Multidimensional Scaling of Synthetic Musical Timbre: Perception of Spectral and Temporal Characteristics

SÉVERINE SAMSON, *Université de Lille 3, France*
ROBERT J. ZATORRE and JAMES O. RAMSAY,
Montreal Neurological Institute and McGill University

Abstract The perceptual correlates of acoustic parameters involved in musical timbre were investigated by examining judgements of timbre dissimilarity. Nine synthetic sounds were created, derived from crossing three levels of spectral and temporal parameters (number of harmonics and rise time, respectively). Two separate conditions were tested, one using single tones, the other using short melodies. Fifteen musically untrained subjects were presented with pairs of stimuli and asked to judge dissimilarity on an 8-point scale. The spatial configuration resulting from multidimensional analysis of the data was best fit by a three-dimensional model, with the first two dimensions accounting for most of the variance. The perceptual space derived from the analysis indicates that these two orthogonal dimensions corresponded closely to the spectral and temporal differences inherent to the stimuli. Similar results were obtained with both melodies and single tones. A second experiment replicated the findings despite the introduction of random loudness variation, indicating that timbre judgements can be made independently of loudness. We conclude that even musically unselected subjects are sensitive to spectral and temporal information in musical tones, and are able to use them independently in making perceptual judgements of musical timbre.

Résumé La perception des paramètres acoustiques impliqués dans le timbre musical a été étudiée en examinant les jugements de dissimilarité de timbre. Neuf sons synthétiques ont été créés, en croisant trois niveaux d'information spectrale (nombre d'harmoniques) et trois niveaux d'information temporelle (temps de montée de l'attaque). Deux conditions, l'une tonale et l'autre mélodique, ont été réalisées. Des paires de stimuli ont été présentées à quinze sujets non-musiciens, qui devaient juger leur degré de dissemblance sur une échelle à 8 points. La configuration spatiale résultant de l'analyse multidimensionnelle des données correspondait à un modèle à trois dimensions, les deux premières dimensions expliquant une proportion importante de la variance.

L'espace perceptif obtenu suggère que ces deux dimensions orthogonales correspondaient aux différences spectrales et temporelles de ces stimuli. Des résultats similaires ont été obtenus avec les sons isolés (condition tonale) et les mélodies (condition mélodique). Ces résultats, répliqués dans le cadre d'une deuxième expérience où des variations d'intensité (sonie) étaient introduites, suggèrent que les jugements de timbre peuvent être réalisés indépendamment de la sonie. L'en-semble de ces données indiquent que même des sujets non-musiciens sont sensibles aux informations spectrales et temporelles des sons musicaux et qu'ils peuvent les utiliser indépen-damment pour percevoir des timbres musicaux.

Musical timbre is an attribute of sound that allows us to distinguish musical instruments when pitch, loudness and duration remain identical. For nearly a century, musical timbre of steady state tones was associated almost exclusively with the spectral energy distribution of a tone, corresponding to the pattern of intensities of the harmonic partials. Studies carried out by Helmholtz (1868) provided experimental evidence to show that this quality of sound depends mainly upon the relative amplitudes of spectral components of complex tones.

However, several lines of evidence suggest that such a definition is too narrow. For instance, it has been shown that the radical changes of the spectrum produced by the distorted output of a transistor radio does not necessarily abolish the recognition of a musical instrument (Eagleson & Eagleson, 1947). Conversely, the timbre of a recorded piano tone is perceived as completely different when it is played backward, even though the original and reversed sound have the same spectra (Berger, 1964). These results indicate that musical timbre does not depend upon a single physical dimension. Several other attributes, such as the amplitude and phase patterns of components and particularly the temporal characteristics of a tone may influence timbre perception as well. Several authors have shown

that the removal or the alteration of the initial part of a sound can also affect the recognition of musical instruments (Berger, 1964; Clark, Robertson & Luce, 1964; Grey & Moorer, 1977; Saldanha & Corso, 1964; Wedin & Goude, 1972). These findings suggest the importance of temporal features in timbre perception, and provide support for the notion of multidimensionality of this psychological attribute (Plomp, 1970; Schouten, 1968).

More recent studies investigating musical timbre have benefited from the development of digital technology. It is now possible to describe very well the physical structure of tones, and produce many kinds of complex sounds by controlling specific acoustical properties. However, the perception of these physical parameters remains poorly understood, and requires the use of an appropriate methodology. An especially well-suited candidate for this purpose is the multidimensional scaling technique (Kruskal, 1964; Shepard, 1962). Multidimensional scaling attempts to represent the dissimilarities between a set of physically different stimuli as distances between points in an n-dimensional space (Ramsay, 1982). By using matrices of perceptual measures, an algorithm operates to maximize a goodness-of-fit function relating the distances between the points to the corresponding dissimilarity ratings between the sounds. This approach provides a powerful tool for the representation of stimulus-attribute relationships.

This type of analysis of perceptual data has been used to gain a better understanding of timbre perception by allowing the identification of the perceptual dimensions along which a listener rates the sounds. Studies in timbre perception have generally settled on two-dimensional (Ehresman & Wessel, 1978; Rasch & Plomp, 1982; Wedin & Goude, 1972; Wessel, 1973; 1979; Wessel & Grey, 1978) or three-dimensional solutions (Grey, 1977; Miller & Carterette, 1975; Krumhansl, 1989; Plomp, 1976; McAdams & Cunibile, 1992).

Initially, multidimensional scaling analysis was used by Plomp (1970) to explore the dependence of timbre on the spectral amplitude pattern. By presenting nine stimuli derived from real musical tones, he showed that differences in frequency spectrum correlated quite well with differences in timbre, suggesting that both the physical analysis and the perceptual judgements were based on the distribution of spectral energy. Several years later, Wessel (1973) also collected perceptual judgements to define the different dimensions involved in timbre perception of musical instrument tones. When applying multidimensional scaling analysis to his data, he obtained a 2-dimensional solution that could be interpreted qualitatively in psychoacoustical terms. The first dimension corresponded to the brightness of the steady-state portion of the sound, distinguishing instruments whose energy was concentrated in the high as opposed to the low portion of the spectrum.

The second dimension was more difficult to interpret but seemed to group the instruments by family. Apparently, this dimension was influenced by the rapidity of the attack as well as by the relative onsets of high and low spectral components.

The importance of spectral and temporal information was examined in more detail in a set of well-controlled studies by Grey (1977; Gordon & Grey, 1978; Grey & Gordon, 1978) in which three dimensions were identified from a set of synthesized imitations of instrumental tones. The first dimension was again related to the spectral energy distribution in the steady-state, but the other two dimensions distinguished more clearly among the temporal properties of tones. One appeared to be related to the presence of low-amplitude high-frequency noise in the attack segment, and the other was interpreted with respect to the synchrony of rise and decay times in the amplitude changes of the spectrum. These authors also examined the effect of exchanging spectral envelopes among four pairs of instrumental tones in timbre perception (Grey & Gordon, 1978; Gordon & Grey, 1978). They found that when spectral envelopes were switched, the position of the instruments in the multidimensional space tended to be exchanged, especially along the first dimension related to the spectral content of the tones. This result suggested that spectral information can be ordered by listeners in a systematic way. This latter finding was subsequently replicated with a slightly different set of hybrid sounds (Krumhansl, 1989; Wessel, 1979). The position of these hybrid sounds in timbral space generally fell at locations between the two stimuli from which they were derived, confirming the results reported by Grey and Gordon (1978).

Consistent results were also reported by Ehresman and Wessel (1978) who carried out another study with a set of resynthesized instrumental tones (taped and analyzed by Grey & Moorer, 1977). The results obtained displayed once more a two-dimensional model, with one dimension corresponding to the spectrum and the other relating to the synchronicity of the frequency components. The originality of this study bears on the idea that there are analogies between timbres and that it seems possible to transpose timbral intervals as we do with pitch intervals. The authors reported preliminary results that were subsequently tested and developed by Wessel (1979) and by McAdams and Cunibile (1992).

All the investigations that we reviewed so far examined real instrumental tones or synthetic imitations of them. In many cases, the use of multidimensional scaling was motivated by the need to assess the perceptual properties of a synthetic musical sound when the goal was to simulate traditional Western orchestral instruments (Grey & Moorer, 1977; Wessel, Bristow & Settel, 1987). Very few

studies have investigated timbre perception in a set of artificial stimuli as Miller and Carterette did in 1975. These authors created stimuli according to specific parameters. In their first experiment, stimuli were synthesized with three different fundamental frequencies, three different amplitude envelopes, and three different onset patterns. The three-dimensional space resulting from the multidimensional scaling showed an overwhelming dimension related to fundamental frequency, and two other dimensions having to do with the amplitude envelopes. To avoid the problem created by the salience of one parameter relative to the others, a new set of stimuli was synthesized for the next experiment. The number of harmonics, the amplitude envelopes and the onset rates were manipulated. The timbral space resulting from this analysis revealed a three-dimensional model in which two dimensions were related to the harmonic structures. One dimension grouped the stimuli by the number of harmonics and the other contrasted sounds with 5 harmonics against sounds with 3 or 7 harmonics. The last dimension was related to the types of amplitude envelopes. It seems therefore that the three dimensions obtained in this analysis corresponded to the harmonic structure. None of them clearly revealed the importance of the temporal features of the tones, leaving unresolved the contribution of temporal characteristics of tones in timbre perception.

The results of all these studies consistently indicate the importance of the spectral energy distribution, but few of them have conclusively demonstrated the role of the temporal variation of tones and their interaction with spectral centroid in timbre perception. With the exception of a single study (Miller & Carterette, 1975), this literature has sought to describe the complex physical parameters of real or synthetic imitations of musical instruments that correlate with the perceptual dimensions of sounds. There is one difficulty inherent to this approach, however, in that the acoustic attributes that contribute to perception must be inferred post-hoc from the obtained solution. Due to the physical complexity of instrumental timbres, it is not always evident how many or which physical parameters correspond to a given perceptual dimension.

The goal of the present study was somewhat different. First, we wanted to verify if listeners would be sensitive to the physical dimensions of a set of unfamiliar stimuli in a systematic fashion. We therefore manipulated the spectral and temporal characteristics of synthetic tones to determine if listeners would recover in their perceptual judgements the unambiguous acoustic dimensions that were physically inherent to the stimuli. A second question raised in the present study was whether timbre perception would be the same when listeners judge isolated tones as compared to melodic sequences. Most studies have investigated only single tones, neglecting perception of sequences

of tones. However, real music usually involves patterns of different notes played over time, in which a more complete picture of a given timbre might emerge. Throughout a musical phrase, different durations and frequencies are heard, and the physical parameters of tones evolve. This evolution could obscure or enhance the importance of certain parameters. It seemed, therefore, ecologically valid to examine timbre perception in melodies as well as in single tones, and to compare the spatial configurations resulting from perception of these two types of stimuli. Finally, unlike most previous studies of timbre perception, we tested musically unsophisticated subjects in order to be able to generalize our findings more broadly, and in order to avoid any pre-existing associations to known musical instruments. Because the synthesized stimuli were unfamiliar to the listeners, any tendencies to use verbalization or semantic categories when making perceptual judgements were also minimized.

For these purposes, we created a set of stimuli by crossing three levels of spectral content and three levels of temporal content resulting in a matrix of nine sounds. Pairs of sounds were presented for rating of dissimilarity. The data were then analyzed via multidimensional scaling. We predicted that the three by three matrix used to create the stimuli would be reflected in a similar three by three pattern in a two-dimensional geometric representation of the dissimilarity ratings. We further predicted that the representation for the melodic stimuli would be more stable than that for the single-tone stimuli.

Experiment 1

METHOD
Subjects
Fifteen normal subjects participated in this experiment. Their mean age was 27 years (range: 20-47). There were 9 females and 6 males. Eleven subjects had no formal musical training, apart from some classroom music instruction in elementary school; the remaining four subjects had had two years or more of music lessons, but were not actively studying music or playing an instrument at the time of testing.

Stimuli
The stimulus sounds were synthesized by selecting three levels of spectral content and three levels of temporal content. The spectral content corresponded to the number of harmonics: either one, four or eight harmonics. The temporal change consisted of the manipulation of the linear rise time, that could be of 1, 100 or 190 ms in length immediately followed by a linear decay time, without a steady-state portion. These parameters were selected on the basis of pilot data indicating that the stimuli would be clearly distinct from one another. The minimum and
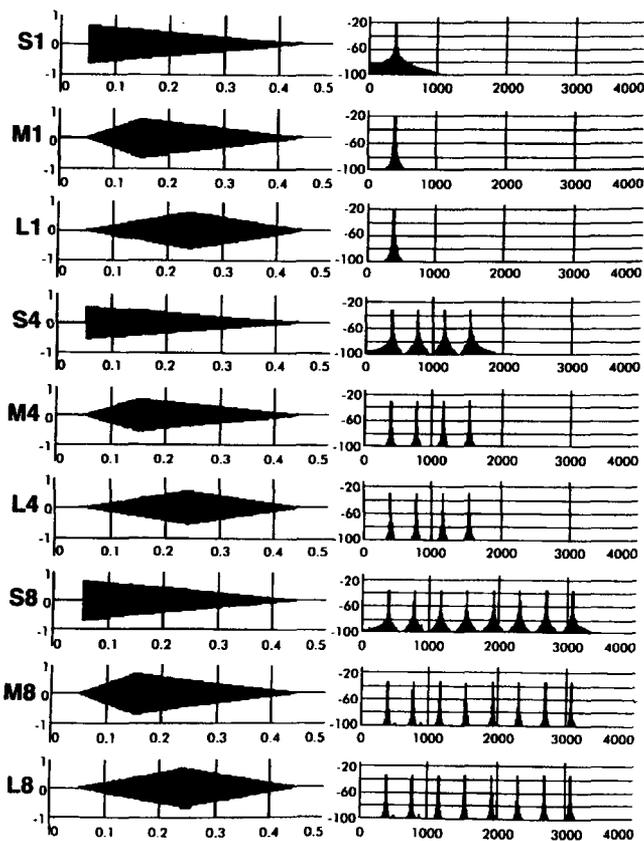
Figure 1. Representations of each of the nine stimuli differing by spectral and temporal parameters. The left panel in each case shows the temporal waveform of the stimulus (abscissa: time, in tenths of seconds; ordinate: amplitude, arbitrary scale). The right panel in each case shows the Fourier spectrum of the stimulus (abscissa: frequency, in Hz; ordinate, energy level, in dB). S, M, and L correspond to short (1 ms), medium (100 ms) and long (190 ms) rise times, respectively. The number 1, 4 or 8 refers to the number of harmonics present in the stimulus. Thus, for example, S1 corresponds to a stimulus containing short (1 ms) onset (note sharp attack on waveform in left panel) and consisting of a single harmonic (note the presence of only one harmonic in the right panel); M4 refers to a stimulus with a medium (100 ms) onset time and containing four harmonics, and so forth.

maximum rise times (1 and 190 msec) were chosen as a function of the total duration of the shortest tone to be used in the study with the melodies (200 msec). Representations of the wave form and spectra of the stimuli are displayed in Figure 1. Nine sounds (resulting from crossing the three levels of spectral composition with the three levels of rise time) were then created by digital synthesis at a 16 kHz sampling rate, using the MITSYN Command Language (Henke, 1976), and stored on the hard disk of a Compaq 386 computer. Stimuli were presented binaurally to each subject through Sennheiser 222 headphones after low-pass filtering at 8 kHz.

Single tones. A 400ms tone played on the musical tone G4 (fundamental of 396 Hz) was created for each timbre,

making therefore 9 tones with the same duration, fundamental frequency and intensity. Each trial was composed of 2 tones separated by a 500ms silence. All tones were presented at a 70dB SPL (A scale), as measured by a GenRad sound pressure meter at the headphone.

Melodies. Eight-note melodies were synthesized, each using one of the 9 musical timbres. The melody consisted of a rising contour, and was constructed with a rhythmic pattern using three different durations (200, 400, and 600 ms); the notes used (C4, E4, G4, A4, B4, and C5) spanned an octave. To prepare each tone within the melody, the rise time was kept constant for each type of temporal manipulation but the decay time was varied according to the tone duration. For instance, a rise time of 1msec was followed by either 199ms, 399ms or 599ms for the three note durations of 200ms, 400ms and 600ms, respectively. These values were chosen on the basis of pilot studies, which indicated that a constant rise time elicited a stable percept, whereas a rise time proportional to the length of the tone yielded different percepts for each value. Each trial was composed of the same two melodies separated by 500 ms silence.

Procedure

The 9 stimuli were presented to each subject in an all-pairs design with no stimulus paired with itself, which resulted in the presentation of 36 stimulus pairs. Two sets of 36 pairs were prepared, the second set consisting of the same pairs of stimuli presented in the reverse order. The identical procedure was used for the single tones and for the melodies. Therefore, we had 2 sets of 36 pairs of tones as well as 2 sets of 36 pairs of melodies.

Subjects were told that the aim of this study was to examine their ability to distinguish different types of sounds. The subject's task was to judge the dissimilarity between two stimuli. On each trial, two stimuli (either tones or melodies) were played with different timbres and immediately after their presentation, a question mark appeared on the computer screen. The subject responded with the keyboard by selecting a number corresponding to the degree of dissimilarity. The dissimilarity rating was made on scale of 1 to 8. The smallest of these is labelled "very similar" and the largest "very different." The program waited until the subject responded before presenting the next trial, which occurred two seconds after response.

The melodic and the tonal versions of the test were presented in the same experimental session. All subjects started with one set of melodies (36 pairs) followed by one set of tones (36 pairs), followed by the other set of melodies (reverse order from the first), and then the other set of tones. This order of presentation was chosen because

TABLE 1
Coordinates for each of the points in the spatial configuration for the single tone and melodic conditions for Experiment 1. Dimensions 1 and 2 correspond to the abscissa and ordinate, respectively (labelled as spectral and temporal), of the graphical representation of the data shown in Figures. 2 and 3.

Stimulus		Single Tone Condition			Melodic Condition		
		Dimension			Dimension		
		1	2	3	1	2	3
S1	1	-28	32	15	-29	41	5
M1	2	-29	-12	8	-34	-8	15
L1	3	-29	-23	2	-38	-17	5
S4	4	-1	26	-15	12	36	7
M4	5	-1	-7	-13	9	-18	17
L4	6	5	-15	-19	7	-29	10
S8	7	27	25	7	25	34	-18
M8	8	30	-9	5	21	-14	-24
L8	9	24	-18	10	27	-25	-18

the melodic version was easier to perform than the tonal one. Half of the subjects started with the first set of melodies and the first set of tones while the other half started with the second set of melodies as well as the second set of tones. The order of presentation of the pairs of stimuli within a set was completely randomized by the computer.

Training sessions. Because subjects were generally unsophisticated musically, prior to the experimental tasks they received training to ensure that they were able to use the rating scale adequately. Preliminary tasks requiring perceptual judgement of tones that differed by pitch and loudness were presented to each subject before the experimental tasks. They consisted of 20 examples followed by 2 series of 36 pairs of tones involving dissimilarity judgements on an 8-point scale. Then, the different timbres were exposed to each subject through 15 examples consisting of pairs of melodies played on different timbre for purposes of familiarization. Examples of the extremes were given by presenting a pair of two very different stimuli and a pair of two very similar ones followed by 13 other samples of different pairs of melodies. Each pair of melodies was followed by a written message on the screen to indicate the type of dissimilarity between the timbres (e.g., "very different," "very similar" or "quite different") and to illustrate the sort of differences the subject should expect to hear. Then, the experimental task per se was demonstrated to the subject through 5 examples consisting

of pairs of melodies. The subject had to rate the dissimilarity between the two members of the pairs on an 8 point scale, identical to the experimental task.

RESULTS
The dissimilarity judgements for each listener were stored as a 9 by 9 matrix of data minus the diagonal. The matrices of dissimilarities for all 15 subjects who participated in the experiment were then processed by a multidimensional scaling program (MULTISCALE II) developed by Ramsay (1982). The spatial configuration obtained from these analyses was the result of maximum likelihood estimation and showed a good statistical relationship between the distances separating the points in the space and the dissimilarity judgements given for the pairs of stimuli. The analyses of the single tones were considered separately from the analyses of melodies. Preliminary analysis showed that there were few meaningful differences in judgement with respect to the order of presentation. We therefore averaged the data obtained during the two orders of presentation used for the single tones as well as for the melodies.

Single tone condition. The multidimensional scaling algorithms were used to generate geometric representations of the sounds in a Euclidian space in which the points correspond to the 9 different tones. Spatial representations were obtained for 2 and for 3 dimensions. It was determined that a three-dimensional scaling solution was most appropriate for the data because the improvement in fit resulting from an extension of this model from 2 to 3 dimensions was significant (p < .001). This was assessed by using the log likelihood ratio, and more specifically the increase in log likelihood multiplied by two. It offers a useful index of improvement in fit. These 3 dimensions accounted for 98% of the variation among the points. However, it was evident that the first two dimensions were far more important than the third. The percentage of variance was evenly distributed between the first two dimensions (48% and 38%), whereas the third dimension only accounted for 12% of the variance. For this reason, the geometric representations for the first two dimensions only are presented in the figures. The coordinates of the 9 stimuli along all three dimensions are presented in Table 1.

Figure 2 shows the scaling solution for all subjects for the first two dimensions. The points corresponding to the sounds that are judged to be very dissimilar are distant while the ones corresponding to the sounds that are judged to be very similar are close to each other. The distribution of points in this plane is systematic and clearly related to the physical characteristics of the sounds. The two dimensions appear to be orthogonal, as well.
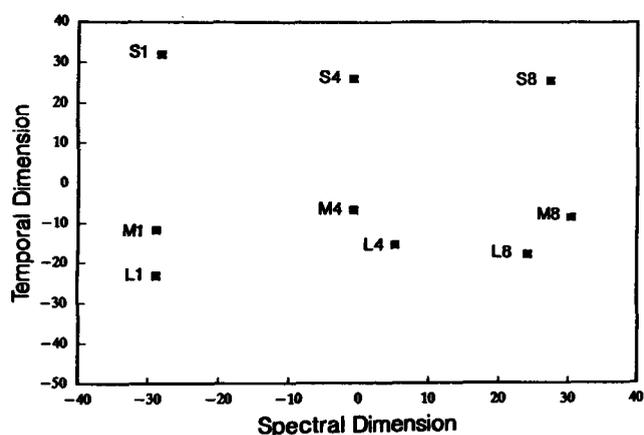
Figure 2. Two-dimensional spatial representation based on the dissimilarity ratings for the nine single tones from Experiment 1, as revealed by multidimensional scaling analysis. Each point represents the position of one stimulus in the perceptual space; the labels (S1, etc.) correspond to the stimuli described in Figure 1. Note that stimuli with the same number of harmonics are mostly organized vertically, while stimuli with the same rise time are primarily organized horizontally.

Along the first dimension, the sounds were grouped according to the number of harmonics. The 1-harmonic tones (S1, M1, and L1) were organized in a vertical column, and were separated from the three 4-harmonic tones by a distance similar to the distance separating the 4- from the 8-harmonic tones. The 4- and 8-harmonic tones were also organized roughly in a vertical fashion. Along the second dimension, tones with 1 ms rise time (S1, S4, and S8) were organized horizontally, and were separated from tones with 100 and 190 ms rise time quite well. The distance between points corresponding to tones with 1 and 100 ms rise time was larger than the distance between points corresponding to tones with 100 and 190 ms rise time. The third dimension (not shown in the figure; see Table 1) distinguished the tones composed of 4 harmonics from those with 1 or with 8 harmonics.

Melodic condition. The first two dimensions of the 3-dimensional solution obtained from perceptual judgement of melodies is visualized in figure 3. Again, it was found that the 3-dimensional model provided better fit to the data than the 2-dimension model, but the first two dimensions were most important. The improvement of fit resulting from an extension of this model from a 2- to a 3-dimensional solution was highly significant ($p < .001$). Modelling in three dimensions accounted for 98% of the variance, the first two dimensions being clearly dominant (accounting for 46% and 39% of the variance, respectively) as compared to the third one (accounting for 13%).

The spatial configuration of the 9 different points is very similar to the one described for the single tones. The
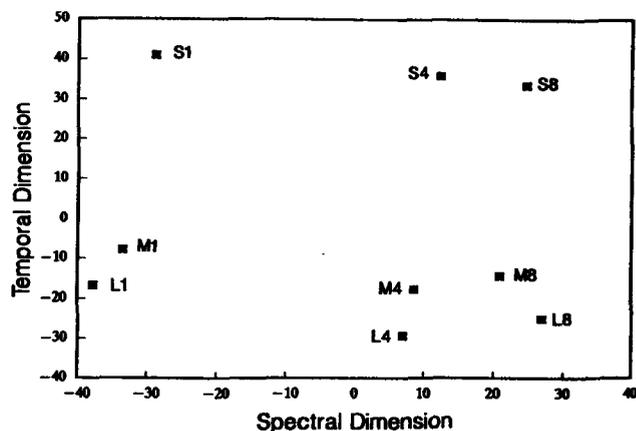

Figure 3. Two-dimensional spatial representation based on the dissimilarity ratings for the nine melodic patterns from Experiment 1, as revealed by multidimensional scaling analysis. Other details as in Figure 2.

distribution of the points in this configuration is also very well ordered: the first dimension, displayed on the horizontal axis, was again related to the three sets of points corresponding to 1, 4 and 8 harmonics, equally distributed along this dimension. Dimension 2 separates points along the vertical axis that corresponds to tones with different rise times. Once again, the 1 ms rise time tones were separated from the 100 and the 190 ms rise time tones, these two latter types of sounds being very close to each other. Finally, the third dimension differentiated points corresponding to 8 harmonic tones from points corresponding to 1 and 4 harmonic tones.

To determine in a more quantitative manner whether the solutions for the single-tone and melodic conditions were similar, we correlated the values obtained for each dimension across the two stimulus types. The agreement for the first (spectral) dimension was good, with a correlation coefficient of $r = .93$. Similarly, the correlation for dimension 2 (temporal) was $r = .83$. The correlation for the third dimension was not significant ($r = -.48$), not surprisingly since in the single-tone condition this dimension distinguished the 4-harmonic tone stimuli from the rest, whereas in the melodic condition this third dimension differentiated the 8-harmonic tones from the other two.

Experiment 2

The results of the first experiment indicate a consistent use of spectral and temporal information in making dissimilarity judgements of timbre. Two issues remain to be clarified, however. First, although the stimuli had been equalized for A-weighted intensity, we cannot exclude the possibility that loudness differences may have influenced the results. We therefore tested a new set of subjects under similar conditions to Experiment 1, but introducing

TABLE 2
Coordinates for each of the points in the spatial configuration for Experiment 2. Dimensions 1 and 2 correspond to the spectral and temporal dimensions, respectively.

Stimulus		Dimension		
		1	2	3
S1	1	-44	28	4
M1	2	-41	-15	16
L1	3	-35	-25	9
S4	4	13	42	14
M4	5	12	-15	26
L4	6	11	-29	6
S8	7	25	49	-21
M8	8	33	-10	-19
L8	9	25	-26	-35

random variations of loudness between stimuli within a pair. We reasoned that the scaling solutions derived from Experiment 1 should be independent of any loudness differences in the stimuli, and we therefore predicted that variation of this parameter should have no effect on the timbre judgements.

Second, in Experiment 1, subjects had been explicitly instructed about the differences and similarities among the stimuli. This was done because we wished to be able to adapt this task to a neuropsychological population in future studies. It can be argued, however, that this procedure introduces a bias, because the subjects are told what to expect and how to respond. In Experiment 2, therefore, subjects were only given a few examples of stimuli prior to testing, with no specific information about the judgements that they should give.

METHOD
Subjects
Twelve new subjects with little or no formal musical training were tested in this experiment.

Stimuli
The stimuli consisted of the same nine sounds as in Experiment 1 except that the intensity was manipulated. Different intensities were used: either the same intensity as in Experiment 1 or a lower intensity level (minus 6 or 12 dB). Only the eight-note melodies were used rather than the single tones.

Procedure
Two sets of 36 pairs of melodies were prepared. The procedure was similar to Experiment 1, except that within a pair of melodies, the intensity of the first or the second stimulus was randomly lowered by 6 or by 12 dB. The subject's task was identical to Experiment 1. Only a few examples of stimuli were given before the task, but no specific information about their judgements was given other than to explain the use of the dissimilarity rating scale. Subjects were informed that there would be intensity differences within each trial, but were told to attend to the timbre of each synthetic instrument rather than to the intensity differences.

RESULTS
The dissimilarity ratings were analyzed in the same fashion as for Experiment 1. The coordinates of the 9 stimuli along the three dimensions are presented in table 2. Again, it was found that the 3-dimensional model provided better fit to the data than the 2-dimensional model. Modeling in three dimensions accounted for 98% of the variance, the first two dimensions being clearly dominant (accounting for 40 and 41% of the variance, respectively) as compared to the third one (accounting for 13%).

The spatial configuration of the 9 different points is very similar to the ones described in Experiment 1. Correlation of the coordinates from the melodic condition of Experiment 1 with those of Experiment 2 yielded highly significant values ($r = .96$ for both spectral and temporal dimensions; $r = .94$ for the third dimension). The first dimension was again related to the three sets of points corresponding to 1, 4 and 8 harmonics. Dimension 2 distinguishes points that corresponds to tones with different rise times, as in Experiment 1. Dimension 3 once again dis-tinguished the tones with 8 harmonics from the others. Thus, despite large differences in intensity level in the sti-muli to be judged, the multidimensional scaling solution was stable, demonstrating that subjects based their judgement on the spectral and temporal information, and that loudness does not influence the ability to extract these features.

General Discussion
The results of these studies clearly demonstrate that musically untrained listeners can systematically order stimuli in a perceptual space that reflects the acoustical structure of the stimuli. Subjects were evidently able to recover two independent perceptual dimensions, which were orthogonal to one another and essentially of equal salience, despite being entirely unaware of the construction of the stimuli, and in the absence of explicit instructions in Experiment 2.

Most importantly, our data indicate that under the present conditions, at least, it is possible to demonstrate the *independent* contribution of temporal features to

timbre perception, in addition to the well-established contribution of harmonic structure. Furthermore, loudness evidently does not seem to influence the structure of the multidimensional space, since the data from Experiment 2 were virtually identical to those of Experiment 1, despite the presence of marked loudness differences between stimuli. The geometrical model resulting from the multidimensional scaling analysis was composed of three dimensions, but the third one was less salient and less clearly defined.

Contrary to our expectations, we did not find a more orderly solution in the melodic condition than in the single tone condition, even though it appears to be a slightly more realistic situation, and presumably offers more information to the listener than the single tones do. It may be that listeners extract sufficient information from the single tones to represent the relevant features of the stimulus, and that the additional information provided by the frequency excursions and different durations of the tones within the melodies is simply not necessary in this particular task. It is still possible, however, that under more difficult task conditions (e.g., more subtle harmonic changes, provided by modifying the spectral envelope), the melodic stimuli would result in a clearer perceptual representation than the single tones.

The first dimension displayed on the spatial configurations distinguishes tones according to their number of harmonics. Moreover, the distance separating the 1- and the 4-harmonic tones seems very similar to the distance separating the 4 and the 8 harmonic tones. This regular pattern was also present in the spatial configuration reported by Miller and Carterette (1975) with 3, 5 and 7 harmonic tones, and suggests that spectral information — in the form of numbers of harmonics — can be perceived in a gradual and ordered way. This finding is also consistent with the important role of the spectral envelope in perceiving timbre previously found with synthesized imitations of instrumental sounds (Ehresman & Wessel, 1978; Grey, 1977; Grey & Gordon, 1978; Krumhansl, 1989; Wessel, 1973; 1979).

The second major dimension resulting from the multidimensional analysis had to do with a temporal characteristic of the tones, which was determined by the rise/decay time. This temporal parameter had not been directly manipulated in past studies, and our results indicate that it can be perceptually isolated. In the spatial configurations we found that the distances separating tones with 1 ms and 100 ms rise time were consistently larger than the one separating tones with 100 ms and 190 ms rise time. These results indicate that subjects were able to differentiate 1 ms from 100 ms rise time clearly, but that the perceptual difference between tones of 100 ms and 190 ms rise time was less salient. Nonetheless, the relative

order of rise time (1, followed by 100, followed by 190 msec) was consistently preserved in both sets of stimuli; a result consistent with recent findings of McAdams and Cunibile (1992). The nonlinearity in perception as a function of rise-time duration could be explained in terms of the underlying psychophysical scale. It is likely that rise time information is ordered perceptually on a logarithmic scale, and that perceived differences among tones that differ in rise time are therefore not linearly related to the physical distances between rise-time durations (see Rosen & Howell, 1987, for further discussion).

The finding that the shortest rise time was perceptually distant from the other two can also be attributed, at least in part, to the presence of additional, non-temporal cues in the perception of tones with 1ms rise time. From the physical point of view, abrupt onsets and offsets produce a spread of energy across the spectrum, which may contribute to the overall percept. It is likely, therefore, that a more even perceptual spacing would be obtained if very short onsets are avoided. However, even if it is acknowledged that spectral integration may be involved in the perception of very short rise times, the results of the present study clearly showed that the temporal and the spectral dimensions are perceptually dissociable, since the two principal dimensions were orthogonal to one another, and did not appear to interact.

The third dimension obtained in these analyses was clearly less salient than the other two. Although the 3-dimensional space provided a statistically better fit of the data than the 2-dimensional space, the third dimension accounted for only a relatively small amount of variance (13%) in both single-tone and melody conditions and in the second experiment as well. Furthermore, this third dimension did not appear to be completely stable, since the ordering of stimuli was different for single tones and for melodies (in one case the 4-harmonic tones were separated from the others by the third dimension, whereas in the other condition it was the 8-harmonic tones which were distinguished by the third dimension). It is interesting, however, that in all conditions the third dimension was related to the harmonic composition of tones. In this respect the findings are similar to those of Miller & Carterette (1975), who also described a dimension that distinguished 5-harmonic tones from those composed of 3 and 7 harmonics. The presence of this dimension, therefore, cannot simply be dismissed, but its meaning and importance remain to be established at this time.

In conclusion, the present study succeeded in demonstrating that scaling of the perceived degree of dissimilarity between different timbres can be related to the underlying acoustical properties of tones. Our results provide evidence that both temporal (rise time) and spectral (number of harmonics) features of a sound are dissociable in a

perceptual space, and that this does not depend on extensive musical training or familiarity with the stimuli. The agreement between the first two dimensions described in the scaling solutions of the single tone, as well as the melodic conditions from Experiments 1 and 2 provides strong support for the stability of this modelling outcome.

The MULTISCALE II program may be obtained free of charge from: J. Ramsay, Dept. of Psychology, McGill University, 1205 Dr. Penfield Ave., Montreal, Quebec, Canada H3A 1B1. Other correspondence regarding this work may be addressed to R. Zatorre, Neuropsychology/Cognitive Neuroscience Unit, Montreal Neurological Institute, 3801 University St., Montreal, Québec, Canada, H3A 2B4, or to S. Samson, whose present address is: Dept. de Psychologie, Université de Lille III, BP 149, 59653 Villeneuve d'Ascq Cedex, France.

References

Berger, K.W. (1964). Some factors in the recognition of timbre. *Journal of the Acoustical Society of America, 36*, 1888-1891.

Clark, M., Robertson, J.P., & Luce, D. (1964). A preliminary experiment on the perceptual basis for musical instrument families. *Journal of the Audio Engineering Society, 12*, 199-203.

Eagleson, H.V., & Eagleson, O.W. (1947). Identification of musical instruments when heard directly and over a public-address system. *Journal of Acoustical Society of America, 19*, 338-342.

Ehresman, D.E., & Wessel, D.L. (1978). *Perception of timbral analogies*. Technical Report No. 13, Paris: IRCAM.

Gordon, J.W., & Grey, J.M. (1978). Perceptual effects of spectral modifications on orchestral instrumental tones. *Computer Music Journal, 2*, 24-31.

Grey, J.M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of Acoustical Society of America, 61*, 1270-1277.

Grey, J.M., & Gordon, J.W. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of Acoustical Society of America, 63*, 1493-1500.

Grey, J.M., & Moorer, J.A. (1977). Perceptual evaluations of synthesized musical instrument tones. *Journal of Acoustical Society of America, 62*, 454-462.

Helmholtz, H. (1868). *Théorie Physiologique de la Musique Fondée sur l'Etude des Sensations Auditives*. Paris: Masson.

Henke, W.L. (1976). *An interactive dialogue language for time signal processing*. Cambridge, MA.: MIT Research Laboratory of Electronics.

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

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

McAdams, S., & Cunibile, J-C. (1992). Perception of timbral analogies. *Philosophical Transactions of the Royal Society of London B, 336*, 383-389.

Miller, J.R., & Carterette, E.C. (1975) Perceptual space for musical structures. *Journal of Acoustical Society of America, 58*, 711-720.

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

Plomp, R. (1976). *Aspect of tones sensation*. New York: Academic Press.

Ramsay, J.O. (1982). *Multiscale II Manual*. Montreal: McGill University, Department of Psychology.

Rasch, R.A., & Plomp, R. (1982). The perception of musical tones. In D. Deutsch (Ed.), *The Psychology of music*. (pp. 1-13) New York: Academic Press.

Rosen, S., & Howell, P. (1987). Auditory, articulatory, and learning explanations of categorical perception in speech. In S. Harnad (Ed.), *Categorical perception*. Cambridge: Cambridge University Press.

Saldanha, E.L., & Corso, J.F. (1964). Timbre cues and the identification of musical instruments. *Journal of Acoustical Society of America, 36*, 2021-2026.

Schouten, J.F. (1968). The perception of timbre. In *Report of the 6th International Congress on Acoustics*, Tokyo, GP6-2.

Shepard, R.N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika, 27*, 125-140

Wedin, L., & Goude, G. (1972). Dimension analysis of the perception of instrumental timbre. *Scandinavian Journal of Psychology, 13*, 228-240.

Wessel, D.L. (1973). Psychoacoustics and music. *Bulletin of the Computer Arts Society, 30*, 1-2.

Wessel, D.L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3*, 45-52.

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

Wessel, D.L., & Grey, J.M. (1978). *Conceptual structures for the representation of musical material*. Paris: IRCAM, Technical Report No. 14.