

The psychomechanics of simulated sound sources: Material properties of impacted bars^{a)}

Stephen McAdams^{b)}

Institut de Recherche et Coordination Acoustique/Musique (SMTS-Ircam-CNRS), 1 place Igor Stravinsky, F-75004 Paris, France and Département d'Etudes Cognitives, Ecole Normale Supérieure (DEC/ENS), 45 rue d'Ulm, F-75230 Paris, France

Antoine Chaigne^{c)}

Département TSI, Ecole Nationale Supérieure des Télécommunications (ENST/TSI), 46 rue Barrault, F-75013 Paris, France

Vincent Roussarie^{d)}

SMTS-Ircam-CNRS, 1 place Igor Stravinsky, F-75004 Paris, France and ENST/TSI, 46 rue Barrault, F-75013 Paris, France

(Received 25 March 2003; accepted for publication 15 December 2003)

Sound can convey information about the materials composing an object that are often not directly available to the visual system. Material and geometric properties of synthesized impacted bars with a tube resonator were varied, their perceptual structure was inferred from multidimensional scaling of dissimilarity judgments, and the psychophysical relations between the two were quantified. Constant cross-section bars varying in mass density and viscoelastic damping coefficient were synthesized with a physical model in experiment 1. A two-dimensional perceptual space resulted, and the dimensions were correlated with the mechanical parameters after applying a power-law transformation. Variable cross-section bars varying in length and viscoelastic damping coefficient were synthesized in experiment 2 with two sets of lengths creating high- and low-pitched bars. In the low-pitched bars, there was a coupling between the bar and the resonator that modified the decay characteristics. Perceptual dimensions again corresponded to the mechanical parameters. A set of potential temporal, spectral, and spectrotemporal correlates of the auditory representation were derived from the signal. The dimensions related to mass density and bar length were correlated with the frequency of the lowest partial and are related to pitch perception. The correlate most likely to represent the viscoelastic damping coefficient across all three stimulus sets is a linear combination of a decay constant derived from the temporal envelope and the spectral center of gravity derived from a cochlear representation of the signal. These results attest to the perceptual salience of energy-loss phenomena in sound source behavior. © 2004 Acoustical Society of America. [DOI: 10.1121/1.1645855]

PACS numbers: 43.75.Cd, 43.66.Jh, 43.75.Kk [NHF]

Pages: 1306–1320

I. INTRODUCTION

The perception of the properties of sound sources is important for their categorization and identification as well as for our own interactions with them in the environment (McAdams, 1993). Recent research has begun to study listeners' sensitivities to auditory cues that communicate aspects of the geometry and material properties of vibrating objects as well as changes in the state of objects (e.g., breaking) or the activity of organisms (e.g., walking). Listeners'

auditory sensitivity to the geometry of vibrating objects has been demonstrated for clapping hands (Repp, 1987), impacted bars (Houix *et al.*, 1999; Lakatos *et al.*, 1997; Lutfi, 2001), impacted plates (Kunkler-Peck and Turvey, 2000), falling and bouncing balls (Guski, 2000), rods that fall and bounce on the floor (Carello *et al.*, 1998), vessels being filled with liquid (Cabe and Pittenger, 2000), the size and velocity of rolling balls (Houben, 2002), or to the geometric integrity or loss thereof in falling jars that bounce or shatter (Warren and Verbrugge, 1984). A similar sensitivity to properties such as density and elasticity that vary with the materials of which an object is composed has also been shown (Avanzini and Rocchesso, 2001; Klatzky *et al.*, 2000; Lutfi and Oh, 1997).

Such studies have used a variety of approaches such as recognition, identification, classification, or categorization of the object (Avanzini and Rocchesso, 2001; Cabe and Pittenger, 2000; Houix *et al.*, 1999; Krotkov *et al.*, 1996; Repp, 1987; Warren and Verbrugge, 1984), discrimination among sounds (Lakatos *et al.*, 1997; Lutfi, 2001; Lutfi and Oh, 1997), scaling sounds according to prespecified properties such as hardness, length, mass, or speed (Carello *et al.*,

^{a)}Preliminary results leading to the present study were first reported at the combined 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America, Seattle, WA and at the 139th Meeting of the Acoustical Society of America, Atlanta, GA.

^{b)}Address correspondence to S. McAdams at DEC/ENS. Electronic mail: mcadams@dec.ens.fr

^{c)}Current address: ENSTA-UME, chemin de la Humière, F-91761 Palaiseau, France. Electronic mail: chaigne@ensta.fr

^{d)}Current address: PSA Peugeot-Citroën, DRJA/SARA/PEFH, route de Gizy, F-78140 Vélizy-Villacoublay, France. Electronic mail: vincent.roussarie@mpsa.com

1998; Freed, 1990; Kunkler-Peck and Turvey, 2000), or estimating relative similarity of sound pairs (Klatzky *et al.*, 2000). The last technique is of particular interest as it lends itself to an exploratory data analysis approach that imposes few *a priori* hypotheses on the nature of the underlying perceptual structure that results from variation in a set of mechanical properties of the objects of interest or the acoustical properties resulting from setting them into vibration.

In most of the studies cited above, the sounds used were either natural (objects set into vibration in the presence of the listeners or digitally recorded) or synthetic (most often by Fourier synthesis). The former have the advantage of representing the richness of the everyday acoustic environment and the disadvantage of clearly establishing what it is in the sound that listeners are using to make their judgments. The latter have the advantage of being deterministic and simple, but often suffer from not truly representing real-life vibrating objects. A judicious compromise is the use of physical models that simulate the vibratory behavior of objects. One can control precisely the mechanical parameters of the synthesis model, all the while obtaining the complex relations among analytic signal parameters that specify the mechanical nature of the object. The perceptual study of such sounds also has the advantage of leading to an understanding of which mechanical properties are the most relevant for simulation in virtual environments.

To study the perceptual representation of material properties of impacted bars, a physical model was used to create the sounds. The model has been presented elsewhere (Chaigne and Doutaut, 1997; Doutaut *et al.*, 1998) and only the aspects that are relevant to the present psychophysical quantification of the relation between some of the model parameters and their corresponding perceptual dimension will be presented here. Listeners performed dissimilarity ratings on pairs of sounds. A multidimensional scaling technique was used to recover a mathematical model that represents the dissimilarities as distances (McAdams *et al.*, 1995). Indeed, multidimensional scaling has been a technique of predilection for the study of musical timbre (Gray, 1977; Hajda *et al.*, 1997; Iverson and Krumhansl, 1993; Lakatos, 2000; McAdams, 1999; McAdams *et al.*, 1995; Plomp, 1970; Wessel, 1979). One then uses the distance model to compare the dimensionality of the “perceptual space” with that of the physical stimulus space and to quantify the relations among mechanical properties, analytic parameters derived from the resulting signals, and the perceptual dimensions. Of particular interest are the nature of timbral and pitch cues that underlie the perceptual representation.

II. PHYSICAL MODELING SYNTHESIS

The software used to build the sound palette for the psychoacoustic experiments is a time-domain sound synthesis program based on a physical model. The program digitally simulates mallet percussion instruments (Chaigne and Doutaut, 1997; Doutaut *et al.*, 1998). The physical equations of the complete model are thus reproduced and solved numerically, providing control of the different mechanical and geometrical parameters of the equations. We will briefly present in order the physical model of the vibrating bar. The

TABLE I. Variables used in the bar model.

Symbols	Definitions (units)
$w(x, t)$	Vertical displacement (m)
E	Young's modulus (N/m ²)
ρ	Density (kg/m ³)
$S(x)$	Cross section (variable) (m ²)
$I(x)$	Moment of inertia of cross-section (m ⁴)
L	Length (m)
b	Width (m)
$h(x)$	Thickness (variable) (m)
γ_B	Fluid damping coefficient (s ⁻¹)
η	Viscoelastic damping coefficient (s)
α	Damping factor (s ⁻¹)
$f(x, x_0, t)$	Force per unit mass (m/s ²)

mechanical equations are taken from Chaigne and Doutaut (1997). They are presented only briefly here to underscore the choices of physical parameters that were made for the synthesis of the test stimuli. For a fuller development of the theory and implementation, in particular the interaction between mallet and bar and the acoustic radiation, see Chaigne and Doutaut (1997) and Doutaut *et al.* (1998). All parameters related to these parts of the model were held constant in our sound stimuli. Experiment 1 used parameter settings of the model for rectangular bars of constant cross-sectional geometry, whereas experiment 2 used parameters for tuned bars of variable cross section as are found in the xylophone and marimba. The notations used to represent the physical parameters in the equations for the bars are shown in Table I.

In this section the mechanical equations are presented, notably taking into account energy loss phenomena that are crucial to our study. Energy loss processes are ubiquitous in physical phenomena. In particular, any excited physical system will dissipate the received energy over the course of time in the form of internal or external exchanges. When a mallet excites a bar, part of the incident energy is dissipated into the air in the form of an acoustic wave, i.e., a sound. This external wave whose acoustic radiation is caused by the vibrating structure is one of the external exchanges. However, this loss is most often negligible compared with the losses due to viscosity within the material (except for metal bars of particular size, around and above the critical frequency). Chaigne and Lambourg (2001), for example, have shown that the fine-grained modeling of damping is complex, but that in the case of bars, an approximate description of damping with two constants (fluid and viscoelasticity) already gives good results for sound synthesis. The energy losses are reflected in the sound's temporal evolution and are consequently perceived by the listener as a progressive attenuation of the sound level and at times as a modification of the timbre of the sound. One of the aims of sound synthesis with physical models is to reproduce as closely as possible the signals arising from mechanically produced events. It is thus imperative for high fidelity synthesis to introduce energy loss phenomena. However, the physical justification and the mathematical formulation of these phenomena can be very complex. It is thus necessary to start with simple, general models. The validity of these models derives from both experimental physical measures and perceptual validation by

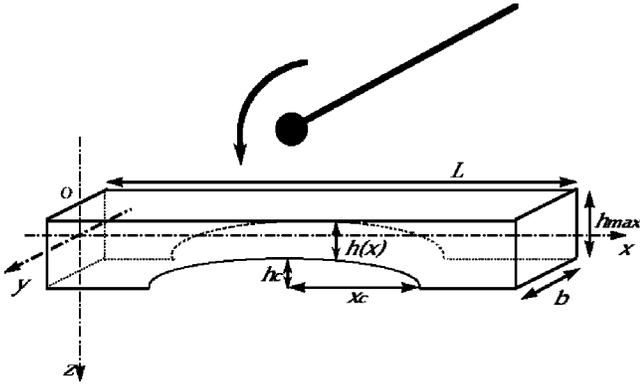


FIG. 1. Geometry of the xylophone bar.

listeners. The present study focuses on the latter aspect.

The physical system under consideration is composed of a bar of constant (or variable) cross section with free edges, of a mallet with a spherical head, and of a resonator. The diagram of the bar is presented in Fig. 1.

The bars are simulated with elastic supports as in a xylophone. One-dimensional Euler–Bernoulli bar equations are used to describe wave propagation in the material, which is presumed to be homogeneous and isotropic. This model is refined by taking into account the damping of the transverse bending waves with a model of viscoelastic loss. The vertical component $w(x, t)$ of the displacement of the bar is governed by the following equation:

$$\frac{\partial^2 w}{\partial t^2}(x, t) = \frac{E}{\rho S(x)} \left(1 + \eta \frac{\partial}{\partial t} \right) \frac{\partial^2}{\partial x^2} \left(I(x) \frac{\partial^2 w}{\partial x^2} \right) - \gamma_B \frac{\partial w}{\partial t} + f(x, x_0, t), \quad (1)$$

where E is Young’s modulus, ρ is the density, $S(x)$ is the rectangular cross section, which varies as a function of the position x along the longitudinal axis of the bar, η is the viscoelastic damping coefficient, $I(x)$ is the moment of inertia of the cross section, γ_B is the fluid damping coefficient, and $f(x, x_0, t)$ is the force per unit mass. As discussed in Chaigne and Doutaut (1997), γ_B determines (asymptotically) the losses in the lowest partials and η determines (asymptotically) the losses in the higher ranking partials. Their combination results in a decay envelope that can be characterized with two decay constants (see Sec. III).

For a rectangular cross section, we have

$$S(x) = bh(x) \quad \text{and} \quad I(x) = \frac{bh^3(x)}{12}. \quad (2)$$

Under the assumption of free-free boundary conditions, one obtains

$$\frac{\partial^2 w}{\partial x^2}(x, t) = 0, \quad \frac{\partial^3 w}{\partial x^3}(x, t) = 0 \Big|_{x=0, L}, \quad (3)$$

which means that force and moment vanish at both ends of the bar. We assume that the sound of an impacted bar is due to the flexural motion of the bar in the vertical x - z plane. This amounts to neglecting the contribution of torsional and longitudinal waves, as well as of flexural waves in the x - y plane. An additional assumption is that this motion can be ade-

quately described by the one-dimensional Euler–Bernoulli equations. Previous measures on xylophone bars show this assumption to be valid in the low-frequency range to a first degree of approximation (Rossing and Russell, 1990), the model leading to a tuning error of less than 2% for fundamental frequencies below 1.2 kHz (Doutaut, 1996).

Although for wood this is an approximation, the material is assumed to be homogeneous and isotropic. Since bars are usually cut in the direction of the fibers, Eq. (1) is acceptable if Young’s modulus, E in the equation, is taken equal to the longitudinal modulus of elasticity (Bucur, 1995).

The damping of the flexural waves is represented in Eq. (1) by two terms. The first one represents viscoelastic losses and is proportional to the coefficient η (in s). The relation between stress σ and strain ε in the material is expressed by

$$\sigma(x, t) = E \left(\varepsilon(x, t) + \eta \frac{\partial \varepsilon}{\partial t}(x, t) \right). \quad (4)$$

The second term is equivalent to a fluid damping. Its magnitude is represented by the coefficient γ_B (s^{-1}). Chaigne and Doutaut (1997) showed that the introduction of these damping terms in Eq. (1) yields a good representation of losses in wooden bars. The coefficients η and γ_B are estimated from the decay times of partials. The solutions of Eq. (1), decomposed on the basis of plane waves, allow us to write the solutions in the form of the product of a spatial function and a temporal function. In order to characterize damping, we define the damping factor α as the inverse of the time it takes for the amplitude to decay by a factor of $1/e$. Assuming damped sinusoidal waves as solutions to Eq. (1) imposes the following relation between damping and frequency:

$$\alpha(f) = \frac{\gamma_B}{2} + 2\pi^2 \eta f^2. \quad (5)$$

This relation allows us to compute the coefficients of the damping model by measuring the value of α for each partial in real bars.

III. DERIVATION OF ANALYTIC SIGNAL PARAMETERS

All sounds were synthesized in 16-bit mono format at a 48-kHz sampling rate. Three classes of analytic signal parameters were derived: temporal, spectral, and spectrotemporal. The temporal parameters were derived from the temporal envelope $[A_{HT}(t)]$, which was determined by taking the imaginary part of the Hilbert transform of the signal and filtering the absolute value of this imaginary part using a third-order Butterworth filter with a low-pass cutoff frequency of 50 Hz.

The spectral and spectrotemporal parameters were derived from one of two time-frequency representations. In the raw signal representation, a 4096-point fast Fourier transform (FFT) with a Hamming window was performed. To estimate the frequencies of the partials composing each sound, the beginning of this time window was positioned just after the attack of the sounds at 23 ms.

The second time-frequency representation simulates the output of the human cochlea. In this representation, the sig-

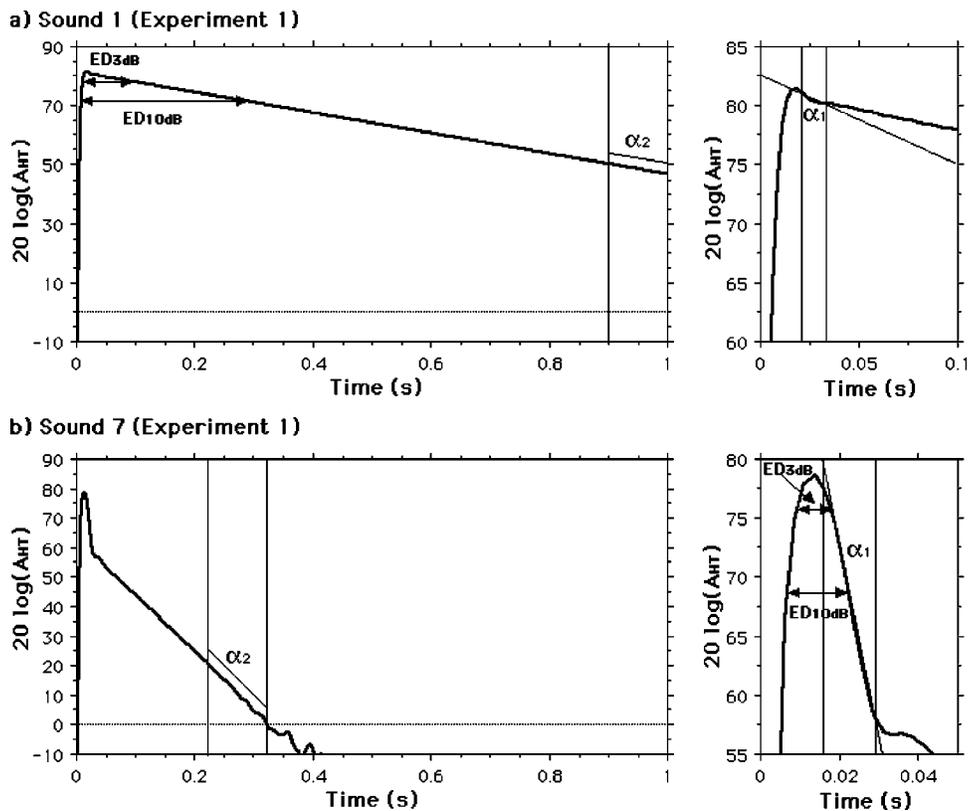


FIG. 2. Temporal envelopes for sounds 1 (a) and 7 (b) from experiment 1, representing the lowest and highest degrees of damping, respectively (see Table III). The derivation of decay constants α_1 , α_2 (slopes of lines fitted to portions of the curve) and equivalent durations ED_{3dB} , and ED_{10dB} from these functions is shown. The full envelope over the 1-s duration is shown on the left and a blowup of the initial portion of each envelope is shown on the right.

nal is first processed in a way that simulates the outer and middle ear filtering. The magnitude transfer function is deduced from behavioral measures of the minimum audible field (absolute auditory threshold; Killion, 1978). It is implemented as a cascade of an IIR high-pass filter to account for the sharp loss of sensitivity at low frequencies and a FIR equalizer (two second-order filters in cascade) to account for the peak of sensitivity around 2 kHz and the loss of sensitivity in the high frequencies. The filtered signal is then processed in several steps to obtain an approximation of the excitation pattern along the basilar membrane, function of the frequency locus (or characteristic frequency, CF) along the membrane and of time. In a first step, the waveform is multiplied by a window shaped as the time-reversed envelope of a gammatone impulse response (Patterson *et al.*, 1995) with an equivalent rectangular duration of about 20 ms. The windowed slices are Fourier transformed to obtain a power spectrum. The frequency resolution of this power spectrum is similar to that of the characteristic frequency with the narrowest cochlear filter. The power spectrum is remapped to get a resolution similar to the cochlea and channels evenly spaced on an equal-resolution scale (Moore and Glasberg, 1983). This scale is derived from psychophysical measures of auditory filter bandwidth (in equivalent rectangular bandwidth, or ERB, units) along the cochlear partition (the ERB-rate scale). Each channel of the new spectrum is the weighted sum of power spectrum coefficients, with weights designed to approximate the shape of a cochlear filter. This new power spectrogram has a frequency resolution and scale similar to that of the cochlea. Its temporal resolution is commensurable with that measured psychophysically (equivalent rectangular duration = 8–13 ms) (Plack and

Moore, 1990). All channels have the same temporal resolution. In this respect, it differs from the gammatone filter-bank which allows better temporal resolution at higher frequencies, but not at the lowest. Different temporal resolution in different channels makes the calculation of parameters such as the instantaneous spectral centroid unstable, so uniform temporal resolution is preferable. As a final step, the power in each channel is raised to the power 0.25 to approximate excitation or partial loudness (Hartmann, 1997).

A. Decay constants (α_1, α_2)

The temporal envelope of percussive sounds is often assimilated to a decaying exponential function of the form $A \exp(-at)$. Damping can thus be practically characterized by the decay constant α . This parameter can be estimated as the slope of the linear regression equation on $\ln[A_{HT}(t)]$. However, as Fig. 2 shows, the form of this envelope has two slopes, related to γ_B and η as discussed previously: one just after the maximum and another anywhere from several tens to several hundreds of ms later, respectively. The first, α_1 , was estimated over a time frame of 15 ms starting 2.5 ms after the envelope maximum. The second, α_2 , was estimated over a 100-ms time frame ending either at the end of the sound for those sounds that were artificially cut off at 1 s, or at the moment at which $\ln[A_{HT}(t)]=0$. The envelope generally becomes erratic below this value.

B. Equivalent durations (ED_{3dB}, ED_{10dB})

Since the damping strongly affects the perceived duration of the sounds, an additional analytic parameter related to the decay constant was computed. This parameter represents

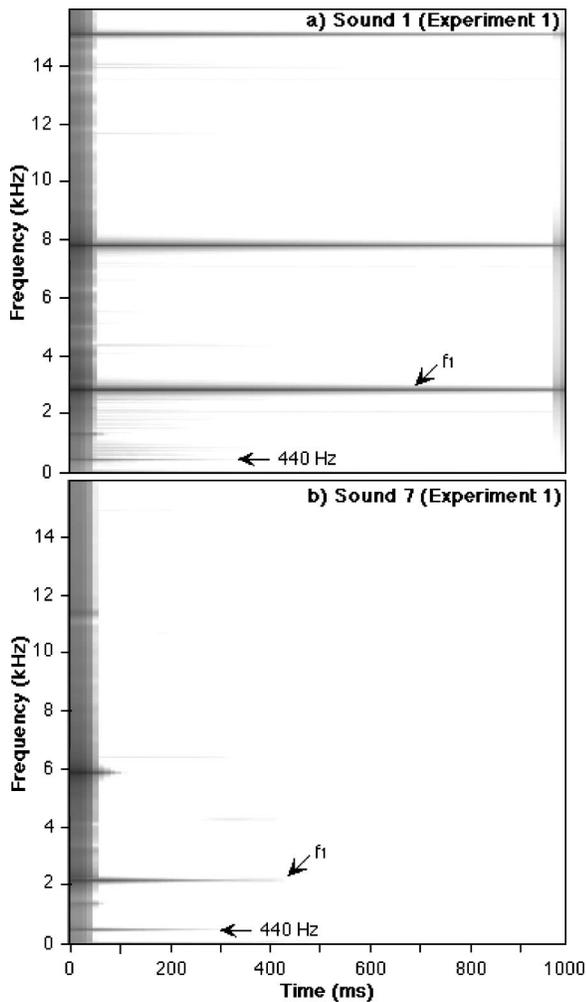


FIG. 3. Spectrograms of sounds 1 and 7 from experiment 1. The 440-Hz component due to the resonator and the lowest partial due to the bar (f_1) are shown.

the amount of time a sound event has a rms amplitude that exceeds a critical value that is expressed in decibels below the maximum (Fig. 2). It has been shown to be strongly correlated with the perceptual dimension that distinguishes continuously and impulsively excited sounds (McAdams *et al.*, 1995). Two critical values were tested here: 3 and 10 dB.

C. Frequency of the first partial (f_1)

The frequency of the first partial, f_1 , was estimated from the lowest prominent component in the FFT, the amplitude of which exceeded that of a constant 440-Hz component present in all stimuli due to excitation of the fixed-length resonator (see Fig. 3). Quite often, f_1 corresponded to the primary pitch heard in the bar sounds.

D. Spectral center of gravity (SCG)

The instantaneous spectral center of gravity is computed on the cochlear representation. For each time frame, the specific loudness-weighted average frequency on the ERB-rate scale is computed. The variation of this quantity over time gives the $SCG(t)$ function (Fig. 4). The mean SCG is com-

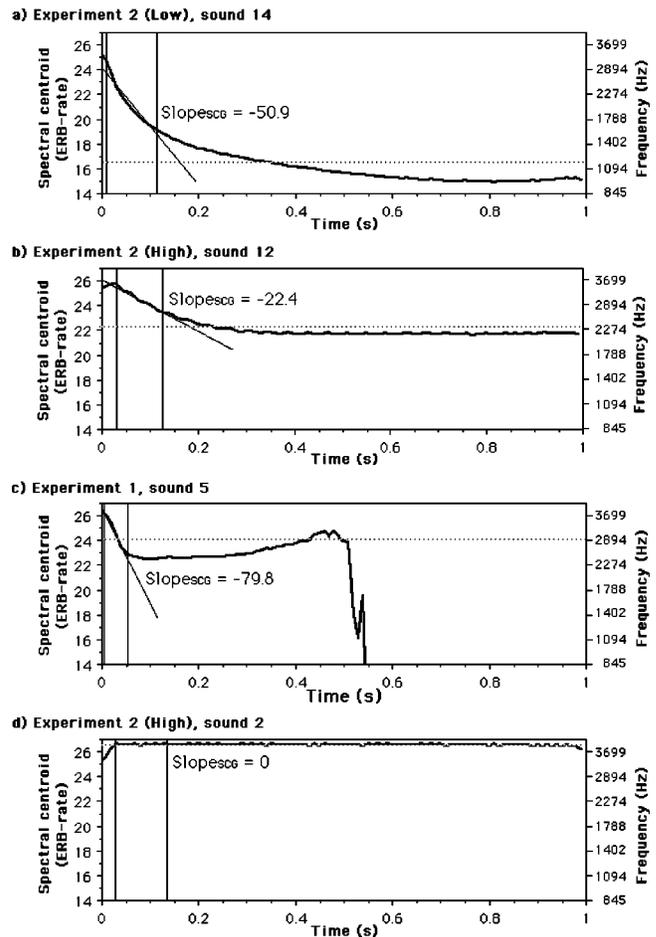


FIG. 4. Variation over time of the spectral center of gravity (SCG). The spectral centroid values are expressed in physiological units (ERB-rate) on the left and physical units (Hz) on the right. $Slope_{SCG}$ is determined from the regression of $SCG(t)$ onto a straight line over the time window between the two vertical lines. The mean SCG for each curve is shown by the horizontal dotted line.

puted on this array of values over that portion of the signal for which $\ln[A_{HT}(t)] > 0$ to avoid including less reliably determined SCG values in low-level parts of the signal essentially represented by fluctuation in the lowest bit.

E. Slope of $SCG(t)$ ($Slope_{SCG}$)

One feature of bar sounds is that damping depends on frequency. Given that the damping increases as the square of the frequency, for a given value of η , the instantaneous SCG of the bar without the resonator decreases monotonically over time and eventually asymptotes, in many cases near the fundamental frequency. The rate of decrease in instantaneous SCG is thus always greater at the beginning of its temporal evolution [Fig. 4(a)]. The $SCG(t)$ curve characterizes this rate of decrease over an initial 100-ms time frame. Due to the short-duration presence at the beginning of all the sounds of the low-amplitude 440-Hz component resulting from the presence of the fixed-length resonator, the $SCG(t)$ curve sometimes rises briefly before beginning its descent, particularly for sounds with higher-frequency partials such as those in experiment 1 and the high sounds in experiment 2 [Figs.

4(b) and (d)]. Since the behavior of this component is not related to the material of the bar, this part of the curve was ignored. The slope was thus measured starting 2.5 ms after the initial maximum or initial plateau. In a few cases, the curve asymptotes before the 100-ms time frame has passed. In such cases, the window over which the slope was estimated was reduced so as to end at the bend in the curve [Fig. 4(c)]. [The fact that the SCG(t) curve increases beyond 0.35s in Fig. 4(c) should be ignored because the amplitude of the signal is close to zero beyond this point.] Finally, in three stimuli in experiment 1, the curves were flat after the initial rise and their slopes were thus set to 0 [Fig. 4(d)].

IV. GENERAL METHOD

A. Procedure

The listener's task was to rate the perceived dissimilarity between two stimuli according to any criteria that seemed perceptually salient. The aim is to discover what listeners hear as being salient, not to impose a particular judgment criterion. This rating was to be performed on all pairs of different sounds within the stimulus set for a given experiment. In order to give the listener a sense of the range of variation to be encountered and thus to make the best use of the rating scale, all of the sounds were presented in random order at the beginning of each experiment. Then ten training trials, chosen at random from the experimental set, were presented to provide practice with the rating scale. On each trial, two sounds were presented, separated by a 1-s silence. In the experimental block, each pair of different sounds was presented once. The order of presentation of the two sounds was randomly chosen for each listener. The dissimilarity rating was indicated by moving a cursor on the computer screen along a horizontal scale labeled "very similar" at the left and "very dissimilar" at the right. The end points of this continuous scale were coded as 0 and 100, respectively. The cursor was moved by clicking on it and dragging it with a mouse. There were also two buttons available on the computer screen that could be activated by clicking on them: one for replaying the sound pair as many times as necessary to reach a satisfactory rating, the other for recording the position of the cursor and moving to the next trial. The ratings were compiled into a triangular matrix without diagonal, which constituted the data set for each listener.

B. Multidimensional scaling of dissimilarity ratings

The analysis of the dissimilarity data was performed with multidimensional scaling, using an extension of the CLASCAL program (McAdams *et al.*, 1995; Winsberg and De Soete, 1993). This analysis represents the relations among sounds with a distance model.

CLASCAL uses an extended Euclidean distance model that includes positions in a Euclidean space shared by all of the sounds, specific dimensions for individual sounds (called "specificities"), and weights on each of the dimensions and on the set of specificities for each latent class of listeners. The distance between sounds i and j for class t is written

$$d_{ijt} = \left[\sum_{r=1}^R w_{tr}(x_{ir} - x_{jr})^2 + \nu_t(s_i + s_j) \right]^{1/2}, \quad (6)$$

where x_{ir} is the coordinate for stimulus i along dimension r , w_{tr} is the weight on dimension r for class t , s_i is the specificity of stimulus i , and ν_t is the weight on the whole set of specificities for class t .

We do not know the number of latent classes in advance. We use a Monte Carlo significance testing procedure proposed by Hope (1968) and first applied in the context of latent class analysis by Aitken *et al.* (1981). [See McAdams *et al.* (1995) for a more detailed description of the procedure as applied to timbre perception.]

For highly nonlinear models such as those used in multidimensional scaling, information-based statistics cannot be used reliably to determine the appropriate spatial model, i.e., the appropriate dimensionality of the Euclidean common space, and the absence or presence of specificities. To choose the model, the following procedure was used:

Step 1: Choose the appropriate number of latent classes, using the Hope procedure on the fully saturated or null model (i.e., the matrix of mean dissimilarities between all pairs of stimuli).

Step 2: Choose two competing spatial models by selecting the two models with lowest BIC statistics from models with generally from about one to five or six dimensions with and without specificities. BIC (Schwarz, 1978) is an information-based criteria for model selection. $BIC = -2 \log L + \log(n_o)n_p$, (where $\log L = \log$ likelihood, n_o = the number of observations, and n_p = the number of model parameters).

Step 3: Using the number of latent classes selected in step 1 above, choose the appropriate spatial model from the top two or three competing models, perhaps also including an additional model that is of interest on theoretical grounds, using the Hope procedure.

Step 4: Verify the appropriate number of latent classes using the Hope procedure as in step 1, but using the spatial model selected in step 3.

Step 5: If the number of latent classes selected in step 4 is identical to the number selected in step 1, STOP; otherwise repeat steps 2 and 3 with the number of classes selected in step 4, and repeat step 4 with the newly selected model.

Since one-class solutions are rotationally invariant, they are subsequently rotated to an extended, weighted Euclidean model (EXSCAL: Winsberg and Carroll, 1988, 1989) in which weights are computed for each subject. This has the effect of orienting the axes along psychologically meaningful dimensions as represented by the directions of maximum variation across subject weights. Note that BIC cannot reliably choose between an N -class model and a one-class model since the degrees of freedom of the former model increase with the number of subjects. So in the one-class case, steps 2, 3, and 4 are reduced to spatial model selection and rotation to the fully weighted model solution.

TABLE II. Parameters held constant in the bar model in experiment 1. The columns represent the name of the parameter, its symbol, its value, and the physical units.

Parameter	Symbol	Value	Units
Bars			
Young's modulus	E	9.29×10^9	N/m^2
Moment of inertia	I	2.67×10^{-5}	m^4
Fluid damping	γ_B	8.0	s^{-1}
Thickness	h	0.02	m
Width	B	0.04	m
Length	L	0.20	m
Cross-section	S	8×10^{-4}	m^2
Mallet			
Mass	m_m	0.012	kg
Radius	R_m	0.012	m
Stiffness	K	4.2×10^8	$\text{Nm}^{-3/2}$

V. EXPERIMENT 1: BARS OF CONSTANT CROSS-SECTION

A. Method

1. Subjects

A group of 19 subjects (9 women and 10 men) participated in the experiment. Their ages ranged from 21 to 40 years. No subject was a professional musician. All reported having normal hearing. They were paid for their participation.

2. Stimuli

Sixteen sounds were synthesized with a bar model with constant cross-section. All sounds were equalized in loudness by an adjustment procedure performed with five listeners. The final levels corresponded to the mean of the adjusted levels.

The geometric and elastic parameters of the bar (E, I, ρ, h, b, L, S) affect the frequencies of the spectral components of the sound and thus the perceived pitch. We varied density (ρ) and held all the other parameters constant (Table II). The mallet and resonator parameters were also held constant. Density varied linearly in the interval $[300, 900 \text{ kg/m}^3]$. The values were chosen so that at least a quarter-tone interval (3% difference in frequency) existed between the lowest partial of adjacent sounds on this continuum, ensuring their discriminability in terms of pitch.

The damping parameters (γ_B, η) affect the spectral distribution and the temporal envelope of the sound. As previously shown in Eq. (5), the effect of the fluid damping coefficient γ_B is to limit the decay time in the low-frequency range (the lowest partial of the bars), whereas the effect of the viscoelastic coefficient η is to reproduce the decrease of the decay time with increasing frequency observed in real bars. We thus chose to vary η as a second parameter, which directly affects the temporal decay as a function of frequency. The damping factor was varied exponentially in the interval $[0.1, 200.0 \text{ ns}]$. This interval of damping coefficients covers a large range of materials from glass to wood, including carbon and some plastic materials. It would not include metals for which the laws of variation of damping as a function of frequency are much more complex. It was determined

in a pilot study that this distribution gave a reasonably homogeneous spread of perceived differences between adjacent values. Beyond this range, the sounds created with the constant cross-section model become perceptually indistinguishable: too highly damped sounds are perceived as clicks, and too lowly damped sounds are perceived as being constant over the duration used in our stimuli.

In order to have 16 different values for each of the physical parameters with only 16 sounds, we chose a homogeneous pseudo-random distribution within a 16×16 matrix in the physical space defined by density and damping (the latter being on a log scale). The distribution allows us to have 16 values on each parameter and ensures that all pairs of sounds are distinguishable while covering the desired physical space. The exact values of the model parameters are listed in Table III as are the corresponding analytic parameters.

3. Procedure

All distinct pairs of the 16 stimuli were presented for a total of 120 experimental trials. The experiment lasted approximately 40–50 min per subject.

4. Apparatus

The experiment took place at Ircam. Listeners were seated in a Soluna S1 double-walled sound-isolation chamber. The sounds were reproduced with a NeXT computer equipped with an ISPW sound-processing card and the Max sound-synthesis software (Lindemann *et al.*, 1991). The digital output was converted to an analog signal with a ProPort DAC equipped with anti-aliasing filters. The signal was then amplified with a Canford stereo amplifier and presented diotically over AKG-1000 open-air headphones. The experimental program, sound presentation, subject interface, and data collection were programmed with the PsiExp (Smith, 1995) software environment.

B. Results

1. CLASCAL analyses

The CLASCAL analysis yielded a model with two dimensions, no specificities, and a single latent class. This model is invariant under rotation and was thus rotated to the N -class model. The distances between the sounds in the model explain 92% of the variance in the mean dissimilarity ratings.

Four main points emerge from this analysis. First, the dimensionality of the perceptual space is identical to that of the physical space. Second, listening to the experimental stimuli in order as projected onto each of the perceptual dimensions reveals a variation in timbre along dimension 1 and in pitch along dimension 2. The timbre variation appears to have a temporal component and a spectral component. At one end of this dimension, sounds ring longer and are brighter. At the other end, they are quickly deadened and duller sounding. Third, since the fit of distances to dissimilarities is performed with a linear metric, we can compare the relative extent of variation along the two dimensions. The range of variation along dimension 1 is 1.75 times larger

TABLE III. Experiment 1: Summary of mechanical parameters used in the physical model (η, ρ) for a constant-section rectangular bar, of descriptors derived from the resulting signals ($\alpha_1, \alpha_2, ED_{3dB}, ED_{10dB}, SCG, Slope_{SCG}$), and of perceptual coordinates derived from the CLASCAL analysis of dissimilarity ratings (Dim1, Dim2).

Sound	Mechanical		Signal descriptors							Perceptual	
	η (ns)	ρ (kg/m ³)	α_1 (s ⁻¹)	α_2 (s ⁻¹)	ED_{3dB} (ms)	ED_{10dB} (ms)	SCG (ERB-rate)	Slope _{SCG} (ERB-rate/s)	f_1 (Hz)	Dim1	Dim2
1	0.1	300	8.6	4.0	72	277	26.7	0.0 ^a	2857	-50.9	32.2
2	70.0	340	35.4	15.8	14	67	24.3	-24.0	2685	22.8	22.9
3	4.0	380	15.5	4.5	41	157	26.2	-1.8	2543	-44.9	19.4
4	40.0	420	50.1	9.8	12	35	25.1	-28.5	2419	0.8	25.1
5	150.0	460	97.3	21.5	9	17	24.1	-79.8	2313	35.5	11.3
6	6.0	500	13.1	4.6	388	140	26.6	-3.3	2220	-41.6	5.8
7	200.0	540	184.0	23.0	9	12	24.5	-109.0	2138	45.9	4.1
8	50.0	580	47.0	9.2	13	37	25.0	-29.3	2062	14.1	-2.0
9	20.0	620	24.7	5.8	21	75	26.0	-7.9	1992	-13.3	-4.5
10	2.0	660	9.7	5.0	58	222	27.9	0.0 ^a	1922	-44.1	-10.5
11	120.0	700	71.4	12.6	12	25	24.5	-45.5	1874	28.3	-11.2
12	1.0	740	9.1	4.4	39	250	27.8	0.0 ^a	1816	-45.8	-14.6
13	30.0	780	27.4	6.2	19	67	25.5	-8.9	1781	-3.3	-23.3
14	100.0	820	62.6	11.6	12	27	24.3	-41.6	1734	26.6	-19.1
15	90.0	860	55.0	10.5	13	32	24.2	-35.7	1693	29.6	-21.5
16	180.0	900	108.3	17.3	10	17	23.8	-78.6	1658	40.2	-14.6

^aThere was some difficulty measuring accurately the value of Slope_{SCG} for some sounds in which SCG(t) was flat or very slightly increasing.

than that along dimension 2, suggesting a perceptual dominance of dimension 1. For this set of sounds, listeners are thus collectively more sensitive to timbral variation than to pitch variation.

2. Correlational analyses

The mechanical and analytic signal parameters as well as the coordinates in the perceptual space are listed in Table III. The acoustic information related to the mechanical parameters is presumed to be carried by the analytic signal parameters (i.e., derivable from the sensory representation of the acoustic signal). Note that in many cases the relation between the parameters is not linear. For example, f_1 is inversely proportional to the square-root of ρ . Over the range of variation of f_1 in this stimulus set, linear and logarithmic relations between f_1 and dimension 2 give the same correlation coefficient. For this reason, only the linear relation is used throughout the paper. The transformation of the objective parameters to achieve a good fit to the perceptual coordinates can be a logarithmic (α, ED) or a power function ($\rho, \eta, Slope_{SCG}$). In the case of the latter, a function of the form $\Psi = a + b\varphi^c$ was used to fit the physical parameters (φ) to the perceptual dimensions (Ψ), leaving a, b , and c as free parameters. In this way, exponents of 0.33 and 0.39 were found for the relation of dimension 1 to η and to Slope_{SCG}, respectively.

The correlations among $\eta^{0.33}, \log(\alpha_1), \log(\alpha_2), SCG, ED_{3dB}, ED_{10dB}, Slope_{SCG}^{0.39}$, and dimension 1, on the one hand, and among $\rho^{-1/2}, f_1$, and dimension 2, on the other, are in most cases very high [$0.87 \leq |r(14)| \leq 1.00, p \leq 0.0001$, in all cases], whereas those between these two groups of parameters are generally low [$0.05 \leq |r(14)| \leq 0.35, p > 0.19$ in all cases], attesting to the perceptual and acoustic independence of the two groups.

Several analytic parameters ($SCG, \alpha_1, \alpha_2, ED_{3dB}, ED_{10dB}, Slope_{SCG}$) are partially correlated with dimension 1

and η . Therefore, a stepwise multiple regression with dimension 1 as dependent variable and the appropriate power or log transforms of these parameters as independent variables was performed. The aim was to evaluate the relative contribution of each analytic parameter to a possible combined perceptual effect represented as dimension 1. The statistic of interest is the coefficient of determination, adjusted for the inclusion of several regressors (R_{adj}^2), which is considered to represent the variance in the dependent variable explained by the linear combination of independent variables. Forward and backward selection procedures were used to verify whether they converged on the same set of regressors. The forward selection starts with no independent variables in the model and adds them in order of their ability to predict the dependent variable based on the partial F -ratio. It stops when the addition of a new parameter is not statistically significant. The backward selection starts with all independent variables in the model and removes the one that is least useful in predicting the dependent variable. It stops when the least useful independent variable still makes a significant contribution to predicting the dependent variable.

For forward selection, once Slope_{SCG} was entered into the regression, no other parameter made a significant enough contribution to enter subsequently ($R_{adj}^2=0.92$). This parameter is also very highly correlated with $\eta^{0.33}$ ($R_{adj}^2=0.96$). Scatter plots of these regressions are shown in Fig. 5(a). The backward selection only removes two of the six analytic parameters leaving Slope_{SCG}, SCG, α_2 , and ED_{10dB} ($R_{adj}^2=0.97$). Note, however, that the increase in explained variance with the addition of the last three parameters is only 5%, due primarily to the fact that the correlations among the predictor variables are quite high, as mentioned previously. This perceptual dimension would thus seem to be primarily explained by a spectrotemporal component that is closely correlated both with the mechanical parameter being varied and the corresponding perceptual dimension, confirming the

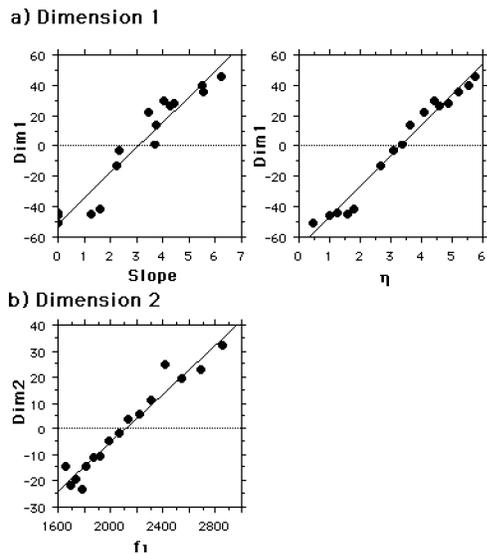


FIG. 5. Experiment 1. Linear regressions of coordinates along one dimension of the perceptual space onto signal descriptors of mechanical parameters. (a) Dimension 1 compared to $\text{Slope}_{\text{SCG}}^{0.39}$ and $\eta^{0.33}$. (b) Dimension 2 compared to f_1 .

triangular relation between mechanical processes, acoustic information, and perceptual result.

For the second perceptual dimension, closely related to the density parameter of the model, and thus to the frequency composition of the bar sounds, a highly significant correlation between f_1 and the coordinates along dimension 2 is found, confirming the impression that this dimension is related to pitch [Fig. 5(b)]. As mentioned previously, pitch variation was rated as occupying a smaller range of perceptual variation than the timbral variation related to the damping parameter. The range of pitches represented by f_1 is situated between 1658 and 2857 Hz (about 9.5 semitones), and adjacent pitches in the stimulus set were separated by intervals ranging from slightly more than a semitone (sounds 1 and 2:107 cents) to slightly less than a quarter-tone (sound 15 and 16:36 cents). It should be noted that no difference in correlation coefficient was found for the comparisons between dimension 2 coordinates and either linear or log frequency scales for f_1 . Indeed, due to the small range of variation of f_1 , its log and linear versions are highly correlated in a linear regression ($R^2=0.99$).

C. Discussion

The multidimensional scaling analysis of dissimilarity ratings, made by listeners on a set of sounds synthesized by a physical model, revealed a perceptual space that had the same dimensionality as the original mechanical space. Transformations of the mechanical parameters were related to potential signal parameters that could be extracted by the auditory system and perceived as perceptual dimensions of the stimuli. Strong correlations were found among all three domains. The nature of the transforms provides important information about the nature of the sensory or perceptual representation of these complex, naturalistic stimuli. This representation would then be the basis upon which stimuli could be compared and perhaps analyzed in order to recog-

nize the material properties of the vibrating objects. One of the surprising results of this experiment is that simultaneous variation in pitch did not annihilate the contribution of timbre to a global impression of dissimilarity among sounds. On the contrary, the timbral variation was much stronger in the resulting perceptual space than was the pitch variation. To pursue further this latter point, a second experiment was conducted on tuned bars in which the range of variation of the damping parameter was reduced and that of the perceived pitch was increased.

VI. EXPERIMENT 2: BARS OF VARIABLE CROSS-SECTION

To achieve the necessary range of f_1 variation in experiment 2, a model of tuned xylophone bars was used. The tuning of the partial structure was achieved by varying the cross-sectional geometry as shown in Fig. 1, and f_1 variation was created by modifying the length (L) of the bar. All of the other geometrical and mechanical properties affecting the frequency content were held constant. Two sets of bars were tested. One set (referred to as high) had modal frequencies in the same range as those of experiment 1. In these sounds, the lowest partial's frequency was at least a major sixth (a ratio of 1.68) above the resonance frequency of the fixed-length tube resonator (440 Hz). The other set (referred to as low) was created by using bar lengths that were four times those of the first set, giving lowest partials in a range near that of the resonator. The aim with this set was to study the effects of bar-resonator coupling on material perception, particularly since such coupling affects the decay characteristics of the sound which were shown in experiment 1 to play a possible role in the perception of sound properties related to the materials.

A. Method

1. Subjects

The experiment was performed in two separate sessions corresponding to the two sets of bar sounds by 21 subjects (13 men, 8 women) aged from 22 to 40 years. None of them had participated in experiment 1. No subject reported having hearing problems and none was a professional musician. All were paid for their participation.

2. Stimuli

Fourteen sounds were synthesized in each set using the same set of viscoelastic damping coefficients (η), but different sets of bar lengths. In light of the results from experiment 1, η was varied in experiment 2 according to a power function in order to obtain a relatively constant perceptual difference between adjacent values along this continuum. The range of variation of η on a log scale is slightly smaller in experiment 2. On the scale of the power function from experiment 1 with an exponent for η of 0.33, the perceptual range for experiment 2 is 66% of that for experiment 1. All other material properties were held constant and took the values specified in experiment 1 except ρ , which was held constant at a value of 782 kg/m^3 .

TABLE IV. Experiment 2: Summary of mechanical parameters used in the physical model (η , L) for a variable-section tuned bar, of descriptors derived from the resulting signals (α_1 , α_2 , ED_{3dB} , ED_{10dB} , SCG, $Slope_{SCG}$), and of perceptual coordinates derived from the CLASCAL analysis of dissimilarity ratings (Dim1, Dim2) for both High (a) and Low (b) sound sets.

Sound	Mechanical		Signal descriptors							Perceptual	
	η (ns)	L (cm)	α_1 (s ⁻¹)	α_2 (s ⁻¹)	ED_{3dB} (ms)	ED_{10dB} (ms)	SCG (ERB-rate)	$Slope_{SCG}$ (ERB-rate/s)	f_1 (Hz) [musical pitch]	Dim1	Dim2
(a) High sound set											
1	0.03	14.68	8.7	3.9	73	280	27.5	-0.1	1659 [G#5]	-8.2	-31.2
2	0.06	13.46	8.0	4.0	74	280	26.5	0.0 ^a	1973 [B5]	-29.3	-13.3
3	0.18	12.34	9.0	4.1	70	270	25.9	-0.3	2347 [D6]	-34.4	2.2
4	0.32	14.26	8.8	4.2	96	263	26.9	-0.5	1757 [A5]	-12.2	-24.5
5	0.56	13.08	9.3	4.6	63	240	26.8	-1.1	2090 [C6]	-24.1	-6.3
6	1.00	9.79	9.6	4.2	63	250	27.5	-1.9	3730 [A#6]	-11.7	43.8
7	1.78	15.55	9.9	4.8	58	220	25.9	-3.0	1478 [F#5]	-14.2	-35.5
8	3.16	10.68	10.5	4.6	56	230	25.9	-6.5	3135 [G6]	-17.1	33.2
9	5.62	11.64	10.5	4.7	56	228	24.5	-7.6	2635 [E6]	0.7	21.7
10	10.00	12.70	25.4	4.9	21	108	23.4	-15.2	2215 [C#6]	17.9	-8.3
11	17.78	13.85	30.9	5.2	17	70	22.2	-22.4	1862 [A#5]	21.7	-19.3
12	23.70	11.31	12.4	6.8	43	168	24.7	-6.5	2792 [F6]	16.2	25.6
13	31.62	11.00	33.9	11.0	13	88	24.9	-7.5	2958 [F#6]	46.7	11.8
14	56.23	12.00	27.2	12.6	16	93	23.8	-11.7	2486 [D#6]	48.1	0.1
(b) Low sound set											
1	0.03	29.36	8.9	4.0	75	285	25.6	-11.2	415 [G#3]	-33.6	-14.9
2	0.06	26.92	8.1	4.0	80	288	24.7	-12.7	493 [B3]	-28.2	-21.3
3	0.18	24.68	7.3	4.1	83	288	24.5	-9.6	587 [D4]	-27.2	7.6
4	0.32	28.52	7.1	4.2	75	273	25.1	-11.1	440 [A3]	-36.9	-11.8
5	0.56	26.16	8.7	4.3	75	265	24.3	-9.2	523 [C4]	-18.9	-9.1
6	1.00	19.02	16.6	4.1	66	255	24.1	-3.0	987 [B4]	-18.5	35.5
7	1.78	31.10	10.2	4.6	53	203	24.3	-15.4	370 [F#3]	-25.0	-31.0
8	3.16	21.36	14.7	4.3	52	213	22.7	-7.2	784 [G4]	4.9	29.0
9	5.62	23.28	12.3	4.1	52	173	21.6	-17.0	659 [E4]	8.8	12.9
10	10.00	25.40	19.0	3.9	35	150	20.3	-24.4	554 [C#4]	20.0	-9.1
11	17.78	27.70	26.7	3.9	19	88	18.5	-36.0	466 [A#3]	29.1	-19.9
12	23.70	22.62	23.8	4.2	23	113	19.5	-25.4	698 [F4]	33.1	12.2
13	31.62	22.00	29.7	4.3	19	108	18.4	-32.5	740 [F#4]	43.6	19.1
14	56.23	24.00	36.1	4.4	15	133	16.5	-50.9	622 [D#4]	48.8	1.0

^aThere was some difficulty measuring accurately the value of $Slope_{SCG}$ for one sound in which $SCG(t)$ was flat.

The width (b) and maximum thickness (h_{max}) were set to the values from experiment 1: 4 cm and 2 cm, respectively. The other geometrical parameters specified a variable cross section to account for the tuning of the bar's partials to a harmonic frequency spectrum (Fig. 1). To change the pitch of the bar sounds, the length of the bar was varied. The lengths (L), corresponding frequencies (f_1), and resulting musical pitches are listed in Table IV. Note that the range of variation of f_1 on a semitone scale is 158% of that in experiment 1 (15 vs 9.5 semitones). Thus, in this experiment the range of pitch variation is greater and the range of variation in the damping factor is less. The height (h_c) and radius (x_c) of curvature (see Fig. 1) vary with L , but the ratios x_c/L and h_c/h_{max} are held constant. This ensures that the tuning of the first three partials of the bar is constant. Their frequencies were in a ratio of 1:4:9 and therefore formed a harmonic series (in contrast to the stimuli of experiment 1 which were inharmonic). These parameters are similar to those used by instrument makers in the manufacture of mallet percussion instruments. The resonator geometry was also held constant. The tube length was $L_T=0.179$ m. The tube radius was $a_T=0.02$ m. The bar-tube distance was $d=0.015$ m.

3. Procedure

For each sound set, all distinct pairs of the 14 stimuli were presented for a total of 105 experimental trials. Each session lasted approximately 30–40 min per subject. The two sessions were performed on the same day in counterbalanced order, separated by a break.

4. Apparatus

The experiment took place at the ENST. Listeners were seated in a sound-isolated room. The sounds were reproduced with a Sun computer equipped with an integrated sound card and DAC. The signal was then amplified with a Rotel stereo amplifier and presented diotically over Sennheiser headphones connected directly to the output of the DACs. The experimental program, sound presentation, subject interface, and data collection were programmed with the Matlab software environment.

B. Results

1. CLASCAL analyses

a. High sound set. The CLASCAL analysis yielded two dimensions without specificities and three latent classes.

Since the orientation is fixed in this model due to the class weights, no rotation was necessary. The perceptual coordinates of each sound are listed in Table IV(a). The model distances explain 85% of the variance in the mean dissimilarity ratings. Five subjects were in class 1, six in class 2, and nine in class 3. Class belongingness as tested by *a posteriori* probabilities with Bayesian statistics was very clear. Posterior probabilities were greater than 0.89 in all cases except one for which classification was ambiguous with respect to classes 2 and 3. The weights were very strong (1.75) on dimension 1 and very weak (0.40) on dimension 2 for class 1, very weak (0.38) on dimension 1 and moderate (1.15) on dimension 2 for class 2, and moderately weak (0.87) on dimension 1 and strong (1.45) on dimension 2 for class 3. The weightings of the two dimensions are thus quite different across classes.

b. Low sound set. The CLASCAL analysis yielded two dimensions without specificities and two latent classes. The model distances explain 86% of the variance in the mean dissimilarity ratings. Class 1 contained 14 of the 21 subjects, the other 7 being in class 2. Class belongingness was very clear, all posterior probabilities being greater than 0.99. The weights on dimensions 1 and 2 were 0.55 and 1.70 for class 1, and 1.45 and 0.30 for class 2, respectively. Listeners in class 1 thus weight dimension 2 much more heavily, while those in class 2 weight dimension 1 more heavily. The perceptual coordinates of each sound are listed in Table IV(b). In spite of the differences in bar length and bar-resonator coupling, the perceptual spaces are very similar for high and low sounds (Dim1 high vs Dim2 low, $R^2=0.90$; Dim2 high vs. Dim1 low, $R^2=0.80$, $df=12$ in both cases).

2. Correlational analyses

The fitting of η and $\text{Slope}_{\text{SCG}}$ to the appropriate perceptual dimensions according to a power function gave exponents of 0.28 for η in both sound sets and of 0.46 and 0.63 for $\text{Slope}_{\text{SCG}}$ in high and low sets, respectively.

a. High sound set. The correlations among $\eta^{0.28}$, $\log(\alpha_1)$, $\log(\alpha_2)$, SCG, $\log(ED_{3\text{dB}})$, $\log(ED_{10\text{dB}})$, $\text{Slope}_{\text{SCG}}^{0.46}$, and dimension 1, on the one hand, and among $\rho^{-1/2}$, f_1 , and dimension 2, on the other, are moderate to high [$0.50 \leq |r(12)| \leq 1.00$, $p \leq 0.07$, in all cases]. The lower correlations are due to $\log(\alpha_2)$ in comparison with SCG ($r = -0.50$, $p = 0.07$) and $\text{Slope}_{\text{SCG}}$ ($r = 0.53$, $p = 0.05$), all others being greater than 0.68 ($p \leq 0.0055$). Correlations across these two groups of parameters are generally low [$0.01 \leq |r(12)| \leq 0.31$, $p > 0.28$ in all cases]. Stepwise regression of the analytic parameters related to damping onto dimension 1 with forward selection yielded a linear combination of $\log(ED_{3\text{dB}})$ and $\log(\alpha_2)$ ($R_{\text{adj}}^2=0.86$), whereas the backward selection yielded a combination of $\log(\alpha_1)$ and $\log(\alpha_2)$ ($R_{\text{adj}}^2=0.88$) [Fig. 6(a)]. Stepwise regression of the parameters on $\eta^{0.28}$ yielded a combination of $\log(\alpha_2)$ and $\text{Slope}_{\text{SCG}}$ ($R_{\text{adj}}^2=0.97$) for both forward and backward selection. Nearly 94% of the variance along dimension 2 is predicted by f_1 [Fig. 6(b)].

b. Low sound set. The correlations among damping-related parameters and dimension 1, on the one hand, and among length-related parameters and dimension 2, on the

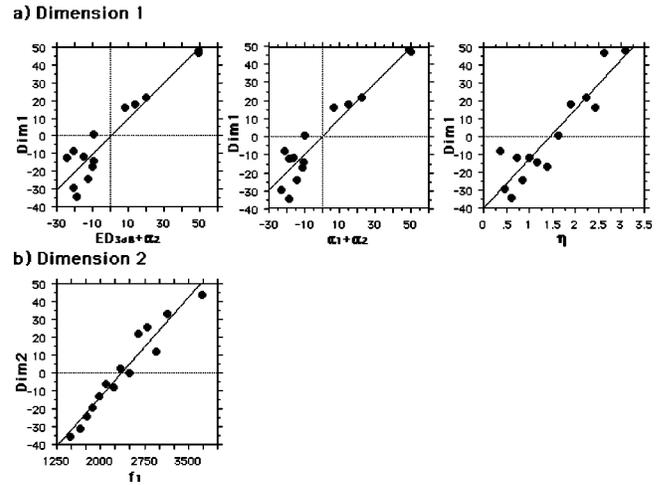


FIG. 6. Experiment 2 (high bars). Linear regressions of coordinates along one dimension of the perceptual space onto signal descriptors of mechanical parameters. (a) Dimension 1 compared to linear combinations of $ED_{3\text{dB}}$ and $\log \alpha_2$ and of $\log \alpha_1$ and $\log \alpha_2$, and to $\eta^{0.28}$. (b) Dimension 2 compared to f_1 .

other, are quite high [$0.83 \leq |r(12)| \leq 1.00$, $p \leq 0.0001$, in all cases], if $\log(\alpha_2)$ is excluded. Due to the prolonged resonance of these sounds, this parameter does not vary much across the sound set. The correlations between the two groups of parameters are generally low [$0.03 \leq |r(12)| \leq 0.42$, $p > 0.13$ in all cases]. Stepwise regression of damping-related analytic parameters on dimension 1 gave the same result with forward and backward selection: a combination of SCG and $\text{Slope}_{\text{SCG}}$ ($R_{\text{adj}}^2=0.98$). Stepwise regression of these parameters on $\eta^{0.28}$ yielded a combination of SCG and $\log(\alpha_2)$ ($R_{\text{adj}}^2=0.98$) for both selection procedures [Fig. 7(a)]. As with the other sound set, the pitch dimension (Dim2) is highly correlated with f_1 ($R^2=0.91$) [Fig. 7(b)].

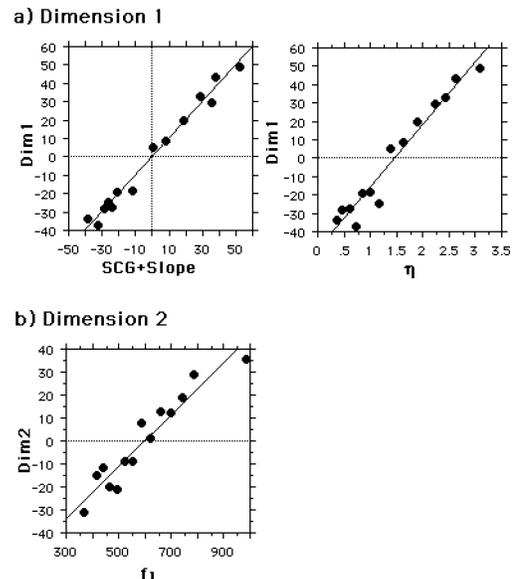


FIG. 7. Experiment 2 (low bars). Linear regressions of coordinates along one dimension of the perceptual space onto signal descriptors of mechanical parameters. (a) Dimension 1 compared to a linear combination of SCG and $\text{Slope}_{\text{SCG}}^{0.63}$, and to $\eta^{0.28}$. (b) Dimension 2 compared to f_1 .

C. Discussion

As in experiment 1, the dimensionality of the perceptual spaces corresponds to that of the physical spaces. A variation in timbre occurs along dimension 1 related to damping and a variation in pitch along dimension 2 related here to bar length. In spite of the smaller variation of the damping parameter and the larger variation of the frequency parameter in the stimulus set, with each sound separated from its neighbor by at least a semitone, the range of values occupied by the pitch dimension in the perceptual space is still less than that occupied by the timbre dimension. The factor has nonetheless increased to 78% for the low sounds and 96% for the high sounds, compared to 57% in experiment 1. In the present experiment, the bars were tuned to a harmonic series by using a variable cross-section model. Further, the sensation of damping was lessened by the fact that the fundamental had a much lower damping than in experiment 1. This effect is even more prominent for the low sounds in which the sound is prolonged by a coupling between the bar and the resonator. Note that the range of the α_2 decay constant was 4.0–23.0 s⁻¹ in experiment 1 compared to 3.9–12.6 s⁻¹ (high) and 3.9–4.6 s⁻¹ (low) in experiment 2, or 60%–75% of the range on a log scale.

A word about the perceptual effects of resonator-bar coupling which affects decay rate are in order here. The mechanical origins of this coupling have been discussed in detail in Chaigne and Doutaut (1997). The same values of η were used for both high and low sound sets. As can be seen in Table IV, α_1 has a similar range of variation for the high and low sound sets, whereas α_2 is nearly constant for the low set and varies only for the five sounds with the highest values of η in the high set. What is of import for the present study is the fact that in spite of the much longer decay times in the low sound set compared to the high set, the main result is essentially the same for the two: the perceptual dimensions related to the damping parameter in the two sets are highly correlated ($R^2=0.80$, $p<0.0001$). It may be, however, that different cues are being used since the correlations between the perceptual dimension and spectral cues (SCG) are higher for the low set, and those with temporal descriptors (ED_{3dB} , α_1) are higher for the high set. We will return to the issue of common or multiple cues in Sec. VII.

VII. GENERAL DISCUSSION

The dimensionality of the three perceptual spaces from both experiments is identical to that of the physical parameter space in which the stimuli were created. Further, there are strong correlations among mechanical and analytic parameters and perceptual dimensions, demonstrating a close relation between mechanical properties and putative parameters derived from the signals that represent the physical stimuli perceptually. Note, however, that many of these signal-derived parameters undergo a monotonic transformation in the perceptual representation that is either a logarithmic or a power function.

Many physical parameters are perceived according to power functions (luminosity, loudness, weight, electric shock, etc.) (Stevens, 1975). The damping factor used in the

bar model, as well as one of its possible signal correlates (Slope_{SCG}), would seem to be additions to this set. This relation is all the more interesting in that a similar exponent for η was found in three different sets of comparable stimuli that nonetheless had many differences in terms of the range of variation of damping factor and either bar length or density, as well as in the spectral content of the sounds (inharmonic in experiment 1, harmonic in experiment 2). We will now consider in turn the relation between mechanical parameters, signal descriptors, and perceptual dimensions across the three sound sets.

A. Damping-related dimension

One dimension in each of the perceptual spaces was related to the shape-invariant damping factor and is clearly related to the sound's timbre. Very strong correlations between a power function of the damping factor and the perceptual dimension were found. The signal characteristics, or combinations of them, that were hypothesized to specify the damping factor seem to vary somewhat according to the stimulus set. A comparison was performed among the eight signal descriptors that were the most highly correlated with the perceptual coordinates and η in the three sound sets. The chosen descriptors include Slope_{SCG}, $\alpha_1+\alpha_2$, α_2+ED_{3dB} , α_2+ED_{10dB} , $\alpha_2+Slope_{SCG}$, SCG+Slope_{SCG}, and α_2+SCG Slope_{SCG}+SCG+ α_2+ED_{10dB} . All eight of these signal descriptors (using the appropriate logarithmic or power functions of each) were fitted by multiple regression to both the damping-related perceptual dimension and η for each of the three spaces. The correlation coefficients and their 95% confidence intervals were determined for each comparison. Two descriptors were considered of equivalent predictive power if the correlation coefficient of one fell within the 95% confidence interval of the other.

For experiment 1, all of the descriptors have equivalent predictive power for both the perceptual dimension and η ($0.93\leq R^2\leq 0.99$). For experiment 2 (High bars), most of the predictors have equivalent predictive power for both the perceptual dimension ($0.85\leq R^2\leq 0.91$) and η ($0.86\leq R^2\leq 0.99$). The exceptions are Slope_{SCG} and SCG+Slope_{SCG} for which R^2 is significantly lower at about 0.55 in relation to the perceptual dimension and 0.72 in relation to η . The correlations of the parameters with η are significantly higher than with the perceptual dimension for this stimulus set. For experiment 2 (low bars), the correlation coefficients of the descriptors are quite dispersed, but the same three descriptors are found at the top, and have equivalent predictive power, for both the perceptual dimension and its mechanical origin. Aside from the four-factor descriptor, the other two include α_2+SCG and SCG+Slope_{SCG}. There is one descriptor that is common to all three spaces (α_2+SCG). Further, the second descriptor for the high bars in experiment 2 is not a good descriptor for the low bars, and vice versa. As mentioned in the discussion on experiment 2, this might lead one to conclude that separate acoustic cues were used by listeners for each sound set, according to what was the most useful for that set, e.g., $\alpha_2+Slope_{SCG}$ for the high bars and SCG+Slope_{SCG} for the low bars, either being a good candidate for the constant cross-section bars of experiment 1. A more

parsimonious approach, and the one that led to this analysis, is to attempt to find a single descriptor that works optimally for all three sound sets. A linear combination of a temporal factor (long-term decay constant, α_2) and a spectral factor (spectral centroid, SCG) is a reasonable candidate as the signal vehicle of the mechanical parameter: it explains 92%–98% of the variance in the perceptual coordinates and 97%–99% of the variance in η .

B. Frequency-related dimension

Dimension 2 in all three spaces was clearly related to pitch perception. Variations in both material density (ρ) and bar length (L) create changes in the component frequencies that vary as $\rho^{-1/2}$ and L^{-2} , respectively. Each perceptual dimension was strongly correlated with the frequency of the first partial, f_1 , which explains 91%–94% of the variance along this dimension. This result is similar to that of Miller and Carterette (1975) (experiment 1) in which both pitch (fundamental frequency) and timbral characteristics (envelope shape, spectral envelope) were varied. Miller and Carterette varied the pitch over a two-octave range. A multidimensional scaling analysis revealed a three-dimensional space with pitch along one dimension, decay rate along another, and a third dimension that seemed to be related to the impulsivity of the beginning of the sound (rapid attack followed by an initial rapid decay that either continued to decay for one sound or attained a sustain level at 40% of the maximum amplitude for another sound). The relative perceptual variation for the three dimensions was equivalent. However, stimulus variation in the spectral envelope (perhaps related to SCG) was not reflected in their perceptual space.

C. Relative perceptual salience of the two dimensions

In the present study, an attempt to increase the pitch variation by 170% (on a log scale) and decrease the variation in damping by 66% (on a scale defined by a power function with an exponent of 0.3) in experiment 2, only created a slight decrease in the dominance of the timbral dimension over the pitch dimension: the range of variation on the second dimension in the perceptual space was 57% of that of the first dimension in experiment 1 and was 78%–96% in experiment 2. This result should be qualified by the relative weights assigned to the two dimensions by the latent subject classes in experiment 2. A quarter of the listeners for the high bars had greater variation on the timbre dimension by a factor of 4.4 compared to the pitch dimension, whereas the other three-fourths had 1.7–3.0 times as much variation on the pitch dimension compared to the timbre dimension. For the low bars, a third of the listeners privileged the timbre dimension by a factor of 4.8, whereas the other two-thirds gave greater weight to the pitch dimension by a factor of 3.0. There would thus seem to be differences among listeners concerning the relative dominance of pitch and timbre when they vary simultaneously.

Lutfi and Oh (1997) found a predominance of frequency over intensity and decay constant in a study in which listeners were to identify which of two sounds came from an object of a particular material (iron versus silver, steel or cop-

per, and glass versus crystal, quartz or aluminum). Their sounds were synthesized sums of damped sinusoids designed to replicate the frequencies, amplitudes, and radiation decay constants of materials with particular elasticity, dimensions, and density. The differences among sets of sounds representing each material were chosen to be around threshold identification performance. It may be that the relative variation of frequency at these small differences was more salient. Closer to the results of the present study, Klatzky *et al.* (2000) presented synthesized sums of sinusoids to represent different materials and asked listeners to judge the proximity of the sounds in terms of the materials they evoked. They varied the sounds according to fundamental frequency and to a frequency-invariant damping parameter. They performed multidimensional scaling and forced the solution to two dimensions (apparently no attempt was made to determine the dimensionality of the perceptual space as was done in the present study). Their two dimensions corresponded clearly to frequency and decay correlates with a predominance being found for the latter, as here. What the present studies add to this result is a statistical verification of the dimensionality, the use of synthesized sounds that more closely resemble those produced by physical impact processes, and the quantification of the signal descriptors that potentially carry information concerning the mechanical properties.

D. Predicting dissimilarities from signal descriptors

The combinations of appropriately transformed analytic parameters can be used to derive an objective distance measure among the sounds, which can in turn be compared with the mean dissimilarities given by the listeners. For the three sound sets used, the resulting distance models combining log α_2 and SCG between sounds i and j are the following: Experiment 1:

$$d_{ij} = \{[0.047(f_{1i} - f_{1j})]^2 + [76.028(\log \alpha_{2i} - \log \alpha_{2j}) - 12.015(\text{SCG}_i - \text{SCG}_j)]^2\}^{1/2}, \quad (7)$$

Experiment 2 (high):

$$d_{ij} = \{[0.037(f_{1i} - f_{1j})]^2 + [118.517(\log \alpha_{2i} - \log \alpha_{2j}) - 5.389(\text{SCG}_i - \text{SCG}_j)]^2\}^{1/2}, \quad (8)$$

Experiment 2 (low):

$$d_{ij} = \{[0.113(f_{1i} - f_{1j})]^2 + [36.68(\log \alpha_{2i} - \log \alpha_{2j}) - 9.973(\text{SCG}_i - \text{SCG}_j)]^2\}^{1/2}. \quad (9)$$

These distances explain 80%, 64%, and 73% of the variance in the mean dissimilarity ratings, respectively. These coefficients of determination are to be compared with those of the distance model derived from the CLASCAL solutions which explain 92%, 85%, and 86% of the variance in the mean dissimilarity ratings. Using the four-factor analytic model does not result in a statistically significant improvement over the two-factor model. It would seem then that while covering a major proportion of the variance in the perceptual data, our signal parameters, adjusted to the individual dimensions, leave a certain proportion of the variance in the data unexplained.

VIII. CONCLUSIONS

Listeners were asked to judge the dissimilarity among synthesized sounds produced with a physical model of a bar. The sounds were varied according to the mass density or the length of the bar and to a viscoelastic damping parameter. Both constant cross-section bars and tuned variable cross-section bars were tested. A multidimensional scaling analysis of the dissimilarity data revealed perceptual spaces with the same dimensionality as that of the original physical parameter space using statistical criteria for the model selection. The results thus validate two aspects of the synthesis model and quantify psychophysically the relations between the mechanical parameters, the potential signal characteristics that carry them, and the perceptual representation. The modal frequencies vary with mass density and length and the likely perceptual correlate is pitch. Variation in the damping coefficient creates spectral and temporal variations in the signal, and thus has several potential signal characteristics that could serve as the acoustic vehicle for the mechanical origin. The most parsimonious signal descriptor that explains a significant amount of variance in the data sets for all stimuli is a linear combination of long-term decay constant and spectral center of gravity. The latter descriptor was derived from a model of the peripheral representation of the sounds in the human auditory system. This is one of the first demonstrations of a perceptual dimension being explained by a combination of temporal and spectral signal descriptors. Further, the perceptual representation of the damping coefficient would seem to have a power function relation to the mechanical parameter. These results demonstrate the perceptual salience of energy-loss phenomena in sound source behavior.

This study has examined only two of the many parameters that characterize the physical model of the impacted bar. It has demonstrated the importance of energy-loss phenomena in the perceptual representation of bar sounds, in coherence with the results of Klatzky *et al.* (2000). The understanding of such phenomena in the perceptual domain are important for the development of high-quality audio synthesis and may be applicable to the creation of sound simulations in virtual reality environments, for example. Future research will need to explore the perception of the other parameters of the synthesis model to quantify further its psychophysical underpinnings.

ACKNOWLEDGMENTS

Portions of this work were funded by a grant from the French Ministry of the Environment to Stephen McAdams and Antoine Chaigne and a doctoral fellowship from the French Ministry of Higher Education and Research to Vincent Roussarie. The cochlear representation and the extraction of SCG(t) and SCG were obtained with a Matlab routine (ERBSpect) developed by Alain de Cheveigné and are used with the permission of the author. The authors thank two anonymous reviewers and Alain de Cheveigné for helpful comments on an earlier version of this paper.

Aitken, M., Andersen, D., and Hinde, J. (1981). "Statistical model of data on teaching styles," *J. R. Stat. Soc., Ser. A* **144**, 419–461.

- Avanzini, F., and Rocchesso, D. (2001). "Controlling material properties in physical models of sounding objects," in *Proceedings of the International Computer Music Conference 2001*, La Habana, Cuba, (ICMA, La Habana, Cuba), pp. 91–94.
- Bucur, V. (1995). *Acoustics of Wood* (CRC, Boca Raton, FL).
- Cabe, P. A., and Pittenger, J. B. (2000). "Human sensitivity to acoustic information from vessel filling," *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 313–324.
- Carello, C., Anderson, K. A., and Kunkler-Peck, A. J. (1998). "Perception of object length by sound," *Psychol. Sci.* **9**, 211–214.
- Chaigne, A., and Doutaut, V. (1997). "Numerical simulations of xylophones. I. Time-domain modeling of the vibrating bars," *J. Acoust. Soc. Am.* **101**, 539–557.
- Chaigne, A., and Lambourg, C. (2001). "Time-domain simulation of damped impacted plates. I. Theory and experiments," *J. Acoust. Soc. Am.* **109**, 1422–1432.
- Doutaut, V. (1996). "Etude expérimentale et simulation numérique d'instruments de percussion à clavier." Ph.D. thesis, Ecole Nationale Supérieure des Télécommunications, Paris.
- Doutaut, V., Matignon, D., and Chaigne, A. (1998). "Numerical simulations of xylophones. II: Time-domain modeling of the resonator and of the radiated sound pressure," *J. Acoust. Soc. Am.* **104**, 1633–1647.
- Freed, D. J. (1990). "Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events," *J. Acoust. Soc. Am.* **87**, 311–322.
- Grey, J. M. (1977). "Multidimensional perceptual scaling of musical timbres," *J. Acoust. Soc. Am.* **61**, 1270–1277.
- Guski, R. (2000). "Studies in auditive kinetics," in *Contributions to Psychological Acoustics: Results of the 8th Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick, M. Meis, and C. Reckhardt (Bis, Oldenburg), pp. 383–401.
- Hajda, J. M., Kendall, R. A., Carterette, E. C., and Harshberger, M. L. (1997). "Methodological issues in timbre research," in *Perception and Cognition of Music*, edited by I. Deliège and J. Sloboda (Psychology, Hove), pp. 253–306.
- Hartmann, W. M. (1997). *Signals, Sound, and Sensation* (AIP, Woodbury, NY).
- Hope, A. C. (1968). "A simplified Monte Carlo significance test procedure," *J. R. Stat. Soc. Ser. B. Methodol.* **30**, 582–598.
- Houben, M. (2002). "The sound of rolling objects: Perception of size and speed," Technische Universiteit Eindhoven, Eindhoven, The Netherlands.
- Houix, O., McAdams, S., and Caussé, R. (1999). "Auditory categorization of sound sources," in *Studies in Perception and Action*, edited by M. A. Grealy and J. A. Thomson (Lawrence Erlbaum Associates, Mahwah, NJ), Vol. V, pp. 47–51.
- Iverson, P., and Krumhansl, C. L. (1993). "Isolating the dynamic attributes of musical timbre," *J. Acoust. Soc. Am.* **94**, 2595–2603.
- Killion, M. C. (1978). "Revised estimate of minimum audible pressure: Where is the 'missing 6 dB'?" *J. Acoust. Soc. Am.* **63**, 1501–1508.
- Klatzky, R. L., Pai, D. K., and Krotkov, E. P. (2000). "Perception of material from contact sounds," *Presence: Teleoperators and Virtual Environments* **9**, 399–410.
- Krotkov, E., Klatzky, R., and Zumel, N. (1996). "Robotic perception of material: Experiments with shape-invariant acoustic measures of material type," in *Experimental robotics IV: Lecture notes in control and information science*, edited by O. Khatib and X. Salisbury (New York, Springer Verlag) Vol. 223, pp. 204–211.
- Kunkler-Peck, A. J., and Turvey, M. T. (2000). "Hearing shape," *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 279–294.
- Lakatos, S. (2000). "A common perceptual space for harmonic and percussive timbres," *Percept. Psychophys.* **62**, 1426–1439.
- Lakatos, S., McAdams, S., and Caussé, R. (1997). "The representation of auditory source characteristics: Simple geometric form," *Percept. Psychophys.* **59**, 1180–1190.
- Lindemann, E., Dechelle, F., Smith, B., and Starkier, M. (1991). "The architecture of the IRCAM Musical Workstation," *Comput. Music J.* **15**(3), 41–19.
- Lutfi, R. A. (2001). "Auditory detection of hollowness," *J. Acoust. Soc. Am.* **110**, 1010–1019.
- Lutfi, R. A., and Oh, E. L. (1997). "Auditory discrimination of material changes in a struck-clamped bar," *J. Acoust. Soc. Am.* **102**, 3647–3656.
- McAdams, S. (1993). "Recognition of sound sources and events," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by S. McAdams and E. Bigand (Oxford U.P., Oxford), pp. 146–198.

- McAdams, S. (1999). "Perspectives on the contribution of timbre to musical structure," *Comput. Music J.* **23**(2), 96–113.
- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (1995). "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," *Psychol. Res.* **58**, 177–192.
- Miller, J. R., and Carterette, E. C. (1975). "Perceptual space for musical structures," *J. Acoust. Soc. Am.* **58**, 711–720.
- Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Patterson, R. D., Allerhand, M., and Giguère, C. (1995). "Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform," *J. Acoust. Soc. Am.* **98**, 1890–1894.
- Plack, C. J., and Moore, B. C. J. (1990). "Temporal window shape as a function of frequency and level," *J. Acoust. Soc. Am.* **87**, 2178–2187.
- Plomp, R. (1970). "Timbre as a multidimensional attribute of complex tones," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 397–414.
- Repp, B. H. (1987). "The sound of two hands clapping: An exploratory study," *J. Acoust. Soc. Am.* **81**, 1100–1109.
- Rossing, T. D., and Russell, D. (1990). "Laboratory observation of elastic waves in solids," *Am. J. Phys.* **58**, 1153–1162.
- Schwarz, G. (1978). "Estimating the dimensions of a model," *Ann. Stat.* **6**, 461–464.
- Smith, B. K. (1995). "PsiExp: An environment for psychoacoustic experimentation using the IRCAM Musical Workstation," *SMPC95: Society for Music Perception and Cognition*, Berkeley, CA (University of California, Berkeley, Berkeley, CA).
- Stevens, S. S. (1975). *Psychophysics: Introduction to its Perceptual, Neural and Social Prospects* (Wiley, New York).
- Warren, W. H., and Verbrugge, R. R. (1984). "Auditory perception of breaking and bouncing events: A case study in ecological acoustics," *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 704–712.
- Wessel, D. L. (1979). "Timbre space as a musical control structure," *Comput. Music J.* **3**(2), 45–52.
- Winsberg, S., and Carroll, J. D. (1988). "A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model," *Psychometrika* **53**, 217–229.
- Winsberg, S., and Carroll, J. D. (1989). "A quasi-nonmetric method for multidimensional scaling of multiway data via a restricted case of an extended INDSCAL model," in *Multi-way Data Analysis*, edited by R. Coppi and S. Bolasco (North-Holland, Amsterdam), pp. 405–414.
- Winsberg, S., and De Soete, G. (1993). "A latent class approach to fitting the weighted Euclidean model. CLASCAL," *Psychometrika* **58**, 315–330.