# Perception of timbral analogies

STEPHEN McADAMS[1,2] AND JEAN-CHRISTOPHE CUNIBLE[1]

[1] *Laboratoire de Psychologie Expérimentale (CNRS URA 316), Université René Descartes, 28 rue Serpente, F-75006 Paris, France*
[2] *IRCAM, 31 rue Saint-Merri, F-75004 Paris, France*

## SUMMARY

Recent studies have investigated the structure of perceptual relations among musical instrument timbres by multidimensional scaling (MDS) techniques. These studies have employed both acoustically produced tones and digitally synthesized imitations and hybrids of acoustic instrument tones. The analyses of dissimilarity ratings for all pairs of a set of tones are usually represented as geometrical structures in a two- or three-dimensional Euclidean space in which the shared 'perceptual' axes are shown to have a qualitative correspondence to acoustic properties such as spectral energy distribution, onset characteristics and degree of change in spectral distribution over the duration of the tone. The present study took as a point of departure a MDS analysis for complex, synthetic tones with the aim of testing whether musician and non-musician listeners used the relations defined by the perceptual space to perform an analogies task of the sort: timbre A is to timbre B as timbre C is to which of two possible timbres, D or D′? A parallelogram model was used to select the D timbres: if the relation between A and B is represented as a vector with both magnitude and direction components, then the appropriate D should form a vector with C having similar magnitude and direction in the timbre space. Aside from conceptual difficulties with the task for both non-musicians and composers, choices for both groups provide support for the parallelogram model indicating a capacity in listeners to perceive abstract relations among the timbres of complex sounds without specific training in such a task.

## 1. INTRODUCTION

One of the properties of the pitch dimension that endows it with its psychological capacity to serve as a vehicle for musical form is the fact that relations between pitches (i.e. intervals) can be perceived as musical qualities in their own right. Musical sequences can be built upon these qualities, and operations on musical material, such as transposition, that maintain them also maintain a strong degree of perceptual similarity between the original and transformed materials. If one were to try to extend the form-bearing possibilities of pitch into the realm of timbre, it would be necessary to determine the kinds of structuring of timbral relations that can be perceived by listeners and reasoned with by composers. Therefore, one of the important issues in research on musical timbre is the way in which listeners might potentially make use of relations among timbres in the perception of musical structure.

The trend toward using timbre in increasingly complex ways in music dates from orchestration practice in the last half of the 19th century (Boulez 1987). This trend has been extended considerably with the advent of analog and digital means of sound generation and processing. These same means provide the researcher with the possibility to generate with precise control sounds of considerable complexity and thus to open the way to the systematic study of timbre perception.

For the psychologist, several interesting questions arise concerning a listener's ability to perceive and remember timbral relations in tone sequences (Krumhansl 1989; McAdams 1989), as well as to build up hierarchical mental representations based these relations (Lerdahl 1987). Research in the past 20 years (cf. Plomp, 1970; Risset & Wessel 1982; Barrière 1990) has attempted to go beyond the loose negative definition of timbre given us by the field of psychoacoustics (i.e. timbre is what distinguishes two tones of identical pitch, loudness and perceived duration). To this end experimental paradigms that reveal the perceptual structure of timbral relations have been employed, and most notably those based on the multidimensional scaling of similarity (or dissimilarity) judgments.

In such a study, a number of tones differing in timbre (and equated for pitch, loudness, and perceived duration) are presented in all possible pairs to listeners who are asked to decide how dissimilar the tones of each pair are and to rate the dissimilarity on a scale of, say, 1 to 8. A multidimensional scaling algorithm is then applied to the matrix of judged dissimilarities. In many types of analyses, the algorithm tries to establish a monotonic relation between the dissimilarity ratings and Euclidean distances among the sounds arranged in a geometric structure in *n* dimensions, each sound being represented as a point. Sounds with similar timbres are thus near one another in the space and those with dissimilar timbres are farther apart. The experimenter tries solutions with varying numbers of dimensions and selects the solution that is a compromise between having a small difference between distances and ratings (which decreases with increasing $n$) and not having more dimensions than can be readily interpreted in terms of their underlying perceptual and psychophysical relevance to the group of listeners tested. Different studies on timbre have generally settled on two (Plomp 1970; Wessel 1973, 1979; Ehresman & Wessel 1978; Rasch & Plomp 1982) or three dimensions (Grey 1977; Krumhansl 1989; Kendall & Carterette 1991). We will focus on the studies that adopted a three-dimensional solution for isolated timbres.

Grey (1977) used 16 digitally recorded, analysed resynthesized musical instrument tones performing a pitch at $E^{b}_{3}$ ($F_0 = 311$ Hz). Krumhansl (1989) used 21 synthetic tones developed by Wessel *et al.* (1987) on a Yamaha frequency modulation synthesizer: some of these tones were imitations of traditional Western orchestral instruments while others were hybrids (e.g. 'vibrone' is a hybrid of vibraphone and trombone, and 'guitarnet' is a hybrid of guitar and clarinet). Both Grey's and Krumhansl's spaces are qualitatively similar in the interpretation of their underlying dimensions, so we will confine our discussion to the latter since these tones were employed in our experiment.

A non-quantitative comparison of acoustic characteristics of the tones with their position along the various perceptual axes gave rise to the following interpretation (see figure 1). Dimension I seems related to the temporal envelope (rapidity of the attack and presence of inharmonic transients at the beginning of the tone) and might be called 'attack quality'. Sharp or biting attacks, such as that of the harpsichord, are found at one end of the dimension and softer, gentler attacks as with the clarinet are found at the other end. Dimension II seems related to a combined spectro-temporal property called 'spectral flux'. Instruments whose spectral envelope evolves relatively little over the duration of the tone (like the oboe) have low spectral flux compared to those whose spectrum changes a great deal (usually brightness increasing and decreasing with intensity as in the brass instruments). Dimension III seems related to the global spectral envelope and is called 'brightness'. Grey & Gordon (1978) have shown brightness to be highly correlated with the centre of gravity of the long-term spectrum represented in terms of specific loudness and critical band rate (Zwicker & Scharf 1965). Bright sounds (like the oboe) have a greater presence of energy in the higher harmonics than do less bright sounds (like the French horn). In most cases the hybrid instruments were situated between the two instruments from which they were derived.

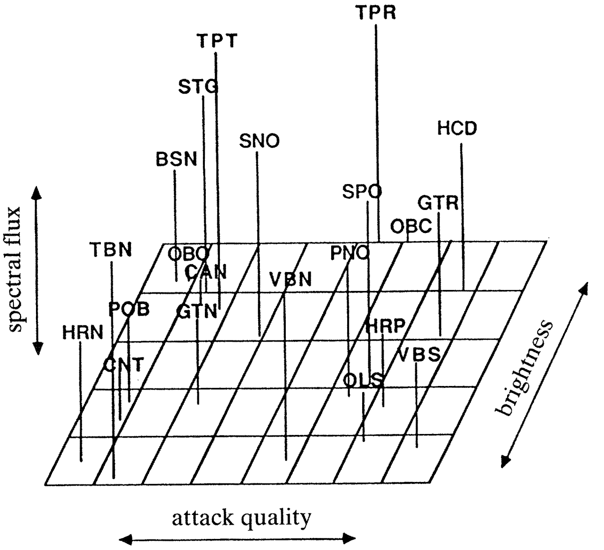

Figure 1. Timbre space derived from a three-dimensional scaling solution for dissimilarity judgments on 21 synthetic instrument tones. BSN = bassoon, CAN = cor anglais, CNT = clarinet, GTN = guitarnet (GTR/CNT), GTR = guitar, HCD = harpsichord, HRN = French horn, HRP = harp, OBC = obochord (OBO/HCD), OBO = oboe, OLS = oboleste (OBO/celeste), PNO = piano, POB = bowed piano, SNO = striano (STG/PNO), SPO = sampled piano, STG = string, TBN = trombone, TPR = trumpar (TPT/GTR), TPT = trumpet, VBN = vibrone (VBS/TBN), VBS = vibraphone. (Adapted from Krumhansl (1989).)

An additional aspect of the Krumhansl (1989) analysis (based on a technique developed by Winsberg & Carroll, 1988)† revealed the existence of unique (although unspecified) perceptual features for certain instrument timbres. These features (called 'specificities' in the analysis technique) are not taken into account by the three common dimensions. Examples of specific features might include the odd-harmonic, hollow tone colour of the clarinet which is not subsumed under brightness, or the 'bump' at the return of the hopper on the end of a harpsichord tone. Eight of the 21 instruments had relatively high specificities (including the clarinet and harpsichord).

Once such a space has been quantified, one might ask whether the structure of the common dimensions is useful as a tool for predicting listeners' abilities to compare relations among the sounds. For example, can one use Euclidean spatial relations to define the properties of an interval formed by two timbres? This idea was initially developed by Ehresman & Wessel (1978) who applied Rumelhart & Abrahamson's (1973) parallelogram model of analogical reasoning in a semantic space to the timbre space composed of the tones used by Grey (1977). Rumelhart & Abrahamson took as a point of departure a three-dimensional space obtained by MDS techniques applied to dissimilarity judgments on animal names (Henley 1969). They were interested in whether the structure of the space would allow them to predict people's choices when presented with an analogy task of the form A is to B as C is to D (or A:B::C:D). In general, if the relation between two objects, A and B, is represented as a vector in the space, the model predicts that subjects will choose an object D which is the closest to the end point of a vector starting at C and having the

---

† In the MDS analysis with specificities, the algorithm tries to find a monotone relation between the dissimilarity ratings and estimated distances, $d_{ij}$, between the tones $i$ and $j$, such that $d_{ij} = \{\Sigma(x_{ik} - x_{jk})^2 + s_i + s_j\}^{1/2}$, where $x_{ik}$ is the coordinate on the $k^{\text{th}}$ dimension for tone $i$ and $s_i$ is the estimated specificity for tone $i$ that is not accounted for by the common dimensions.

same magnitude and direction as ***AB*** (vectors are denoted in italic boldfaced type). They called this the ideal solution point, I. ***AB*** and ***CI*** thus form a parallelogram in the space. In their experiment, subjects were presented with analogies of the form A:B::C:{$D_1$, $D_2$, $D_3$, $D_4$}, where the $D_i$'s varied according to their distance from I. The probability of choosing $D_i$ as the best solution was found to be a monotonically decreasing function of the absolute distance of $D_i$ from I, thus supporting the parallelogram model. Ehresman & Wessel proceeded in analogous fashion with musical instrument tones. The underlying assumption behind the definition of a timbre interval as a vector is that processes exist for the encoding and processing of relations between timbres that are isomorphic with those for representing and processing vector quantities. While the results were not as strongly supportive of the parallelogram model as in the Rumelhart & Abrahamson study, they were better predicted by this model than a number of other models. This early paper is encouraging as (i) it formalizes the notion of a timbre interval as being composed of both distance and degree of change along important perceptual dimensions, and (ii) it shows that this definition is correlated with listeners' judgments across intervals. The weakness of the study is that timbral vectors were computed from only a two-dimensional solution and that only relative vector magnitude was tested, ignoring the direction components. Our study systematically selected pairs of timbre vectors to be compared in an analogy task to test both magnitude and direction components.

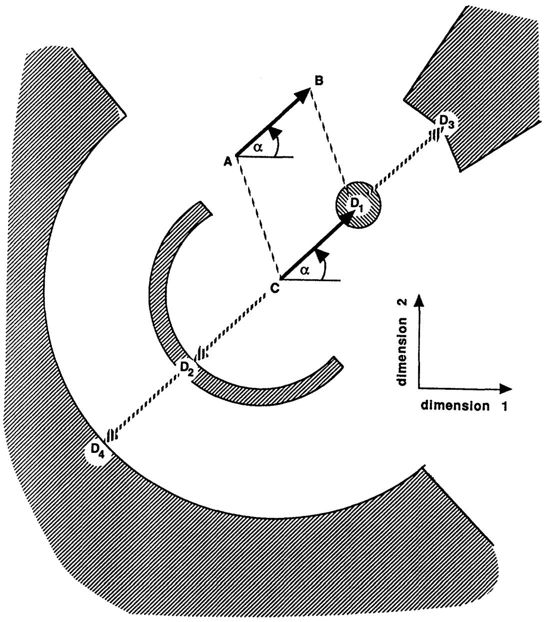

Figure 2. Two-dimensional representation of the different sequence types. The angle $\alpha$ is with respect to dimension 1. The angle $\beta$ would be with respect to dimension 2 if the vectors were three-dimensional and coming out of the page. The hashed areas represent the constraint space for the end points of $\boldsymbol{CD_i}$ vectors and are labeled $D_1$, $D_2$, $D_3$ or $D_4$, accordingly. The ideal point I would be at the tip of the arrow-head for $\boldsymbol{CD_i}$. For the three-dimensional case, the area would be a sphere for $D_1$, a shell for $D_2$, part of a cone for $D_3$, and a solid with a spherical hollow for $D_4$.

## 2. METHOD

### (*a*) *Stimuli*

Tones were derived from the set of 21 synthetic instruments described above. Each tone was realized playing an $E^{b}_{3}$ at mezzo forte (MIDI velocity 70) on a Yamaha TX802 FM Tone Generator. All sounds had been equalized for pitch and loudness by Krumhansl (1989). There were some significant differences in duration, however, certain plucked and struck sounds lasting longer than sounds imitating forced vibration instruments (winds and bowed strings). The nominal duration for each tone was 300 ms.

The magnitude and direction components of a vector between any pair of sounds in the three-dimensional perceptual space derived by Krumhansl for these tones can be computed as follows (e.g. for A and B).

*Magnitude*

Magnitude, $d$ (corresponds to the estimated perceived dissimilarity):

$|\boldsymbol{AB}| = \{\Sigma(x_{\mathrm{A}k} - x_{\mathrm{B}k})^2\}^{1/2}$, where $x_{\mathrm{A}k}$ is the coordinate on the $k$th dimension for timbre A, and $k = 1, 2, 3$.

*Direction angles*

Direction angles, $\alpha$ (degree of change on dimension I) and $\beta$ (degree of change on dimension II; the angle $\gamma$ for dimension III is complementary to the other two by the relation $\cos^2\alpha + \cos^2\beta + \cos^2\gamma \equiv 1$):

$$\alpha_{\boldsymbol{AB}} = \cos^{-1}((x_{\mathrm{A}1} - x_{\mathrm{B}1})/|\boldsymbol{AB}|),$$

$$\beta_{\boldsymbol{AB}} = \cos^{-1}((x_{\mathrm{A}2} - x_{\mathrm{B}2})/|\boldsymbol{AB}|).$$

Vectors can then be compared in terms of $d$, $\alpha$ and $\beta$. Accordingly, four classes of four-tone sequences were constructed to be of the form A:B::C:$D_i$. Constraints were established for the selection of four different kinds of $D_i$, such that the magnitude and direction components of ***AB*** and $\boldsymbol{CD_i}$ were either similar or quite different. These constraints are schematically illustrated (for the two-dimensional case only) in figure 2. They can be formalized as follows.

*Sequence 1* – A:B::C:$D_1$ (right magnitude, right direction on ***CD*** with respect to ***AB***); $D_1$ close to I with small error ($\varepsilon$) on $d$, $\alpha$, and $\beta$:

$$|\boldsymbol{CD}_1| = |\boldsymbol{AB}| \pm \varepsilon_d,$$

$$\alpha_{\boldsymbol{CD}1} = \alpha_{\boldsymbol{AB}} \pm \varepsilon_\alpha,$$

$$\beta_{\boldsymbol{CD}1} = \beta_{\boldsymbol{AB}} \pm \varepsilon_\beta.$$

*Sequence 2* – A:B::C:$D_2$ (right magnitude, wrong direction); small error on $d$, but at least one of $\alpha_{\boldsymbol{CD}2}$ or $\beta_{\boldsymbol{CD}2}$ must differ by at least 90° from $\alpha_{\boldsymbol{AB}}$ or $\beta_{\boldsymbol{AB}}$, respectively:

$$|\boldsymbol{CD}_2| = |\boldsymbol{AB}| \pm \varepsilon_d,$$

$|\alpha_{\boldsymbol{CD}2} - \alpha_{\boldsymbol{AB}}| \geqslant 90^\circ$ and/or $|\beta_{\boldsymbol{CD}2} - \beta_{\boldsymbol{AB}}| \geqslant 90^\circ$.

Table I. *The possible sequence comparison types and the effects they were designed to test*

(Sequence labels are abbreviated: e.g. $D_2 = A{:}B{::}C{:}D_2$. Comparison $D_2/D_3$ was not included in the experiment because no pairs of sequences satisfying the appropriate constraints could be found in the chosen stimulus set.)

| comparison type | vector component tested | origin of effect |
|---|---|---|
| $D_1/D_2$ | direction | right magnitude in both cases<br>right direction on $D_1$<br>wrong direction on $D_2$ |
| $D_1/D_3$ | magnitude | right direction in both cases<br>right magnitude on $D_1$<br>wrong magnitude on $D_3$ |
| $D_1/D_4$ | magnitude and direction | right magnitude and direction on $D_1$<br>wrong magnitude and direction on $D_4$ |
| $D_2/D_3$ | magnitude vs. direction | right magnitude and wrong direction on $D_2$<br>wrong magnitude and right direction on $D_3$ |
| $D_2/D_4$ | magnitude under wrong direction | wrong direction in both cases<br>right magnitude on $D_2$<br>wrong magnitude on $D_4$ |
| $D_3/D_4$ | direction under wrong magnitude | wrong magnitude in both cases<br>right direction on $D_3$<br>wrong direction on $D_4$ |

*Sequence 3* – $A{:}B{::}C{:}D_3$ (wrong magnitude, right direction) small error on $\alpha$ and $\beta$, but $|\boldsymbol{CD}_3|$ must be larger than $|\boldsymbol{AB}|$:

$$|\boldsymbol{CD}_3| \geqslant 1.8|\boldsymbol{AB}|,$$

$$\alpha_{CD3} = \alpha_{\boldsymbol{AB}} \pm \varepsilon_\alpha,$$

$$\beta_{CD3} = \beta_{\boldsymbol{AB}} \pm \varepsilon_\beta.$$

*Sequence 4* – $A{:}B{::}C{:}D_4$ (wrong magnitude, wrong direction):

$$|\boldsymbol{CD}_4| \geqslant 1.8|\boldsymbol{AB}|,$$

$$|\alpha_{\boldsymbol{CD}4} - \alpha_{\boldsymbol{AB}}| \geqslant 90° \text{ and/or } |\beta_{CD2} - \beta_{\boldsymbol{AB}}| \geqslant 90°.$$

In the above equations, the maximum allowed value of the error terms was fixed as follows: $|\varepsilon_d| \leqslant 0.35$, $|\varepsilon_\alpha| \leqslant 22.9°$, $|\varepsilon_\beta| \leqslant 22.9°$. These values were determined empirically to be as small as possible while giving a reasonable number of sequences for each type listed above.‡ The range of *d* for timbre pairs used in the experiment was 2.5–14.6 with a mean of 7.60. The range of angles was 14.2°–177.7° (mean = 95.7°) for $\alpha$ and 7.7°–164.6° (mean 104.8°) for $\beta$.

### (*b*) *Procedure*

Ideally, we want to find appropriate $D_1$, $D_2$, $D_3$, and $D_4$ for any given set of A, B, and C tones and ask listeners to rank order them with respect to their relative success in fulfilling the analogy as was done in Ehresman & Wessel (1978). This would allow us to test directly for the relative importance of magnitude and direction components of the timbral vectors. With the given space however, this was impossible since sets of seven timbres (A, B, C, $D_1$, $D_2$, $D_3$, $D_4$) satisfying the constraints could not be found. We were obliged to settle on an experimental paradigm in which pairs of sequences were presented and subjects were to compare them and determine which best satisfied the analogy $A{:}B{::}C{:}D$. This reduced the stimulus search constraints to finding sets of five timbres (A, B, C, D, D′). The comparison types and the effect each was designed to test are listed in Table I. The following is an example of a D1/D4 comparison, where 'oboleste' is a hybrid of oboe and celeste:

D1: harp is to harpsichord as oboleste is to guitar, or
D4: harp is to harpsicord as oboleste is to clarinet.

At least five versions of each of the six possible pairs of sequence types were found with the exception of $A{:}B{::}C{:}D_2/A{:}B{::}C{:}D_3$ (subsequently referred to simply as $D_2/D_3$). This comparison was thus dropped from the experiment. Each version of a comparison was composed of different timbres while still satisfying the stimulus constraints for the two sequence types. The use of multiple versions allowed us to test the generality of the analogy task across different sets of timbres.

In each trial, listeners heard two sequences of four timbres with the following time structure, where the durations indicate silent intervals between the 300 ms tones: A–500 ms–B–900 ms–C–500 ms–D–1300 ms–A–500 ms–B–900 ms–C–500 ms–D′. After a pause of 2700 ms, the eight tone sequence was repeated once.

A complete block of 50 trials included the five sequence comparison types ($D_1/D_2$, $D_1/D_3$, $D_1/D_4$, $D_2/D_4$, $D_3/D_4$) each being presented in five versions with different timbres and with the order of presentation of the sequences counterbalanced.

Two groups of subjects were tested: 18 psychology students from René Descartes University without any formal musical training (non-musicians) and seven professional composers participating in a workshop on

---

‡ It should be noted that the accumulated error in $\varepsilon_\alpha$ and $\varepsilon_\beta$ leads in some cases to an $\varepsilon_\gamma$ as large as 86° which results in the D for that sequence being farther removed from I. The mean $|\varepsilon_\gamma|$ in $D_1$ and $D_3$ sequences was 36.3° with a standard deviation of 22.6°.

computer music at IRCAM. The non-musicians were tested individually over headphones in a single-walled soundproof chamber and entered their responses on the computer keyboard. The composers were tested in a group, listening to loudspeakers in a sound treated studio and entered their responses on a numbered answer sheet. The non-musicians completed two blocks of trials whereas the composers completed a single block. The sounds were presented at a comfortable listening level.

Subjects were given an instruction sheet that explained the analogy task using a semantic and a visual example. The correct solutions to each example were explained. Six practice trials were given using a randomly selected set of experimental trials. No feedback was given on either the practice or the experimental trials. After completing the practice trials, any further questions the subject(s) had were answered before proceeding to the first block of trials.

### (*c*) ***Hypotheses***

The following sets of hypotheses were tested in the experiment. Each hypothesis refers to a separate aspect of the data. They are thus not mutually exclusive.

1. Subjects will prefer $D_1$ over $D_2$, $D_3$, and $D_4$ as a solution to the analogy, as it is the best fit to the parallelogram model. A corollary to this hypothesis would predict that the preference of $D_1$ over $D_4$ be stronger than that over $D_2$ or $D_3$ as $D_4$ is the farthest removed in all respects from the ideal point.
2. $D_2$ will be preferred over $D_4$: listeners prefer the right magnitude even though the direction is wrong in both ***CD*** intervals.
3. $D_3$ will be preferred over $D_4$: listeners prefer the right direction even though the magnitude is wrong in both ***CD*** intervals.
4. There will be no differences among the different versions of each comparison type since the analogy judgment is based on a perception of abstract relations among the timbres of the stimulus tones.
5. The effects of hypotheses 1–3 will be stronger for composers than for non-musicians as the activity of reasoning with sound and making timbre judgments in composition will have allowed the former group to develop more consistent judgment strategies.

An additional point of interest concerns the missing $D_2/D_3$ condition. In the absence of this condition, a comparison between $D_1/D_2$ and $D_1/D_3$ preferences will indicate something of the relative effect of distance and direction. We have no a *priori* hypothesis about this result based on the parallelogram model.

## 3. RESULTS AND DISCUSSION

The data consisted of percent choices of one of the paired sequences over the other for each version of each comparison type collected across order of presentation. An effect of block of trials was only found for the $D_1/D_2$ comparison in the nonmusicians' data: the percent choice of $D_1$ was greater in the second block (two-tailed, $t_{(17)} = 3.01$, $p < 0.01$). In the subsequent analyses, the data are grouped across blocks for the nonmusicians.

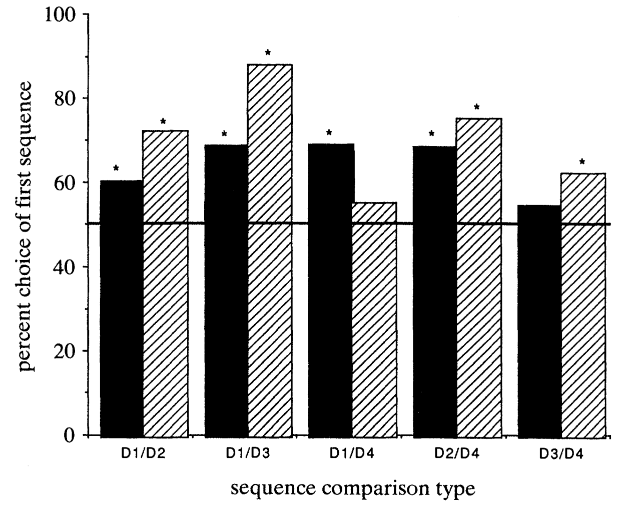

Figure 3. Global means (across versions, presentation orders and listeners) for five sequence comparison types. The comparison type is labelled on the horizontal axis. The two groups of subjects (18 non-musicians and seven composers) are shown with solid and hashed bars, respectively. The horizontal line is positioned at 50% (chance choice). The asterisks over certain bars indicate that the mean is significantly different from chance.

The means for the experimental conditions are highly correlated between subject groups ($r = 0.65$, $p < 0.01$). Although composers tend to express stronger preferences (one-tailed, $t_{(24)} = 1.66$, $p = 0.055$), the patterns of both data sets are qualitatively similar. Thus hypothesis 5 is at most only weakly supported by the data.

### (*a*) ***Global effect of comparison type***

The means for each comparison type obtained for each subject group are shown in figure 3. To test for differences from chance choice (50%), one-group *t*-tests were performed on means for each comparison type across versions for each of the subject groups. The Bonferoni-adjusted criterion was 0.005 (ten tests). All means except for the $D_3/D_4$ comparison were significantly different from chance for nonmusicians, and all except for the $D_1/D_4$ comparison were different from chance for composers.

Hypothesis 1 which predicted that $D_1$ would be preferred over all other sequences is confirmed in all cases for non-musicians and in all cases except $D_1/D_4$ for composers. This latter result is quite surprising, because according to the parallelogram model, $D_4$ should be the farthest from the ideal point and $D_1$ the closest. Examination of the means for the five versions of $D_1/D_4$ for composers shows that three hover around chance, one is significantly higher than chance (preference for $D_1$), and one is quite lower than 50% (preference for $D_4$) although this latter mean just misses being significantly different from 50%. In general, however, the results suggest that the parallelogram model captures a significant portion of subjects' judgment strategies since the timbre closest to the ideal point is preferred over other more distant timbres. The corollary to hypothesis 1 is not confirmed, i.e. preferences for $D_1$ over $D_4$ are not higher than those of $D_1$ over $D_2$ or $D_3$. This will require further reflection since the parallelogram model of Rumelhart & Abrahamson (1973) predicts monotonically decreasing preference with increasing distance from the ideal point.

That relative distance between timbre pairs can be evaluated by listeners even though the directions are dissimilar is suggested by the fact that the mean preference for $D_2$ over $D_4$ is reliably above chance for both subject groups. Hypothesis 2 is thus confirmed indicating that the distance component of the timbral change is perceptually important in perceiving timbral relations.

Hypothesis 3 ($D_3$ is preferred over $D_4$) is confirmed for composers but not for non-musicians. This result suggests that the former group can evaluate relative direction of timbral change even though the distances of the two D timbres from C are quite different from that between A and B. Examination of the five versions of $D_3/D_4$ for the non-musicians reveals that two had means reliably above 50% (preference for $D_3$) and one was significantly below 50% (preference for $D_4$).

In the absence of a $D_2/D_3$ condition, a comparison of $D_1/D_2$ means with those for $D_1/D_3$ suggests that distance change across timbre pairs ($D_1/D_3$) is more easily noticed than direction change ($D_1/D_2$), because $D_1$ is preferred more over $D_3$ than over $D_2$. This difference is not statistically significant, however, indicating that neither magnitude nor direction predominates over the other component.

Overall the results are encouraging, witnessing an ability to make judgments on timbral relations. However, some of these global effects need to be qualified by a closer look at the different versions grouped under each comparison type.

### (*b*) ***Effects of individual versions of each comparison type***

To test for effects of individual versions within comparison type, one-way analyses of variance with repeated measures on version were performed. The results are shown in table 2. For both subject groups, four out of five comparison types have significant overall differences between versions. This indicates that not every version of each comparison had the same perceptual result and was thus not judged in a similar way. In particular, one notes a great dispersion of means for certain comparisons ($D_3/D_4$ for non-musicians and $D_1/D_4$ for composers). This dispersion on either side of 50% results in the global mean being not different from random choice. Globally, we must reject hypothesis 4 which predicted equal performance for all versions of a comparison type.

Table 2. *One-way analyses of variance with repeated measures on version for each comparison type and subject group*

(Sequence labels are abbreviated: e.g. $D_2 =$ A:B::C:$D_2$. For non-musicians $n = 18$, and for composers $n = 7$.)

| | non-musicians | | composers | |
|---|---|---|---|---|
| comparison type | $F_{(4,68)}$ | $p$ | $F_{(4,24)}$ | $p$ |
| $D_1/D_2$ | 3.58 | $< 0.01$ | 8.26 | $< 0.005$ |
| $D_1/D_3$ | 2.36 | $> 0.05$ | 3.12 | $> 0.10$ |
| $D_1/D_4$ | 4.49 | $< 0.005$ | 5.14 | $< 0.005$ |
| $D_2/D_4$ | 9.20 | $< 0.001$ | 7.10 | $< 0.001$ |
| $D_3/D_4$ | 9.00 | $< 0.001$ | 3.88 | $< 0.05$ |

### (*c*) ***Effect of the relative distance of $D_i$s from the ideal point***

According to the Rumelhart & Abrahamson model, the choice of one sequence over another should be a monotonically increasing function of the distance between the ideal point and D. Therefore, for each comparison type, these distances were calculated and the mean percent choices for each comparison type were regressed onto the difference between these distances. This analysis indicates the degree to which judgments may have been based purely on the relative distance of D from I in each sequence. The regression was performed independently for nonmusician and composer groups. For non-musicians the regression yielded a significant fit between mean data and distances ($R = 0.48$; $F_{(1,23)} = 6.80$, $p < 0.05$). Although the fit is not bad, the regression only accounts for 23% of the variance in the data indicating that other factors are entering into the judgments that are unaccounted for by a simple distance-from-ideal-point model. For composers, the fit between mean data and distances is not significant ($R = 0.04$; $F_{(1,23)} = 0.04$). In spite of the strong correlation between the means for non-musicians and composers, there appears to be no relation between relative distance from the ideal point and the sequence preferred as best completing the analogy for the composers.

Another possibility is that listeners made judgments based on the relative degree of change along the individual perceptual dimensions. Accordingly, we performed a multiple regression of the differences in change along each dimension between ***AB*** and ***CD*** or ***CD'*** vectors onto mean percent choice for each group of listeners. For nonmusicians, the fit was not significant ($R = 0.46$; $F_{(3,21)} = 1.93$) whereas for composers the fit was significant ($R = 0.57$; $F_{(3,21)} = 3.33$, $p < 0.05$). The partial $F$s for the multiple regression show that differences in change along dimensions I and II (attack, spectral flux) are largely responsible for this fit. Taken together, these two regression analyses may indicate differences in listening and judgment strategies between the two groups.

## 4. CONCLUSIONS

A number of experimental conditions were designed within the framework of a Euclidean distance model of timbre space (Krumhansl, 1989) in order to test listeners' abilities to perceive timbral relations and to judge their similarity in terms of magnitude and direction of timbre change. These results support and extend those of Ehresman & Wessel (1978). They are also coherent with work by Kendall & Carterette (1991) who have shown that a vector model of timbre can account for perceived similarities among simultaneously sounding wind instrument dyads.

A vector model of timbre intervals was fairly successful at predicting the choice of one type of sequence over another, where the sequences varied in the degree to which the magnitude and direction components of the timbral vectors match across pairs of timbres. In general, timbres close to the ideal point predicted by the vector model were preferred as best fulfilling an analogy of the form A:B::C:D than were timbres that were at some distance from that point (conditions D1/D2, D1/D3, D1/D4). We have also shown that in some cases the model even predicts preference when both Ds in a sequence comparison are quite far removed from I, indicating an ability to appreciate the appropriate vector magnitude under conditions of wrong direction (D2/D4) and of appropriate direction under conditions of wrong magnitude (D3/D4), though the latter condition is quite weak. What the model does not do is make predictions about the relative contributions of magnitude and direction of the comparison timbre vector. This is a subject for future research.

The strong effect of the timbre set chosen to realize each comparison type suggests a relative lack of generalizability of timbral interval perception across different timbres. This result may be due to a number of factors that were not controlled in this study: (i) there may be a relative instability of judgment strategies, since most of the listeners have never encountered a listening situation in which focusing on, or comprehending, abstract timbral relations was appropriate; (ii) there may be effects of the relative magnitude of a given vector and the distance between to-be-compared vectors: very large vectors may be difficult to compare with precision and small vectors that are very far apart in the space may also be difficult to compare; (iii) there may be effects of the degree of change along different common dimensions: the perceptual weights of change along individual dimensions may not be equivalent in this kind of listening task; and (iv) there may be effects of specific features of individual timbres that are not taken into account by the common dimensions of the timbre space, but which influence the perceived distances between timbres and thus the timbre intervals that are to be compared.

Portions of this study were realized in partial fulfillment of the requirements for J.-C. Cunibile's Master's thesis at the Laboratoire de Psychologie Expérimentale, Université René Descartes (Cunibile 1991). This research was supported in part by a grant from the French Ministry of Culture.

## REFERENCES

Barrière, J.-B. (ed.) 1990 *Le timbre, métaphore pour la composition.* Paris: Christian Bourgois/IRCAM.

Boulez, P. 1987 Timbre and composition—timbre and language. *Contemp. Music Rev.* **2**, 161–172.

Cunibile, J.-C. 1991 Perception des analogies de timbre. Master's thesis, Laboratoire de Psychologie Expérimentale, Université René Descartes, Paris.

Ehresman, D. & Wessel, D.L. 1978 Perception of timbral analogies. *Rapp. IRCAM*, no. 13. Paris: IRCAM.

Grey, J.M. 1977 Multidimensional perceptual scaling of musical timbre. *J. acoust. Soc. Am.* **61**, 1270–1277.

Grey, J.M. & Gordon, J.W. 1978 Perceptual effects of spectral modifications on musical timbres. *J. acoust. Soc. Am.* **63**, 1493–1500.

Henley, N.M. 1969 A psychological study of the semantics of animal terms. *J. verb. Learn. verb. Behav.* **8**, 176–184.

Kendall, R.A. & Carterette, E.C. 1991 Perceptual scaling of simultaneous wind instrument timbres. *Music Percept.* **8**, 369–404.

Krumhansl, C.L. 1989 Why is musical timbre so hard to understand? In *Structure and perception of electroacoustic sound and music* (ed. S. Nielzen & Olsson), pp. 43–53. Amsterdam: Elsevier (Excerpta Medica 846).

Lerdahl, F. 1987 Timbral hierarchies. *Contemp. Music Rev.* **2**, 135–160.

McAdams, S. 1989 Psychological constraints on form-bearing dimensions in music. *Contemp. Music Rev.* **4**, 181–198.

Plomp, R. 1970 Timbre as a multidimensional attribute of complex tones. In *Frequency analysis and periodicity detection in hearing* (ed. R. Plomp & G. F. Smoorenburg), pp. 397–414. Leiden: Sijthoff.

Rasch, R. & Plomp, R. 1982 The perception of musical tones. In *The psychology of music* (ed. D. Deutsch), pp. 1–24. New York: Academic Press.

Risset, J.-C. & Wessel, D.L. 1982 Exploration of timbre by analysis and synthesis. In *The psychology of music* (ed. D. Deutsch), pp. 25–58. New York: Academic Press.

Rumelhart, D.E. & Abrahamson, A.A. 1973 A model for analogical reasoning. *Cogn. Psychol.* **5**, 1–28.

Wessel, D.L. 1973 Psychoacoustics and music. *Bull. comput. Arts Soc.* **30**, 1–2.

Wessel, D.L. 1979 Timbre space as a musical control structure. *Comput. Music J.* **3**(2), 45–52.

Wessel, D.L., Bristow, D. & Settel, Z. 1987 Control of phrasing and articulation in synthesis. *Proceedings of the 1987 International Computer Music Conference*, pp. 108–116. San Francisco: Computer Music Association.

Winsberg, S. & Carroll, J.D. 1988 A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. *Psychometrika* **53**, 217–229.

Zwicker, E. & Scharf, B. 1965 A model of loudness summation. *Psychol. Rev.* **72**, 3–26.