The dependency of timbre on fundamental frequency[a)]

Jeremy Marozeau,[b)] Alain de Cheveigné,[c)] Stephen McAdams, and Suzanne Winsberg
Institut de Recherche et Coordination Acoustique/Musique (Ircam-CNRS), 1, place Igor Stravinsky, F-75004 Paris, France

(Received 15 August 2002; revised 25 June 2003; accepted 25 August 2003)

The dependency of the timbre of musical sounds on their fundamental frequency (F_0) was examined in three experiments. In experiment I subjects compared the timbres of stimuli produced by a set of 12 musical instruments with equal F_0, duration, and loudness. There were three sessions, each at a different F_0. In experiment II the same stimuli were rearranged in pairs, each with the same difference in F_0, and subjects had to ignore the constant difference in pitch. In experiment III, instruments were paired both with and without an F_0 difference within the same session, and subjects had to ignore the variable differences in pitch. Experiment I yielded dissimilarity matrices that were similar at different F_0's, suggesting that instruments kept their relative positions within timbre space. Experiment II found that subjects were able to ignore the salient pitch difference while rating timbre dissimilarity. Dissimilarity matrices were symmetrical, suggesting further that the absolute displacement of the set of instruments within timbre space was small. Experiment III extended this result to the case where the pitch difference varied from trial to trial. Multidimensional scaling (MDS) of dissimilarity scores produced solutions (timbre spaces) that varied little across conditions and experiments. MDS solutions were used to test the validity of signal-based predictors of timbre, and in particular their stability as a function of F_0. Taken together, the results suggest that timbre differences are perceived independently from differences of pitch, at least for F_0 differences smaller than an octave. Timbre differences can be measured between stimuli with different F_0's.
 [DOI: 10.1121/1.1618239]

PACS numbers: 43.75.Cd, 43.66.Jh, 43.66.Hg [NFV] Pages: 2946–2957

I. INTRODUCTION

The word ''timbre'' has several meanings. In a musical context it designates aspects of sound that allow an instrument to be identified and distinguished from others. In the context of psychoacoustic experiments, it designates an elementary sound quality akin to pitch or loudness (the ''Klangfarbe'' of Helmholtz, 1885). In the next paragraph we shall use the words ''identity'' and ''quality'' to distinguish these two meanings, respectively. The identity of a musical instrument obviously depends in some way on the quality of the sounds it produces (their ''timbre'' in a psychoacoustic sense). However, this dependency may be complex.

For certain instruments, quality varies as a function of the note played, the intensity at which it is played, and time. This is obvious from casual listening, and corroborated by measurements or calculations that show variations of signal properties that are known to affect sound quality (spectral centroid, harmonicity, etc.) (Martin, 1999). For example, notes of the trumpet become brighter with increased intensity (Luce and Clark, 1967), while those of the violin are subject to complex interactions between body resonances and the harmonic spectrum of string vibration (Fletcher and Rossing, 1998). The latter varies with fundamental frequency (F_0) and thus with the note played. The timbre of a wind instrument may change abruptly between the low register (with the register hole closed) and the high register (with the hole open), a characteristic revealed only if the instrument is played over a range of notes (Risset and Wessel, 1999). Sound qualities produced by a particular instrument follow a particular ''trajectory,'' and indeed we could formulate the hypothesis that this in part determines its identity. In other words, timbre (identity) might depend on the *pattern of variation* of timbre (quality) specific to an instrument. To test such a hypothesis experimentally requires comparing timbre (quality) across time, intensity, or F_0. The purpose of the present study was to characterize variations of timbre (quality) as a function of F_0.

The standard methodology for studying timbre is multidimensional scaling (MDS) (Grey, 1977). Typically, subjects are presented with pairs of sounds and asked to rate their dissimilarity on a continuous scale. Dissimilarity scores are processed by an MDS algorithm to produce models of ''timbre space'' that give insight into the nature of the timbre percept. It is usually found that the timbre space involved in a task is of small dimensionality (two to four dimensions), that different subjects may weight dimensions differently, and that these dimensions can usually be predicted by signal-based ''descriptors.'' The relevant dimensions (and corresponding descriptors) tend to vary between experiments, no doubt as a function of the set of sounds included in each experiment. Nevertheless certain dimensions (e.g., ''brightness,'' predicted by a ''spectral centroid'' descriptor) tend to recur in all. MDS seems the appropriate tool to study variations of timbre with F_0.

There are potential problems however. A difference in

[a)]Portions of these results were presented at the 141st meeting of the Acoustical Society of America.
[b)]Electronic mail: jeremy.marozeau@ircam.fr
[c)]Electronic mail: alain.de.cheveigne@ircam.fr

F_0 produces a difference in *pitch* that adds to the dissimilarity between sounds. Even if this extra term is constant, its contribution sets a lower limit to every dissimilarity score, and so the method could be insensitive to small variations in timbre. Worse, if the F_0-dependent term varies, these variations would be confounded with variations of timbre-related dissimilarity and affect the validity of cross-F_0 timbre comparisons. Past studies that allowed F_0 to vary generally found that pitch dominated dissimilarity at the expense of timbre (Miller and Carterette, 1975), with the result that MDS solutions were relatively insensitive to F_0-induced variations of timbre.

One solution is to instruct subjects to *ignore* pitch when making timbre judgments. Unfortunately we are not *a priori* certain that they can do so. Pitch and timbre might not be *separable*, that is, timbre comparisons may be possible between sounds with the same pitch, but not between sounds with different pitch. This worry is reinforced by the scarcity of F_0-dependent timbre studies in the past. A first aim of our study was to determine whether subjects can reliably make cross-F_0 comparisons of timbre while ignoring differences in pitch.

If they can, we may hope to bring an empirical answer to a question such as: how does timbre change with F_0? Two sorts of change are to be expected: first, instrument-specific changes such as evoked earlier, for example due to changes of resonator geometry as a function of the note played, and second, hypothetical changes of a more basic nature, due to a perceptual interaction between pitch and timbre, or the presence of F_0 as a cofactor in the relation between signal descriptors and psychophysical dimensions of timbre. A second aim of our study was to measure F_0-dependent timbre changes, in particular of a basic, noninstrument-specific nature.

There are reasons to expect interactions between pitch and timbre. While pitch is defined as ''that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high'' (ANSI, 1960), more complex structures have been proposed such as a spiral involving both a linear dimension of tone height and a circular dimension of chroma (Shepard, 1964; Ueda and Nimmo-Smith, 1987). Chroma is related to fundamental periodicity, while tone height depends more on the spectral envelope (Patterson *et al.*, 1993). The envelope also determines timbre, so it could be that timbre and pitch are not entirely distinct. This might result in nonseparability (if a pitch difference degrades comparisons between timbre) or a systematic shift (if pitch and timbre are partly colinear).

To a first approximation, the spectral envelope of a vowel does not change with variations of F_0, and vowel identity (another usage of ''timbre'') is likewise relatively invariant. Small systematic variations have nevertheless been observed (see de Cheveigné and Kawahara 1999, for a review). Slawson (1968) asked subjects to adjust the formant frequencies of different-F_0 vowels so that they had the same timbre. The best match was obtained for a 10% increase of formant frequencies for a one-octave increase in F_0. This suggests that envelope-related dimensions of timbre might depend on F_0 in addition to their dependency on envelope characteristics. In other words, F_0 might need to be included as a cofactor in the formulas of signal-based descriptors that predict those dimensions.

A third aim of our study was to test the validity of signal-based descriptors across F_0. Signal-based measures that correlate well with perceptual dimensions revealed in MDS studies (such as spectral centroid, log attack time or spectral flux) have been proposed as ''descriptors'' for applications such as the retrieval of multimedia data (Misdariis *et al.*, 1998; Peeters *et al.*, 2000). Such applications involve data at a wide range of F_0's, yet these descriptors have been tested only with a restricted set of F_0's (often only one). There is clearly a need to verify their generality, and if necessary to modify them to improve their generality. This might entail adjustment of the formulas to remove a spurious F_0 dependency, or inclusion of an F_0-dependent corrective term or, in the extreme, establishment of an array of F_0-dependent formulas.

It is worth discussing the forms of dependency of timbre on F_0 that we expect to find. Supposing a ''timbre space'' such as revealed in MDS studies, three hypotheses can be distinguished: (1) invariance of instrument positions with changes in F_0, (2) isometric displacement keeping relative positions invariant, and (3) non-isometric displacement.

According to hypothesis (1), variations of timbre with F_0 are negligible compared for example to between-instrument differences. Hypothesis (2) allows for a rotation or drift in timbre space common to all instruments. Hypothesis (3) allows that timbres of individual instruments change in arbitrary ways. The experiments were designed to decide between these hypotheses.

We used recordings of natural musical instrument sounds as stimuli. By doing so we confounded two sorts of F_0-dependent timbre changes: those specific to instruments, and those of a non-instrument-specific nature. We reasoned that natural instrumental sounds would guarantee the musical relevance of our sampling, while instrument-specific effects could be interpreted by a *posthoc* analysis of the waveforms of the stimuli.

II. EXPERIMENTS

A. Experiment I

Experiment I consisted of three sessions labeled a, b, and c. In each, subjects rated the dissimilarity between stimuli with the same F_0. This F_0 varied from session to session.

1. Methods

a. Stimuli. Ten natural and two synthetic instruments were used. Each instrument was played at three notes: B3 (247 Hz), C♯4 (277 Hz) and Bb4 (466 Hz), chosen to explore the effects of a small difference (two semitones: B3–C♯4) and a moderate difference (11 semitones: B3–Bb4) of F_0. Natural instrument samples were extracted from the Studio On Line (SOL) database of Ircam (IRCAM, 2000): *guitar* (B3 was played on the E string, C♯4 on the A string, and Bb4 on the D string), *harp, violin pizzicato* (B3 and C♯4 on the G string, Bb4 on the D string), *bowed violin* (strings were the same as for the violin pizzicato), *bowed double bass* (all

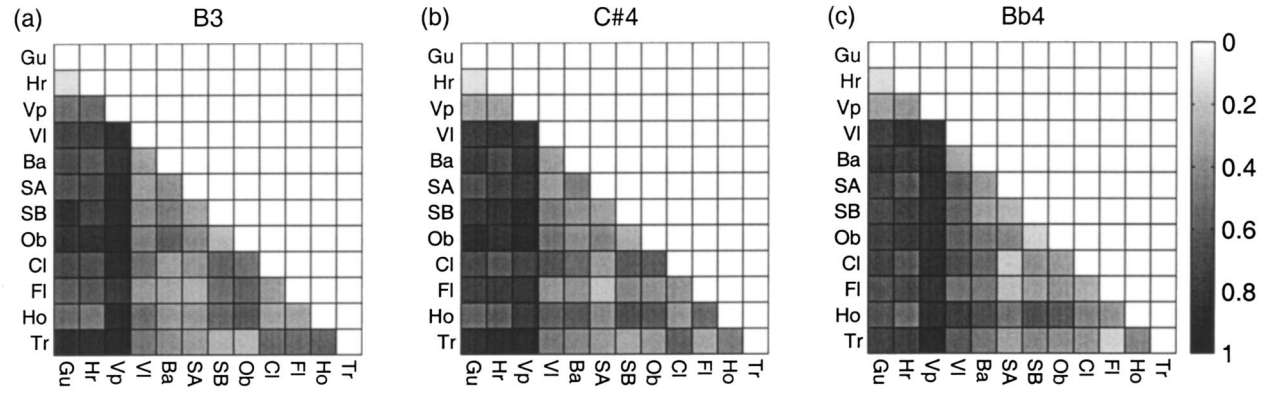

FIG. 1. Dissimilarity matrices for the three sessions of experiment I. Within each matrix, each square represents the dissimilarity between two instruments. Darker means greater dissimilarity. The first three columns of each matrix correspond to impulsive instruments (Gu, Hr, Vp). Dissimilarities are greater for pairs that associate an impulsive instrument with a sustained instrument than for pairs of instruments within either group. Patterns of dissimilarity are similar at each F_0.

notes were played on the G string), *oboe, clarinet, flute, horn in F, trumpet in C*. In the following, these instruments will be abbreviated as Gu, Hr, Vp, Vl, Ba, Ob, Cl, Fl, Ho, and Tr, respectively. In addition to these natural instruments, synthetic instruments SA and SB were created using fixed spectral envelopes derived from that of the saxophone.

Stimuli were clipped to a duration of 1.5 s by applying a 200-ms cosinusoidal offset ramp. Amplitudes were determined by asking six subjects (who did not participate in the main experiments) to adjust levels of stimuli presented at approximately 60 dB SPL for equal loudness. Stimuli were sampled at a rate of 44 100 Hz with 16-bit resolution, and presented diotically over Sennheiser 520 II headphones.

b. Subjects. Twenty-seven subjects aged 22 to 30 (14 men and 13 women, 15 musicians and 12 nonmusicians), participated in the experiment. Musicians were defined as having played an instrument for at least 3 years.

c. Procedure. Before the experiment, the subjects were informed that the goal of the experiment was to estimate the similarity of timbre between sounds. Timbre was defined as ''the fourth component of sound quality, the first three being pitch, loudness and duration.'' For each pair, they were instructed to judge whether the timbres were similar or different, using the entire scale of the cursor. Eventual differences of pitch, loudness, duration or ''recording noise'' were to be ignored. The identity of the instrument, if recognized, was also to be ignored. Subjects sat inside an audiometric booth. Presentation software was based on the PsiExp environment (Smith, 1995). The screen comprised a mouse-controlled cursor labeled from ''similar'' (coded 0) to ''different'' (coded 1), and two buttons (one to listen to the pair again, the other to validate the response). The experiment consisted of three sessions that were performed on the same day, separated by 5-min breaks. Before each session the subjects were presented with each of the 12 stimuli in random order to acquaint them with the range of timbre differences in the set of instruments. They were then presented with the full set of 66 pairs of different stimuli. The order within pairs and the order of pairs were random (a different randomization was used for each session and subject). Data for this and the following experiments are available at http://www.ircam.fr/pcm/archive/timbref0.

2. Results

a. Outliers, effect of musical experience. Correlation coefficients between dissimilarity scores were calculated for all pairs of subjects. These scores were submitted to a hierarchical cluster analysis based on the nearest-neighbor (complete linkage) algorithm (Kaufman and Rousseeuw, 1990). On the basis of this and a similar analysis for experiment II, three subjects were discarded for both experiments. Analysis was performed on data of the remaining 24 subjects.

To reveal an eventual effect of musical experience, an analysis of variance (ANOVA) was performed for each session with between-subjects factor musical experience (2) and within-subjects factor instrument pair (66), taking into account the fact that experience levels were represented by variable numbers of subjects (Abdi, 1987; Wonnacott and Wonnacott, 1990). No effect of musical experience was found, either as a main effect $[F(1,22)<1]$ or as an interaction $[F(65,1430)<1]$. Data for both groups are subsequently combined.

b. Dissimilarity matrices. Dissimilarity scores for each subject and session were placed in a matrix of dimension nXn, where n is the number of stimuli and the ijth entry $(i>j)$ is the dissimilarity between stimuli i and j. Since order was not distinguished, only the lower triangle was filled. Matrices averaged over subjects are plotted in Fig. 1 for the three sessions. Averaged over F_0's and subjects, dissimilarities ranged from 0.146 between the guitar and the harp to 0.872 between the trumpet and the violin pizzicato. One can distinguish two groups of instruments: impulsive (Gu,Hr,Vp) and sustained (Vl,Ba,SA,SB,Ob,Cl,Fl,Ho,Tr). Dissimilarities tended to be small within each group (upper left and lower right triangles) and large between groups (lower left rectangle), a pattern that was stable across F_0's.

To quantify the effects of F_0, a repeated-measures ANOVA was performed with factors $F_0(3) \times \text{instrument pair}(66)$. Results are shown in Table I. The

TABLE I. ANOVA table for experiment I. S: subjects, F_0: fundamental frequency, I: instrument pairs, SS: sum of squares, MS: mean square, F: F-values, ϵ: Greenhouse–Geisser correction factor applied to the degrees of freedom, p: corrected p-value, R^2: percentage of total variance accounted for by each effect. Adding a total of 48.5% due to intersubject differences, variance scores sum to 100%.

Source	df	SS	MS	F	ϵ	p	R^2
S	23	32.02	1.39				
F_0	2	0.93	0.46	5.72	0.90	0.008	**0.3**
F_0*S	46	3.73	0.08				
I	65	167.65	2.58	53.97	0.10	0.0001	**46.6**
I*S	1495	71.45	0.05				
F_0*I	130	16.53	0.13	5.65	0.122	0.0001	**4.6**
F_0*I*S	2990	67.29	0.02				

effects of both main factors were significant, as was their interaction. It is instructive to consider effect sizes. The percentage of total variance accounted for by each effect is indicated by the R^2 coefficient (last column in Table I) (Wonnacott and Wonnacott, 1990). The main effect of instrument pair represents the part of interinstrument dissimilarity that is constant across F_0. It accounts for about 47% of the variance. The interaction and main effect of F_0 together represent the part of dissimilarity that varies across F_0. They account for only about 5%. In agreement with the relatively small interaction, correlation coefficients between matrices (averaged over subjects, considering only the lower triangular parts) are relatively large: 0.88 between "a" and "b," 0.81 between "a" and "c," and 0.89 between "b" and "c" (df=64, $p<0.001$ for all three coefficients).

It could be argued that F_0-related effects are dwarfed by the contrast between impulsive and sustained instruments. Table II shows the percentage of variance accounted for by each effect for the full data set (column 2), or when dissimilarity scores are restricted to pairs of impulsive, sustained, or impulsive and sustained instruments (columns 3–5). After removing this major source of F_0-independent variance, as expected, F_0-independent effects represent a smaller proportion of total variance. However they still are larger than F_0-related effects.

To summarize the results of experiment I, interinstrument *timbre dissimilarities* varied significantly with F_0, but the variation was relatively small. It would be nice to conclude that *timbres* themselves were stable to the same degree [hypothesis (1) of the Introduction]. Unfortunately the results of experiment I do not allow us to draw that conclusion. As

TABLE II. Percentage of variance accounted for by each effect in experiment I, for the complete data set (All) or for data restricted to pairs of impulsive or sustained instruments, or to mixed pairs (impulsive and sustained). The first line (F_0-related) represents the sum of the F_0 effect and its interactions. The last line (other) represents variance due to disagreement between subjects. Each column sums to 100.

	R^2(%)			
Source	All	Impulsive	Sustained	Mixed
F_0-related	4.9	10.2	9.5	3.8
I	46.6	26.2	16.4	17.1
Other	48.5	63.6	74.1	79.1

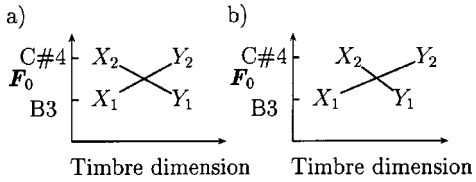

FIG. 2. Experiment II. Illustration of the hypothetical effect of F_0 on a dimension of timbre space (abscissa) along which two instruments X and Y differ. The instruments are represented at two different F_0's as X_1, Y_1 and X_2, Y_2 respectively, and it is supposed that $\overline{X_1Y_1}=\overline{X_2Y_2}$, as found in experiment I. The left plot illustrates hypothesis (1) (invariance) and the right hypothesis (2) (isometric shift). The latter implies $\overline{X_1Y_2}\neq\overline{X_2Y_1}$.

comparisons were made only at constant F_0, an eventual shift or rotation of the entire set of instruments in timbre space [hypothesis (2)] could not be detected. Furthermore, as subjects were instructed to use the full scale of dissimilarities in each session, an eventual compression or expansion also could not be detected. The next experiment allows for a shift, rotation, compression, or expansion to be detected.

B. Experiment II

In experiment II subjects rated the dissimilarity between stimuli with a constant *difference* of F_0 (ΔF_0) of either 2 semitones or 11 semitones. In contrast to experiment I, the response matrices were full, as each instrument pair was compared using both F_0 orders, and same-instrument pairs were included. Subjects were instructed to ignore differences in pitch which, contrary to experiment I, were salient. Supposing they can do so, this experiment allows us to refine the conclusions of experiment I, and in particular to decide between hypotheses (1) (invariance) and (2) or (3) (isometric or non-isometric deformation).

If hypothesis (1) is true, dissimilarity matrices should show three features. First, values on the diagonal should be zero. Second, the matrix should be symmetric: the lower triangular part should be the mirror image of the upper triangular part. Third, the lower triangular part should be identical to that observed at each F_0 in experiment I. To understand why the matrix should be symmetric, consider two instruments (X and Y) that differ along some dimension of timbre space (abscissa of Fig. 2). The positions of X and Y along this dimension at two F_0's are represented by X_1, Y_1 and X_2, Y_2, respectively. From experiment I we know that distances $\overline{X_1Y_1}$ and $\overline{X_2Y_2}$ are approximately equal. If additionally the timbres themselves are stable along this dimension, then we must have $\overline{X_1Y_2}=\overline{X_2Y_1}$ [Fig. 2(a)]. If instead they shift along this dimension, then $\overline{X_1Y_2}\neq\overline{X_2Y_1}$ [Fig. 2(b)]. Equality thus means either that timbres of X and Y did not shift with F_0, or that the shift was in a direction *orthogonal* to the dimension along which X and Y differ. Supposing that this holds for all instrument pairs, it follows that timbres did not move in the timbre space that spans the instrument set. Symmetry of the dissimilarity matrix, if observed, implies timbre invariance with respect to F_0 [hypothesis (1)].

1. Methods

Stimuli were those of experiment I, paired with a constant ΔF_0 of 2 semitones (B3–C♯4, session "a") or 11

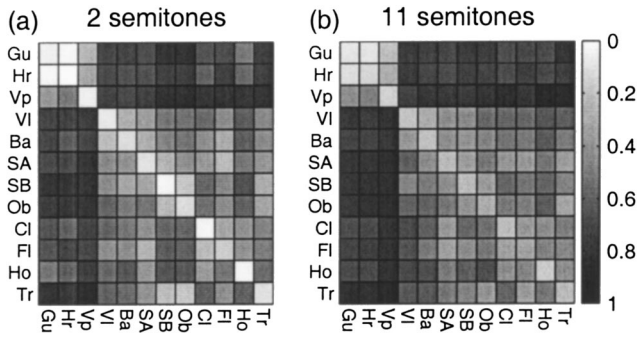

FIG. 3. Dissimilarity matrices for the two sessions of experiment II, each corresponding to a different F_0 pair. The lower triangular part corresponds to pairs for which the instrument on the abscissa was on the lower F_0 and the instrument on the ordinate on the higher. The upper triangular part corresponds to the opposite order. The diagonal represents instruments compared to themselves with an F_0 difference.

semitones (B3–Bb4, session ''b''). All pairs were included, resulting in 144 stimulus pairs per session. Within a session, the order of F_0's (low-high or high-low) was always the same, so as to make it easier for subjects to ignore the difference in pitch. Subjects were the same 27 that participated in experiment I. The three subjects that were eliminated from experiment I were also eliminated here. The remaining subjects were divided into four groups of approximately the same size (six to eight subjects) that differed in the order of presentation of sessions (''ab'' *versus* ''ba''), and in the order of F_0's within each session (low-first versus high-first). The proportion of musicians and nonmusicians was approximately the same in each group. Subjects performed both sessions on the same day (approximately one week after experiment I), separated by a 10-min pause.

2. Results

a. Dissimilarity matrices. Matrices averaged over subjects are plotted in Fig. 3 for the two sessions. Several features are obvious. First, ratings along the diagonals are relatively small. Second, the matrices appear fairly symmetrical. Third, the two matrices resemble each other. Fourth, the lower triangular parts of each matrix resemble the three matrices of experiment I.

To quantify these effects, the upper and lower triangular parts of both matrices were excised, ignoring the diagonals. The upper triangular parts were reflected with respect to the diagonal so as to have the same shape as the lower triangular parts, and the data were submitted to a repeated-measures ANOVA with factors instrument pairs $(66)\times\Delta F_0(2)\times F_0$ orders (2). Results are shown in Table III. The F_0 order factor (upper versus lower triangular parts) is not interpretable in itself as it depends on the arbitrary way in which instrument pairs were combined with F_0 pairs. It is included so as to allow F_0-dependent variance to be quantified.

Main effects and two-way interactions were significant, the three-way interaction was not. Effect sizes are quantified by R^2 scores (last column of Table III). The pair effect represents 43.8% of total variance, whereas F_0-related effects together sum up to a total of only 2.8%. As in experiment I, it appears that timbre dissimilarity depends on F_0 differences

TABLE III. ANOVA table for experiment II. S: Subjects, I: Instrument pairs, ΔF_0: F_0 difference, O: F_0 order, SS: sum of squares, MS: mean square, F: F-values, ϵ: Greenhouse–Geisser correction factor applied to the degrees of freedom, p: corrected p-Value, R^2: percentage of total variance accounted for by each effect (intersubject differences amounted to 53.4%).

Source	df	SS	MS	F	ϵ	p	R^2
S	23	54.5	2.37				
ΔF_0	1	1.01	1.01	5.31	1	0.03	**0.2**
ΔF_0*S	23	4.37	0.19				
O	1	0.69	0.69	10.48	1	0.003	**0.1**
O*S	23	1.52	0.07				
I	65	221.4	3.41	51.5	0.08	0.0001	**43.8**
I*S	1495	98.9	0.07				
ΔF_0*O	1	0.23	0.23	7.54	1	0.012	**0.04**
ΔF_0*O*S	23	0.68	0.03				
ΔF_0*I	65	5.58	0.09	3.62	0.21	0.0001	**1.1**
ΔF_0*I*S	1495	35.45	0.02				
O*I	65	4.14	0.06	2.3	0.2	0.006	**0.8**
O*I*S	1495	41.33	0.03				
ΔF_0*O*I	65	2.47	0.04	1.71	0.19	0.06	**0.5**
ΔF_0*O*I*S	1495	33.28	0.02				

to a limited degree. Taking the average over lower and (reflected) upper parts of the matrix for each session, the correlation coefficient between sessions is 0.95 (df=64, $p<0.001$). Averaging over sessions within experiment II and within experiment I, the correlation coefficient between experiments is 0.98 (df=64, $p<0.001$).

Table IV shows the percentage of variance accounted for by each effect for the full data set (column 2), or when dissimilarity scores are restricted to pairs of impulsive, sustained, or impulsive and sustained instruments (columns 3–5). The ratio of F_0-invariant effects (I) to F_0-dependent effects (ΔF_0, etc.) is smaller for restricted sets (particularly pairs of sustained instruments) than for the full set. Nevertheless, for each subset F_0-invariant effects remain larger than F_0-dependent effects.

When the diagonals of the matrices (not included in the previous analysis) were averaged over instruments, dissimilarity was 0.11 at 2 semitones and 0.20 at 11 semitones. Single sample t-tests show that the mean is significantly different from zero [$t(287)=10.7$, $p<0.0001$ for 2 ST; $t(287)=15.4$, $p<0.0001$ for 11 ST]. Further, in a repeated-measures ANOVA with factors Instrument $(12)\times\Delta F_0(2)$, the main factors were significant [$F(11,253)=5.3$, $\epsilon=0.43$, $p=0.0001$ and $F(1,23)=16.1$, $p=0.0005$, respectively], but their interaction was not. These results suggest that the pitch difference affected the dissimilarity judgments and that the

TABLE IV. Percentage of variance between dissimilarity scores accounted for by each effect of experiment II, for the entire data set (All) or for data restricted to pairs of impulsive, sustained, or impulsive and sustained (mixed) instruments. The last line (Other) represents variance due to disagreement between subjects. Each column sums to 100.

	R^2(%)			
Source	All	Impulsive	Sustained	Mixed
ΔF_0-dependent	2.8	6.3	4.6	2.8
I	43.8	29.2	15.2	9.1
Other	53.4	64.5	80.2	88.1

effect increased with increasing pitch difference. The effect was independent of instrument, however.

Supposing timbre invariance, we expected the diagonals to be zero. To some degree, the nonzero values observed can be attributed to an edge effect due to the fact that the response range had a lower bound of zero (variability of responses then necessarily results in a nonzero mean). However, given the significant effects of instrument and ΔF_0 this explanation is at best incomplete: we must admit a shift of timbre with F_0 (or a contamination of dissimilarity responses with pitch dissimilarity). The values on the diagonal are nevertheless small. Averaged over ΔF_0's, same-instrument dissimilarities were smaller (mean: 0.16) than different-instrument dissimilarities (mean: 0.59). The largest same-instrument dissimilarity (0.25 for the flute) was smaller than every different-instrument dissimilarity score except one (0.1 for Gu/Hr).

To summarize the results of experiment II, a first outcome is that subjects can compare timbre across F_0 despite salient pitch differences. Subjects apparently performed the tasks of experiments I and II in similar fashion. As a second outcome, we can rule out the hypothesis of a large global shift of timbre space with F_0, as dissimilarity matrices were symmetrical and their diagonals small. This extends the conclusion of experiment I that instruments retain their relative positions as F_0 changes: they also do not shift *as a group*. However, beyond these conclusions valid in the first approximation, both experiments revealed effects that were significant, albeit small. It would be nice to infer from these effects the nature of shifts of individual intruments. Unfortunately, each score reflects the timbre of *two* instruments, and it is not obvious which of the two determined a change in dissimilarity. Experiment III introduces a new form of analysis that reveals timbre changes of individual instruments with F_0.

C. Experiment III

In experiment III subjects rated timbre dissimilarity between pairs of instruments with and without a difference in F_0. The aim was to extend and generalize the results of experiments I and II, and in particular to see whether subjects could make reliable timbre dissimilarity judgments between sounds that differed by a variable amount along the pitch dimension.

1. Methods

To keep the stimulus set size within reasonable limits, 9 instruments were selected among the 12 used in experiments I and II. These were Gu, Hr, Vp, Vl, SA, Ob, Cl, Ho, and Tr. Each was played at two F_0's, resulting in a set of 18 sounds that were paired (excluding comparisons between the same instrument at the same F_0) to produce 153 pairs that were presented in a single session with a 5 min pause half-way. There were two sessions: ''a'' with notes B3 and C♯4 (2 semitones), ''b'' with notes B3 and Bb4 (11 semitones). Within a session, different-F_0 pairs were presented in the same order, low-high or high-low (depending on the subject). Otherwise, presentation conditions and instruments were the same as for experiment II.

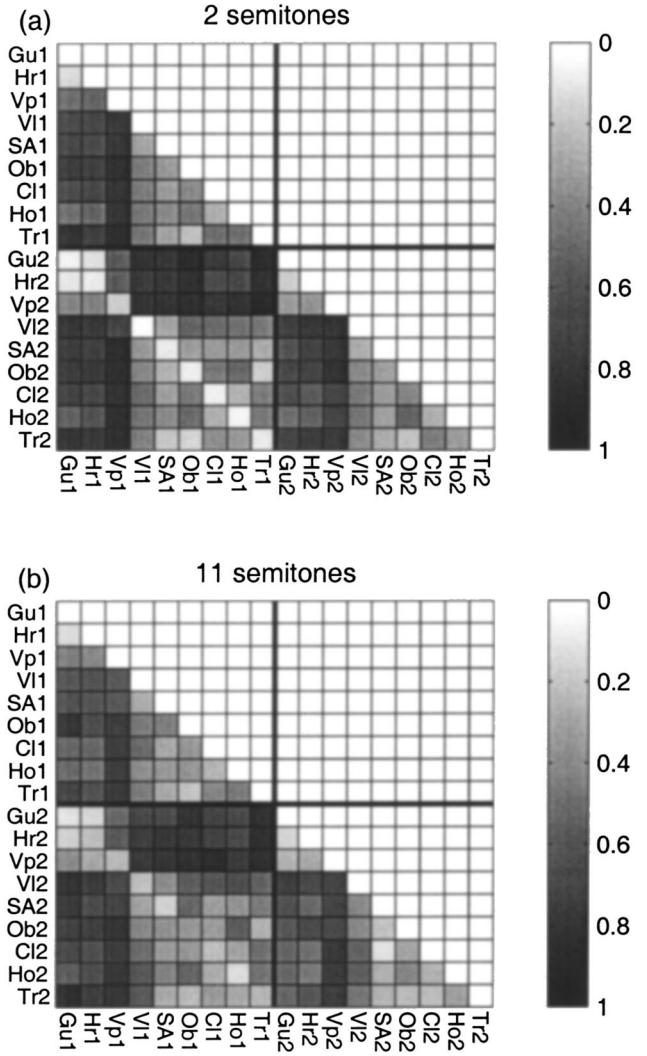

FIG. 4. Dissimilarity matrices for experiments IIIa and IIIb. For the axis labels, indices 1 and 2 mean that instruments were played at B3 and C♯4, respectively (Bb4 in experiment IIIb).

Session ''a'' involved 25 subjects aged 19 to 30, 15 men and 10 women, 13 musicians and 12 nonmusicians. None had participated in experiments I or II. Session ''b'' involved 18 subjects (11 of which had taken part in session ''a''), aged 19 to 30, eight women and ten men, nine musicians and nine nonmusicians.

2. Results

a. Outliers. Among the 25 subjects of session ''a,'' three gave answers that were poorly correlated with the rest [$r < 0.33$] and were excluded from the analysis. None were excluded from session ''b.''

b. Dissimilarity matrices. Dissimilarity scores averaged over subjects were placed in the lower triangular part of a matrix as shown in Fig. 4(a) for session ''a.'' This matrix has three parts: an upper-left triangle (instruments compared at B3), a lower-right triangle (instruments compared at C♯4), and a 9×9 square (instruments compared across F_0's). The two triangles are analogous to the matrices of experiments Ia and Ib, the square to that of experiment IIa.

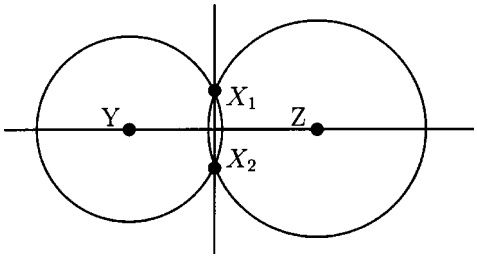

FIG. 5. Experiment III. Schema illustrating the anchor method of analysis of timbre change. The plane represents a hypothetical two-dimensional timbre space, *Y* and *Z* are ''anchor'' instruments, and *X* is an instrument whose displacement with F_0 is being considered. If the dissimilarity of *X* with respect to *Y* does not change when the F_0 of *X* is changed, X_1 and X_2 must be on a circle centered on *Y*. Similarly, if dissimilarity with respect to *Z* does not change, X_1 and X_2 must also belong to a circle centered on *Z*. The displacement of *X* is therefore on a line orthogonal to *YZ*. If the same is true for all anchor pairs, the displacement of *X* is orthogonal to the space they span (or else it is zero).

TABLE V. Effect size (R^2) of factors F_0 and $F_0\times$anchor for each instrument at both ΔF_0's. Only effects significant at the $p<0.05$ level are shown. These figures quantify the magnitude of displacement of each instrument in timbre space as a function of F_0.

	R^2(%)			
	2 semitones		11 semitones	
Instrument	F_0	$F_0\times$anchor	F_0	$F_0\times$anchor
Gu	⋯	⋯	⋯	⋯
Hr	⋯	⋯	⋯	⋯
Vp	2.79	2.51	⋯	4.28
Vl	⋯	⋯	1.17	⋯
SA	⋯	⋯	0.55	1.52
Ob	⋯	⋯	1.23	1.92
Cl	⋯	⋯	0.59	⋯
Ho	⋯	⋯	⋯	⋯
Tr	0.40	⋯	1.66	1.92

To compare results with those of experiments I and II, a triangular matrix similar to the one just described was populated with scores from corresponding conditions of experiment Ia (upper left triangle), Ib (lower right triangle) and IIa (square). The correlation between this composite matrix and that obtained from experiment IIIa was 0.95 (df=151, $p<0.0001$). Similarly, a triangular matrix was populated with scores of experiments Ia, Ic, and IIb. The correlation between this composite matrix and that obtained from experiment IIIb was 0.92 (df=151, $p<0.0001$). This indicates a high degree of similarity between data sets despite the difference in task and subjects. Overall ANOVAs are not reported here (they support conclusions similar to experiments I and II). Instead, a different analysis is presented that assigns effects to timbre changes of individual instruments.

c. Instrument-specific ANOVAs. Each of the nine instruments was analyzed in turn. For each, the eight other instruments were used as ''anchors'' with respect to which to measure its timbre changes.

To illustrate the principle, take an instrument *X* and denote its timbre at two different F_0's as X_1 and X_2, assimilated to two points in timbre space. We wish to know if X_1 and X_2 are distinct, and for this we use a second instrument *Y* as anchor. We ignore eventual shifts of *Y* itself for the moment. The displacement of *X* towards or away from *Y* can be estimated by comparing $\overline{X_1Y}$ and $\overline{X_2Y}$. In geometric terms, the equality

$$\overline{X_1Y}=\overline{X_2Y} \tag{1}$$

implies that *X* has followed a hypersphere centered on *Y* (illustrated as a circle in Fig. 5). If a similar equality holds for another anchor instrument *Z*, then X_1 and X_2 belong to the intersection of two hyperspheres. In the plane (Fig. 5), the intersection consists of two points on a line perpendicular to *YZ*. In three dimensions it would be a circle in a plane orthogonal to *YZ*, and in higher dimensions a sphere or hypersphere in a hyperplane orthogonal to *YZ*. In every case the displacement is *orthogonal* to the timbre dimension along which *Y* and *Z* differ. If Eq. (1) holds for every anchor, taking them two by two, it follows that the displacement of *X* is orthogonal to the subspace that contains the anchors. Supposing that the anchors together span the whole of timbre space, *X* did not move in this space.

Actually, each instrument has two positions, e.g., Y_1 and Y_2 for *Y*. Either could be used as the anchor, but there is a difficulty. Testing for $\overline{X_1Y_1}=\overline{X_2Y_1}$, instruments on the left have the same F_0 but those on the right differ. The comparison is thus sensitive to eventual effects of an F_0 difference *per se* (for example, if subjects failed to completely ignore pitch). Using Y_2 instead as the anchor we have a similar problem in the other direction. However, by adding term to term,

$$\overline{X_1Y_1}+\overline{X_1Y_2}=\overline{X_2Y_1}+\overline{X_2Y_2}, \tag{2}$$

F_0-related effects apply equally to both sides and thus balance out. Equation (2) can be used in place of Eq. (1) for the previous analysis. To summarize, if Eq. (2) holds when an instrument is compared to each of the eight others, we may assume that that instrument's timbre did not change with F_0.

d. Two semitones. For each instrument *X*, terms of Eq. (2) were compared using a repeated-measures ANOVA with factors anchor $(8)\times F_0\,(2)$. To be precise: the F_0 factor contrasted timbres X_1 and X_2 at two different F_0's by comparing $\overline{X_1Y_1}+\overline{X_1Y_2}$ to $\overline{X_2Y_1}+\overline{X_2Y_2}$. The anchor factor contrasted the various anchor instruments *Y*. Nine such ANOVAs were performed. The main effect of anchor was, as expected, highly significant for all instruments, and will not be considered further. For seven instruments (Gu,Hr,Vl,SA,Ob,Cl,Ho), the effect of F_0 and its interaction with anchor were not significant. For the trumpet, the main effect of F_0 was significant but tiny ($R^2=0.4\%$, as compared to 65.5% for anchor). For the violin pizzicato, both the main effect of F_0 and its interaction with anchor were significant and relatively large. The other instruments remained essentially stable when F_0 changed from B3 to C♯4. These results are summarized in columns 2 and 3 of Table V.

e. Eleven semitones. Similar ANOVAs were performed for session ''b.'' Effect sizes are summarized in the last two columns of Table V. These effects were nonsignificant for Gu, Hr, and Ho, and very small for Cl. They were significant and larger for Vp, Vl, SA, Ob, and Tr.

To summarize the results of experiment III, subjects succeeded in making timbre dissimilarity judgments while largely ignoring a difference in pitch that was present on some trials and not on others. Experiments I and II had found *timbre dissimilarity* to be fairly stable with F_0 changes. Experiment III refined this conclusion: *timbre* itself was stable for some instruments (eight for 2 semitones, four for 11 semitones, out of nine intruments). Timbres of others appeared to change slightly. The next section presents a MDS analysis that allows these changes to be interpreted in terms of displacement within a model of perceptual timbre space.

III. MDS ANALYSIS

For each session of experiment III, the dissimilarity matrices for all subjects were processed by the EXSCAL MDS program (Winsberg and Carroll, 1989). We chose a two-way MDS model without individual differences, as this model is rotationally invariant, allowing solutions to be rotated and their dimensions compared to physical descriptors, as well as compared across experiments. The two-way EXSCAL model postulates that the distance, d_{ij}, between the ith and jth stimuli, is given by

$$d_{ij}=\left[\sum_{r=1}^{R}(X_{ir}-X_{jr})^2+(S_i+S_j)\right]^{1/2}, \tag{3}$$

where X_{ir} is the coordinate of the ith stimulus on the rth dimension and R is the number of dimensions. In this model, in addition to R common dimensions, the stimuli have unique dimensions not shared by other stimuli. The specificity or uniqueness of the ith stimulus is denoted S_i. Since a maximum likelihood criterion is used to estimate the fit of the model to the data, BIC statistics (Schwarz, 1978) can be used to choose the dimensionality R and decide whether additional unique dimensions should be included.

The BIC criterion suggested three- and two-dimensional models without specificities for experiments IIIa and IIIb, respectively. For nonlinear models like MDS, BIC statistics have a heuristic value and do not preclude consideration of other models. We therefore also examined two-, three-, and four-dimensional models in search of a model interpretable in terms of dimensions related to signal descriptors. In each case, the solution was rotated with a procrustean procedure to a target matrix of signal descriptors described in Sec. IV. Only the four-dimensional solutions will be described in detail.

Solutions for sessions ''a'' (2 semitones) and ''b'' (11 semitones) are illustrated in the upper and lower parts of Fig. 6, respectively. For each instrument, the position at B3 is represented by the symbol and that at the other F_0 (C♯4 or Bb4) by the extremity of the line. The first three dimensions of the spaces are well correlated between sessions (0.99, 0.95, and 0.94, respectively). A large score for dimension 1 is to be expected because the salient contrast between impulsive and sustained instruments is unlikely to depend on F_0. However, the good scores for dimensions 2 and 3 suggest that additional dimensions of timbre are stable across F_0. The fourth dimension was poorly correlated between sessions (0.31).

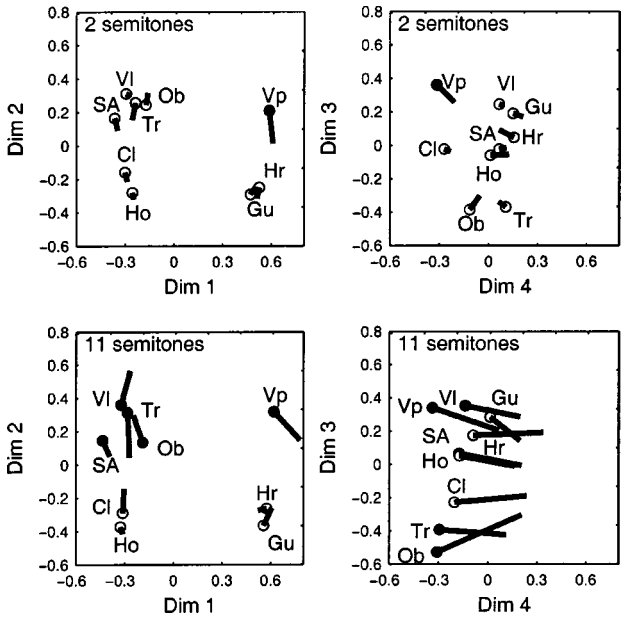

FIG. 6. Timbre spaces for experiment IIIa (top) and IIIb (bottom). The symbol represents the position of the instrument at the lower F_0 (B3), the end of the line represents the position of the same instrument at the other F_0 (C♯4 or Bb4). The symbol is filled for instruments for which a significant timbre change was found in Sec. II C and open for others.

Filled symbols in Fig. 6 are instruments for which we know (on the basis of the ANOVAs of the previous section) that their timbre changed. We would expect the lines to be of nonzero length (in at least one projection) for them and of zero length for open symbols. Such is not always the case. A possible explanation for this discrepancy is that the data used for ANOVAs excluded dissimilarities between the same instrument at different F_0's, whereas the MDS included them. Even if the timbre of an instrument did not change across F_0, the measured dissimilarity was likely to take a nonzero value as a result of an edge effect (Sec. II B 3 a) or a residual pitch dissimilarity that the subjects failed to ignore. This has the effect of ''pushing apart'' the corresponding points of the MDS solution. Whatever the explanation, this discrepancy weakens the usefulness of interpreting the detailed pattern of F_0-induced shifts we observe in Fig. 6.

IV. COMPARISON WITH SIGNAL DESCRIPTORS

In the spirit of previous studies on timbre, this section attempts to relate perceptual dimensions revealed by MDS to descriptors of the signal (sometimes called ''physical dimensions''). A feature of the present study is that this relation is tested over several fundamental frequencies. On the basis of our data we can formulate three constraints for a signal-based descriptor: (1) at each F_0, the descriptor should predict the corresponding perceptual dimension; (2) for instruments whose timbre did not vary across F_0, descriptor values should not vary; and (3) for instruments whose timbre did vary across F_0, and to the degree that this variation is reliably described in terms of change along a perceptual dimension, we should observe a corresponding change of the descriptor. We consider only data for experiment IIIb (11 semitones).

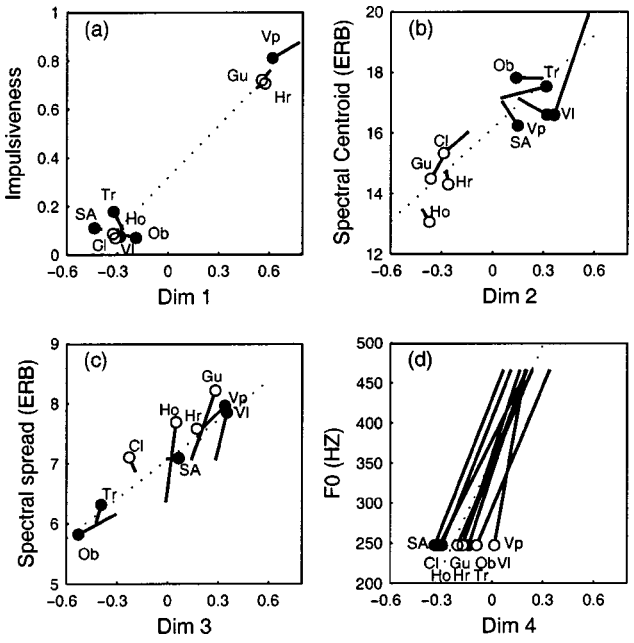

FIG. 7. Experiment IIIb (11 semitones). Scatter plots relating each signal descriptor to the MDS dimension that it explains best. (a) Impulsiveness versus dimension 1. (b) Spectral centroid versus dimension 2. (c) Spectral spread versus dimension 3. (d) F_0 versus dimension 4. For each instrument, the symbol represents its position at note B3, and the opposite end of the line represents its position at note Bb4. Filled symbols indicate instruments for which the timbre changed significantly according to the analysis of C. Open symbols represent instruments for which it did not. Dotted lines represent regression lines.

A. Dimension 1

To predict the first dimension we use a measure of impulsiveness proposed by Susini (1996), defined as follows. The instantaneous power s_n^2 is smoothed by convolution with an 8-ms square window. The duration during which the smoothed power is above 40% of its maximum value is then divided by the duration for which it is above 10%, and one minus this ratio is taken as the measure of impulsiveness. It is close to one for impulsive sounds and to zero for sustained sounds.

Figure 7(a) plots this descriptor as a function of dimension 1 for experiment IIIb. The descriptor does a good job of predicting the clustering of impulsive and sustained instruments into well-separated groups. Correlation coefficients are 0.98 for experiment IIIa and 0.99 for experiment IIIb (df=16, $p<0.01$ in both cases). As a comparison, previous studies (e.g., Krimphoff *et al.*, 1994) suggested the log of attack time as a descriptor for impulsiveness. That descriptor gave correlation scores of 0.95 and 0.94 (df=16, $p<0.01$ in both cases) for experiments IIIa and IIIb, respectively, when the MDS solutions were rotated towards values determined by it.

B. Dimension 2

To predict dimension 2 we used a spectral centroid descriptor similar in spirit to the definition of sharpness (Zwicker and Fastl, 1990; Hartmann, 1997). The waveform was first filtered to model the drop in sensitivity at low and high frequencies due mainly to outer and middle ear filtering (Killion, 1977). Then it was filtered by a gammatone filterbank (Patterson *et al.*, 1995; Slaney, 1993) with channels spaced at half-ERB intervals on an ERB-rate scale (z) calculated according to the formula $z=21.3\log(0.00437f+1)$ between 25 Hz and 19 kHz (Hartmann, 1997). Instantaneous power was calculated within each channel and smoothed by delaying it by $1/4f_c$ (where f_c is the characteristic frequency of the channel), adding it to the undelayed power, and convolving the sum with an 8-ms window corresponding roughly to the equivalent rectangular duration of power integration measured by Plack and Moore (1990). Smoothed power was then raised to the power 0.3 to obtain a rough measure of ''partial loudness'' for each channel. The partial-loudness-weighted average of ERB rate was taken over channels, the result being an ''instantaneous spectral centroid'' function of time according to

$$\overline{z}(t)=\sum_z z\psi_z(t) \Big/ \sum_z \psi_z(t), \tag{4}$$

where $\psi_z(t)$ is the ''partial loudness'' of the channel z at instant t. Finally, the instantaneous centroid $\overline{z}(t)$ was weighted by ''instantaneous loudness'' (sum over channels of partial loudness) and averaged over time to obtain a single descriptor value, $\overline{z}$, to characterize the entire signal.

Figure 7(b) shows the value of spectral centroid as a function of dimension 2. Data points are relatively well aligned. We expect the displacements of those instruments that significantly changed in timbre (filled symbols) to follow this trend. Such is roughly the case for Vl and Tr, but not for SA, Vp, or Ob. The descriptor thus predicts the overall trend but not all details. The correlation between descriptor values and projections along dimension 2 is 0.93 and 0.90 for experiment IIIa and IIIb, respectively (df=16, $p<0.01$ in both cases).

Our definition of spectral centroid is one of many that have been proposed. A common definition is the following:

$$\overline{k}=\sum_k ka_k \Big/ \sum_k a_k, \tag{5}$$

where k is the rank of a partial and a_k is its amplitude (on a linear, power, or log scale). If the spectral envelope remained constant when F_0 varies (approximately the case for most of our instruments), this definition would lead to an *inverse* dependency of k with F_0, a variation of a factor 1.9 between B3 and Bb4. Since timbre was instead rather stable, this definition can be ruled out, as concluded earlier by Slawson (1968) or Plomp (1976). A better definition defines the centroid as a weighted sum of frequencies (e.g., Kendall *et al.*, 1999), for example:

$$\overline{f}=\sum_k f_k a_k \Big/ \sum_k a_k, \tag{6}$$

where k is the rank of a partial or discrete Fourier transform coefficient, f_k is its frequency and a_k is its amplitude (on a linear, power or log scale). For a constant spectral envelope this definition leads to values of $\overline{f}$ that are approximately constant as F_0 varies. However, there are several ways of implementing this definition according to whether a_k desig-

nates the linear, power, or log amplitude, whether k designates the rank of a partial, a DFT coefficient, or a filter band, whether the frequency scale is linear or warped (log or ERB-rate scale), whether a nonlinearity is applied after summing coefficients within channels, etc. Our definition of spectral centroid was chosen to make all operations and parameters explicit in a psychoacoustically reasonable way, and avoid hidden parameters such as window size or sampling rate, or the implicit assumption of a line spectrum needed to apply Eq. (6).

As a comparison, the definition of Eq. (6) implemented according to Peeters *et al.* (2000) gave correlation coefficients of 0.95 and 0.85 for experiments IIIa and IIIb, respectively, when the MDS solutions were rotated using that definition ($df=16$, $p<0.01$ in both cases).

C. Dimension 3

Dimension 3 was found to be relatively well correlated with a measure $\tilde{z}$ of spectral spread defined as

$$\tilde{z}=\sqrt{\sum_z (z-\overline{z})^2\psi_z \Big/ \sum_z \psi_z}. \tag{7}$$

Figure 7(c) shows the value of spectral spread as a function of dimension 3 for experiment IIIb. Data points are roughly distributed along a line. Two instruments that significantly changed in timbre (Ob and Vp) move roughly along this line, as expected. However, two instruments that did not change timbre (Ho and Gu) also show relatively large changes in descriptor value. Such is also the case for Vl, which did change timbre but (according to the MDS analysis) not along this dimension. The descriptor would be better if such changes could be avoided. Overall, the correlation between descriptor values and projections along dimension 3 was 0.94 and 0.87 for experiments IIIa and IIIb, respectively ($df=16$, $p<0.01$ in both cases).

As a comparison, the definition of spectral spread of Peeters *et al.* (2000), analogous to the spectral centroid definition of Eq. (6), gave correlation scores of 0.83 and 0.65, respectively, when the MDS solutions were rotated to that descriptor. Our descriptor was also applied to the stimuli used by McAdams *et al.* (1995) and compared to the coordinates along the third dimension of their MDS space. The correlation found was 0.87, as opposed to 0.54 for the spectral flux descriptor used in that study.

D. Dimension 4

Dimension 4 was found to be well correlated with F_0 (0.90) for experiment IIIb. For experiment IIIa the correlation with F_0 was poor and no better descriptor was found. Figure 7(d) plots F_0 as a function of dimension 4 for experiment IIIb. Displacements of all instruments are roughly parallel with the regression line, consistent with the good correlation. Subjects thus based their timbre dissimilarity judgments in part upon a dimension related to F_0. This is the only evidence we found of a pitchlike dimension.

To summarize, the signal descriptors reviewed roughly satisfy the constraint of F_0-invariance for instruments that did not change timbre. For instruments that did change timbre, the minor changes in descriptor value with F_0 were in some cases consistent with the minor changes in timbre, in other cases not. Overall, the descriptors did a very good job of predicting perceptual dimensions. They compared favorably with previously proposed descriptors, but variability of data is such that we cannot reliably conclude on this basis alone that one given descriptor is superior to another.

V. DISCUSSION

A first outcome of this study, not obvious from the start, is that timbres of instruments played at different notes can be compared. Classic techniques such as MDS can be applied, and this opens the perspective for more detailed and extensive studies of timbre variations of specific instruments across their register. Subjects performed the task in a very similar fashion with or without F_0 differences between stimuli, and had little difficulty ignoring the very salient pitch differences that accompanied them. Timbre behaved as if it were separable from pitch, and there was only slight evidence of a small perceptual interaction between pitch and timbre dimensions.

Cross-F_0 timbre comparison being possible, a second outcome is the relative stability of timbre with respect to F_0 changes. For several instruments there was no measurable change in timbre, so we can exclude the hypothesis of a basic, non-instrument-specific dependency of timbre upon F_0. The hypothesis that such a dependency does exist, but was balanced by opposite changes of instrument characteristics, is unlikely to be simultaneously true for four out of nine instruments across all F_0's, and eight between B3 and C♯4. Lack of measurable change is not due to lack of sensitivity of our methods: for other instruments we demonstrated significant timbre changes of relatively small size.

The ''anchor-based'' analysis technique introduced in Sec. II C 2 revealed small but significant timbre changes for certain instruments. The MDS analyses provided an interpretation of the changes in terms of displacement along particular dimensions of timbre space. However, relatively large displacements were also observed for instruments known *not* to have changed timbre significantly, so we must not give too much weight to such detailed features, as argued in Sec. III. MDS solutions were generally stable across experiments and conditions, and the correlations between their dimensions and physical descriptors was high, as found in previous studies.

Stability of timbre as a function of F_0 for certain instruments puts a strong constraint on signal descriptors for predicting timbre: they too must demonstrate the same degree of stability. Such was the case for the descriptors we used, but other methods proposed in the literature may not be so stable. These conclusions are very important for applications that use signal descriptors for content-based indexing of audio and multimedia data. So far, such descriptors had been validated only at particular F_0's. Our results demonstrate that they generalize well to the F_0's we tested, although the question remains open for the wider range of F_0's.

We used relatively small F_0 steps because we expected the task of comparing timbre while ignoring pitch to be difficult (Miller and Carterette, 1975). The pitch differences are

nevertheless quite salient. The smaller step (two semitones, a major second) is one-third the maximum distance along the chroma circle. The larger step (11 semitones, a major seventh) is both larger in terms of tone height and smaller in terms of chroma, and thus offered the opportunity to tease apart the eventual contributions of each. It is also about one-third of the range of typical instruments such as the violin, and thus probes instrument-specific variations to some extent. Obviously, a wider range of notes is needed for a more complete study of instrument-specific timbre variations. The present study showed that such a study is in principle possible. There is, however, evidence that instrument identification performance degrades beyond an octave (Handel and Erickson, 2001).

The generality of our results is also limited by our choice of instruments. Previous studies found that criteria vary according to the stimulus set, leading to MDS solutions that correlate with rather different physical dimensions. One or both of our first two MDS dimensions were usually also salient in those studies, but one cannot exclude that for certain stimulus sets, other dimensions might be relevant that are more sensitive to F_0 changes.

VI. CONCLUSIONS

(1) Subjects made timbre dissimilarity judgments between natural musical instrument sounds that differed in F_0 by 0, 2 or 11 semitones. Results were orderly, even when the stimulus set comprised both same-F_0 and different-F_0 sound pairs. The salient pitch difference produced by the F_0 differences did not prevent timbre comparisons.

(2) As a first approximation, timbre dissimilarities depended little on F_0. Dissimilarity scores varied more between different-instrument pairs than across F_0's. Experiment I (same F_0) showed that instruments kept their relative positions in timbre space at different F_0's, and experiments II and III showed further that they did not shift as a group.

(3) As a second approximation, small but significant timbre changes were observed. Instrument-specific ANOVAs in experiment III found that the changes affected certain instruments and not others. It is likely that these timbre changes were due to instrument-specific changes in, for example, resonator geometry.

(4) The lack of significant effects for certain instruments in the instrument-specific ANOVAs of experiment III, together with the symmetry of dissimilarity matrices in experiment II, suggest the absence of any basic, non-instrument-specific change of timbre with F_0.

(5) Multidimensional scaling yielded low-dimensional linear models of perceptual timbre spaces (four-dimensional without specificities). After an appropriate rotation, dimensions were found to be well correlated with a set of signal-based descriptors. Projections on each of the first three dimensions were relatively stable with F_0. The projection on the fourth was correlated with F_0 for an 11-semitone (but not 2-semitone) step. This is the only evidence we found for a pitchlike dimension in a timbre space.

(6) Signal-based descriptors ''impulsiveness,'' ''spectral centroid,'' ''spectral spread,'' and F_0 were used. The first describes the temporal envelope. The second two describe the spectral envelope in terms of the first two moments of a ''partial loudness'' spectrum (cubic root of power within channels of a cochlear filter bank). These three descriptors appeared to be good predictors of the first three timbre dimensions over the range of F_0's used, while the fourth (F_0) is known as a good predictor of pitch.

This study opens the way for more extensive studies of timbre change with F_0, such as instrument-specific timbre changes across their register. The anchor method applied in experiment III seems particularly promising to distinguish timbre changes from fluctuations due to experimental noise.

ACKNOWLEDGMENTS

The authors thank Stephen Handel for useful comments on a previous version of this paper. This work is part of the first author's Ph.D project which is funded by the Swiss National Science Foundation. The research project is funded in part by the European Union Project CUIDADO and was conducted within the Music Perception and Cognition team at IRCAM, and at CNMAT, University of California, Berkeley.

Abdi, H. (**1987**). *Introduction au Traitement Statistique des Données Expérimentales (Introduction to Statistical Processing of Experimental Data)* (PUG, Grenoble).

ANSI (**1960**). ''USA Standard Acoustical Terminology'' (American National Standards Institute, New York).

de Cheveigné, A. and Kawahara, H. (**1999**). ''Missing data model of vowel perception,'' J. Acoust. Soc. Am. **105**, 3497–3508.

Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Intruments*, 2nd ed. (Springer-Verlag, New York).

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

Handel, S., and Erickson, M. L. (**2001**). ''A rule of thumb: The bandwith for timbre invariance is one octave,'' Music Percept. **19**, 121–126.

Hartmann, W. (**1997**). *Signals, Sound, and Sensation* (AIP, New York).

Helmholtz, H. (**1885**). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (from 1877 trans. by A. J. Ellis of 4th German ed., republ. 1954 by Dover, New York).

IRCAM (**2000**). ''Studio On Line'' http://www.ircam.fr/.

Kaufman, L., and Rousseeuw, P. J. (**1990**). *Finding Groups in Data. An Introduction to Cluster Analysis* (Wiley-Interscience, Brussel).

Kendall, R., Carterette, E., and Hajda, J. (**1999**). ''Perceptual and acoustical features of natural and synthetic orchestral instrument tones,'' Music Percept. **16**, 265–294.

Killion, M. C. (**1977**). ''Revised estimate of minimum audible pressure: Where is the 'missing 6 dB'?'' J. Acoust. Soc. Am. **63**, 1501–1508.

Krimphoff, J., McAdams, S., and Winsberg, S. (**1994**). ''Caractérisation du timbre des sons complexes. Analyses acoustiques et quantification psychophysique'' (Characterization of the timbre of complex sounds. Acoustical analyses and psychophysical quantification), J. Phys. I **4**, 625–628.

Luce, D., and Clark, M. (**1967**). ''Physical correlates of brass-instrument tones,'' J. Acoust. Soc. Am. **42**, 1232–1243.

Martin, K. D. (**1999**). ''Sound-Source Recognition: A Theory and Computation Model,'' Massachusetts Institute of Technology, unpublished doctoral dissertation.

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, C. (**1975**). ''Perceptual space for musical structures,'' J. Acoust. Soc. Am. **58**, 711–720.

Misdariis, N., Smith, B., Pressnitzer, D., Susini, P., and McAdams, S. (**1998**). ''Validation of a multidimensional distance model for perceptual dissimilarities among musical timbres,'' J. Acoust. Soc. Am. **103**, 2812.
Patterson, R. D., Milroy, R., and Allerhand, M. (**1993**). ''What is the octave of a harmonically rich note?,'' Contemp. Music Rev. **9**, 69–81.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1995**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.
Peeters, G., McAdams, S., and Herrera, P. (**2000**). ''Instrument sound description in the context of MPEG-7,'' Proc. Int. Conf. Comput. Music, Berlin, pp. 166–169.
Plack, C. J., and Moore, B. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.
Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).
Risset, J., and Wessel, D. (**1999**). ''Exploration of timbre by analysis and synthesis,'' in *The Psychology of Music*, 2nd ed., edited by D. Deutsch (Academic, New York), pp. 113–169.
Schwarz, G. (**1978**). ''Estimating the dimensions of a model,'' Ann. Stat. **6**, 461–464.
Shepard, R. N. (**1964**). ''Circularity in judgments of relative pitch,'' J. Acoust. Soc. Am. **36**, 2346–2353.
Slaney, M. (**1993**). ''An efficient implementation of the Patterson-Holdsworth auditory filter bank,'' Apple Computer Technical Report 35.
Slawson, A. (**1968**). ''Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency,'' J. Acoust. Soc. Am. **43**, 87–101.
Smith, B. (**1995**). ''PsiExp: an environment for psychoacoustic experimentation using the IRCAM Musical Workstation,'' in *Society for Music Perception and Cognition Conference '95*, edited by D. Wessel (Univ. of California, Berkeley).
Susini, P. (**1996**). ''Analyses acoustiques des sons'' (Acoustical analyses of sounds), IRCAM Technical Report.
Ueda, R. D., and Nimmo-Smith, I. (**1987**). ''Perceptual components of pitch: Spatial representation using a multidimensional scaling technique,'' J. Acoust. Soc. Am. **82**, 1193–1200.
Winsberg, S., and Carroll, J. D. (**1989**). ''A quasi-nonmetric method for multidimensional scaling via an extended euclidean model,'' Psychometrika **54**, 217–229.
Wonnacott, T. H., and Wonnacott, R. J. (**1990**). *Introductory Statistics for Business and Economics*, 4th ed. (Wiley, New York).
Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics: Facts and Models* (Springer-Verlag, New York).

Erratum: "The dependency of timbre on fundamental frequency" [J. Acoust. Soc. Am. 114, 2946–2957 (2003)]

Jeremy Marozeau, Alain de Cheveigné, Stephen McAdams, and Suzanne Winsberg
Institut de Recherche et Coordination Acoustique/Musique (Ircam-CNRS), 1, place Igor Stravinsky, F-75004 Paris, France

(Received 26 November 2003; accepted for publication 28 November 2003)

[DOI: 10.1121/1.1642764]

PACS numbers: 43.75.Cd, 43.66.Jh, 43.66.Hg, 43.10.Vx [ADP]

The citation ''Ueda and Nimmo-Smith (1987)'' and corresponding reference are incorrect. The correct reference is: Ueda, K., and Ohgushi, K. (1987). ''Perceptual components of pitch: Spatial representation using a multidimensional scaling technique,'' J. Acoust. Soc. Am. **82**, 1193–1200.

