Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones[a)]

Anne Caclin,[b)] Stephen McAdams,[c)] Bennett K. Smith, and Suzanne Winsberg
Institut de Recherche et Coordination Acoustique/Musique (STMS-IRCAM-CNRS), 1 place Igor Stravinsky, F-75004 Paris, France

(Received 13 September 2004; revised 18 April 2005; accepted 18 April 2005)

Timbre spaces represent the organization of perceptual distances, as measured with dissimilarity ratings, among tones equated for pitch, loudness, and perceived duration. A number of potential acoustic correlates of timbre-space dimensions have been proposed in the psychoacoustic literature, including attack time, spectral centroid, spectral flux, and spectrum fine structure. The experiments reported here were designed as direct tests of the perceptual relevance of these acoustical parameters for timbre dissimilarity judgments. Listeners presented with carefully controlled synthetic tones use attack time, spectral centroid, and spectrum fine structure in dissimilarity rating experiments. These parameters thus appear as major determinants of timbre. However, spectral flux appears as a less salient timbre parameter, its salience depending on the number of other dimensions varying concurrently in the stimulus set. Dissimilarity ratings were analyzed with two different multidimensional scaling models (CLASCAL and CONSCAL), the latter providing psychophysical functions constrained by the physical parameters. Their complementarity is discussed. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1929229]

PACS number(s): 43.75.Cd, 43.66.Jh [NHF] Pages: 471–482

I. INTRODUCTION

Human abilities to recognize sound sources (the sound of a door slamming, footsteps, musical instruments, voices, etc.) are essential to our everyday behavior. Their sophistication allows for complex tasks such as recognizing the emotions embedded in speech or enjoying music. Such abilities are presumed to rely largely on a capacity to perceive and process timbre differences, making timbre analysis a fundamental task of the auditory system (see McAdams, 1993; Handel, 1995).

By definition, timbre is the *perceptual* attribute that distinguishes two tones of equal pitch, loudness, and duration (ANSI, 1973). Typically musical timbre is what distinguishes perceptually a piano from a clarinet playing the same note (e.g., A4), with the same intensity (e.g., mezzo forte), and for the same duration. This example highlights the relationship between timbre and sound source identification. Instrument identification on the basis of timbre has been investigated after various modifications of the original acoustic signals (Berger, 1964; Saldanha and Corso, 1964; Hajda, 1999). Hajda (1999) has shown that for brief tones (from impulsive instruments), the integrity of the temporal structure is crucial for identifying the instrument. However for long tones, the sustained part (played forward or backward) is sufficient to allow recognition. It implies that instrument recognition relies upon temporal and spectral or spectrotemporal information, suggesting that different perceptible acoustical parameters might be grouped under the term timbre.

Timbre is indeed usually described as a *multidimensional* perceptual attribute of complex tones. It has been hypothesized that contrary to pitch, which relies primarily on the tone's fundamental period, and loudness, which depends on tone intensity, timbre relies on several parameters (acoustical dimensions) of the sound. The holy grail of timbre studies has been to uncover the number and nature of these dimensions. Two major strategies have been used: ratings on verbal scales (see Kendall and Carterette, 1993, for a review) and most often, multidimensional scaling (MDS) of (dis)similarity ratings (see McAdams, 1993; Hajda *et al.*, 1997, for reviews).

First applied to musical timbre by Plomp (1970), MDS offers the advantage that it does not make any assumptions regarding the underlying acoustical dimensions. In such a study, listeners rate the dissimilarity between the two stimuli of all possible pairs of sounds from a set of stimuli. The resulting dissimilarity matrices are subjected to multidimensional scaling. MDS is a procedure in which dissimilarity data arising from N subjects ($N \geqslant 1$) are modeled to fit distances in some type of space, usually Euclidean of low dimensionality R. Several MDS models and techniques have been developed, such as INDSCAL (Carroll and Chang, 1970) or CLASCAL (Winsberg and De Soete, 1993). Some MDS models (e.g., Torgerson, 1958; Kruskal, 1964a, b) produce rotationally invariant solutions. Weighted Euclidean models in which the salience of each dimension is different for each subject (INDSCAL) or for each latent class of subjects (CLASCAL) produce solutions with axes oriented in a psychologically meaningful way. In the latent class approach, each of the N subjects is assumed to belong to one and only one of

[a)]Portions of this work were presented in "A confirmatory analysis of four acoustic correlates of timbre space," Acoustical Society of America, Cancun, Mexico, 2002.

[b)]Electronic mail: anne.caclin@chups.jussieu.fr. Present address: LENA-CNRS UPR 640, Hôpital Pitié-Salpétrière, 47 bd de l'Hôpital, 75013 Paris, France.

[c)]Electronic mail: smc@music.mcgill.ca. Present address: CIRMMT, Faculty of Music, McGill University, 555 rue Sherbrooke ouest, Montréal, Québec H3A IE3, Canada.

a small number T ($T \ll N$) of latent classes, and all the subjects in the same class are assumed to weight all dimensions identically. It is not known in advance to which latent class a particular subject belongs. This latent class approach, incorporated in CLASCAL, drastically reduces the number of parameters of the INDSCAL model. An extension of the Euclidean or weighted Euclidean model proposed by Winsberg and Carroll (1989) postulates that stimuli differ not only with respect to common dimensions, but also with respect to specific or unique dimensions possessed by each stimulus (these specificities could of course be zero). In the CLASCAL model, the distance between two objects i and j for class t is given by

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

where R is the number of dimensions of the model, x_{ir} the coordinate of the ith stimulus on the rth dimension, s_i the square of the specificity value for object i, w_{tr} the weight assigned to the rth dimension by class t, and v_t the weight assigned to the whole set of specificity values by class t.

CONSCAL (Constrained Scaling) is a more recent development in MDS modeling (Winsberg and De Soete, 1997). The CONSCAL models are useful in situations in which a small number of known physical parameters may be used to describe the stimuli, and it is likely that these are the very attributes upon which the subjects make their dissimilarity ratings. The CONSCAL models constrain the axes of the distance model to be monotone transformations of these physical attributes. Two CONSCAL models have been proposed by Winsberg and De Soete (1997), where the distances between object i and object j are given by

$$d_{ij} = [(\mathbf{f}_i - \mathbf{f}_j)^t \mathbf{I} (\mathbf{f}_i - \mathbf{f}_j)]^{1/2}, \quad (2)$$

$$d_{ij} = [(\mathbf{f}_i - \mathbf{f}_j)^t \mathbf{A} (\mathbf{f}_i - \mathbf{f}_j)]^{1/2}, \quad (3)$$

where there are R physical dimensions, $\mathbf{f}_i$ is the vector of monotone transformations for the ith object, its rth component being $f^{(r)}(x_i^{(r)})$, $x_i^{(r)}$ is the coordinate for the ith object on the rth dimension, and $f^{(r)}$ is the monotone transformation for the rth dimension. $\mathbf{A}$ is an $R \times R$ symmetric matrix with ones on the diagonal and $\mathbf{I}$ is an $R \times R$ identity matrix. In the first model [Eq. (2)] the axes are assumed to be orthogonal, but in the second model [Eq. (3)] they are not.

The monotone transformation $f^{(r)}$ for the rth dimension is represented by a monotone spline function. Splines are piecewise polynomials of a given degree joined at a number of interior knots. Splines have the advantage of great flexibility, ease of computation, and local support. Note that in the CONSCAL models, instead of estimating a coordinate for each stimulus for each dimension, a function is estimated for each dimension. In the latent class extension of CONSCAL, a different function is estimated for each dimension for each latent class of subjects. Thus, the CONSCAL models have the advantage for psychophysical research of producing psychophysical functions, and the shapes of those functions are informative (see McAdams and Winsberg, 2000).

When applied to timbre dissimilarity ratings, MDS yields a perceptual space commonly known as a *timbre space*. This space is only derived from listeners' dissimilarity ratings, except when using a model like CONSCAL. In all cases, the distances between timbres in the space are perceptual distances. In using general (as opposed to constrained) MDS models the last step of such a timbre space study is to propose an interpretation for each of the perceptual dimensions, usually in terms of an underlying acoustical dimension.

Depending on the set of stimuli and the group of subjects, two- to four-dimensional timbre spaces have been reported in MDS studies using natural sounds from musical instruments or synthetic tones, generally made to imitate instruments of the orchestra. When acoustical correlates are proposed, most of the studies so far (e.g., Grey, 1977; McAdams *et al.*, 1995; Marozeau *et al.*, 2003) have emphasized the role of spectral center of gravity [(SCG), amplitude-weighted mean frequency of the energy spectrum] and attack time, the latter separating impulsive from sustained tones. Other parameters have been less systematically reported, e.g., spectral flux, a measure of the fluctuation of the spectrum over time (McAdams *et al.*, 1995), spectral spread (Marozeau *et al.*, 2003), spectral irregularity (Krimphoff, McAdams, and Winsberg, 1994), harmonic onset asynchrony (Grey, 1977). Along with these shared dimensions, a number of instruments are characterized by rather high specificity values (Krumhansl, 1989). It has not always been possible to propose an interpretation for these specificity values (McAdams *et al.*, 1995), although for some instruments there are good candidates (e.g., return of the hopper for the harpsichord).

From a psychophysical point of view, dissimilarity rating methods are known to be sensitive to judgment bias, in the sense that listeners usually pay attention to only a limited number of parameters, which is interpreted as reflecting the perceptual salience of the parameters (Miller and Carterette, 1975). MDS studies are thus presumed to highlight the most perceptually salient timbre parameters that are likely to be of importance in a variety of situations (voice recognition, music listening). Nevertheless they have a common drawback: given the multiplicity of acoustical parameters that could be proposed to explain perceptual dimensions, one can never be sure that the selected parameters do not merely covary with the true underlying parameters. This is especially true for three-dimensional (3D) or four-dimensional spaces, where for some dimensions only weak correlations between perceptual and acoustical dimensions are found. It turns out to be difficult in this case to select the appropriate acoustical parameter(s), particularly if the proposed parameters are correlated among themselves to some extent within the stimulus set. There are also concerns regarding whether or not it is valid to use MDS when some of the dimensions might be categorical in nature, which is the case when the stimulus set includes impulsive and sustained instruments. Therefore two questions remain open: (1) is a continuous space a good model of the perceptual relationships among timbres? and (2) if such a model is appropriate, are the proposed underlying acoustical dimensions correct?

The experiments reported here intend to deal with these issues, according to the following approach: we start with the

construction of spaces of synthetic sounds varying along continuous acoustical dimensions that have been chosen on the basis of previous MDS studies of timbre. We then use stimuli sampling these acoustical spaces in classical dissimilarity rating experiments. If the timbre space model is correct, there should be a good match between the physical space and the resulting perceptual space. We believe that this confirmatory approach is relevant both in the context of musical acoustics and for auditory perception studies in general, considering timbre perception as a model of complex sound perception. A similar approach has already been used (Miller and Carterette, 1975; Samson, Zatorre, and Ramsay, 1997). Miller and Carterette (1975) have shown that subjects can use fundamental frequency, amplitude envelope, and the number of harmonics when making dissimilarity ratings, but not harmonic structure or the pattern of harmonic onsets. A few points warrant mention here: First, no attempt was made *a priori* to equalize the perceptual ranges of variation of the different parameters in their study. Therefore, because fundamental frequency variations were very salient, they could have prevented subjects from using harmonic structure in their ratings. Second, their approach was to explore possible musical spaces, and they did not model explicitly the timbre dimensions arising from dissimilarity rating studies of natural instrument sounds. Samson, Zatorre, and Ramsay (1997) confirmed that attack time and spectral center of gravity values are used in timbre dissimilarity ratings. Note that in these two studies only a limited number of values (3) along each dimension were used, so the results could be explained in terms of categorical perception, and not as reflecting the perception of attributes varying along continuous axes. Further it is impossible to characterize adequately the corresponding psychophysical functions.

For the present study, the parameters varying between the sounds were selected on the basis of previous timbre dissimilarity studies (in particular Grey, 1977; Krimphoff *et al.*, 1994; McAdams *et al.*, 1995). Along each acoustical dimension, there were 16 different values to avoid creating unwanted categories. In the first experiment, we tested a space where attack time, spectral centroid, and spectral flux varied between the timbres. Based on the results of this first experiment, further testing was conducted to investigate contextual effects on the salience of spectral flux (Experiment 2). Finally a 3D space where attack time, spectral centroid, and spectral irregularity varied was tested (Experiment 3). Analyses were conducted both with CLASCAL and CONSCAL models. CLASCAL analyses aim to select freely the model best fitting the data, to uncover the most salient acoustical parameters. CONSCAL analyses allow for more refined interpretation, with an estimation of the shape of psychophysical functions for the different axes. CONSCAL has been used only once in the context of timbre studies (McAdams and Winsberg, 2000). The data sets of the present study offered a valuable opportunity to test the relative adequacy of the CLASCAL and CONSCAL models.

II. EXPERIMENT 1: ATTACK TIME, SPECTRAL CENTROID, AND SPECTRAL FLUX

A. Method

1. Participants

Thirty listeners (aged 19–51 years, 14 female) participated in Experiment 1. None of them reported any hearing loss, and 18 of them had received musical training. Musicians had been practicing music for 12 years on average (ranging from 4 to 27 years). Listeners were naive as to the purpose of the experiment and were paid for their participation.

2. Stimuli

Sixteen tones having 20 harmonics were created by additive synthesis, using MAX/ISPW (Lindemann *et al.*, 1991) on a NeXT workstation (sampling rate=44 100 Hz, resolution =16 bits). The fundamental frequency was 311 Hz (E♭4). Three parameters varied among the stimuli (Fig. 1): attack time, SCG, and fluctuation of the spectral envelope over the first 100 ms (spectral flux). The amplitude envelope was always composed of a linear rise (attack time), followed by a plateau and an exponential decay [Fig. 1(a)]. The spectrum was harmonic, and at any time point, the amplitude spectrum was a power function of harmonic rank n [Fig. 1(b)]:

$$A_n = k \times 1/n^{\alpha} \tag{4}$$

with A_n the amplitude of the nth harmonic. The exponent of this power function completely determined the value of the instantaneous spectral center of gravity.

A preliminary adjustment experiment run with eight listeners (who did not participate in the main experiment) allowed us to uncover formulas to keep perceived duration constant when changing attack time and loudness constant when changing SCG. The reference tone (T_0) had 15 ms attack, 400 ms plateau, and 200 ms decay, with a SCG of 933 Hz, and no spectral flux. Its level was approximately 80 dB SPL (A-weighted). Listeners had to adjust the plateau duration for a test tone with a 200 ms attack, so that it appeared as long as the reference tone. Listeners were also asked to adjust the SCG and spectral flux of test tones so that they would appear as different perceptually from the reference tone (T_0) as was the 200 ms attack tone. When SCG was adjusted, attack time was fixed at 15 ms and spectral flux at 0 Hz, and when spectral flux was adjusted, attack time was fixed at 15 ms and SCG at the third harmonic (933 Hz). The ranges of variation of the three parameters for the main experiment were chosen accordingly. Finally participants of this preliminary experiment had to adjust the levels of the test tones to match the loudness of the reference.

In the main experiment, attack time ($t1$) varied between 15 and 200 ms, in logarithmic steps, as it has often been proposed that the logarithm of attack time explains the corresponding timbre dimension better than the attack time itself (cf. Krimphoff *et al.*, 1994; McAdams *et al.*, 1995). The duration of the plateau ($t2$) was fixed at 400 ms when the attack lasted 15 ms, and was adjusted to keep perceived

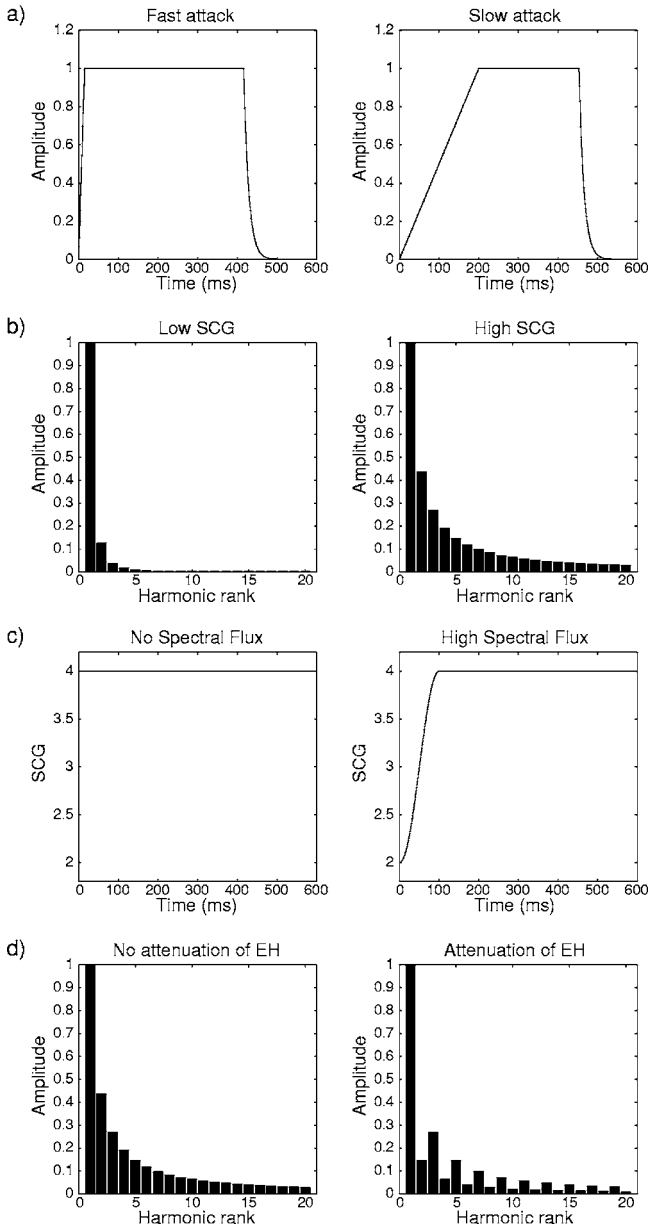

FIG. 1. Stimulus construction. (a) Attack time (Experiments 1, 2A, 3). (b) Spectral center of gravity (SCG, Experiments 1, 2B, 3). (c) Spectral flux with SCG plotted in units of harmonic rank (Experiments 1, 2A, 2B, 2C). (d) Even harmonic (EH) attenuation (Experiment 3). Amplitudes are given on arbitrary linear scales.

sound duration constant using a formula derived from the preliminary adjustment experiment: $t2=412-0.8\times t1$ ($t1$ and $t2$ in ms).

The spectral center of gravity was computed as follows:

$$\mathrm{SCG}=\frac{\sum_n n\times A_n}{\sum_n A_n}. \tag{5}$$

SCG was varied by changing the value of the exponent α in Eq. (4). SCG varied in linear steps between 933 and 1400 Hz (i.e., from 3 to 4.5 in harmonic rank units). After equalization of rms levels of T_i and T_0, the loudness of tone T_i was adjusted according to

$$20\log\frac{A(T_i)}{A(T_0)}=-1.9\times(\mathrm{SCG}(T_i)-\mathrm{SCG}(T_0)). \tag{6}$$

Spectral flux consisted of a half-cycle sinusoidal variation of SCG in the first 100 ms. It models the progressive expansion of the spectrum toward the higher harmonics that exists in some instruments, such as the trumpet. This variation of instantaneous SCG is equivalent to an asynchrony in the rise of the harmonics, the higher ones appearing later. Grey (1977) has proposed it as a correlate of one dimension of timbre. More generally, this model was chosen because it represents a variation of the spectrum over time that is not perceived as vibrato (see Hajda, 1999, for a detailed analysis of vibrato perception) and corresponds roughly to the kind of spectral envelope variation found in the attack portions of some musical instruments, although its evolution is necessarily independent of attack time in our stimuli. The parameter that varied between sounds was the difference between the SCG value in the steady portion of the tone and the initial value of the SCG. This difference ranged from 0 to 560 Hz (i.e., 0–1.8 harmonic ranks) with equally spaced steps.

The coordinates of the points sampling the physical space were chosen among a set of computer-generated random spaces complying with the following rules: each of the three parameters took 16 different values as described earlier, and each value was used only once, thus leading to a total of 16 different sounds. The retained distribution of stimuli in the 3D space was chosen as a compromise between two constraints: first, any two stimuli should not be too close to each other; second, the stimuli should be homogeneously distributed, avoiding "empty" regions in the space. In this way we obtained a good sampling of the physical space with no two sounds sharing a value on any of the three dimensions of interest [see Fig. 2(a) and Table I].

3. Procedure

All experimental routines were programmed using PSIEXP (Smith, 1995) on a NeXT workstation. The experiment took place in a soundproof booth, and the sounds were played to the subject via Sennheiser HD 520 II headphones after digital-to-analog conversion and amplification (Yamaha P2075 power amplifier). Participants were first asked to listen to the 16 sounds of the experiment to become familiarized with the range of variation. After ten randomly chosen practice trials, they were to rate the dissimilarity for all possible pairs of the 16 sounds (136 pairs). Dissimilarity ratings were made with the mouse on a scale presented on the computer screen with end points labeled "same" and "different" (in French). Scale values were digitized on a 0 to 1 scale. Listeners were allowed to listen to the pairs as many times as they wanted prior to making their ratings, and they were requested to keep their rating strategy as constant as possible.

4. Statistical analyses

First, correlations were computed among subjects' dissimilarity ratings for pairs of nonidentical sounds ($N=120$). Hierarchical clustering (average linkage algorithm) was per-

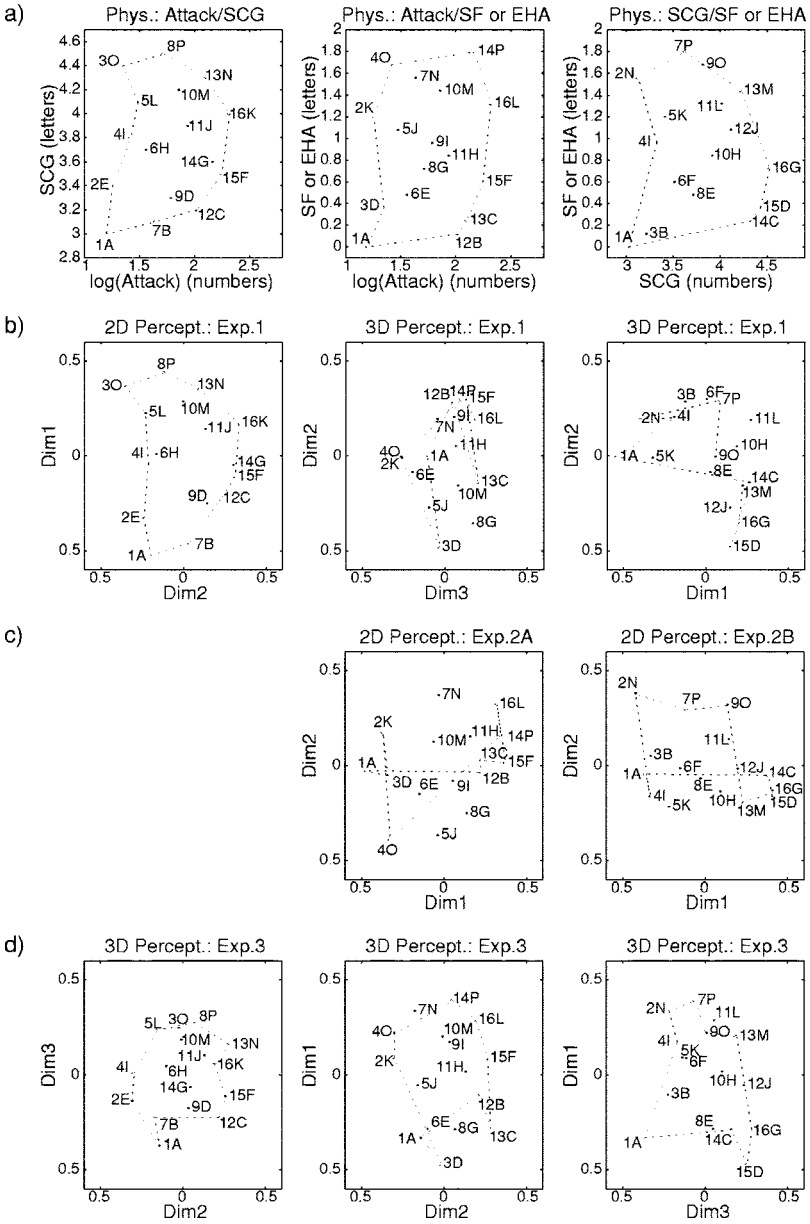

FIG. 2. Physical and perceptual spaces. (a) Physical space. (b) CLASCAL spaces for Experiment 1, left panel =2D space, middle and right panels=3D space. (c) 2D CLASCAL spaces for Experiment 2A (middle panel) and 2B (right panel). (d) 3D CLASCAL space for Experiment 3. In the top row, for each panel, the numbers refer to values for the physical dimension along the abscissa, and the letters to values for the physical dimension along the ordinate. The same labels are used in the other rows (per column). Note that any number–letter pairing only refers to the same sound within a column. Dotted lines connect the outer stimuli in the physical space to help appreciate visually the structural relations between physical and perceptual spaces.

formed on the correlation matrix in order to detect the subjects whose ratings systematically differed from the others. More precisely the values that were entered in the analysis were one minus the correlation coefficient (r), in order to transform the correlations into distances. Data from outlying subjects were discarded from subsequent analyses. The interest of this preliminary step in the analysis is practical: removing the subject(s) that is (are) the most different from all the others allows us to obtain more stable MDS solutions, making model selection an easier task.

Dissimilarity ratings from the remaining subjects were analyzed using CLASCAL (Winsberg and De Soete, 1993) and CONSCAL models (Winsberg and De Soete, 1997). The coordinates of the stimuli on the three physical dimensions tested (logarithm of attack time, SCG, and spectral flux values) were entered into the CONSCAL models. In both CLASCAL and CONSCAL analyses, model selection involves choosing the appropriate number of latent classes and the appropriate number of dimensions. Analyses were made in three steps as outlined in detail in Winsberg and De Soete (1993). A first choice is made for the number of latent classes (from one to five) using Hope's (1968) Monte Carlo test (using 100 repetitions) on the null model, i.e., a model with no spatial structure whatsoever. Bayesian Information Criterion (BIC statistic, see Winsberg and De Soete, 1993, for example), which reflects the fit of the model with a penalty for models with increasing number of parameters, is then used to select candidates for the best spatial model: for CLASCAL the number of dimensions with or without specificities; for CONSCAL the number of dimensions and the classes of spline functions (see the following for more details on spatial model selection). Finally, the appropriate number of latent classes is chosen using Hope's procedure on the spatial model. If the number of classes selected in the last step is not equal to that chosen on the null model, the last two steps (selection of spatial model and of number of classes) are repeated until they converge. After selection of the best CLASCAL model

TABLE I. Physical and CLASCAL perceptual coordinates for Experiment 1. Attack time is given in ms; SCG and spectral flux (SFl) are given in harmonic rank. Perceptual coordinates are given on an arbitrary scale. The superscripts indicate the perceptual dimensions significantly correlated ($p < 0.05$) with Attack (+) and SCG (●).

Physical space			2D perceptual space		3D perceptual space		
Attack^{+}	SCG$^{●}$	SFl	Dim 1$^{●}$	Dim 2^{+}	Dim 1$^{●}$	Dim 2^{+}	Dim 3^{+}
15	3.00	0.00	−0.53	0.21	−0.55	0.01	0.11
42	3.10	1.56	−0.46	−0.03	−0.41	−0.19	0.04
100	3.20	0.12	−0.23	−0.22	−0.13	−0.28	−0.04
59	3.30	0.96	−0.25	−0.14	−0.19	−0.21	−0.06
17	3.40	1.20	−0.33	0.25	−0.32	0.01	0.26
168	3.50	0.60	−0.11	−0.31	0.01	−0.30	−0.13
141	3.60	1.80	−0.05	−0.30	0.08	−0.29	−0.09
35	3.70	0.48	0.01	0.16	0.03	0.09	0.19
25	3.80	1.68	−0.04	0.22	0.06	0.00	0.26
84	3.90	0.84	0.14	−0.13	0.19	−0.05	−0.07
199	4.00	1.32	0.17	−0.33	0.27	−0.19	−0.18
29	4.10	1.08	0.23	0.23	0.15	0.27	0.10
70	4.20	1.44	0.29	0.00	0.22	0.16	−0.08
119	4.30	0.24	0.36	−0.08	0.26	0.14	−0.20
21	4.40	0.36	0.37	0.36	0.14	0.48	0.04
50	4.50	0.72	0.44	0.12	0.19	0.36	−0.17

and the best CONSCAL model, a parametric bootstrap procedure is used to compare them.

a. CLASCAL spatial model selection. For a given number of latent classes, the three models that best fit the data were determined on the basis of the BIC statistic. Two- to five-dimensional models, with or without specificity values, were considered. The best model among those three was selected according to the results of three Hope's tests. The CLASCAL algorithm also produces *a posteriori* probabilities for the belongingness of each subject to each of the latent classes. A subject was assigned to a particular class when any of these probabilities was greater than 0.9. Correlations between the perceptual dimensions and the original acoustical dimensions were computed and their significance was tested using Fisher's *r*-to-*z* test.

b. CONSCAL spatial model selection. For a given number of latent classes, we estimated several models, with spline orders (the degree of the piecewise polynomial plus 1) from two to four, and one to three interior knots (the number of polynomial pieces minus 1 in this case). We started modeling with the same kind of spline for all the dimensions, using the CONSCAL model assuming orthogonal dimensions. The best two models among these nine models were chosen according to the BIC statistics. We then looked at models with combinations of the two types of splines selected and models with lower-order splines or splines with less interior knots on one dimension. We selected the model with the smallest BIC. Finally, with Hope's test we compared the retained model with the model having the same types of splines as the selected one, but allowing dimensions to be nonorthogonal.

c. Comparison of CLASCAL and CONSCAL models. Monte Carlo samples were generated using the best CONSCAL model plus normal error. The log likelihood ratio for the real data obtained for the two models (the best CONSCAL model and the best CLASCAL model) was then compared to the distribution of log likelihood ratios obtained for the Monte Carlo samples. The null hypothesis was rejected when the log likelihood ratio for the real data exceeded the 0.95 level for the bootstrapped distribution under the null model (the best CONSCAL model in this case).

B. Results

The data from two subjects were removed after the hierarchical clustering analysis. The mean coefficient of determination between these two subjects and other subjects' ratings were $r^2(118)=0.14$ and $r^2(118)=0.15$, respectively. After removing those subjects, the mean coefficient of determination between subjects' dissimilarity ratings was $r^2(118)=0.23$ (SD=0.04).

1. CLASCAL space

MDS with CLASCAL on the data from the 28 remaining subjects yielded a two-dimensional (2D) perceptual space without specificity values and two latent classes of subjects [Fig. 2(b), left panel, Table I]. These two dimensions were correlated with SCG [$r^2(14)=0.96, p<0.0001$] and attack time [$r^2(14)=0.81, p<0.0001$], respectively (note that r^2 is the percentage of variance explained). For the second dimension, the correlation was even better using the logarithm of the attack time [$r^2(14)=0.87, p<0.0001$]. Neither of these two dimensions was correlated with spectral flux ($p>0.5$ in both cases). Even when considering the 3D solution [again with two latent classes of subjects and no specificity values, Fig. 2(b), middle and right panel, Table I], the additional dimension did not correlate significantly with spectral flux values [$r^2(14)=0.06, p=0.38$]. Adding a dimension actually tended to disorganize the attack-time dimension [Fig. 2(b), middle panel]. When considering the 2D solution with specificity values, sounds with high spectral flux values did not have greater specificity values than the rest of the sounds.

The subjects were classified *a posteriori* into two classes of 11 and 17 subjects. Subjects in the second class weighted more heavily the two dimensions (weights for dimensions 1 and 2: 1.20 and 1.36) than those in class 1 (weights: 0.80 and 0.65), suggesting that they used a larger range of the available dissimilarity scale. It is important to note that the ratios between the weights for the two dimensions are different for the two classes, with the first class weighting more on the first dimension (corresponding to SCG) and the second weighting more on the second dimension (corresponding to attack time). This means the dimensions do not have the same relative salience for each class. Both classes contained musically trained subjects (3 out of 11 in class 1 and 7 out of 17 in class 2).

2. CONSCAL space

MDS with CONSCAL selected a model with one class of subjects and orthogonal dimensions. A cubic spline with one interior knot best modeled the attack-time dimension and

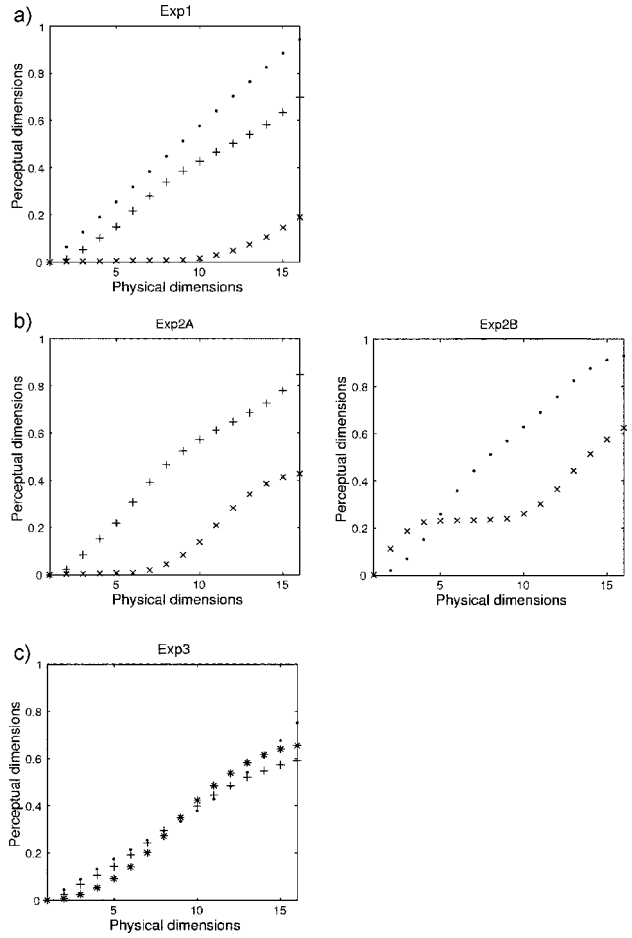

FIG. 3. Psychophysical functions as determined with CONSCAL. (+) Log(attack), (●) SCG, (×) spectral flux, and (*) even-harmonic attenuation. (a) 3D space for Experiment 1. (b) 2D spaces for Experiment 2A (left panel) and Experiment 2B (right panel). (c) 3D space for Experiment 3. The numbers on the abscissa are for the rank of the parameter value (from 1 to 16) in the stimulus set. Log(attack), SCG, spectral flux, and attenuation of even harmonics increase from left to right.

quadratic splines with one interior knot were sufficient for the other two dimensions (note that the monotone splines we are using are integrated B-splines as described in Winsberg and De Soete, 1997, and that the degrees reported here are those of the monotone splines, which correspond to the degrees of the B-splines plus 1). Given that CLASCAL analyses selected a 2D model with two dimensions correlated with the logarithm of attack time and SCG, we also compared the 3D CONSCAL model with the 2D CONSCAL model obtained when eliminating the axis corresponding to spectral flux. The 3D model was retained (Hope's test with 100 Monte Carlo simulations, $p<0.01$ for rejection of the 2D model). The 3D one-class CONSCAL model was finally compared to the 2D two-class CLASCAL model, and the CONSCAL model was not rejected ($0.09<p<0.10$ for rejecting CONSCAL). The CONSCAL spaces for all three experiments are represented in Fig. 3 using psychophysical functions. To help compare CLASCAL and CONSCAL models, an example of two such spaces represented in the same way (scatterplots) is provided in Fig. 4 (data from Experiment 2A).

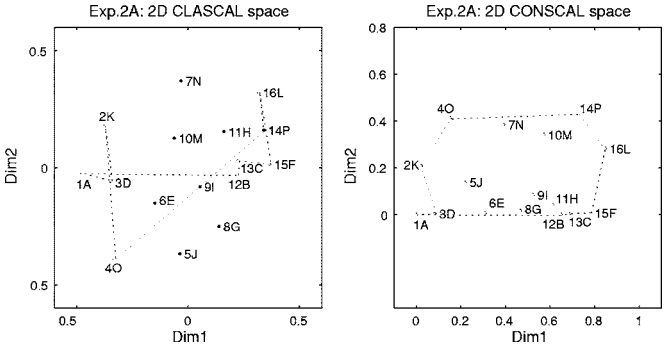

FIG. 4. Experiment 2A: 2D CLASCAL and CONSCAL spaces for comparison. Numbers refer to attack time, letters to spectral flux, as in the middle column of Fig. 2.

C. Discussion

These data confirm the perceptual significance of attack time as well as SCG, as revealed by dissimilarity ratings. It is also clear that listeners use a continuous range of attack times to perform their dissimilarity ratings. At the same time spectral flux is *not* used by listeners to perform their dissimilarity ratings according to the CLASCAL analysis. The CONSCAL analysis suggests that they might actually use spectral flux to some extent to perform their dissimilarity ratings, but this concerns only the sounds with the highest spectral flux values, and their contribution to the dissimilarity ratings is much smaller than for attack time and SCG [see the rightmost part of the psychophysical curves, Fig. 3(a)]. These MDS results are consistent with participants' subjective reports. Indeed they reported using attack-time information and "sound-color" information in their ratings but did not report using any *variation* of "sound-color."

The data from this first experiment contradict the idea that listeners take spectro-temporal variations into account in their dissimilarity ratings of timbre pairs. There are several possible explanations: either our model of spectral flux is not perceptually relevant, or it is a less salient parameter than the other two. Two facts must be mentioned here: First, participants of the preliminary adjustment experiment did not report any difficulty in adjusting the spectral flux value of the test tone, highlighting the fact that this parameter is salient when it is the only physical dimension to vary. Second, the values of the three physical parameters were chosen to correspond approximately to an equal range of perceived dissimilarity in unidimensional contexts. We therefore hypothesized that spectral flux salience might be sensitive to contextual effects, and particularly to the number of other dimensions varying concurrently. We investigated this hypothesis in the second experiment using physical spaces with only one or two parameters varying among the sounds.

III. EXPERIMENT 2: ONE- AND TWO-DIMENSIONAL SPACES WITH SPECTRAL FLUX

A. Method

1. Participants

Thirty-one listeners (aged 17–48 years, 22 females) were recruited to participate in this experiment. None of

them had participated in Experiment 1. None of them reported any hearing loss, and 15 of them had received musical training. Musicians had practiced music for 13 years on average (2–25 years). Listeners were naive as to the purpose of the experiment and were paid for their participation. They were randomly assigned to three groups (condition A: 11 subjects, conditions B and C: 10 subjects each).

2. Stimuli

Three different sets (A, B, and C) of 16 stimuli were constructed as in Experiment 1, except that for condition A SCG was kept constant (933 Hz, i.e., at the third harmonic), for condition B attack time was kept constant (15 ms), and for condition C both SCG and attack time were kept constant (933 Hz and 15 ms, respectively). Thus the stimulus set for condition A was the projection of the 3D space of Experiment 1 onto the SCG=3 plane, for condition B its projection onto the t1=15 ms plane, and for condition C its projection onto the spectral flux axis.

3. Procedure

Each of the three groups of subjects performed dissimilarity ratings on only one of the three sets of stimuli. The experimental procedure and statistical analyses were identical to those in Experiment 1 for conditions A and B, except for the number of latent classes in the MDS analyses, which was only allowed to vary between one and three given the smaller number of subjects in each condition. For the CONSCAL analyses, only the two physical variables used to create the sounds were entered in the analyses. For condition C, the experimental procedure was as in Experiment 1, but the analyses were different, since one may not perform unidimensional analyses with CLASCAL and CONSCAL. We simply looked at the correlation between dissimilarity measures and the distances predicted by the single physical axis.

B. Results

In each of the three conditions, the data from one subject were discarded after hierarchical clustering of the correlation matrix. The mean coefficients of determination with other subjects' ratings were $r^2(118)=0.13$, $r^2(118)=0.18$, and $r^2(118)=0.16$, for conditions A, B, and C, respectively. After removing these data, the mean coefficients of determination between subjects ratings were $r^2(118)=0.17$ (SD=0.03), $r^2(118)=0.23$ (SD=0.04), and $r^2(118)=0.20$ (SD=0.01), for conditions A, B, and C, respectively. In conditions A and B, CLASCAL analyses yielded 2D perceptual spaces, with two latent classes of subjects and no specificities. One-class CONSCAL models were retained in both cases. CLASCAL and CONSCAL spaces are represented in Figs. 2(c) and 3(b), and CLASCAL perceptual coordinates are reported in Table II.

1. Condition A: 2D attack and spectral flux

a. CLASCAL space. Dimension one of the perceptual space was correlated with attack time $[r^2(14)=0.75, p<0.0001]$, and again, the correlation was even better with its logarithm $[r^2(14)=0.91, p<0.0001]$. The second dimension was not significantly correlated with spectral flux values $[r^2(14)=0.05, p=0.4]$. A more detailed analysis revealed that the second dimension separated sounds with low spectral flux (data points collapsed in the center of the axis) from sounds with higher spectral flux, but within these two groups the sounds are not ordered according to spectral flux.

TABLE II. Physical and CLASCAL perceptual coordinates for Experiment 2. Units are as in Table I and the superscripts indicate the perceptual dimensions significantly correlated ($p<0.05$) with Attack (+), SCG (●), and SFI (×).

Experiment 2A (SCG=third harmonic)				Experiment 2B (Attack=15 ms)			
Physical space		Perceptual space		Physical space		Perceptual space	
Attack$^+$	SFl	Dim 1$^+$	Dim 2	SCG$^●$	SFl$^×$	Dim 1$^●$	Dim 2$^×$
15	0.00	−0.48	−0.03	3.00	0.00	−0.37	0.04
42	1.56	−0.03	0.37	3.10	1.56	−0.42	−0.38
100	0.12	0.22	−0.03	3.20	0.12	−0.33	−0.05
59	0.96	0.05	−0.08	3.30	0.96	−0.33	0.16
17	1.20	−0.38	0.18	3.40	1.20	−0.22	0.22
168	0.60	0.37	0.02	3.50	0.60	−0.15	0.01
141	1.80	0.34	0.16	3.60	1.80	−0.12	−0.29
35	0.48	−0.15	−0.15	3.70	0.48	−0.03	0.07
25	1.68	−0.33	−0.39	3.80	1.68	0.14	−0.32
84	0.84	0.16	0.15	3.90	0.84	0.09	0.14
199	1.32	0.31	0.32	4.00	1.32	0.14	−0.14
29	1.08	−0.04	−0.37	4.10	1.08	0.20	0.02
70	1.44	−0.06	0.13	4.20	1.44	0.22	0.19
119	0.24	0.22	0.03	4.30	0.24	0.38	0.05
21	0.36	−0.35	−0.06	4.40	0.36	0.40	0.16
50	0.72	0.14	−0.25	4.50	0.72	0.41	0.13

The two classes of subjects were composed of five subjects each. The first class weighted more heavily the first dimension corresponding to attack time than the second dimension (weights: 0.42 and 0.30), and the reverse was true for the second class (weights: 1.58 and 1.70). Both classes contained musically trained subjects (three out of five in class 1, two out of five in class 2).

b. CONSCAL space. The one-class model with orthogonal dimensions and having a cubic spline with one interiot knot for the attack-time dimension and a quadratic spline with two interior knots for the spectral-flux dimension was selected. This model was compared with the two-class 2D CLASCAL model. At the 5% and 1% levels, CLASCAL was selected ($p<0.01$ for rejecting CONSCAL).

2. Condition B: 2D SCG and spectral flux

a. CLASCAL space. Dimension one of the perceptual space was correlated with SCG $[r^2(14)=0.98, p<0.0001]$. Dimension two was weakly correlated with spectral flux values $[r^2(14)=0.26, p=0.04]$. Dimension two was more strongly correlated with the exponential of the spectral flux values $[r^2(14)=0.40, p=0.007]$.

The two classes of subjects were composed of three and six subjects. The second class weighted more heavily the first dimension corresponding to SCG than the second dimension (weights: 1.51 and 1.04), and the reverse was true for the first class (weights: 0.49 and 0.96). The first class was composed of three musically untrained subjects. Four out of six subjects in class 2 were musically trained.

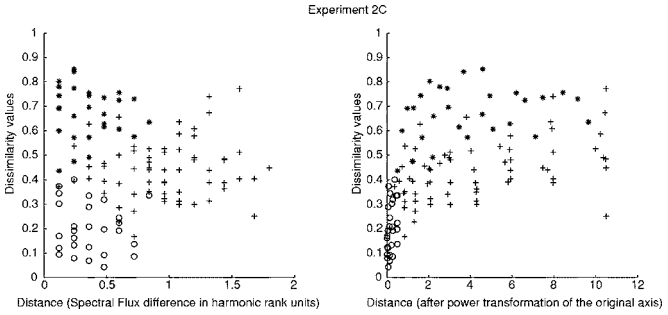

FIG. 5. Mean dissimilarities versus physical distances for Experiment 2C. Left panel: original physical distances along the abscissa in units of harmonic rank; right panel: physical distances computed after power transformation of the original physical axis. Dissimilarities along the ordinate are on an arbitrary scale [0-1]. (○) pairs of sounds with low spectral-flux values; (+) pairs of sounds with one having little spectral flux and one having a high spectral flux; (∗) pairs of sounds with high spectral flux values.

b. CONSCAL space. The one-class model with nonorthogonal dimensions having quadratic splines with three interior knots on both dimensions was selected ($0.01<p<0.02$ for rejecting the model with orthogonal dimensions). The off-diagonal coefficient in the 2×2 transformation matrix **A** [see Eq. (3)] was equal to −0.341. This means that in this model, the physical axes (SCG and spectral flux) are at an angle of 110° instead of being orthogonal (90°), suggesting that the perceptual effect of a high spectral flux is larger when the SCG in the steady portion of the tone is high (or conversely that the perceptual effect of a high SCG is larger when the spectral flux value is high). This one-class CONSCAL model was compared with the two-class 2D CLASCAL model, and CONSCAL was not rejected ($0.08<p<0.09$ for rejecting CONSCAL).

3. Condition C: One-dimensional spectral flux

As can been see in Fig. 5, dissimilarities were not well predicted by the distances along the spectral flux axis $[r^2(118)=0.0008, p=0.76]$. At first glance only a disappearance of very low dissimilarity ratings for the largest differences in spectral flux can be observed (compare the left and right halves of Fig. 5). A closer look at the results shows that even if dissimilarities and physical distances are not significantly correlated, there is some structure in the data. Because in Experiments 2A and 2B there was a trend for sounds with high spectral flux values to be separated from those with little or no spectral flux in subjects' ratings, we separated the 16 sounds into two groups of eight, according to spectral flux values (see Fig. 5). When the two sounds in a pair have little spectral flux (28 pairs), the perceived dissimilarity is small: 0.20 on average (SD=0.10), whereas it is bigger when one of the sounds has little spectral flux and the other a high spectral flux value: 0.44 on average (SD=0.12, 64 pairs). The average dissimilarity rating is even higher between two sounds with high spectral flux values: 0.67 (SD=0.11, 28 pairs). This confirms that sounds with high spectral flux values are separated from those with little spectral flux in dissimilarity ratings, and further that sounds with high spectral flux values are distinguished from one another.

Because CONSCAL analyses in Experiments 1, 2A, and 2B suggested that the psychophysical functions for the spectral flux axis were very compressive for low spectral flux values and expansive for high spectral flux values, we considered power transformations of the original physical scale. Instead of taking as explanatory physical distances the differences between spectral flux values (i.e., SF_j-SF_i as the distance between sounds i and j), we took the differences between power transformations of those values ($SF_j^a-SF_j^a$). In order to choose the most appropriate value for a, we computed the correlation coefficients between these transformed physical distances and dissimilarity ratings for increasing values of a. There was a sharp elbow in the correlation coefficient versus a curve for $a=4$, so we retained this value. The dissimilarity ratings were better predicted by these transformed physical distances than by the original ones $[r^2(118)=0.28, p<0.0001]$. For comparison, CONSCAL models explained more than 70% of the variance in dissimilarity ratings for the other experiments. So even if subjects used spectral flux in their dissimilarity ratings in this unidimensional context, spectral flux only explained part of the variance in the data, suggesting that there is a fair amount of uncertainty in listeners' ratings as far as this parameter is concerned. It may also be that the power transformation of the physical axis is not the best estimate of the psychophysical function for spectral flux.

C. Discussion

The perceptual significance of attack time and SCG was confirmed as in Experiment 1. The results of Experiment 2 shed new light on the relevance of spectral flux for timbre dissimilarity ratings. First, both in uni- and bi-dimensional contexts, listeners used spectral flux in their dissimilarity ratings, as shown in conditions A and B with the CONSCAL results, and in condition C by the correlation between transformed physical distances and dissimilarity ratings. In bi-dimensional contexts, as in Experiment 1, the range of dissimilarities accounted for by spectral flux is much smaller than that accounted for by either attack time or SCG. As compared with the first experiment (see Fig. 3), spectral flux information was used to a larger extent when variation only occurred along one concurrent dimension instead of two, and was more strongly inhibited by attack time than by SCG. This confirmed our hypothesis about the relationship between spectral flux salience and the number of concurrent variable acoustical dimensions.

These results raise a number of issues. First, it might be difficult to use a dynamic parameter when making dissimilarity ratings, especially in our spectral flux case, which values may not easily be ordered on a perceptual scale (see Experiment 2C). Second, there might exist interactions in the processing of spectral flux and other timbre dimensions. In particular, the preliminary adjustment data used to set the range of variation of spectral flux in the set of stimuli were collected with constant attack time (15 ms) and SCG (933 Hz), but the perceptual effect of a given spectral flux value might depend on attack time or SCG. The nonorthogonality of the dimensions of the CONSCAL space in Experiment 2B favors the existence of interactions of this type between

SCG and spectral flux. Finally, the spectral flux model chosen might have a limited perceptual salience. It is possible that modeling spectral flux by a fluctuation of the spectrum during the plateau of the sound might be more salient. The perceptual effects of such fluctuations have been studied by Hajda (1999), who was able to relate them to vibrato perception. We propose that subjects might be more sensitive to spectral variations when they are pseudoperiodic and when several cycles of variations are present in the sound.

The question remains open concerning whether or not it is possible to find a third dimension of timbre that is salient enough to be used in dissimilarity ratings when both attack time and SCG vary. In Experiment 3 we tested an additional dimension to address this issue.

IV. EXPERIMENT 3: ATTACK TIME, SPECTRAL CENTROID, AND SPECTRAL IRREGULARITY

A. Method

1. Participants

Thirty listeners (aged 19–45 years, 18 female) participated in this experiment. None of them participated in either of the previous experiments. None of them reported any hearing loss, and 15 of them had received musical training. Musicians had practiced music for 11 years on average (3–38 years). Listeners were naive as to the purpose of the experiment and were paid for their participation.

2. Stimuli

Sixteen sounds were created as in Experiment 1, except that the spectral flux dimension was replaced by a selective attenuation of even harmonics relative to odd harmonics. A preliminary adjustment experiment (with six listeners not included in the main experiment) was performed, as in Experiment 1, in order to choose a range of variation of this parameter that was perceptually equivalent to that for the other two parameters. In the main experiment, attenuation of even harmonics ranged from 0 to 8 dB, and could take 16 different values separated by equal steps (in dB, see Table III). In the preliminary experiment, subjects were also requested to adjust the level of a tone with attenuated even harmonics to match the loudness of a reference tone without attenuated even harmonics. We found that it was not necessary to adjust the level of the attenuated-even-harmonic tones.

3. Procedure

The experimental procedure and statistical analyses were as in Experiment 1.

B. Results

On the basis of hierarchical clustering of the correlation matrix, the data from two subjects were discarded from subsequent analyses. The mean coefficient of determination between these two subjects' and other subjects' ratings were $r^2(118)=0.08$ and $r^2(118)=0.10$, respectively. After removing those subjects, the mean coefficient of determination between subjects' dissimilarity ratings was $r^2(118)=0.23$ (SD $=0.04$).

TABLE III. Physical and CLASCAL perceptual coordinates for Experiment 3. Units are as in Table I and the superscripts indicate the perceptual dimensions significantly correlated ($p<0.05$) with Attack (+), SCG (●), and EHA (∗). Even harmonic attenuation (EHA) is given in dB relative to odd harmonics.

Physical space			Perceptual space		
Attack^{+}	SCG$^{\bullet}$	EHA*	Dim 1*	Dim 2^{+}	Dim 3$^{\bullet}$
15	3.00	0.00	−0.33	0.14	−0.37
42	3.10	6.93	0.34	0.18	−0.23
100	3.20	0.53	−0.10	−0.21	−0.23
59	3.30	4.27	0.17	−0.03	−0.17
17	3.40	5.33	0.09	0.31	−0.14
168	3.50	2.67	0.09	−0.26	−0.11
141	3.60	8.00	0.39	−0.05	−0.07
35	3.70	2.13	−0.28	0.10	0.05
25	3.80	7.47	0.22	0.30	0.01
84	3.90	3.73	0.02	−0.13	0.10
199	4.00	5.87	0.29	−0.19	0.05
29	4.10	4.80	−0.05	0.16	0.24
70	4.20	6.40	0.20	0.01	0.19
119	4.30	1.07	−0.28	−0.28	0.16
21	4.40	1.60	−0.47	0.02	0.25
50	4.50	3.20	−0.29	−0.07	0.28

1. CLASCAL space

MDS analysis showed that a 3D model without specificity values and two latent classes of subjects best fit the data. The three perceptual dimensions correlated with the three physical dimensions that were included in the synthetic space [Fig. 2(d) and Table III]. Dimension one was correlated with the degree of attenuation of even harmonics [$r^2(14)=0.74, p<0.0001$]. Dimension two was correlated with attack time [$r^2(14)=0.63, p=0.0001$], and with its logarithm [$r^2(14)=0.72, p<0.0001$]. Finally dimension three was correlated with SCG [$r^2(14)=0.95, p<0.0001$].

The two classes of subjects were composed of 8 and 19 subjects, respectively. One subject could not be assigned with certainty to either of the classes. The first class weighted more heavily the first dimension corresponding to even-harmonic attenuation (weights: 0.84, 0.67, and 0.60), and the second class weighted more heavily the third dimension corresponding to SCG (weights: 1.16, 1.13, and 1.40), suggesting that subjects in the different classes may favor one of the spectral dimensions over the other. Both classes contained musically trained subjects (5 out of 8 in class 1 and 9 out of 19 in class 2).

2. CONSCAL space

The one-class model with orthogonal dimensions having quadratic splines with one interior knot on all three dimensions was retained. When compared with the two-class 3D CLASCAL model, CLASCAL was selected at the 5% level ($0.04<p<0.05$ for rejecting CONSCAL), but at the 1% level, CONSCAL would not have been rejected. It is interesting to note that here [see Fig. 3(c)], the ranges of dissimilarities accounted for by the different dimensions are about equal, which contrasts with the results of Experiments 1 and 2.

C. Discussion

As in the two previous experiments, attack time and SCG appeared as two major determinants of timbre. Spectral irregularity (modeled with an attenuation of even-harmonic amplitudes) was also confirmed as a salient parameter of timbre. All three dimensions appear to have the same perceptual status: they contribute along a continuous scale to the dissimilarity ratings. Notice in Fig. 3 that the psychophysical functions are slightly S-shaped for the three parameters, but that there is no gross distorsion of the physical scale. This confirms that the perceptually relevant parameters are the logarithm of attack time, SCG on a linear frequency scale, and attenuation of even harmonics relative to odd ones on a linear dB scale. It is likely that with larger ranges of variation of these physical parameters saturation may occur.

V. GENERAL DISCUSSION

A. Timbre space dimensions

The experiments reported here were designed as direct tests of timbre space models. Synthetic timbre spaces were constructed using acoustical dimensions derived from natural and simulated instrument studies, and their perception was characterized using multidimensional scaling of dissimilarity ratings. We sought to confirm whether the structure of the perceptual spaces would closely parallel that of the acoustical spaces.

The present study intended in particular to confirm that attack time, spectral centroid, spectral flux, and attenuation of even harmonics explained timbre space dimensions, using fully controlled synthetic timbres. Overall it appeared to be the case, with one important restriction: spectral flux was hardly used in dissimilarity ratings when attack time and SCG varied concurrently in the stimulus set. Its contribution to timbre dissimilarity ratings was only evident when a single other attribute varied in the stimulus set. Hence, the salience of spectro-temporal parameters such as spectral flux might be more context-dependent than that of other parameters, because it decreases when the number of other parameters that vary in the stimulus set increases.

All the experiments reported here confirm that attack time and spectral centroid are salient timbre parameters, and can be used in a continuous fashion for dissimilarity ratings. As suggested earlier (e.g., McAdams *et al.*, 1995), it appears that the logarithm of attack time and not linear attack time is used in dissimilarity ratings. Spectral centroid seems to be ordered perceptually on a linear frequency scale. Attenuation of even harmonics was also confirmed as a salient timbre dimension. Attenuation in dB of even harmonics was used linearly in listeners' dissimilarity ratings. We propose that this dimension of timbre can be more generally interpreted as a model of the degree of spectral irregularity or spectrum fine structure. It is worth mentioning that people were able to use two different types of spectral information simultaneously in their ratings, one related to the global shape of the spectrum, modeled by SCG, and one related to the local shape of the spectrum, modeled by even-harmonic attenuation.

All together the present study fits well with the commonly accepted idea that attack time and spectral centroid are major determinants of timbre (Grey, 1977; Krimphoff, McAdams, and Winsberg, 1994; McAdams *et al.*, 1995; Marozeau *et al.*, 2003). The experiments reported here shed new light on possible third dimensions of timbre spaces. As stated in Sec. I, there is no established consensus on that topic. Proposed third dimensions mostly fall into two categories: spectro-temporal parameters (e.g., spectral flux) or spectral parameters (e.g., spectrum fine structure, spectral spread) related to the shape of the spectrum. Progressive expansion of the spectrum toward higher harmonics, a possible source of spectral flux, was not very influential in dissimilarity ratings when attack time and SCG also varied in the stimulus set (Experiment 1). Conversely, participants were able to use attenuation of even harmonics in dissimilarity ratings when attack time and SCG varied (Experiment 3). Therefore among all the previously proposed acoustic correlates of timbre dimensions, some might be more adequate than others. In particular, the type of spectro-temporal information that listeners can use in dissimilarity ratings remains unclear and warrants further testing.

The contrast between the respective contribution of spectral flux (Experiment 1) and attenuation of even harmonics (Experiment 3) to timbre dissimilarity ratings raises a number of questions related to dimension salience. In the current study, we have taken the contribution of a given parameter to dissimilarity ratings as an index of its perceptual salience, as in Miller and Carterette (1975). One might ask whether this index would predict performances in other tasks, such as timbre analogy judgments (see McAdams and Cunibile, 1992, for example), or speeded classification tasks (see Krumhansl and Iverson, 1992, for example). Furthermore, the relationship between dimension salience as defined here and sound source recognition remains to be explored.

B. Comparison of CLASCAL and CONSCAL models

We were able to directly compare CLASCAL and CONSCAL models in this study. Overall both of them appear adequate to model dissimilarity ratings. Out of four cases, CONSCAL was selected twice (Experiments 1 and 2B), CLASCAL once (Experiment 2A), and in one case (Experiment 3) it was not possible to decide with certainty. CONSCAL offers the theoretical advantage of requiring a much smaller number of parameters to be estimated than CLASCAL, and the psychophysical functions produced by CONSCAL analyses can provide additional information concerning the mapping of signal properties in the auditory system. Furthermore, the contribution to dissimilarity ratings of less salient dimensions such as spectral flux was better captured using CONSCAL than was the case with CLASCAL. Nevertheless, it is not possible to use such a parcimonious model without *a priori* knowledge about the underlying acoustical dimensions. Interestingly, when in all four cases CLASCAL analyses yielded two latent classes of subjects, only one class was necessary for CONSCAL models. This suggests that the shapes of psychophysical functions are rather similar across subjects.

C. Inter-individual differences

The present data suggest that inter-individual differences exist when making timbre dissimilarity ratings. Indeed, in all experiments, subjects were classified by CLASCAL into different classes of subjects that weighted timbre space dimensions differently. Several conclusions emerge from Experiments 1 and 3, where there are sufficient numbers of subjects to draw conclusions about inter-individual differences. First, the difference between subject classes depends partly upon the range of the rating scale used effectively. Another important difference in the classes concerns the different relative saliences of the axes for each class. In the first experiment, some subjects favored temporal information (attack time) over spectral information (SCG), and for other subjects the reverse was true. In the third experiment, the two classes of subjects favored a different type of spectral information. Whether these different weights reflect differences at a perceptual level or at a decisional level remains an open question. Finally, these inter-individual differences do not seem to be related to musical training. All together these results are consistent with previous findings (e.g., McAdams *et al.*, 1995), although it might be that the differences between the weights on the different dimensions are smaller in the present case. For example, in the McAdams *et al.* (1995) study the ratios between the weights on two axes could be as big as 3, whereas they never exceed 2 in our case. This could result from the greater homogeneity of synthetic timbre spaces as opposed to spaces of simulated natural instruments. In the latter case, it is likely that subjects select a limited number of acoustical parameters among all possibilities to make their ratings. Such a selection step is not necessary in the present context of low-dimensional synthetic timbre spaces.

D. Complex sound perception

Timbre space studies raise general questions about complex sound perception. In particular, one might hypothesize that the salient parameters found in these studies correspond to parameters that are extracted along the auditory pathway. There could either be pathways dedicated to the processing of each of these parameters or a more global mechanism might be involved. The present experiments were not designed to answer such a question, but using calibrated perceptual spaces such as those described here will prove to be a powerful tool for answering questions about the processing of complex sound dimensions.

ANSI (**1973**). *American National Standard—Psychoacoustical Terminology* (American National Standards Institute, New York).

Berger, K. W. (**1964**). "Some factors in the recognition of timbre," J. Acoust. Soc. Am. **36**, 1888–1891.

Carroll, J. D., and Chang, J. J. (**1970**). "Analysis of individual differences in multidimensional scaling via an *n*-way generalization of Eckart-Young decomposition," Psychometrika **35**, 283–319.

Grey, J. M. (**1977**). "Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. **61**, 1270–1277.

Hajda, J. M. (1999). "The effect of time-variant acoustical properties on orchestral instrument timbres," Ph.D. thesis, University of California, Los Angeles.

Hajda, J. M., Kendall, R. A., Carterette, E. C., and Harschberger, M. L. (**1997**). "Methodological issues in timbre research," in *Perception and Cognition of Music*, edited by I. Deliège and J. Sloboda (Psychology, Hover), pp. 253–307.

Handel, S. (**1995**). "Timbre perception and auditory object identification," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 425–461.

Hope, A. C. (**1968**). "A simplified Monte Carlo significance test procedure," J. R. Stat. Soc. Ser. B. Methodol. **30**, 582–598.

Kendall, R. A., and Carterette, E. C. (**1993**). "Verbal attributes of simultaneous wind instruments timbres. I. von Bismarck's adjectives," Music Percept. **10**, 445–468.

Krimphoff, J., McAdams, S., and Winsberg, S. (**1994**). "Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique [Characterization of the timbre of complex sounds. II. Acoustical analyses and psychophysical quantification]," J. Phys. (Paris) **C5**, 625–628.

Krumhansl, C. L. (**1989**). "Why is musical timbre so hard to understand?," in *Structure and Perception of Electroacoustic Sound and Music*, edited by S. Nielsen and O. Olsson (Elsevier, Amsterdam), pp. 43–53.

Krumhansl, C. L., and Iverson, P. (**1992**). "Perceptual interactions between musical pitch and timbre," J. Exp. Psychol. Hum. Percept. Perform. **18**, 739–751.

Kruskal, J. B. (**1964a**). "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," Psychometrika **29**, 1–27.

Kruskal, J. B. (**1964b**). "Nonmetric multidimensional scaling: A numerical method," Psychometrika **29**, 115–129.

Lindemann, E., Puckette, M., Viara, E., De Cecco, M., and Dechelle, F. (**1991**). "The architecture of the IRCAM musical workstation," Comput. Music J. **15**, 41–49.

Marozeau, J., de Cheveigné, A., McAdams, S., and Winsberg, S. (**2003**). "The dependency of timbre on fundamental frequency," J. Acoust. Soc. Am. **114**, 2946–2957.

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 University Press, Oxford), pp. 146–198.

McAdams, S., and Cunibile, J. C. (**1992**). "Perception of timbral analogies," Philos. Trans. R. Soc. London, Ser. B **336**, 383–389.

McAdams, S., and Winsberg, S. (**2000**). "Psychophysical quantification of individual differences in timbre perception," in *Contributions to Psychological Acoustics 8*, edited by A. Schick, M. Meis, and C. Reckhardt (BIS, Oldenburg), pp. 165–181.

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.

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.

Saldanha, E. L., and Corso, J. F. (**1964**). "Timbre cues and the identification of musical instruments," J. Acoust. Soc. Am. **36**, 2021–2026.

Samson, S., Zatorre, R. J., and Ramsay, J. O. (**1997**). "Multidimensional scaling of synthetic musical timbre: Perception of spectral and temporal characteristics," Can. J. Exp. Psychol. **51**, 307–315.

Smith, B. (**1995**). "PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation," in *Society for Music Perception and Cognition*, edited by D. Wessel (University of California Press, Berkeley).

Torgerson, W. S. (**1958**). *Theory and Methods of Scaling* (Wiley, New York).

Winsberg, S., and Carroll, J. D. (**1989**). "A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model," Psychometrika **54**, 217–229.

Winsberg, S. and De Soete, G. (**1993**). "A latent-class approach to fitting the weighted Euclidean model, CLASCAL," Psychometrika **58**, 315–330.

Winsberg, S., and De Soete, G. (**1997**). "Multidimensional scaling with constrained dimensions: CONSCAL," Br. J. Math. Stat. Psychol. **50**, 55–72.