notice the existence of some patterns that, we conjecture, can be read as the *signature* of a particular shape.

It is interesting to notice that the cube has more than one vertical alignment of peaks, this suggesting that expert listeners would be in principle enabled to hear more than one pitch. Moreover, we observed that the curved pattern of the sphere becomes more evident as long as size is increased. Conversely, for small spheres it is barely noticeable: this is confirmed by corresponding fair results in the discrimination of shape for small cavities.

Even though we should be cautious in concluding that these results reflect the kind of preprocessing that is used by our hearing system to discriminate between shapes, nevertheless the correlogram proved to be a useful tool for detecting shapes from sound.

Our analysis also suggests that only few resonances are necessary to form the patterns seen in Figure 6.5. It is likely that no more resonances must be

properly located, if we want to convey a sense of shape.

Continuous rendering of morphed shapes

Here we present a method for *sonifying morphing shapes*. By this method we can test whether listeners perceive the *roundness* of morphing shapes, i.e., whether specific cues exist that characterize shape morphing, and whether such cues are conveyed by cartoon models of morphing resonators.

We devised a method for sonifying objects having intermediate shapes between the sphere and the cube, then realizing a model that runs in real time on inexpensive hardware. This model can be used for rendering morphing shapes in 3D virtual scenarios.

The model must be informed about the shapes to render. A specific set of shapes was selected that could be driven by a direct and meaningful morphing parameter, in a way that the control layer relied on a lightweight mapping strategy.

We chose *superquadrics* [15] to form a set of geometries that could be mapped using a few parameters. In particular, a sub-family of superquadrics described by the equation

$$|x|^\gamma + |y|^\gamma + |z|^\gamma = 1 \quad (6.7)$$

(it can be seen that they represent a set of ellipsoidal geometries) fitted our investigation. In fact, changes in shape are simply determined by varying γ , that acts for that family like a morphing parameter:

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

To implement this 3D resonator model we used a bank of second-order bandpass (peaking) filters, each filter being tuned to a prominent peak in the magnitude response of the superellipsoid. Such a filter bank satisfies the requirement of simplicity and efficiency, and allows “sound morphing” by smoothly interpolating between superellipsoids of different geometry. This is done, for each specific geometry, by mapping γ onto the corresponding filter coefficients, holding condition $2 < \gamma < \infty$.

Since the modal positions are known by theory only in the limit cases of the cube and the sphere, we first calculated the responses of several superellipsoids

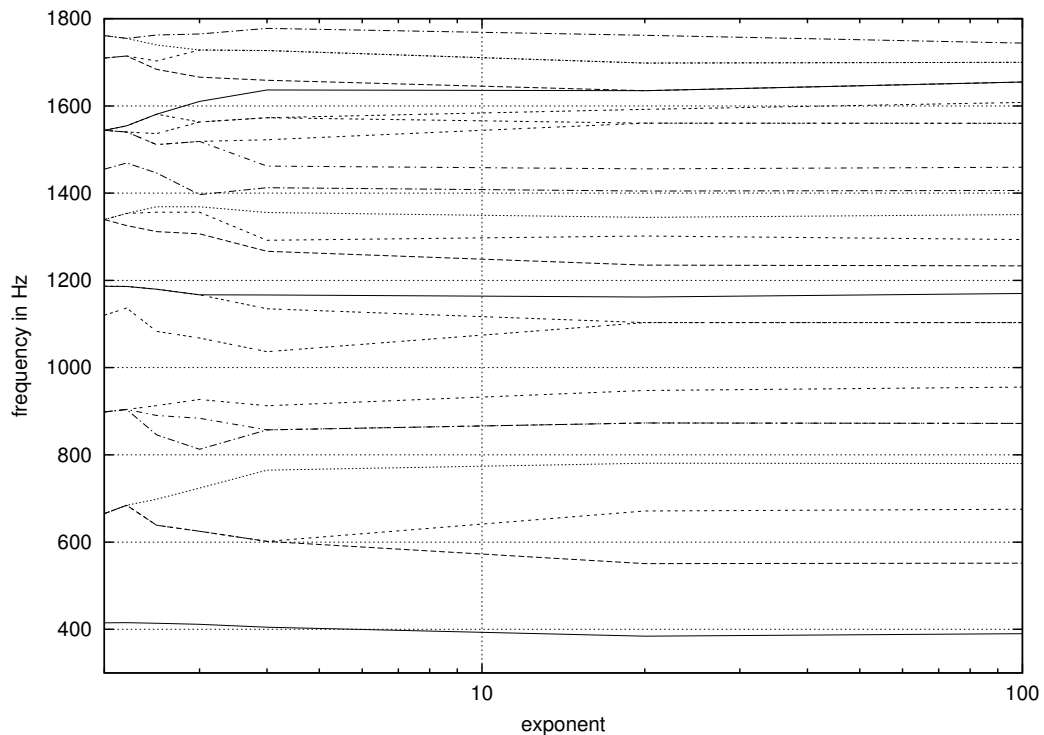


Figure 6.6: Trajectories for the lowest resonance frequencies of a superellipsoid morphed from a sphere to cube.

off-line, with enough precision [78]; then we approximated those responses using filter banks properly tuned. Morphing between two adjacent responses was obtained by linearly interpolating the filter coefficients between the values producing the frequency peaks in the two cases, which were previously calculated off-line.

Similarly, different acquisition points in a given resonator were reproduced by smoothly switching on and off the resonance peaks (again calculated off-line) that were present in one acquisition point and not in the adjacent one; otherwise, when one peak was present (absent) in both acquisition points then the corresponding filter was kept active (off) during the whole migration of the acquisition point. Though, we finally decided to pay less attention to the peak amplitudes as they in practice depend on several factors, such as the excitation and acquisition point positions and the internal reflection properties of the resonator.

For what we have said, the resonances thus follow trajectories obtained by connecting neighbor peak positions. These trajectories occasionally split up and rejoin (Figure 6.6).

Moreover, for reasons of symmetry some resonant frequencies can disap-

pear or be strongly reduced while moving the excitation/acquisition points toward the center of the resonator. We accordingly lowered the related peak amplitudes, or cleared them off.

It must be noticed that the frequency trajectories cannot accurately describe the morphing process; they are rather to be taken as a simple (yet perceptually meaningful) approximation, from a number of “sample” shapes, of the continuously changing magnitude response of a morphing object.

During the morphing process the volume of the superellipsoids was kept constant, so respecting the volume-pitch relation [208]. This volume constancy leads to a slight decrease of the fundamental mode position when moving toward the cube. Simultaneously, the overall perceived brightness (related to the frequency centroid of the spectrum) decreases when using a fixed number of filters having constant amplitudes.

Several strategies could be conjectured for compensating for this change in brightness. Though, we did not reputed this correction to be necessary, since the effect does not seem to be strong enough to mask other sound features related to the shape morphing.

6.2 Material

Rendering an impression of object material is not always possible or cost effective in graphics. On the other hand, physics-based sound models give the possibility to embed material properties with almost no computational overhead.

Wildes and Richards [253] developed theoretical considerations about the relationship between the damping (equivalently, decay) characteristics of a vibrating object and the auditory perception of its material. Two recent studies provided some experimental basis to the conjectures formulated in [253], but results were not in accordance. On the one hand, Lutfi and Oh [163] found that changes in the decay time are not easily perceived by listeners, while changes in the fundamental frequency seem to be a more salient cue. On the other hand, Klatzky et al. [137] showed that decay plays a much larger role than pitch in affecting judgment, and therefore confirmed predictions by Wildes and Richards. Both of these studies used synthetic stimuli generated with additive synthesis of damped sinusoids.

A physically-based approach was taken by Djoharian [65], who developed viscoelastic models in the context of modal synthesis and showed that finite difference models of resonators can be “dressed” with a specific material qual-

ity. Indeed, the physical parameters of the models can be chosen to fit a given frequency-damping characteristic, which is taken as the sound signature of the material. Sound examples provided by Djoharian convinced many researchers of the importance and effectiveness of materials in sound communication.

Any sound practitioner is aware of the importance of the attack in percussive sound timbres. Nonetheless, none of the above mentioned works made use of impact models, instead using simple impulse responses with no attack transients. It remains to be proved to what extent is material perception affected when realistic and complex impacts are used.

Using a method for artifact-free simulation of non-linear dynamic systems [24], we developed a digital impact model that simulates collision between two modal resonators. This digital hammer is based on a well known model in impact mechanics [167], and is described in detail in chapter 8. We produced synthetic auditory stimuli with this model, and used the stimuli for investigating material perception through listening tests. In order to keep the number of model parameters low, the modal resonator in the synthesis algorithm was parametrized so to have only one mode.

Two acoustic parameters were chosen for controlling synthesis of stimuli, namely pitch f_0 and quality factor q_o . Using our resonator model, this latter parameter relates to decay characteristics via the equation $q_o = \pi f_0 t_e$, where t_e is the time for the sound to decay by a factor $1/e$. We synthesized 100 stimuli using five equally log-spaced pitches from 1000 to 2000 Hz and 20 equally log-spaced quality factors from 5 to 5000; these extremal q_o values correspond to typical values found in rubber and aluminum, respectively. In a recent study on plucked string sounds, Tolonen and Järveläinen [234] found that relatively large deviations (between -25% and $+40\%$) in decay time are not perceived by listeners. Accordingly with the values we chose, the relative lower/upper spacings between q_o values are $-31\%/ +44\%$.

In the listening test, subjects were asked to listen to these 100 sounds and to indicate the material of the sound sources, choosing among a set of four material classes: rubber, wood, glass and steel. Each sound was played only once. The subjects were 22 volunteers, both expert and non-expert listeners, all reporting normal hearing.

Figure 6.7 summarizes test results: it shows the proportion of subjects responses for each material, as a function of the two acoustic cues (pitch and quality factor). The inter-subject agreements (proximity of the response proportions to 0 or 1) are qualitatively consistent with indications given by [253], namely (1) q_o tends to be the most significant cue and (2) q_o is in increasing order for rubber, wood, glass and steel. A slight dependence on pitch can be

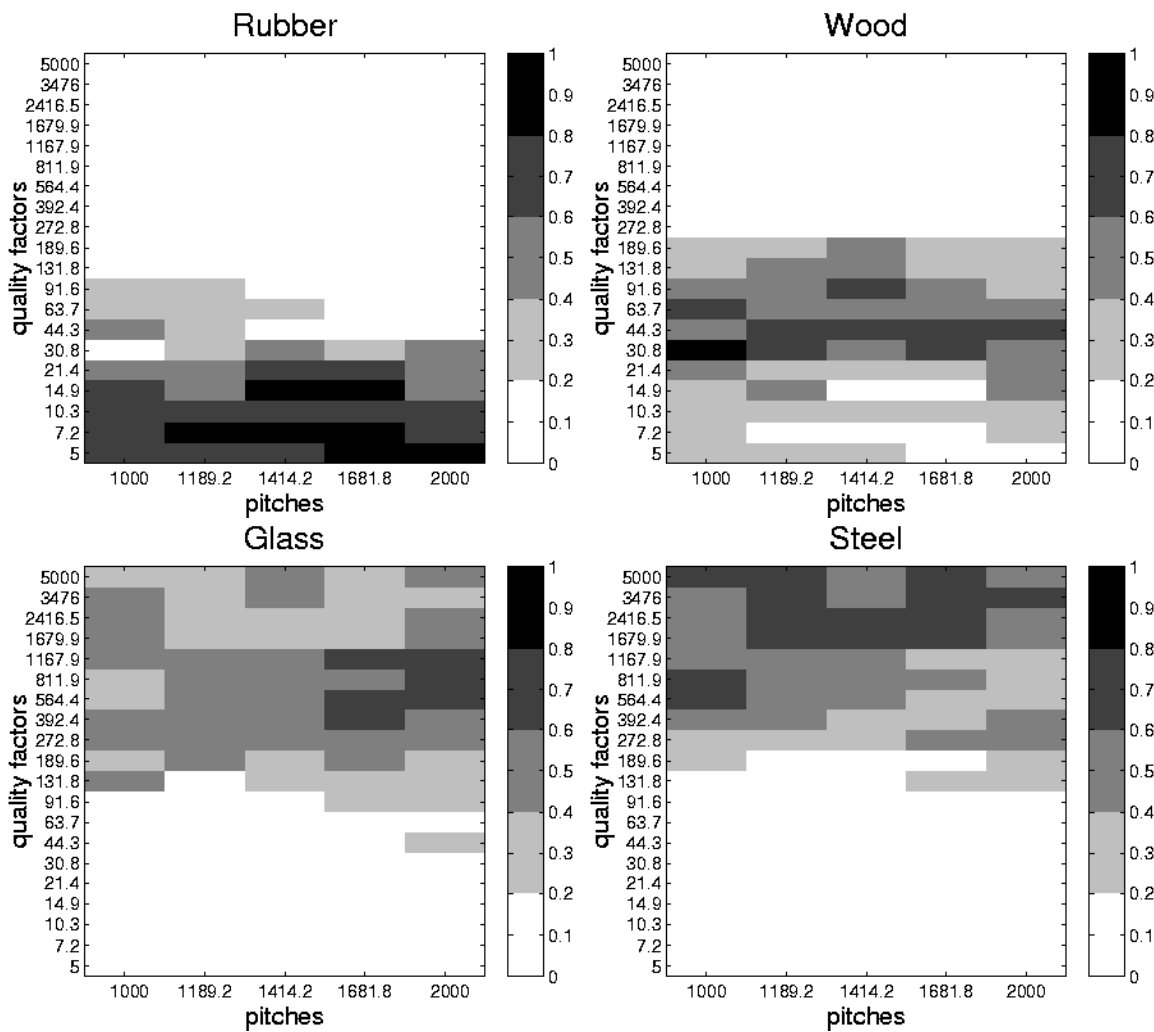


Figure 6.7: Proportion of subjects who recognized a certain material for each stimulus.

noticed: rubber and glass tend to be preferred at high pitches, while wood and steel are more often chosen at low pitches. It appears clearly that the upper and lower halves of the q_0 range are well separated, while materials within each of these two regions are less easily discriminated. In particular, it is evident from the figure that the regions corresponding to glass and steel are largely overlapping, while ranges for rubber and wood are better delimited. The attribution of material qualities to sound is more problematic when recordings are used. Chapter 5 extensively covers this case.