In J. Beauchamp (Ed.), *Analysis, Synthesis, and Perception of Musical Sounds* (pp. 272-319), 2007. New York, NY: Springer.

8

Mental Representation of the Timbre of Complex Sounds

SOPHIE DONNADIEU

"Un des paradoxes les plus frappants à propos du timbre est que, lorsqu'on en savait moins sur lui, il ne posait pas beaucoup de problèmes . . . "
[One of the most striking paradoxes concerning timbre is that when we knew less about it, it didn't pose much of a problem . . .]
Philippe Manoury (1991)

1 Timbre: A Problematic Definition

Timbre, in contrast to pitch and loudness, remains a poorly understood auditory attribute. Persons attempting to understand it may be confused as much by its nature as its definition. Indeed, timbre is a "strange and multiple" attribute of sound (Cadoz, 1991, p. 17), defined by what it is not: it is neither pitch, nor loudness, nor duration. Consider the definition proposed by the American National Standards Institute (1973, p. 56): "Timbre is that attribute of auditory sensation in terms of which a subject can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar." Therefore, timbre is that perceptual attribute by which we can distinguish the instruments of the orchestra even if they play the same note with the same dynamics.

The absence of a satisfactory definition of timbre is primarily due to two major problems. The first one concerns the multidimensional nature of timbre. Indeed, it is timbre's "strangeness" and, even more, its "multiplicity" that make it impossible to measure timbre along a single continuum, in contrast to pitch (low to high), duration (short to long), or loudness (soft to loud). The vocabulary used to describe the timbres of musical instrument sounds indicates the multidimensional aspect of timbre. For example, "attack quality," "brightness," and "clarity" are terms frequently used to describe musical sounds. The second problem concerns timbre as a concept that refers to different levels of analysis. Schaeffer (1966, p. 232) observed that one can talk about "the timbre of a sound without attributing it to a given instrument, but rather in considering it as a proper characteristic of this sound, perceived per se." He noted that "we shouldn't confuse two notions of timbre: one related to

the instrument, an indication of the source that is given to us by ordinary listening, and the other related to each of the objects provided by the instrument, appreciation of the musical effects in the objects themselves, effects desired by musical listening as well as by musical activity. We have even gone further, attaching this word timbre to an element of the object: timbre of the attack, distinguished from its stiffness." So, the concept of timbre is much more general than the ability to distinguish instruments. The problem is that only one term refers to many different notions: Timbre can be described in terms of (1) a set of sounds of an instrument and also of the specific timbre of each sound of a particular instrument, (2) an isolated sound, (3) a combination of different instruments, (4) the composition of a complex sound structure, or (5) in the case of timbres produced by analysis/resynthesis, hybrid timbres or chimeras, sounds never heard before, which can be associated with no known natural source. For the purposes of this chapter, we refer to timbre in terms of sound sources or multidimensional perceptual attributes.

Timbre conveys the identity of a sound source. In other words, the timbre of a complex sound comprises the relevant information for identifying sound sources or events, even in a musical context. As Schaeffer (1966) said: "It is denying the evidence to believe that pure music can exempt the ear from its principal function: to inform humans about the events that are occurring" (cited by Cadoz, 1991, p. 17). In the same way, we do not have any difficulty knowing that someone is playing a violin in the neighboring room or that a car has suddenly arrived behind us. This capacity to identify sound objects is necessary to our survival. Indeed, when we hear a motor noise while crossing a street, our reaction is to immediately step back onto the sidewalk to avoid an accident. Most certainly, in everyday life, we use all the sensory systems at the same time. However, the events mentioned above can be identified even if they occur outside our visual field and outside any context likely to facilitate our interpretation of the sound objects (McAdams, 1993).

Most studies of musical timbre have used single, isolated instrument tones, which are easy to manipulate for experimentation. Our discussion of these studies is organized by the theoretical models adopted by the researchers. The first model is information processing (Lindsay and Norman, 1977), which describes the perceptual dimensions of timbre in terms of abstract attributes of sounds. In other words, the acoustical parameters (spectral, temporal, and spectrotemporal) of the signal are processed by the sensory system, and the perceptual result is the timbre of complex sounds. Multidimensional scaling has been fruitful in determining these different perceptual dimensions of timbre. The second approach, based on *ecological theory* proposed by Gibson (1966, 1979), has only recently resulted in systematic experimentation in auditory perception. According to this viewpoint, timbre perception is a direct function of the physical properties of the sound object. The aim of these studies is to describe the physical parameters that are perceptually relevant to the vibrating object.

2 The Notion of Timbre Space

2.1 Continuous Perceptual Dimensions

Multidimensional scaling (MDS) has been a effective tool for studying the timbral relationships among stimuli possessing multiple attributes. The principal advantage of this exploratory technique is that *a priori* hypotheses concerning the number of dimensions and their psychophysical nature are not required. Generally, MDS is used in an auditory study in the following manner: A set of sound stimuli—in this case, the sounds of musical instruments—are presented in all possible pairs. The listener's task is to judge the dissimilarity between the timbres for each pair of sounds. The dissimilarity is measured generally on a numerical scale (for example, 1 to 9, with 1 being very similar and 9 being very dissimilar) or on a bounded, continuous scale (for example, indicated with a cursor varied continuously on a scale between "very similar" and "very dissimilar," which is subsequently coded numerically). The pitch, subjective duration, and loudness of all the sounds are usually equalized so that the subject's ratings concern only timbral differences. At the end of the experiment, a dissimilarity matrix is tabulated. The aim of MDS is to produce a geometric configuration that best represents, in terms of metric distances, the perceptual dissimilarities between the timbres of the sounds. So, two timbres judged on average to be very similar should appear close together in the space, and two timbres judged to be very dissimilar should appear far apart in the space. The number of dimensions required for the spatial solution is determined by using a goodness-of-fit measure or statistical criterion.

The last step in the MDS analysis is the psychophysical interpretation. The goal is to find a relationship between some acoustical parameters and the perceptual dimensions of the MDS solution. Typically, we measure a number of physical parameters, such as spectral envelope, temporal envelope, and so on, for all of the stimuli. Then we compute correlations between the positions of the timbres relative to the perceptual axes and the physical parameters.

2.1.1 Spectral Attributes of Timbre

Scientists have devoted themselves to the psychophysical analysis of musical sounds for several decades. These studies showed that spectral characteristics have an important influence on timbre. The influence of such spectral factors is revealed by multidimensional analyses. Plomp (1970, 1976) used multidimensional techniques to study synthesized steady-state spectra derived from recordings of musical instrument tones. He found a two-dimensional solution for a set of synthetic organ-pipe stimuli and a three-dimensional solution for a set of wind and bowed-string stimuli. He did not give a psychoacoustical interpretation of the individual MDS axes, but he showed that the spectral distances (calculated as differences in energy levels across a bank of 1/3-octave filters) were similar to those for the dissimilarity ratings for each stimulus set. This result suggests that global activity level present in the human auditory system's array of frequency-specific nerve fibers

may constitute a sufficient sensory representation from which a small number of perceptual factors related to the spectral envelope may be extracted. De Brujin (1978) found a correlation between the spectral envelope of synthesized tones and dissimilarity judgments.

Preis (1984) asked listeners to judge the degree of dissimilarity between synthetic and original musical instrument tones. In this case, a correlation was observed between the metric distances separating the tones and a measure of the degree of dissimilarity between the tones' spectral envelopes. In the same way, Wedin and Goude (1972) observed that spectral-envelope properties explained the three-dimensional perceptual structure of similarity relations among musical instrument tones (winds and bowed strings). In one of their experiments on synthesized tones, Miller and Carterette (1975) varied the number of harmonics, a spectral property. This spectral property corresponded with two of three perceptual dimensions. The remaining acoustical variables employed corresponded with the third perceptual dimension. These were the amplitude-vs-time envelope (temporal) and the pattern of onset asynchrony of the harmonics (spectrotemporal). The results of this study suggested a perceptual predominance of spectral characteristics in timbre judgments. In the same way, Samson et al. (1996) observed a two-dimensional space in which the organization of timbres reflected spectral and temporal differences. Nine hybrid synthetic sounds were created, derived from crossing three levels of spectral change corresponding to a change in the number of harmonics. (The tones were comprised of one, four, or eight harmonics.) The authors observed that the positions of tones along one of the dimensions corresponded closely to the number of harmonics. These results suggest that the manipulation of certain parameters influences subjects' perception of complex sounds.

Grey (1975, 1977) and Wessel (1979) observed similar multidimensional spaces with relatively complex synthesized tones meant to imitate conventional musical instruments (winds, bowed strings, plucked strings, or mallet percussion). Figure 8.1 shows the timbre space constructed by Grey (1975). The first axis is interpretable in terms of the spectral energy distribution. At one extreme, instruments like the French horn or the cello had low spectral bandwidths and concentrations of low-frequency energy. At the other extreme, the oboe has a very wide spectral bandwidth and less concentration of energy in the lowest harmonics.

Grey and Gordon (1978) were the first to propose a quantitative interpretation of spectral energy distribution. They found that the centroid of a loudness function based on time-averaged amplitudes of stimulus harmonics correlated strongly with the first dimension of MDS models for tones interpolated acoustically between Grey's (1975, 1977) original acoustic instrument tones and their spectral modifications of some of these tones. Iverson and Krumhansl (1993), using complete synthetic tones, those with attack portion only, and those with attacks removed, gave a similar interpretation of the second dimension of their three spaces.

Krimphoff (1993) and Krimphoff et al. (1994) conducted acoustical analyses on the set of 21 sounds created by Wessel et al. (1987) and used by Krumhansl (1989) in an MDS timbre study. Most of these synthetic sounds imitated traditional

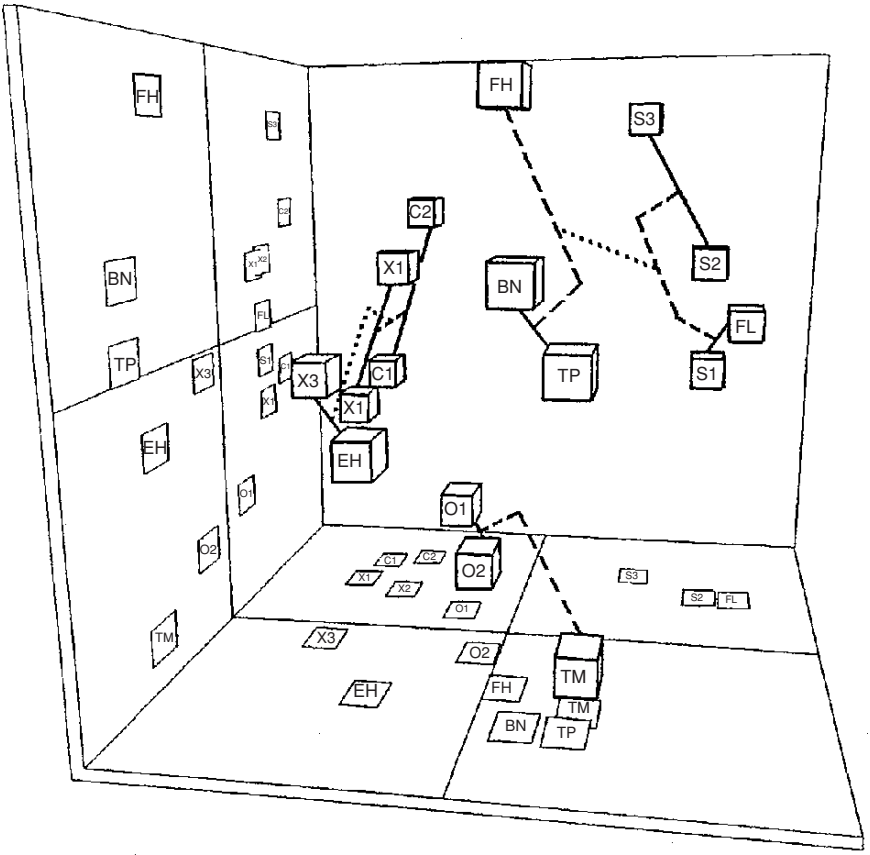


FIGURE 8.1. Three-dimensional INDSCAL solution derived from similarity ratings for 16 musical instrument tones. Two-dimensional projections of the configuration appear on the wall and the floor. Abbreviations for the instruments: O1 and O2, two different oboes; C1 and C2, E$^{\flat}$ and bass clarinets; X1 and X2, alto saxophone playing softly and moderately loud, and X3, soprano saxophone, respectively; EH, English horn; FH, French horn; S1, S2, and S3, cello playing with three different bowing styles: *sul tasto, normale, sul ponticello*, respectively; TP, trumpet; TM, muted trombone; FL, flute; BN, bassoon. Dimension I (top-bottom) represents spectral envelope or brightness (brighter sounds at the bottom). Dimension II (left-right) represents spectral flux (greater flux to the right). Dimension III (front-back) represents degree of presence of attack transients (more transients at the front). Hierarchical clustering is represented by connecting lines, decreasing in strength in the order: solid, dashed, and dotted. [From Grey (1977), Fig. 1, used by permission of Acoustical Society of America.]

instruments, but some were chimerical hybrids (e.g., a "trumpar" created by combining spectrotemporal characteristics of the trumpet and the guitar). Krumhansl (1989) did not attempt to give a quantitative interpretation of her MDS solution, but she intuitively interpreted each of its axes according to the positions

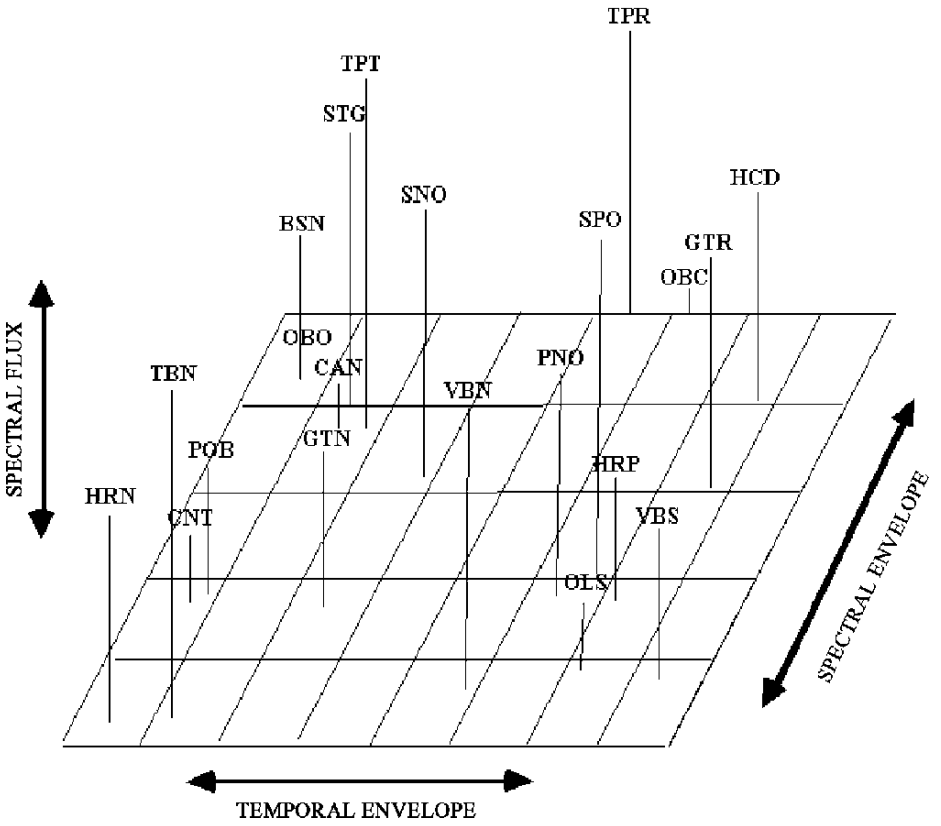


FIGURE 8.2. Three-dimensional EXSCAL solution derived from dissimilarity ratings for 21 synthesized musical instrument tones. Abbreviations for the instruments: BSN, bassoon; CAN, English horn; CNT, clarinet; GTN, guitarnet (hybrid between GTR and CNT); GTR, guitar; HCD, harpsichord; HRN, French horn; HRP, harp; OBC, obochord (hybrid between OBO and HCD); OBO, oboe; OLS, oboleste (hybrid between OBO and celeste); PNO, piano; POB, bowed piano; SNO, striano (hybrid between STG and PNO); SPO, sampled piano; STG, string; TBN, trombone; TPR, trumpar (hybrid between TPT and GTR); TPT, trumpet; VBN, vibrone (hybrid between VBS and TBN); VBS, vibraphone. Dimension I (left-right) represents the Temporal Envelope or attack quality of the sounds (blown-bowed sounds at the right and plucked-struck sounds on the left). Dimension II (front-back) represents the Spectral Envelope of the sounds (brighter sounds at the back). Dimension III (top-bottom) represents Spectral Flux (more spectral flux on the top). [From Krumhansl (1989), Fig. 1, used by permission of Excerpta Medica]

of the different timbres (see Fig. 8.2). Krimphoff aimed to find the acoustic parameters that correlated most strongly to the three dimensions that Krumhansl qualitatively referred to as Temporal Envelope, Spectral Envelope, and Spectral Flux. Thus, two of the three dimensions were expected to correlate with spectral characteristics. Krimphoff found that Dimension 2 (Spectral Envelope) correlated

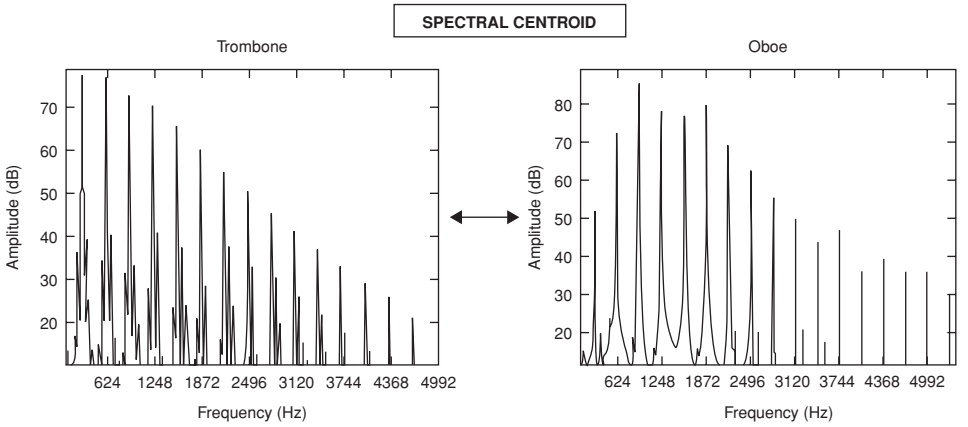


FIGURE 8.3. Spectra of two extreme sounds positioned along the second perceptual dimension of timbre spaces in Figs. 8.1 and 8.2 illustrating the "spectral centroid" parameter. On the left a trombone spectrum has a lower spectral centroid value, and on the right an oboe spectrum has a higher spectral centroid value.

very strongly ($r = 0.94$) with the *spectral centroid* (measured as the time-average of the instantaneous spectral centroid over the duration of the tone. A comparison of spectra with low and high spectral centroids is shown in Fig. 8.3.) However, as discussed further in Section 2.1.3, none of Krimphoff's several measures of spectral variation over time corroborated Krumhansl's suggestion that the third dimension could be interpreted in terms of "spectral flux," a variation of the spectrum over time. Krimphoff's best measure of spectral flux explained only 34% of the variance ($r = 0.59$).

In an attempt to quantify the acoustic nature of Krumhansl's third dimension, Krimphoff proposed two new acoustic parameters related to the spectral envelope. First, he tested an acoustic parameter proposed by Guyot (1992) that measures the ratio between the amplitudes of even and odd harmonics. The clarinet, for example, has a high value for this parameter, because its odd-numbered spectral components have higher energy than its even-numbered ones. On the other hand, the trumpet's value for this parameter is low, because its spectrum is more homogeneous with regard to the amplitudes of the various harmonics. Krimphoff found that the odd/even parameter explained 51% ($r = -0.71$) of the MDS variance for the third dimension. However, a second parameter corresponding to a measure of the *spectral irregularity* of the spectrum (taken as the log of the standard deviation of component amplitudes from a global spectral envelope derived from a running mean of the amplitudes of three adjacent harmonics) yielded a stronger correlation, explaining 73% ($r = -0.85$) of the variance along Krumhansl's third dimension. (A comparison of spectra with low and high spectral irregularity is shown in Fig. 8.4.) Krimphoff's spectral envelope result suggested a new interpretation of the third dimension.

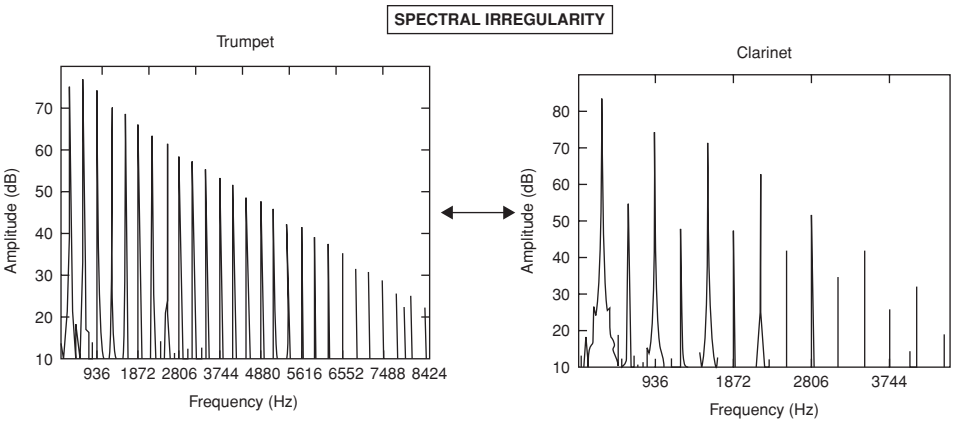


FIGURE 8.4. Spectra of two extreme sounds positioned along the second perceptual dimension of the timbre space in Fig. 8.2 illustrating the "spectral irregularity" parameter. On the left a trumpet spectrum has a lower spectral irregularity value, and on the right a clarinet spectrum has a higher spectral irregularity value (i.e., a more jagged spectral envelope).

One of the aims of a study by McAdams et al. (1995) was to replicate the Krumhansl (1989) study with a large set of listeners having varying degrees of musical training and to check whether any of the acoustic correlates described by Krimphoff (1993) and Krimphoff et al. (1994) could explain the resulting dimensions of the timbre space. Figure 8.5 shows the three-dimensional timbre space produced by McAdams et al. (1995). They correlated several acoustical parameters with derived MDS dimensions for 18 sounds (drawn from the 21 sounds used by Krumhansl and Krimphoff). They found that spectral centroid accounted for 88% of the variance ($r = -0.94$) along Dimension 2 of the figure. However, spectral irregularity did not correlate well with Dimension 3 (only $r = 0.13$), whereas spectral flux gave the highest Dimension 3 correlation ($r = 0.54$).

Grey and Moorer (1977) and Charbonneau (1981) used a different approach, where controlled modifications of acoustical analyses of instrument tones were used as the basis for resynthesis. Grey and Moorer used a computer resynthesis technique based on a heterodyne-filter analysis method to first produce a set of intermediate data for additive synthesis consisting of time-varying amplitude and frequency functions for the set of partials of each tone. Then, from those data they produced synthetic musical instrument stimuli that were used to evaluate the perceptual discriminability of original and resynthesized tones taken from a wide class of orchestral instruments. Sixteen versions of each tone were presented to listeners: (1) original tones; (2) tones resynthesized with line-segment approximations of the amplitude and frequency variations; (3) line-segment approximations with deletion of initial transients; and (4) line-segment approximations with flattening of the frequency variations. Instrument tones from the string, woodwind,

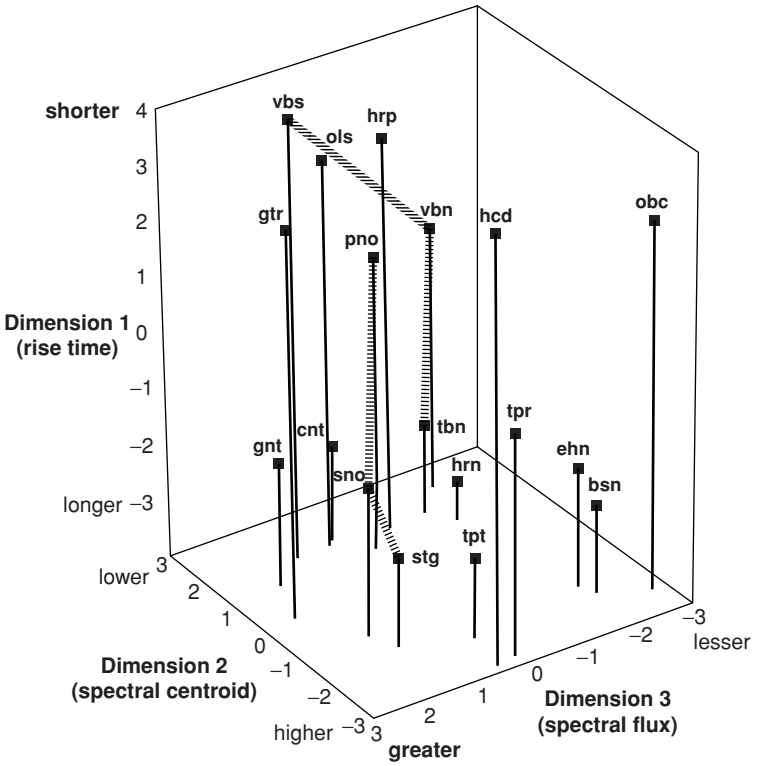


FIGURE 8.5. Three-dimensional CLASCAL solution with specificities and five latent classes derived from dissimilarity ratings on 18 timbres by 88 subjects. The acoustic parameters correlated to the dimensions are indicated in parentheses. Hashed lines connect two of the hybrid timbres (vbn and sno) to their progenitors. [From McAdams et al. (1995), Fig. 1, used by permission of Springer-Verlag.]

and brass families were modified. The pitch, subjective duration, and loudness of these tones were equalized.

Three identical tones and one different tone were presented in an AA-AB vs AB-AA discrimination procedure. Musically trained listeners were asked to discriminate which tone pair was "different" and to rate how different it was on a numerical scale. The data showed that: (1) simplifying the pattern of variation of the amplitudes and frequencies of individual components in a complex sound had an effect on discrimination for some instruments but not for others; (2) tones in which the attack transients were removed were easily discriminated from the originals; and (3) tones in which frequency variations were suppressed were easily discriminated as well. These results suggest that microvariations in frequency and intensity functions are not always essential to timbre and that a reduction of the data can be applied without affecting the perception of some sounds.

Charbonneau (1981) extended Grey and Moorer's study by constructing instrumental sounds that maintained their original global structure, while simplifying the microstructure of the amplitude and frequency envelopes of each harmonic partial. Listeners were asked to evaluate the timbral differences between original sounds and three types of simplifications: (1) replacing the harmonics' amplitude-vs-time envelopes so that each had the same amplitude shape (calculated as the average harmonic-amplitude envelope) but scaled to preserve its original peak value and start- and end-times; (2) replacing the frequency-vs-time envelopes so that each had the same relative frequency variation as the fundamental, meaning that the sound remained perfectly harmonic throughout its duration; and (3) fitting the start- and end-time data to fourth-order polynomials. Results indicated that the amplitude-envelope simplification had the greatest effect. However, as with the Grey and Moorer study, the strength of the effect depended on the instrument. These studies showed that simplifications performed on temporal parameters, and specifically on time-varying functions of amplitude and frequency, influence to a greater or lesser degree the discrimination of musical sounds.

McAdams et al. (1999) attempted to determine the extent to which simplified spectral parameters, without the use of straight-line approximations, affected the perception of synthesized instrumental sounds produced by instruments of various families of resonators (air column, string, or bar) and types of excitation (bowed, blown, or struck). Listeners were asked to discriminate sounds resynthesized with full data from sounds resynthesized with six basic data simplifications: (1) harmonic-amplitude variation smoothing; (2) coherent variation of harmonic-amplitudes over time; (3) spectral-envelope smoothing; (4) coherent harmonic-frequency variation; (5) harmonic-frequency variation smoothing; and (6) harmonic-frequency flattening. (Methods 2 and 4 were similar to Charbonneau's methods 1 and 2.) The results showed very good discrimination for spectral-envelope smoothing and coherent harmonic-amplitude variation, demonstrating, in a negative way, the importance of spectral-envelope detail and spectral flux. However, for coherent harmonic-frequency variation, harmonic-frequency variation smoothing, harmonic-frequency flattening, and harmonic-amplitude variation smoothing, discrimination was moderate to poor in decreasing order.

These techniques appear to be important for the study of timbre perception because they allow modification of the different spectrotemporal parameters of sound in order to reveal which are most important for timbre perception.

2.1.2 Temporal Attributes of Timbre

The classical point of view associates timbre with the spectrum of a sound signal. However, this point of view remains limited because it ignores the importance of temporal factors in timbre. Indeed, instrumental tones physically and perceptually evolve over time. Moreover, the classical conception runs into serious obstacles because musical instruments can be recognized or identified even when their spectra are seriously distorted. This happens in the case of mediocre recordings and when instruments are performed in normal reverberant rooms, where the spectra

of sounds vary a great deal throughout the space. Indeed, when we move about in a room, timbres are not transformed as much as we would expect if they depended exclusively on the precise structure of the source spectra. Nevertheless, spectral factors are undeniably important in timbre, while temporal factors seem to play a role only in certain contexts or for certain instruments.

Let us first examine the extent to which temporal factors are important in the timbre of musical sounds. We often consider musical sounds as composed of three parts: an initial attack portion, a middle sustain portion, and a final decay (the sustain portion being absent, of course, in resonant percussion sounds). The temporal shape of the sound of a piano is an important factor in the definition of its timbre. This is proven by listening to a sound presented in reverse-time. While its long-term average spectrum is identical to that of the original sound, the time-reversed version is often totally unrecognizable (George, 1954; Schaeffer, 1966). In the same way, Berger (1964) showed that suppressing the initial portion of sounds perturbs their recognition. Listeners were asked to discriminate between original musical instrument tones and modified versions with either their initial portions (attacks) or their final ones (decays) removed. Identification was poorest for sounds without attack. Also, Saldanha and Corso (1964) evaluated the relative importance of onset and offset transients, spectral envelope of the sustain portion, and vibrato for identifying musical instruments playing isolated notes. Identification was particularly affected when the attack portions were removed. However, identification was affected less if the instruments were performed with vibrato than if they were performed without vibrato. These results suggest that the attack plays a major role in the identification of instruments, but in the absence of the attack, additional information still exists in the sustain portion (McAdams, 1993). The studies of Grey and Moorer (1977) and Charbonneau (1981) described above also demonstrated the importance of such temporal factors. For example, tones with the attack transients removed were easily discriminated from originals in their studies.

As part of a multidimensional analysis, Samson et al. (1996) produced a two-dimensional space in which each dimension corresponded to temporal factors. The authors observed that the duration of the attack (1, 100, or 190 ms) correlated strongly with one of the perceptual dimensions. Grey (1977) and Wessel (1979) also observed a dimension of this nature. Wessel (1979) determined that the second dimension of his perceptual space corresponded to "attack rapidity." Grey (1977) interpreted two of his three dimensions to be related to attack features, the second of which corresponded to the "presence of inharmonic transients in the high frequencies just before the onset of the main harmonic portion of the tone." Strings, flutes, and clarinet, for example, have low-amplitude, high-frequency energy near their tone onsets, contrary to those of the bassoon or the English horn. Krimphoff (1993) and Krimphoff et al. (1994) confirmed this finding in their interpretation of the "Temporal Envelope" dimension of Krumhansl's (1989) space (see Fig. 8.2). The positions of timbres along this axis were strongly correlated ($r = 0.94$) with the logarithm of the rise time of the temporal envelope (where rise time was measured as the difference between the time at which the amplitude reaches a threshold of 2% of the maximum amplitude to the time it attains maximum

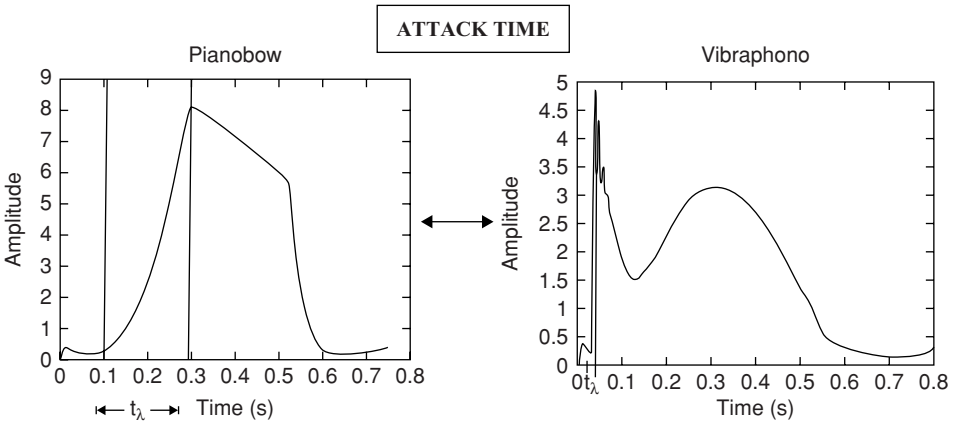


FIGURE 8.6. Temporal envelope of two extreme sounds positioned along the first perceptual dimension of the timbre spaces shown in Figs. 8.2 and 8.5 illustrating the "log-attack time" parameter. On the left, the pianbow has a long attack time (about 190 ms) similar to those for wind and bowed string instruments. On the right, the vibraphone has a short attack time (about 4 ms) similar to those for the set of struck and plucked instruments.

amplitude). The first dimension of McAdams et al.'s (1995) timbre space (Fig. 8.5) is also strongly correlated to this acoustical parameter with 88% of the variance ($r = -0.94$) explained by it. Examples of the measurement of rise time are shown in Fig. 8.6.

2.1.3 Spectrotemporal Attributes of Timbre

Multidimensional scaling in timbre studies has often revealed three perceptual dimensions. While two of these are often easily characterized by acoustical parameters, the third one remains poorly defined. This lack of satisfactory interpretation is probably due to the variability in stimulus sets or listener characteristics across studies. Not all studies have found a valid third dimension (e.g., Wessel, 1979), and those that have interpreted this perceptual axis differently from one study to the next. Some authors have proposed that this dimension corresponds to a spectral factor other than spectral centroid (Krimphoff et al., 1994; McAdams et al., 1995), and others have proposed that it corresponds to a temporal variation in the spectral envelope (Grey, 1977; Krumhansl, 1989) (see Figs. 8.1, 8.2, and 8.5).

Up to this point, this chapter has presented the influence of temporal and spectral factors, considered independently, on timbre. Nevertheless, these factors are not generally independent, and their association may also play a role in musical timbre. Risset and Mathews (1969) notably observed that synthesized trumpet sounds with static spectra and a common amplitude-vs-time envelope, applied synchronously to all frequency components, did not give a satisfactory perceptual result. They demonstrated the necessity of taking into account the variations of the different spectral components over time for certain timbres. Grey (1977) also suggested

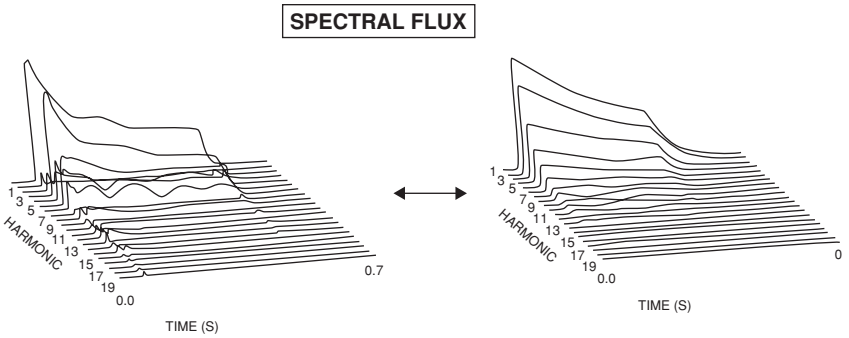


FIGURE 8.7. Time-frequency perspective plots illustrating the "spectral flux" parameter for two extreme sounds positioned along the third perceptual dimension of the timbre space of Fig. 8.2. On the left, the obochord (hybrid between the oboe and the harpsichord) has a high spectral flux value and on the right, the striano (hybrid between a string and a piano) has a low value.

that the physical nature of one of the perceptual dimensions of timbre could be a spectrotemporal factor. The interpretation of the second dimension of his solution was a combination of the degree of fluctuation in the spectral envelope over the duration of a tone and the synchrony of onset of its different harmonics. The woodwinds were at one extreme and tended to have upper harmonics that reached their maximum during the attack but were often in close alignment during the decay. Also, their spectra tended to have little fluctuation over time contrary to the strings or brass situated at the other extreme of this axis.

As mentioned in Section 2.1.1, Krumhansl (1989) named the first two dimensions obtained in her MDS study Temporal Envelope and Spectral Envelope, and the third dimension was called Spectral Flux, because the distribution of timbres along this dimension was presumed to correspond to the degree of spectral variation over time. (Time-variant spectra with high and low spectral variation are compared in Fig. 8.7). This interpretation agreed for the most part with the one proposed by Grey (1977) for simplified, resynthesized instrument sounds. The psychophysical interpretation proposed by Krimphoff (1993) and Krimphoff et al. (1994) for the first two dimensions agreed with the qualitative interpretation of Krumhansl, as previously discussed.

For analysis of Krumhansl's third dimension, Krimphoff (1993) tested three acoustical parameters that quantified spectral fluctuation over the duration of a sound. These parameters were: (1) "spectral flux," defined, in this case, as the rms variation of instantaneous spectral centroid around the mean spectral centroid; (2) "spectral variation," defined as the average of correlations between amplitude spectra in adjacent time windows (note that the smaller the degree of variation of the spectrum over time, the higher the correlation); and (3) "coherence," defined as the standard error of the onset times across all harmonics. Correlations observed between these three parameters and the third dimension of Krumhansl's (1989)

solution were not significant, except for spectral flux, which only explained 34% ($r = 0.59$) of the variance along this dimension. Krimphoff (1993) and Krimphoff et al. (1994) found that spectral irregularity, a spectral rather than spectrotemporal parameter, best explained Krumhansl's third dimension. On the contrary, spectral irregularity was not best correlated to the third dimension of the McAdams et al.'s (1995) timbre space, which used 18 of the same 21 sounds in Krumhansl's study. Indeed, spectral variation was the only acoustical parameter in McAdams et al.'s (1995) study that significantly correlated with the third dimension, even though it accounted for only 29% ($r = 0.54$) of the variance along this dimension. When four of the timbres (clarinet, trombone, guitarnet, and vibrone) were removed, the variance increased to 39%, and their removal did not affect the correlations of attack time and spectral centroid with Dimensions 1 and 2.

2.2 The Notion of Specificities

The degree of variability in similarity data from the early scaling studies on timbre leads us to think that two or three common dimensions are not enough to describe the perception of timbre. Moreover, one may question the validity of the assumption that two or three dimensions can explain all the differences among extremely complex sounds like musical instrument tones. To take into account this complexity, some authors suggest that each timbre may also be defined by unique characteristics (Krumhansl, 1989; McAdams, 1993). On the other hand, it will be important to take these specificities into account in the modeling of the mental structure of timbre because they might play a major role in the identification of musical instruments. For example, when the spectral envelope is unique (e.g., clarinet vs trumpet), it seems to contribute more to identification than when the temporal envelope is distinguished (e.g., flute vs trombone). This suggests that listeners use characteristics that specify the instrument with the least ambiguity and that they are not constrained to listening for a single cue across all possible sources (Strong and Clark, 1967a,b). For example, in a study on string instruments, Mathews et al. (1965) found an initial inharmonic frequency component corresponding to the irregular vibration that appears when the bow first sets the string into vibration. Such details can be characteristic of particular sound sources, and the auditory system seems to be sensitive to these identifying details.

Timbre may thus be defined by not only two or three common, continuous dimensions but also by distinguishing features or dimensions that are specific to a given sound. To test this notion, Krumhansl (1989) applied an extended Euclidean model developed by Winsberg and Carroll (1989). By postulating the existence of unique features for certain timbres, this model was designed to provide an explanation of the variability in similarity judgments that could not be attributed to the three principal MDS dimensions derived from dissimilarity judgments based on 21 synthesized imitations and hybrids of conventional Western musical instruments. Globally, 60% of the timbres yielded non-zero specificity values. Specific examples are the harpsichord, the clarinet, and some of the hybrid timbres such as the "pianobox" (bowed piano), the "guitarnet" (guitar/clarinet hybrid), and the

"vibrone" (vibraphone/trombone hybrid) that yielded high values of specificity. While no attempt was made to interpret these specificity values by systematically relating them to acoustic properties, Krumhansl conjectured that the specificities of certain instruments reflected specific mechanical characteristics that could be important for their identification. For example, the return of the jack in the harpsichord mechanism or the cylindrical geometry of the air column of clarinet could have important perceptual ramifications.

McAdams et al. (1995) attempted to find a qualitative interpretation of the specificities captured by their model on the same set of sounds. First, the authors noted a monotonic relationship between the specificity values and the perceptual strength of the specificities. However, this relationship was not tested systematically. Second, the authors distinguished: (1) continuous features that varied by degree (such as "raspiness" of attack, inharmonicity, "graininess" deviation of pitch glide, and "hollowness" of tone color); and (2) discrete features that varied by perceptual strength (such as a high-frequency chiff on the onset, a suddenly damped or pinched offset, or the presence of a clunk or thud during the sound). The authors concluded that such specificities may account for both additional continuous dimensions and discrete features of variable perceptual salience.

Another hypothesis was that specificities may reflect unfamiliarity of sounds to listeners, and, therefore, hybrid timbres should yield a high value of specificity. However, on average, in the two models (Krumhansl, 1989; McAdams et al., 1995), hybrid timbres did not yield higher specificities than those of conventional instruments. Actually, half of the hybrid timbres tested yielded lower specificities than the average value. Moreover, this hypothetical relationship between specificity and familiarity was not supported by the (very familiar) piano timbre, which yielded a high value in both studies. In fact, the piano is probably one of the most familiar instruments to the primarily European listeners who participated in these studies.

To conclude, these results suggest that structural sound characteristics influence dissimilarity judgments made by subjects. These characteristics may be common to all the timbres within a stimulus set or specific to some timbres. A classical Euclidean model could not take these specific features into account and an extended model is, therefore, more appropriate. Acoustical analyses must still be conducted in order to give a psychoacoustical interpretation of the specificities that were found.

2.3 Individual and Group Listener Differences

Most of the timbre spaces described above were derived exclusively from musician listeners (Grey, 1977; Wessel, 1979; Krumhansl, 1989). A few studies have tried to determine whether perceptual differences between auditory classes correspond to biographical factors, such as the level of musical training or cultural origin, but they have found no systematic differences related to musical training (Miller and Carterette, 1975; Wedin and Goude, 1972). However, whereas most of us, musician or not, can distinguish a guitar from a clarinet, we might suppose that

the mental structure of the perceptual relations among different timbres would not be the same depending on the musical competence of the listener.

Musical competence potential differences might be found by analysis of weight patterns attributed to the different dimensions and specificities of a common space. The weights' interpretation could be based on biographical factors such as musical experience. The INDSCAL (INdividual Differences SCALing) model, proposed by Carroll and Chang (1970), can account for such individual perceptual differences. Serafini (1993) used individual-differences scaling to test two groups of Western musician listeners on a set of Javanese percussion sounds (xylophones, gongs, and metalophones) and a plucked-string sound. One group was familiar with Indonesian gamelan music (they had played Javanese Gamelan music for at least two years), and the other was unfamiliar with this type of music. The task was to judge the dissimilarity between pairs of isolated notes and pairs of melodies played by these instruments. Stimulus and subject INDSCAL two-dimensional solutions yielded one dimension (Dimension 1) corresponding to the spectral centroid of the attack portion of tones and a second dimension (Dimension 2) to the mean amplitude level of the resonant portion of the tone (a dimension related to loudness). For isolated tones (see Figs. 8.8a and 8.8b), no differences were found between the two groups of listeners. However, for melodies, the group unfamiliar with gamelan music gave equal weight to the two dimensions, whereas gamelan players weighted the attack dimension more heavily (see Figs. 8.8c and 8.8d).

McAdams et al. (1995) conducted a study on a large number of listeners of varying levels of musical training with an analysis of *latent-class structure*. The aim was to examine whether listeners could be sorted into different classes according to their perceptual data and whether a relation between the class structure and musical training of the listeners could be found. For musical pitch, Shepard (1982) had observed a dimensional structure that was different for musicians and non-musicians. The structure was richer, i.e., had higher dimensionality, for musicians than for non-musicians. This result led McAdams et al. (1995) to hypothesize that the same type of result could be observed for timbre perception: either the number of dimensions would be greater for the musicians' dimensional structure, or the weights on the dimensions would be more evenly distributed. However, in fact, musicians, amateurs, and non-musicians did not fall into separate latent classes even if some differences were observed in the proportional distribution of biographical factors. The analysis of the different weights across dimensions and specificities showed that two of the five classes observed among 98 listeners gave roughly equal weights across dimensions and specificities, while the other classes gave high weights on two dimensions, or on one dimension and the specificities, and low weights on the others, respectively (see Fig. 8.9).

Two different interpretations of this weight pattern were proposed by the authors. First, the weight pattern observed could reflect a strategy difference between subjects over the course of the experimental session. Equal weights across the three dimensions and specificities observed for two of the five classes could be due to the subjects in these classes shifting their attention among the different dimensions and specificities, while the subjects in the other classes may have adopted more

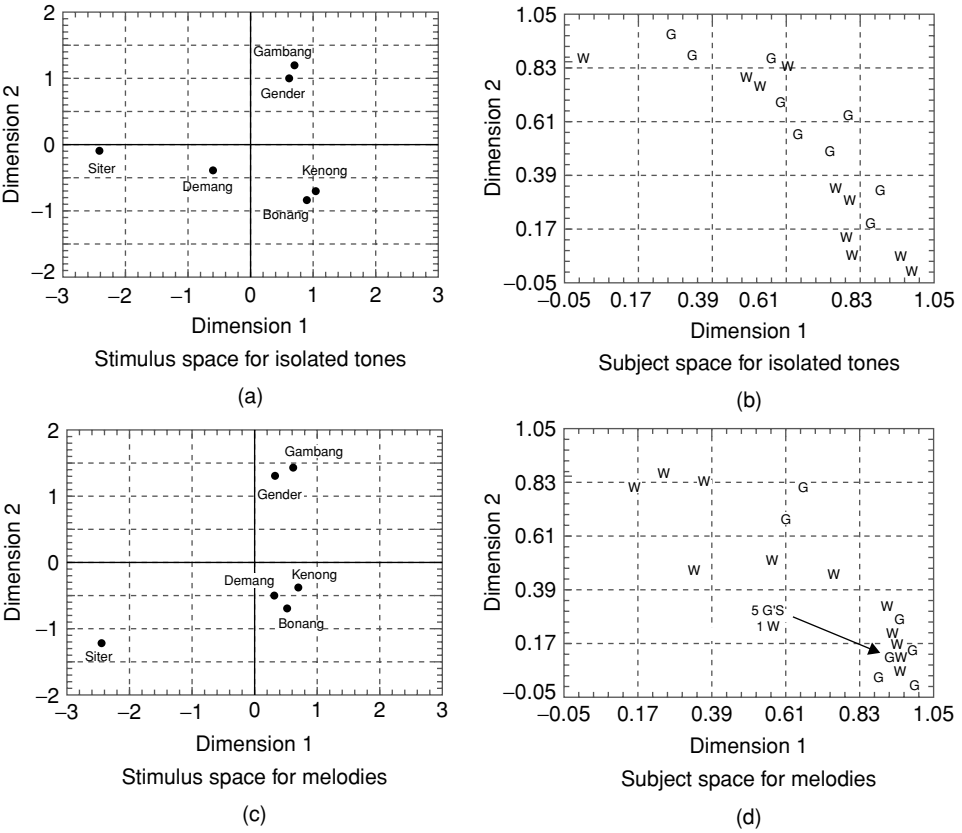


FIGURE 8.8. (a) Two-dimensional stimulus space derived from an INDSCAL analysis on similarity ratings on six isolated gamelan sounds. (b) Subject space observed for the six isolated gamelan sounds. (c) Two-dimensional stimulus space derived from an INDSCAL analysis on similarity ratings on six melodies played by six gamelan sounds. (d) Subject space observed for the six melodies. ("G" refers to listeners familiar with Indonesian gamelan music.) "W" refers to "Western" listeners unfamiliar with gamelan music. [From Serafini (1993), adapted with permission of Waterloo University.]

consistent strategies of judgment that focused on a smaller number of dimensions and stuck to them throughout the experimental session. The second interpretation suggested a difference between subjects in different classes in their cognitive capacity to process different aspects of sounds in parallel. According to this interpretation, subjects in the two classes who equally weighted the dimensions and specificities were able to focus on more dimensions at a time than could members of the other classes, and one might predict *a priori* that these would be principally musicians. However, the authors observed that both musicians and non-musicians were able either to equally weight all dimensions or to give special attention to some dimensions like the attack time or the spectral centroid. Thus, the distribution of

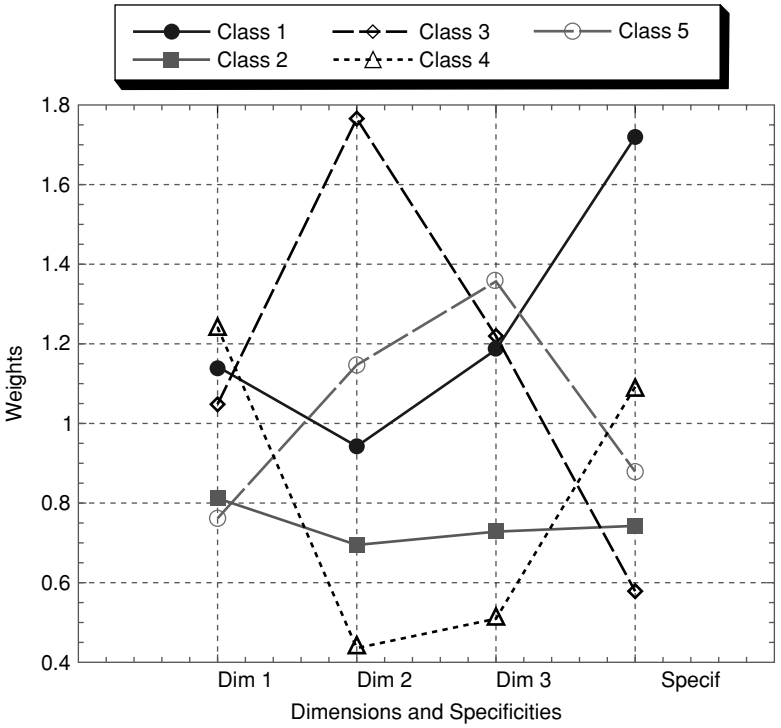


FIGURE 8.9. Class weights (mean weights across dimensions and specificities for each class) for spatial model plotted for each of five classes. Weights were estimated in a three-dimensional space for five latent classes. [Derived from McAdams et al. (1995), Table 4 used by permission of the author.]

the three original classes of listeners (musician, amateur, and non-musician) within each latent class was roughly equivalent to their distribution in the whole subject population employed. The pattern of weighting of a given subject cannot be predicted simply from the biographical data related to that subject concerning their degree of musicianship or their years of music training, performing, or listening.

The only differences observed between musicians, amateurs, and non-musicians were the variances about the model distances, observed for the solutions computed separately for musician and non-musician groups: The variance for non-musicians and amateurs combined was greater than that for musicians. However, the variances observed for individual latent classes, composed of musicians, amateurs, and non-musicians, were less than the variance of the musician group, suggesting that the inclusion of class weights in the dimensional models is justified in terms of model fit because it reduces the overall variance. This pattern of results suggests that the effect of musicianship is, among other things, one of variance. Latent classes do not differ with respect to variance, but musicians and non-musicians do. So musicianship appears to affect judgment precision and coherence.

2.4 *Evaluating the Predictive Power of Timbre Spaces*

In some studies that attempted to evaluate timbre space as a predictive model, the explicit aim was to determine the validity of the model. In others, the idea was to see if timbre space could be used to test other hypotheses. Four types of research will be discussed that support the validity and utility of such models.

2.4.1 Perceptual Effects of Sound Modifications

An assumption of the timbre space model is that specific acoustic properties underlie the continuous perceptual dimensions. If we modify the acoustic properties for a single perceptual dimension in a systematic way, we should observe perceptually interpretable changes of the positions of stimuli along that dimension. A study conducted by Grey and Gordon (1978) confirmed this assumption. They exchanged the spectral envelopes of pairs of instruments drawn from the Grey (1975, 1977) study, while trying to preserve other properties, and conducted a new multidimensional study with half of the original sounds modified and the other half intact. The hypothesis was that the positions of the original and hybrid sounds should change along the dimension that best correlated with a measure of the spectral envelope. The results demonstrated that in all cases the tones exchanged places along the "brightness" (or spectral-centroid) dimension, although in some cases displacements along other dimensions also occurred. These displacements still respected the nature of the perceptual dimensions: Temporal-envelope changes resulting from the way the spectral envelope varied with time resulted in appropriate changes along the dimension that best correlated with spectral flux.

On the other hand, the most natural way to move in a timbre space would be to attach the handles of control directly to the different dimensions of the space. Wessel and colleagues (1979, 1983, 1987) examined such a control scheme in a real-time context. A two-dimensional timbre space was represented on a computer graphics terminal allowing control of a digital processor. One dimension of the space was used to manipulate the shape of the spectral-energy distribution. This was accomplished by appropriately scaling line-segment spectral envelopes according to a shaping function. The other axis of the space was used to control either the attack rate or the extent of synchronicity among the various components. Overall, the timbral trajectories in these spaces were reported by the author to be smooth and otherwise perceptually well-behaved.

All of these results and observations suggest that some intermediate regions of the timbre space could be filled in and that regular, finely graded transitions are conceivable, thus supporting the hypothesis that timbre perception can be modeled by continuous physical dimensions that underlie a small number of perceptual dimensions.

2.4.2 Perception of Timbral Intervals

Classical musical structures are based on the separation and the grouping of sound events according their relative differences in pitch (melody), intensity (dynamics),

duration (rhythm), and timbre (instrument). Research on timbre tries to expand this conception of the organization of musical sequences. Indeed, transposing timbral sequences may be heard by listeners and used consciously by composers. The aim of the following studies was to test the idea of the composer Arnold Schoenberg (1911) that musical phrases can be formed by notes which differ only in timbre. Once a timbre 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 different timbres.

Ehresman and Wessel (1978) were among the first to apply Rumelhart and Abrahamson's (1973) parallelogram model of analogical reasoning with a two-dimensional space composed of traditional musical instrument sounds (Grey, 1977). This model predicts that if the relation between two objects A and B is represented as the vector A-B in the space, another vector C-D will be perceived as analogous if it has the same magnitude and orientation as A-B. In the analogy task, vector A-B is presented and a series of vectors C-D_i are presented. According to the model, the subjects will choose the D_i that is closest to the end point of a vector starting at C and having the same magnitude and direction as A-B. This ideal point is called I and the vectors A-B and C-I thus form a parallelogram in the space. Analogies of the form A, B, C (D_1, D_2, D_3, D_4), where D_i was varied according to its distance from I, were constructed. 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 and Wessel proceeded in analogous fashion with musical instrument tones. The two perceptual dimensions of their space corresponded to (1) "spectral energy distribution" of the tones and (2) "nature of the onset transients." The results were better predicted by this model than a number of other models. In addition, timbral vectors were computed from a two-dimensional solution and only relative vector magnitude (corresponding to the estimated perceived dissimilarity) was tested, ignoring the direction components.

McAdams and Cunibile (1992) tested a similar geometric model for the three-dimensional space observed by Krumhansl (1989) taking into account separately the magnitude and orientation of the different timbral vectors. Sequences of four timbres of the perceptual space (five different sets for each experimental condition) were constructed according to four experimental conditions differing in the degree to which they corresponded to the "good" analogy defined by the model: (1) good magnitude, good orientation; (2) good magnitude, bad orientation; (3) bad magnitude, good orientation; and (4) bad magnitude, bad orientation (see Fig. 8.10). Two sequences of four timbres, where only the last varied between the two sequences, were presented to listeners (musicians and non-musicians). The task was to choose the sequence that best corresponded to an analogy of the form: timbre A is to timbre B as timbre C is to timbre D. The hypotheses were: (1) sequences in which the A-B and C-D vectors formed a parallelogram would be preferred; (2) sequences in which the C-D vector had a good magnitude but a bad orientation would be preferred over those with bad magnitude and orientation; (3) sequences in which the C-D vector had a bad magnitude but a good orientation would be

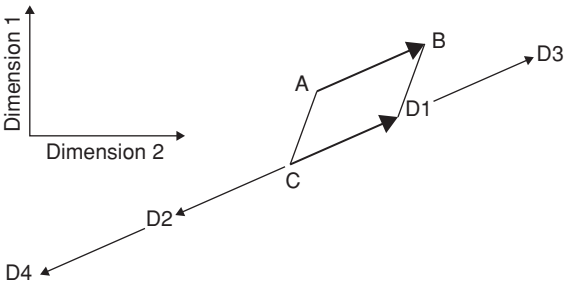


FIGURE 8.10. Parallelogram model of timbre analogies. (The two-dimensional case is shown.) A to B is a given change in timbre; C to D is a desired timbral analogy, with C given. D1, D2, D3, D4 are the different analogies offered to the listeners with D1 corresponding to the ideal point according to the model.

preferred over those with bad magnitude and orientation; and (4) there would be no differences among the different versions of each comparison type because the analogy judgment is based on a perception of abstract relations among the timbres of the stimulus tones. The results showed that: (1) the listeners preferred sequences with good magnitude and orientation; (2) sequences with either good magnitude and bad orientation or bad magnitude and good orientation were preferred significantly more often than those with bad magnitude and bad orientation; (3) however, the judgments for the different versions of each comparison differed significantly from one another. According to the authors, these latter differences may have been due to the presence of specificities that were not taken into account in computing the vectors in this experiment. Indeed, if we consider that certain timbres had specificities, this would distort the vector established on the basis of the common dimensions alone.

Overall the results were encouraging, indicating an ability to perceive timbral analogies on the basis of a timbre space describing the dissimilarity relations among different timbres. However, these studies lacked control for specificities, which can influence the dissimilarity between timbres and thus the distances separating them in the space. They also needed to control for the positions of vector pairs in the perceptual space by using a synthetic space in which the timbres are distributed in a homogeneous fashion.

2.4.3 The Role of Timbre in Auditory Streaming

We know now that many aspects of sound are important in auditory streaming: intensity (Van Noorden, 1975), fundamental frequency (Bregman, 1990; Bregman et al., 1990; Miller and Heise, 1950; Singh, 1987; Van Noorden, 1975), spectral factors (Hartmann and Johnson, 1991; McAdams and Bregman, 1979), and temporal factors (Hartmann and Johnson, 1991). Even if the majority of reseachers consider timbre to be an attribute of sound processed after auditory grouping, it

seems that the spectral, temporal, or spectrotemporal properties giving rise to timbral attributes may also contribute to auditory stream segregation. A hypothesis may be made that sequential groupings of complex sounds are based on the spectral or temporal similarity of the sounds. In these cases, the auditory system would organize sound events in the same stream when they are sufficiently similar.

Several researchers (McAdams and Bregman, 1979; Wessel, 1979; Iverson, 1993; Gregory, 1994; Bey and McAdams, 2003) have studied streaming by musical timbre. Wessel (1979) conducted an early demonstration of streaming employing 16 synthetic instrument tones. In a previous experiment he had subjects rate the similarity of these tones and used MDS to fit the judgments to a two-dimensional space with one dimension corresponding to spectra and the other to onset transients. To test the relationship between similarity judgments and streaming, he constructed repeated sequences of three ascending notes with alternate notes differing in timbre, but otherwise the pitch sequence and rhythmic timing remained fixed. When the timbral distance between the adjacent notes was small along the spectral dimension, a repeating, ascending pitch line was heard. However, when the timbral distance was enlarged along this same dimension, listeners heard two streams with one stream comprised of timbre A and the other of timbre B. This phenomenon is called "melodic fission" or "auditory stream segregation." On the other hand, a different effect was obtained when the note timings were modified. In this case, a single stream with perceptually irregular rhythm was perceived regardless of the timbral distance separating the different notes. This result suggested that the spectral dimension influenced auditory streaming but the temporal dimension did not.

Iverson (1993) also conducted a series of experiments to test the relationship between similarity judgments and auditory streaming. In a previous study, the author examined 16 tones using a standard similarity-scaling technique and found a two-dimensional MDS space where the 16 tones were represented. The second experiment assessed the relationship between similarity judgments and streaming. Pairs of sequences, constructed with the same 16 tones used in the first experiment, were presented to listeners. The task was to rate the degree of streaming of each sequence on a continuous scale, resulting in a triangular matrix giving the relative streaming of each pair of tones. The streaming ratings were used as a similarity metric for MDS, so tones that formed one stream were closer in the space than tones that formed two streams. A two-dimensional space was obtained similar to those observed with the similarity ratings on single tones. The first dimension corresponded to attack quality and the second to the perceived brightness of the sounds. Acoustical attributes were identified and correlated with the judgments. Iverson showed that sounds with similar spectral or temporal envelopes were integrated into one stream and sounds with different spectral or temporal envelopes were segregated into different streams. This result showed the importance of temporal factors in auditory streaming, contrary to the results observed by Wessel (1979) and Hartmann and Johnson (1991).

Gregory (1994) tested the influence of each perceptual dimension of timbre in auditory streaming. The three dimensions of his MDS space were "relative

percentage of energy in the first three partials," "decay duration," and "relative strengths of odd to even partials." Listeners were tested to determine their abilities to separate streams according to the perceptual distances of timbres observed in the timbre space. When the timbral difference was increased, auditory streaming was not based on the pitch difference but on timbral difference. Moreover, the temporal dimension seemed more important than the two other spectral dimensions in auditory streaming.

A study conducted by Bey and McAdams (2003) confirmed the role of temporal factors. The subjects' task corresponded to a recognition of interleaved melodies. Sequences were composed of two melodies with timbres that were more or less distant in Krumhansl's (1989) perceptual space. Results showed that differences along the spectral and spectrotemporal dimensions were not sufficient to separate the two melodies, and recognition of the embedded melodies was thus not possible. However, if sounds also varied on the temporal dimension, listeners could separate the two melodies, and recognition performance was improved. Furthermore, the authors showed that a timbre difference combined with a pitch difference led listeners to separate the two melodies even more than if only a timbre or pitch difference distinguished them.

The studies conducted by Gregory (1994), Iverson (1993), and Bey and McAdams (2003) illustrate the contribution of temporal factors to listeners' abilities to separate sound streams and counter the idea that only spectral factors are significant in auditory streaming (Bregman, 1990; Bregman et al., 1990; McAdams and Bregman, 1979; Miller and Heise, 1950; Singh, 1987; Van Noorden, 1975; Wessel, 1979).

2.4.4 Context Effects

While spectral factors seem to systematically influence timbre perception, depending on context, temporal factors are not always as salient. Wedin and Goude (1972) observed that the presence or absence of attack transients did not influence the perceptual representation of a set of musical timbres. The mean dissimilarity of the two tested conditions was highly correlated ($r = 0.92$).

Miller and Carterette (1975) attempted to demonstrate the perceptual importance of temporal parameters of timbre. Their stimuli were synthetic tones with variable harmonic spectra, variable amplitude-vs-time envelope, and variable onset delays for the harmonics (temporal properties). They obtained a three-dimensional MDS solution that accounted only for harmonic structure (in dimensions 1 and 2) and the amplitude-vs-time envelopes (in dimension 3), so that the contribution of harmonic onset delay pattern on timbre perception remained in doubt. However, their temporal properties were indeed organized and combined along the third dimension.

According to Iverson and Krumhansl (1991), when sounds are isolated, the attack seems essential to their recognition but does not seem to be the determining factor in similarity judgments. Iverson and Krumhansl (1993) confirmed these results showing that attributes on which listeners based their dissimilarity judgments

among different timbres were present in the duration of the sound. Indeed, they observed similar multidimensional spaces for sounds in which only the first 80 ms were presented, sounds where only the first 80 ms were removed, as well as original sounds.

Many studies have shown that temporal aspects of sounds are perceptually less pertinent when situated in a musical context. Grey (1978) used simplified sounds of three instruments: bassoon, trumpet, and clarinet. He first created notes of different pitches by transposing each instrument's spectrum to higher or lower frequencies. He then asked listeners to distinguish simplifications applied for isolated instrument sounds or for the same sounds placed in different musical configurations, differing in the number of simultaneous melodic lines, rhythm variety, and temporal density. The musical context effect was measured by noting the difference in discrimination ability for the various conditions. While for the bassoon no effect of musical context was observed on discrimination between the original and modified versions, discrimination performance was found to decrease with musical context for the clarinet and trumpet. An acoustical analysis of the original and modified bassoon sounds showed that the simplification involved changes in the spectral envelope, which was not the case for the other instruments. For the bassoon, the changes were described by listeners as brightness differences, which corresponded to spectral envelope changes. On the other hand, changes described for the trumpet and clarinet were located in the "attack" or in the articulation. Small spectral differences were thus slightly enhanced in single-voice contexts compared with isolated tones and multivoiced contexts, although discrimination remained high. Articulation differences, on the other hand, were increasingly disregarded as the complexity and density of the context increased.

Similarly, Kendall (1986) conducted an experiment in which tone modifications were made by time-domain editing. Two different note sequences were presented to listeners whose task was to decide which instrument in the second sequence corresponded to the instrument sounded in the first sequence. The first sequence was an edited version of the same melody played by one of the three instruments used: clarinet, trumpet, or violin. The second sequence consisted of the melody played in unedited form in random order by each of the three instruments under the following conditions: (1) normal tones, (2) sustain portion only (cut attacks and decays), or (3) transients only (with either a silent gap in the sustain portion or an artificially stabilized sustain portion). The results suggested that transients in isolated notes enhance instrument recognition when they were alone or coupled with a natural (time-varying) sustain portion but were of little value when coupled with a static sustain part. They were also of less value in continuous musical phrases where the information present in the sustain portion (probably related to the spectral envelope) was more important. This conclusion confirmed Grey's (1978) discrimination study and was verified by McAdams's (1993) study, which utilized stimuli with more realistic variations.

In comparison to studies on isolated sounds (Berger, 1964; Charbonneau, 1981; Grey and Moorer, 1977; Saldanha and Corso, 1964), these results suggest that

attack transients play a less important perceptual role for musical phrases than they do for isolated tones.

2.5 Verbal Attributes of Timbre

2.5.1 Semantic Differential Analyses

One approach to the study of timbre perception of complex sounds is the analysis of verbal attributes used to describe them. Some authors (Lichte, 1941; Solomon, 1959; Terhardt, 1974; Vogel, 1974; von Bismark, 1974) have hypothesized that timbre can be described by semantic scales. For example, scales can be presented to listeners in which the extremities are two opposing verbal attributes such as "smooth–rough" or "light–dark." They are asked to rate each timbre on each scale. A factor analysis is used to identify a number of factors or scales contributing to explaining variance in the judgments. The remaining scales are considered to describe the different timbres used.

Semantic studies began with Lichte (1941) study of the "bright/dull" and "thin/full" scales using synthetic harmonic tones. Solomon (1959) investigated seven timbral attributes of sonar recordings and the contribution of each spectral region made to each attribute. Terhardt (1974) and Vogel (1974) both examined the notion of "roughness" for the steady-state portion of synthetic sounds.

One of the most complete psychophysical timbre studies was performed by von Bismark (1974) in which subjects had to rate 35 speech sounds (having equal loudness but different spectral envelopes) on 30 verbal scales such as "brilliant-dull" or "wide-narrow." A factor analysis showed that four orthogonal factors were sufficient to account for 90% of the variance. Timbre would have, according to this study, four dimensions: (1) thick/thin; (2) compact/diffuse; (3) colorful/colorless; and (4) full/empty. A major problem with this type of study is that the choice of the verbal attributes characterizing the scales does not always correspond to scales that subjects would choose spontaneously. A timbral dimension correlating with a specific acoustic parameter such as spectral fine structure cannot be revealed by such a study. Moreover, the meaning of certain terms is likely to vary according to the musical culture of the subject.

2.5.2 Relations between Verbal and Perceptual Attributes or Analyses of Verbal Protocols

To eliminate some problems posed by semantic differential studies, Faure et al. (1996) and Faure (2000) used subjects' free verbalizations analyzed by a paradigm developed by Samoylenko et al. (1996). Free verbalization does not impose a vocabulary on the listener. The aim was to define the verbal correlates of the different perceptual dimensions of timbre. The listeners (musicians, non-musicians, and amateurs) were asked to judge the degree of dissimilarity of pairs of timbres [a subset of Krumhansl's sounds (1989)] and then to describe all the dissimilarities and similarities between the timbres. The listeners could modify their dissimilarity judgment after their verbalization.

Two different multidimensional analyses were performed on the ratings given before and after the verbalization. The two resulting timbre spaces were similar suggesting that the verbalization process did not affect the mental structure of timbre. This result allowed a comparison of the dissimilarity judgments to the verbalization.

To find verbal correlates, 22 descriptors were extracted from expressions of the form: "sound 1 is more (or less) X than sound 2". These descriptors were /high/-/sharp/-/shrill/, /low/-/deep/, /long/, /clean/-/distinct/, /mussed/-/dull/, /round/, /clear/-/light/, /resonant/, /nasal/, /metallic/, /vibrated/, /strong/-/loud/, /dry/, /soft/, /rich/, /high/, /low/, /wide/, /diffuse/, /brilliant/-/bright/, /plucked/ and /blown/. Some descriptors were correlated with one MDS dimension while others were correlated with more than one dimension.

Coefficients from multiple regressions were used to project verbal vectors in the multidimensional timbre space. If a descriptor's vector was correlated to only one dimension, it was aligned along the axis of this dimension. If a descriptor was partially correlated to two different dimensions, the vector formed an angle with the two dimension axes, the slope reflecting the ratio between the regression coefficients. Only a few descriptors were correlated to only one dimension. These descriptors were /dry/—correlated with the log of the attack time dimension; /round/—correlated to the spectral centroid; and /brilliant/-/bright/—correlated to the spectral flux. The other descriptors were correlated to more than one dimension: /metallic/, for example, was correlated with three perceptual dimensions. The authors explained these multiple correlations by the fact that sounds characterized as /metallic/ generally have a fast attack, a resonance with much energy in the high frequencies, and a spectral evolution that reflects more rapid damping of high frequencies. On the other hand, the descriptor /mussed/-/dull/ was very often the antonym of /metallic/. Indeed, its vector formed a 180° angle with that descriptor. This result suggests that Faure's approach may be very useful for research to determine verbal antonyms and synonyms describing the timbre of complex sounds.

3 Categories of Timbre

A different view of perceptual activity will now be presented. According to this view, when we are subjected to multiple physical stimulation coming from the environment, we experience multiple sensations. In order to behave coherently when faced with this environment, we need to classify the stimuli. How is this done and what is the structure of our mental representation when this classification is accomplished? Numerous authors have proposed the existence of categorization processes which could be at the origin of a categorical structure of the perceptual representation of most stimuli. An example is the categorical perceptual phenomenon of speech, in which the capacity to discriminate differences between speech sounds is determined by the capacity to differently categorize these kinds of sounds. In this case, we seem to transform initial continuous information into a discrete form. Some authors postulate that the conversion from a continuous variation of a stimulus to a discrete form is based on a late stage of the recognition

process, while others postulate that this conversion occurs during low-level stages of the perceptual process.

Besides spatial models used to determine the mental structure of the timbre of complex tones presented in the last section, there are non-spatial models in which each object is described in terms of its common and distinctive features and represented by discrete clusters. Tversky (1977) proposed a "feature matching model" based on the idea that when faced with a set of objects, subjects often sort them into clusters to reduce information load and facilitate further processing. This model is based on a similarity relation that is very different from that of the geometric models. According to Tversky, each object is represented by a set of features or attributes. Thus, the degree of similarity s(A, B) between objects A and B, for all distinct A and B, is defined by a matching function between the common and distinctive features of the two compared objects. This function is composed of three arguments: (1) A intersect B: the features that are common to the two compared objects A and B; (2) A minus B: the features belonging to A but not B; and (3) B minus A: the features belonging to B but not A.

This approach is formalized by cluster analysis. In a cluster analysis, objects that are similar belong to the same cluster and objects that are dissimilar belong to different clusters. Clustering of objects can be hierarchical or nonhierarchical. In the case of nonhierarchical clustering, objects can belong to one and only one cluster. However, with hierarchical clustering, objects can belong to more than one cluster as long as they are hierarchically nested; i.e., all members of a lower-level cluster belong to a higher-level cluster. One way to represent hierarchical clustering is with a tree, a graph in which the similarity between two objects is represented by the length or the height of the link joining the two objects. In a hierarchical representation obtained by the application of the HICLUS model proposed by Johnson (1967), objects that are most similar are joined at lower levels in the tree, whereas dissimilar objects are joined together only at higher levels in the tree. Also, the ADCLUS model proposed by Shepard and Arabie (1979) provides a representation that allows partial overlapping of clusters. Finally, there are additive trees in which similarity between objects is given by the lengths of links between nodes in the trees.

Other authors (Gibson, 1966, 1979; Rosch, 1973a,b), have postulated that the physical world that surrounds us has discontinuities, which eliminates the problem of determining the level at which such a categorization occurs. According to Gibson (1966, 1979), all information necessary for visual perception is present in the environment and the perceiving subject has only to pick it up. This conception leads us to consider only natural situations (from which the term "ecological" is derived) and to reject the general validity of laboratory experiments. Gibson's theory is opposed to all constructivist positions according to which information is extracted from sense systems (visual, auditory, and the like) by computational procedures and processes. All these processes are judged to be useless because information given by the physical environment to the perceiving subject is already structured and organized in a coherent manner. Perception is thus direct because the information is presorted and does not need to be processed.

For auditory nonverbal perception, this conception suggests that the physical nature of the sound object, the means by which it is set into vibration, and its function for the listener are perceived directly, without intermediate processes. In other words, there is no analysis of the individual elements that comprise a sound event; nor is there a reconstruction of an auditory image that is compared to a representation in memory (McAdams, 1993). Thus, the approach for ecological psychologists is to describe the structure of the physical world in order to understand perceived properties as invariants. Note that we can usually recognize a saxophone or a piano played on the radio even if the signal is modified by bad transmission. If invariants can be isolated, then the task of the psychologist is to determine how the listeners detect these properties. This approach allows us to evoke a mechanism of "causality inference": Received data are indices considered as effects of a causality, which is the perceived object. This conception thus suggests a strong relation between the mental representation of a sound event, its production mode, and its perceptual identity.

The question whether perception of timbre is categorical is not neutral with respect to causality. Historically, the relation between the physical production of a sound event and its auditory result has been obvious. Indeed, at one time the term "timbre" designated a particular instrument, a sort of drum with stretched strings that gave a characteristic "color" to its sound (Dictionnaire de l'Acadamie Francaise, 1835). But the predominance of pitch in most musical cultures has relegated timbre to a secondary role. Classical instruments, excluding some percussion instruments, were constructed so that anything that disturbed pitch recognition was eliminated. In the absence of an explicit musical function, it is natural that "timbre" tends to no longer refer to a particular sound source or instrument. However, even today, classical instruments are categorized, and if categorization appears at a perceptual level, it is likely to be due to the type of sound source. While it is difficult to physically construct an intermediate instrument between a percussive instrument and a sustained instrument, electronic synthesis allows us to create hybrid timbres and place perception outside of the mechano-acoustical instrument categories. Even so, the perception of timbre as revealed by multidimensional space analysis, where continua of timbre are theoretically possible, seems partially categorical. According to Grey (1975) "the scaling for sets of naturalistic tones suggests a hybrid space, where some dimensions are based on low-level perceptual distinctions made with respect to obvious physical properties of tones, while other dimensions can be explained only on the basis of a higher level distinction, like musical instrument families" [cited by Risset and Wessel (1982, p. 48)]. The intervention of cognitive processes, such as familiarity with or recognition of an instrument, shows that it is perhaps impossible to obtain a totally continuous timbre space.

3.1 Studies of the Perception of Causality of Sound Events

An alternative approach for studying timbre perception is to consider that musical instruments are often grouped on the basis of their belonging to resonator and/or exciter categories and that the mechanical properties of sound sources could

influence dissimilarity judgments among different timbres. Indeed, some categorization processes may be likely to influence listeners' dissimilarity judgments on which the notion of timbre space is based. The aim of an experiment performed by Donnadieu (1997) was to examine such categorization processes by a classification task and to specify the relation that could exist between a multidimensional representation of timbre and a categorical representation. In other words, the study's goal was to determine the perceptual categories underlying 36 digitally recorded musical instrument sounds selected from the McGill University Master Samples compact disk (Opolko & Wapnick, 1987). These included tones produced by traditional pitched sustained instruments (e.g., flute, trumpet, piano), tones of strongly pitched percussion instruments (e.g., celesta, marimba, vibraphone bowed, vibraphone struck, tympani), weakly pitched (e.g., bowed cymbal, log drum), and unpitched (e.g., tam-tam, bamboo chimes), representing most of the types of exciters and resonators used in the orchestra. The objective was to determine whether listeners based their classifications on instrument families or on certain physical attributes of sound objects. A multidimensional representation of the categorical structure was used in order to define how timbre categories are partitioned in a timbre space and to evaluate the influence of the physical functioning of instruments on perceptual categorical structure.

Sixty subjects were asked to perform a free classification task. Two advantages of this type of task are that it is easily performed by listeners and it can help to determine the kinds of sound properties that are worth investigating more systematically. All stimuli were first presented to the subjects, and they were asked to create their own categories and to assign similar stimuli to the same category and dissimilar stimuli to different categories. In a free classification task, subjects can create as many categories as they want and can assign as many stimuli as they wish in each category. To determine the categorical structure of this set of stimuli, an ADTREE analysis (Barthélemy and Guénoche, 1988) was used, which allowed the development of an additive tree, a graph in which the similarity between any two nodes, corresponding to the objects, is given by the length of the link between those nodes. The observed tree for the 36 orchestral instruments is represented in Fig. 8.11. According to Tversky's model (1977), the nodes can be interpreted as the prototype of a category which corresponds to the object that shares common features with the objects belonging to this category, while the length of the link between two nodes corresponds to the weight of the features belonging to class A but not to class B, for example. This last model was used because it was particularly easy to interpret this type of representation according to the model proposed by Tversky. Trees were established for all the subjects. An attempt was made to establish a relation between different perceptual categories and the stimulus properties, most of the time by seeking structural similarities among stimuli classed together and differences between stimuli classed in different categories. Such a classification was observed with all impulsive excitation (percussion) instruments in one category and all sustained excitation instruments in another category. Classifications were also observed according to each instrument's resonator type, with strings, plates, and bowed membranes placed in different categories. Influence of resonator type was particularly evident when two types of vibraphone sounds were examined:

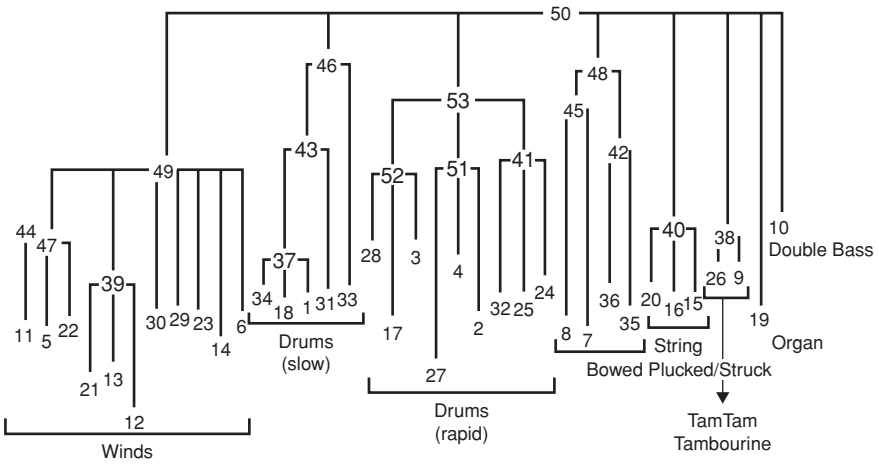


FIGURE 8.11. Hierarchical representation of the timbral classification tree structure for 36 acoustic sounds derived from an ADTREE analysis of free classifications by 60 subjects. Numbers from 1 to 36 correspond to the 36 sounds used and numbers from 37 to 53 are nodes computed by the algorithm.

They fell in the same category even if the type of excitation was very different [e.g., see Fig. 8.11, numbers 33 (vibraphone bowed) and 34 (vibraphone struck)].

In summary, we can conclude that the notion of categories of timbre corresponds to a perceptual reality and that these categories seem in most cases to be based on the physical functioning of the different instruments. This result suggests that an implicit knowledge of the physical functioning of instruments has a strong correlation with perceptual classification of their corresponding sounds. An idea worth exploring would be to systematically define these perceptual categories within a timbre space. This space could probably be continuous in that the boundaries between categories could be fuzzy. Still, discrimination within and between those boundaries would be possible.

3.2 Categorical Perception: A Speech-Specific Phenomenon

3.2.1 Definition of the Categorical Perception Phenomenon

The *categorical phenomenon*, first described by Liberman (1957), refers to a situation where it is possible to identify and discriminate two objects belonging to two distinctive categories, but not possible to discriminate two objects belonging to the same category. A procedure generally employed to demonstrate such a phenomenon is the following: A continuum of N stimuli is constructed (N is typically equal to 10) by variation of a control parameter. This stimuli continuum is composed of two distinct end-point stimuli (e.g., two distinct phonemes) at the extremes of the continuum and different intermediate synthetic stimuli between the extremes. Listeners are asked to perform two tasks: (1) an *identification* task for them to identify each stimulus according to two contrasted categories; and

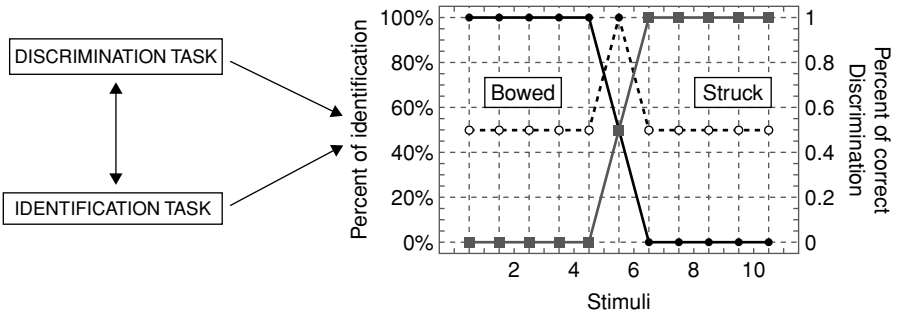


FIGURE 8.12. Theoretical discrimination and identification functions are shown for stimuli gradually changing from "bowed" (left) to "struck" (right). Stimuli 1–5 are categorically identified as "bowed," whereas stimuli 7–11 are categorically identified as "struck." However, percent correct discrimination between adjacent stimuli is 50% (guessing) except at the category boundary corresponding to stimulus 6, where discrimination is 100%.

(2) a *discrimination* task where stimuli are presented in pairs and subjects are asked to respond whether the stimuli are identical or not.

According to Studdert-Kennedy et al. (1970), three criteria are necessary to conclude that categorical perception exists in a continuum: (1) "peaks," regions of high discriminability in the discrimination function; (2) "troughs," regions where discrimination is near the chance level; and (3) a correspondence between the peaks and troughs and the shape of the identification function, with peaks occurring at the identification boundaries and troughs within each category. In other words, in contrast to continuous perception, categorical perception occurs when intracategorical discrimination is absent: Subjects discriminate two neighboring stimuli only if they (or their control parameters) are situated on either side of the boundary separating the two categories. Fig. 8.12 represents hypothetical results of such a categorical perception experiment.

3.2.2 Musical Categories: Plucking and Striking vs Bowing

It has been asserted that one of the most important differences between speech stimuli and non-speech stimuli is that the former are categorically perceived whereas the latter are not. However, it seems possible to observe this categorical perception phenomenon for non-speech stimuli. Miller et al. (1976) used noise and buzz sounds, with the onset of the noise varying from -10 to $+80$ ms with respect to the onset of the buzz. Discrimination was best when the noise led the buzz by about 16 ms, which was about the same amount of delay as the category boundary in a labeling task. Pisoni (1977) and Pastore (1976) also observed such a phenomenon for two-tone stimuli and critical-flicker fusion, respectively. Locke and Kellar (1973) and Siegel and Siegel (1977), as well as others, have observed categorical perception for musical intervals.

Cutting and Rosner (1974, 1976) used sawtooth waveforms varying in rise time from 0 to 80 ms in increments of 10 ms and found that best discrimination occurred between tones whose rise times straddled 40 ms, the position of the plucked/bowed perception category boundary between where subjects had identified rapid (0 to 30 ms) rise-time stimuli as plucked strings and slower (50 to 80 ms) rise-time stimuli as bowed strings. While discriminations between the plucked and bowed category regions were easy to make, subjects were not able to discriminate rise time differences very well within the bowed and plucked category regions. Also, Remez (1978) created a plucked-to-bowed continuum by tailoring natural tokens of musical sounds played on a bass viol. These, too, were perceived categorically. However, his continuum was a rise-time-by-amplitude-at-onset continuum rather than simply a rise-time continuum. Macmillan (1979), using analog-generated stimuli of considerably lower fundamental frequency, also found categorical perception. However, the boundaries fell at 25 ms rather than 40 ms for the discrimination and identification functions.

Cutting et al. (1976) extended their previous findings (Cutting and Rosner, 1974, 1976) by demonstrating selective adaptation effects with the same stimuli. Subjects had to categorize stimuli before and after repeated exposure to an *adaptor*, which corresponded to either a stimulus with the same spectral envelope, a stimulus with the same frequency, or a stimulus with the same spectral envelope and frequency. The boundary between the two categories shifted as expected after the exposure to the adaptor, and the greatest shift was observed when the adapting stimulus shared all dimensions with the test continuum. Remez et al. (1980) found reliable adaptation by using end-point adaptors on a plucked-bowed continuum. More recently, Pitt (1995) found such an effect on identification and reaction-time performance using a trumpet-to-piano continuum of acoustic sounds. He showed that after the exposure to adaptors corresponding to the end-points of the continuum, the categorization boundary indicated by the identification function shifted significantly, and reaction times were significantly faster for stimuli situated near the end-points of the continuum. Direct comparison of the identification results with those from previous timbre adaptation studies is not possible because different measures of adaptation magnitude were used. However, visual comparison of identification functions suggests that the trumpet-to-piano continuum produced a larger boundary shift than the pluck-to-bow continuum of Cutting et al. (1976).

Such results suggest that categorical perception of music-like sounds may be explained by a theory based on feature detection. Indeed, according to such a theory, the repetitive presentation of a stimulus belonging to a perceptual category would lead to a decrease in the rate response of the detector for other stimuli in the same category.

3.2.2.1 Are the Same Feature Detectors Used for Speech and Nonspeech Sounds?

In the light of these results, one might ask whether specific detectors involved in the processing of speech sounds and nonverbal sounds are the same. Some authors

have indeed demonstrated a similar adaptation for verbal stimuli using nonverbal stimuli as adaptors (Diehl, 1976; Kat and Samuel, 1984; Samuel, 1988; Samuel and Newport, 1979), although other authors (Remez et al., 1980) have not observed such a phenomenon. For example, Diehl (1976) showed that the spectrum of a plucked string could influence the perception of a continuum from /ba/ to /wa/, but that the spectrum of a bowed string did not influence the result. Samuel and Newport (1979) conducted this experiment with continua from /ba/ to /wa/ and from /tʃ a/ to /ʃ a/. They used four types of nonverbal adaptors: two periodic sounds (where the fundamental frequency was different from that of the verbal sounds) both imitating either a plucked or a bowed string. Results showed that periodic sounds with rapid attack times had an influence if they shared properties with the /ba/ sound but not with the /tʃ a/ sound, while the sounds with slow attack times had an effect if they shared a property with the /ʃ a/ sound but not with the /wa/ sound.

Nonetheless, results observed by Remez et al. (1980) argue against the hypothesis of common specific detectors for nonspeech and speech stimuli. In their study, the authors crossed adaptor stimuli, which could be either verbal or nonverbal, with test stimuli that were either verbal or not. Adaptor stimuli were either extreme stimuli of the two types of tested continua or difference stimuli according to their acoustical properties. Adaptor stimuli differed from the continuum neither in attack time, nor in fundamental frequency, but only in terms of their spectral envelopes. This difference gave rise to a difference in source identity. Results showed adaptation only when the type of the adaptor (e.g., verbal vs nonverbal) corresponded to that of the test stimuli; i.e., adaptation occurred for verbal test stimuli with a verbal adaptor and for nonverbal test stimuli with a nonverbal adapter, while nonverbal and verbal adaptor stimuli did not influence the verbal and nonverbal test stimuli, respectively. These results thus suggest that specific detectors involved in the processing of verbal sounds and those involved with nonverbal sounds are different in nature, confirming the idea formulated by Cutting et al. (1976) that the importance of the adaptation effect is a function of the number of auditory attributes shared by the continuum and the adaptor.

These results suggest that nonspeech stimuli could be categorically perceived and could be explained by a feature-detector theory (Cutting et al., 1976; Pitt, 1995; Remez et al., 1980). However, it is difficult to make any conclusions about existence of common feature detectors for speech and nonspeech stimuli (Diehl, 1976; Remez et al., 1980; Samuel and Newport, 1979).

3.2.2.2 Categorical Perception in Young Infants

Infants, like adults, seem to perceive nonspeech stimuli in a categorical manner. Jusczyk et al. (1977) used a high-amplitude sucking technique to explore 2-month-olds' perception of rise-time differences for the same stimuli used by Cutting and Rosner (1974, 1976). The authors observed that the sucking rate did not vary if the change was within one of the two categories, but that it was significantly higher when the change was across the two categories: "bowed" vs "plucked." More

specifically, infants seemed to perceive a difference between stimuli with 30-ms to 60-ms rise times, which corresponded to the boundaries observed by Cutting and Rosner (1974) for adults, but not for stimuli between 0 to 30 ms and 60–90 ms rise times. Like adults, infants discriminated rise-time differences between the two category boundaries but not equal differences within either category. The presence of such categorical perception in 2-month-old infants suggests that it is relatively independent of auditory experience. To account for similar results in infants for verbal stimuli, Eimas (1975) proposed the hypothesis that newborns are equipped with specific detectors which respond to relevant acoustical properties of verbal sounds. Results observed for nonverbal sounds lead to a similar interpretation. However, research on prenatal audition (Granier-Deferre & Busnel, 1981; Granier-Deferre & Lecanuet, 1987; Lecanuet et al., 1988; Lecanuet et al., 1992) has shown that newborn infants do not begin their auditory experience at birth, but actually several months before, therefore providing several months of auditory experience during which perceptual learning can take place.

3.2.2.3 The McGurk Effect for Timbre

McGurk and MacDonald (1976) and MacDonald and McGurk (1978) showed that the perception of an acoustic syllable could be affected by the simultaneous presentation of visual information specifying a speaker's articulatory movement of a different syllable. For example, if the auditory syllable is a /ba/ and if the subjects see a video tape of a speaker producing a /ga/, they report having heard a /da/. This /da/ syllable is an intermediate syllable, the place of articulation of which is between those of /ba/ and /ga/. This effect, called the "McGurk effect," clearly shows that visual and auditory information can be integrated by subjects, the response being a compromise between normal responses to two opposing stimuli. Moreover, Kuhl and Meltzoff (1982) observed that young infants show a preference for pairs of stimuli in which auditory and visual information are matched. Infants look longer at a mouth which presents the articulatory movement of the heard syllable than at one whose articulatory movement does not correspond to the sound. This result suggests a predisposed functional relationship between the perception and the production of language.

For nonverbal sounds, Rosenblum and Fowler (1991) observed that visual information could have an influence on auditory judgment. They showed, for example, using the McGurk paradigm, that loudness judgments of syllables or hand-clapping could be influenced by visual information. More recently, Saldana and Rosenblum (1993) observed the same type of effect for plucked and bowed string sounds. In their first experiment, they presented each sound along a continuum between a plucked and a bowed string. At each presentation of a sound, the subject had to estimate whether the sound was plucked or bowed on a continuous scale. The instructions were to use the middle of this scale if the sound was ambiguous. In one condition, the sound was presented simultaneously with a video tape showing a player plucking or bowing a string. Results showed that subjects' responses were greatly influenced by the visual information. Indeed, the identification function for

judgments based only on the auditory presentation of the sound was significantly different from that based on the audiovisual presentation. In fact, the authors observed that the identification function corresponding to the audiovisual condition shifted to the plucked response scale and inversely for the condition where the video tape presented a bowed string. However, this study did not include the opposite possibility of allowing the subjects to identify a plucked string as a bowed string when the visual information described a bowed string. The hypothesis of the authors was that the effect could be explained by the ecological theory according to which the influence of the visual information would be in direct relation with the production mode of the sound event. To test this last hypothesis they replaced the visual information by a visual presentation of the two words "plucked" and "bowed." In this last case no effect was observed.

3.2.3 Is There a Perceptual Categorization of Timbre?

In contrast to the above discussion, some researchers argue that nonspeech sounds are not categorically perceived. Van Heuven and van den Broecke (1979) measured the variability of settings in a rise-time reproduction task. They found that the standard deviation of adjustments was an increasing linear function of rise time. They felt that the differences between their results and Cutting and Rosner's could be attributed to differences in stimulus generation techniques. Rosen and Howell (1981) synthesized a new continuum of sawtooth waves differing in linear increments of rise time, analogous to the array reported by Cutting and Rosner (1974). In order to test the hypothesis that a different generation technique could produce different results, Van Heuven and van den Broecke (1979) included a condition in which stimuli were recorded before presentation to the subjects. They did not obtain results consistent with categorical perception. Although they obtained a similar identification function, they did not observe a peak in the discrimination function. Instead, they found a discrimination function that might be predicted better on the basis of a Weber fraction for rise time. So, the method of stimulus generation and presentation was not responsible for the discrepancies between the results. Using the original tapes of Cutting and Rosner (1974) to replicate their results, they found categorical perception of these stimuli. To reconcile the difference in the two findings, they measured the original stimuli and found that the rise times differed from those reported in the Cutting and Rosner paper. Moreover, the discrepancies were such that they predicted nonlinearities in the discrimination results. Thus, they concluded that plucked and bowed music-like sounds are not categorically perceived.

According to Hary and Massaro (1982), categorical perception results do not necessarily imply categorical perception. Indeed, they showed that a bipolar continuum of increasing and decreasing onset times yielded traditional categorical results but that when only half of this continuum was tested, the same sounds were perceived continuously. On the other hand, contradictory results for the identification function have also been found. Smurzynski (1985) asked subjects to learn envelopes by rise-time value and later to identify them. Analysis of responses

showed that trained subjects did not classify a continuum of sawtooth waveforms varying in rise time into two sharply defined categories, but were able to resolve rise-time values with much greater accuracy than would be achieved by simply dividing the continuum into two categories such as "plucked" and "bowed." To conclude, Cutting (1982) found that stimuli with equal linear increments of rise time were not categorically perceived, but stimuli with logarithmic increments of rise time were categorically perceived. The stimuli and the results observed by Rosen and Howell (1981) are shown in Figs. 8.13a and 8.13b, and the results found by Cutting and Rosner (1974) are represented in Fig. 8.13c.

Donnadieu and McAdams (1996), Donnadieu et al. (1996), and Donnadieu (1997) confirmed the idea of noncategorical perception of rise time using two continua of attack time constructed on the basis of two original vibraphone sounds. One sound resulted from a vibraphone struck by a hard mallet, and the other came from the vibraphone bowed with a violin bow on its edge. (Both sounds were taken from a McGill University Master Samples compact disk.) Most studies on categorical perception of rise time have utilized synthesized sounds. We chose to use acoustic sounds even though in this case the definition of attack time is somewhat arbitrary. For both sounds, 10 stimuli were constructed in which only the rise time of the amplitude-vs-time envelope differed. Utilizing a phase-vocoder analysis/resynthesis program (Beauchamp, 1993), a "struck" continuum was constructed by successively modifying the rise time of the struck vibraphone sound so that it started at 0.13 s, increasing from step to step by a factor of 1.29, and ending at 1.30 s. A corresponding "bowed" continuum was constructed by decreasing the rise time of a bowed vibraphone sound, starting at 0.35 s, decreasing by factors of 0.76, and ending at 0.03 s. For each continuum, the rapid-onset stimuli tended to sound like a struck instrument and the slower onset stimuli like a bowed instrument.

Subjects were asked to perform three tasks: (1) discriminate pairs of stimuli along the two continua, (2) identify (or categorize) them as one of the end-points ("struck" or "bowed"), and (3) judge the perceptual dissimilarity of the stimulus pairs. The stimulus pairs were separated by two steps on each continuum. The three tasks were performed separately for each of the two continua ("reduced contexts") and for the combined set of these two continua stimuli ("extended context"). The discrimination task was of type AX. For this task, the subjects heard each stimulus pair with a 1-s interstimulus interval (ISI) and were asked to judge if the stimuli were "same" or "different." Raw discrimination scores were adjusted by subtracting "false alarm" scores (responding "different" when the sounds were identical). For the identification task, subjects were asked to label the stimuli as "struck" or "bowed." For the dissimilarity task, subjects judged the degree of dissimilarity (on a scale varying from very similar to very dissimilar) of stimuli pairs (1-3, 2-4, etc.). A scale was presented on a computer screen and listeners had to push the button at the desired position. For each task, subjects participated in three sessions corresponding to the three contexts: (1) stimuli from the "bowed" continuum ("reduced context"), (2) stimuli from the "struck" continuum ("reduced context"), and (3) stimuli from the union of these two continua ("extended context"). Subjects completed all three types of tasks (discrimination, identification, and dissimilarity)

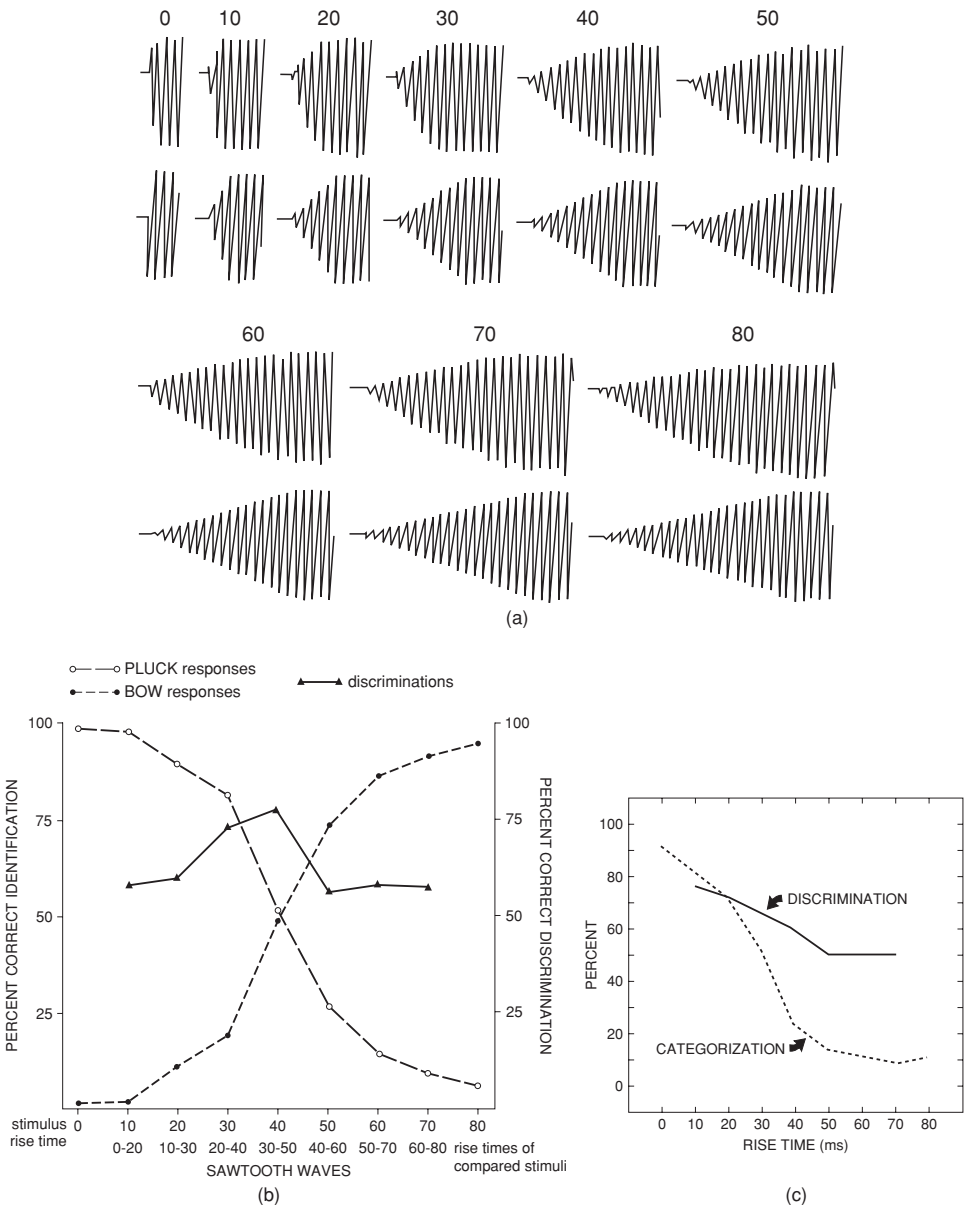
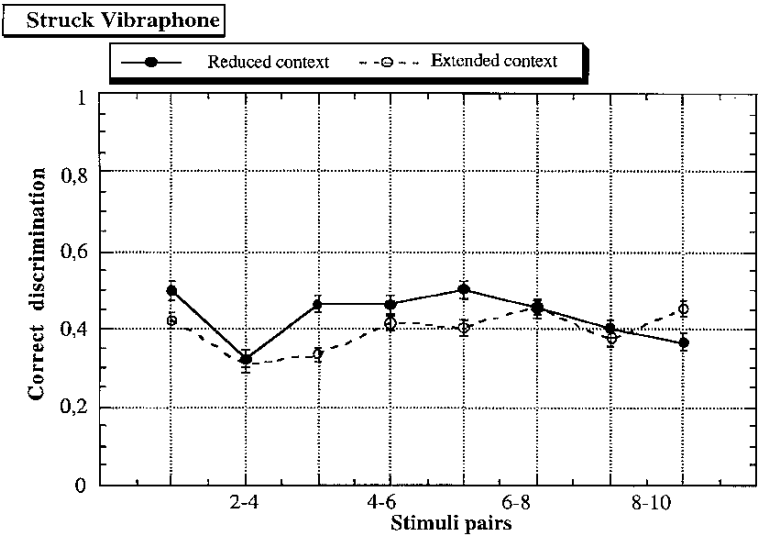


FIGURE 8.13. Categorical perception of a rise-time continuum. (a) Oscillograms for a nine sawtooth-wave stimuli continuum used by Cutting and Rosner (1974) and by Rosen and Howell (1981). (b) Identification and discrimination functions observed by Cutting and Rosner (1974). (c) Results observed by Rosen and Howell (1981) on the Cutting and Rosner stimuli. [From Cutting and Rosner (1974), Fig. 2 and Rosen and Howell (1981), Figs. 3 and 4, adapted by permission of the Psychometric Society.]

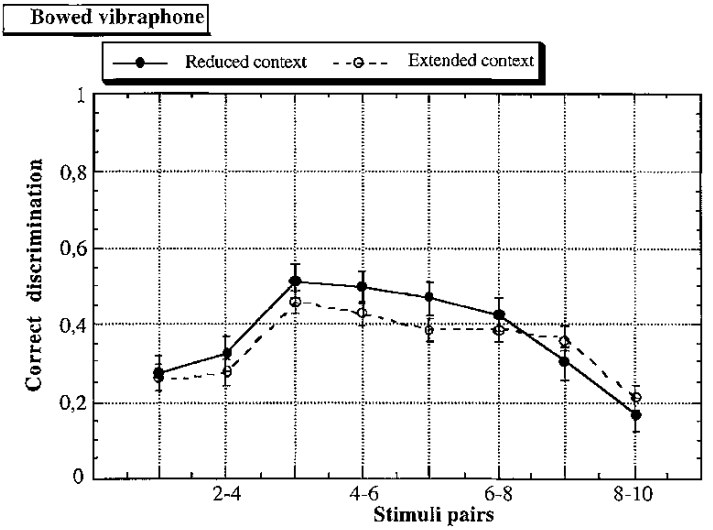
in each session. From the results, it was clear that although subjects on average gradually changed their classification from bowed vibraphone to struck vibraphone and vice versa along the two continua, discrimination performance was fairly constant along the continua. Figures 8.14 and 8.15 give the results of the discrimination and identification experiments, respectively. Note that two sets of data were extracted from the extended context session: one which focused on the listener's ability to correctly identify or discriminate the "struck" continuum data in the presence of the "bowed" continuum data, and vice versa.

These results were not consistent with the numerous studies that have shown categorical perception of rise time (Cutting, 1982; Cutting and Rosner, 1974, 1976; Cutting et al., 1976; Jusczyk et al., 1977; Macmillan, 1979; Miller et al., 1976; Pitt, 1995; Remez, 1978; Remez et al., 1980). However, Rosen and Howell (1981) observed that discrimination performance for equally spaced stimuli is always best for shortest rise times. Our results were partially consistent with these results because, although we did not observe a categorical perception of the rise time of acoustic struck or bowed vibraphones, we did observe that discrimination performance was relatively constant across the two continua tested. The difference in our results could be due to the fact that our continua were constructed by logarithmic rather than linear rise time increments. We chose logarithmic increments first because Cutting's last results showed that only in this case is categorical perception observed for the attack time of nonspeech sounds and second because the first dimension of timbre is generally more correlated with a measure of the logarithm of the attack time than with linear rise time of the temporal envelope (McAdams et al., 1995). On the other hand, category boundaries observed in previous studies were very different from our category boundaries. This difference could be due to the fact that our stimuli corresponded to resynthesized transformations of recorded acoustic sounds. Moreover, in this experiment we used a bowed bar (a vibraphone) rather than a bowed string. The modes of vibration of a metal bar are very different than that of a string and take more time to be set into vibration when bowed.

We also noted during the construction of the stimuli that there was a large difference between the attack times of the two continua endpoints. Indeed, to induce a perception of the bowed vibraphone we had to considerably augment the rise time of the temporal envelope of the original struck vibraphone beyond that of the rise time of the original bowed vibraphone. Moreover, the boundary between the "struck" and "bowed" categories was quite different for the struck and bowed continua. This calls into question the definition of the attack as being characterized uniquely by rise time and suggests that other factors in addition to the logarithm of the attack time of the sounds contributed to the identity of the type of excitation. Indeed, the attack epochs of these sounds may include many characteristics, such as the presence of a high-frequency component or the presence of noise produced by the contact between mallet and bar. These characteristics could be used by the auditory system to identify an instrument's resonator (e.g., as a bar or a string) or the type of exciters used. These aspects would correspond to the structural invariants of the ecological approach.

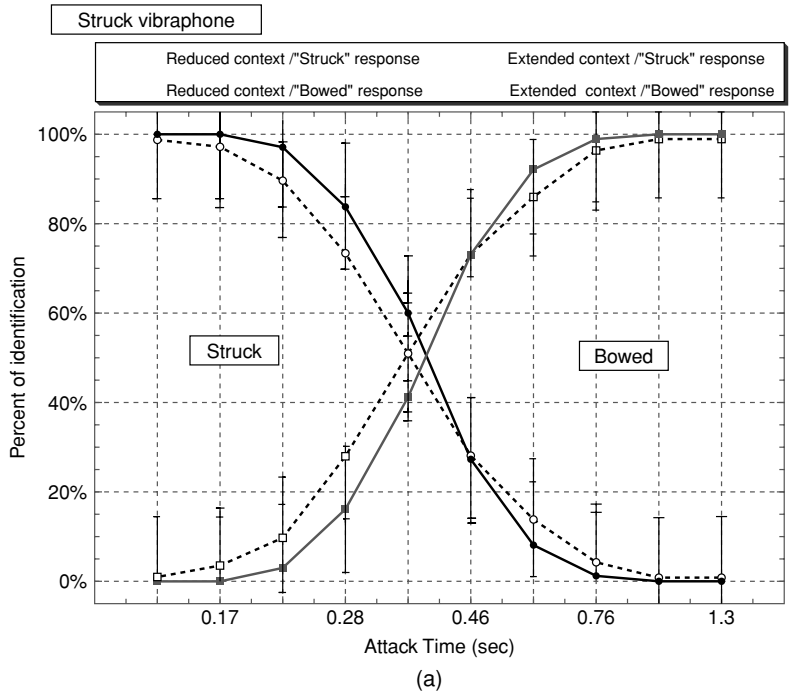


(a)



(b)

FIGURE 8.14. Mean discrimination functions for "struck" and "bowed" vibraphone continua stimuli presented in "reduced context" (each continuum alone) and "extended context" (continua stimuli combined). (a) Discrimination of stimuli pairs (1 and 3, 2 and 4, etc.) along continua of gradually increasing rise time of struck vibraphone sounds. (b) Discrimination of stimuli pairs along continua of gradually decreasing rise time of bowed vibraphone sounds. 0% discrimination corresponds to the guessing level. Note that in the "extended context" responses to the same stimuli are scored as in the "reduced context," but in the former case the stimuli are intermixed with stimuli from the other continuum.



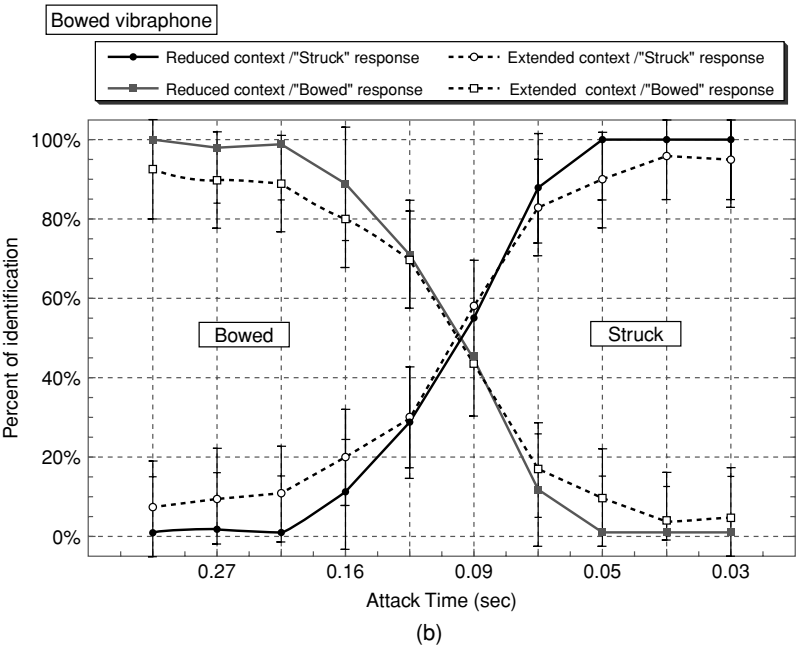


FIGURE 8.15. Mean identification functions for identifying the same stimuli as given in Fig. 8.14 as either "struck" or "bowed" vs attack time of the stimuli for the "reduced" and "extended" contexts. (a) Identification of struck vibraphone continua stimuli. (b) Identification of bowed vibraphone continua stimuli.

The fact that attack quality of sounds seems to be perceived along a continuum suggests that the mental structure of musical timbre could be represented in terms of several perceptually continuous dimensions. However, the fact that we can classify the timbres of vibraphone as struck or bowed, as the identification function shows, suggests that this type of representation could be influenced by a higher-level process of categorization.

4 Conclusions

This chapter describes studies on the perception of timbre of complex tones. Two approaches were presented corresponding to two ways of understanding the perceptual representation of musical timbre.

The first approach describes different perceptual dimensions of timbre in terms of abstract properties. It seeks to determine which acoustical parameters of the complex signal are processed by the auditory system and in the end contribute to the perception of timbre. Multidimensional scaling has been particularly fruitful for this type of study. Results suggest that essentially three dimensions can be used to describe the timbres of a given set of musical complex tones. The physical correlates of the different dimensions seem to be well identified, corresponding to spectral, temporal, and spectrotemporal aspects of the acoustical signal (Grey, 1975, 1977; Krumhansl, 1989; Miller and Carterette, 1975; Plomp, 1970, 1976; Samson et al., 1996; Wedin and Goude, 1972; Wessel, 1979). However, if the contribution of spectral aspects of the sounds is clear today, the influence of temporal aspects is less clear. Indeed, some multidimensional studies call into question the perceptual importance of such temporal aspects in the perception of timbre. Wedin and Goude's (1972), Miller and Carterette's (1975), and Iverson and Krumhansl's (1991) results suggest a predominance of spectral factors. Moreover, it seems that the perceptual salience of temporal judgments depends largely on context, and, in particular, on the musical context in which the sounds are presented (Grey, 1978; Kendall, 1986). Nevertheless, for the case where sounds are presented in isolation, we cannot doubt the importance of such factors. Results from experiments based on deletion of parts of sounds (Berger, 1964; Saldanha and Corso, 1964), those based on spectral modifications associated with discrimination tasks (Charbonneau, 1981; Grey and Moorer, 1977), and observations from some multidimensional studies (Grey, 1977; Krumhansl, 1989; Wessel, 1979) support the importance of the influence of temporal aspects on timbre perception.

Results from multidimensional studies suggest a continuous representation of the timbre of complex sounds. In the same way, studies based on sound modifications (Saldanha and Corso, 1964; Grey and Moorer, 1977; Grey, 1978; Kendall, 1986) have shown that the capacity of listeners to identify sounds diminishes when acoustical parameters are manipulated. This degradation of identification efficacy may depend on whether an auditory stimulus varies continuously along dimensions related to specific acoustical parameters and whether the categories involved have fuzzy boundaries. Moreover, it appears from studies of timbral analogies

(Ehresman and Wessel, 1978; McAdams and Cunibile, 1992) and studies involving modification of sounds to examine their consequences on timbre space (Grey and Gordon, 1978; Wessel, 1979, 1983), that intermediate areas of a timbre space can be filled in and that regular perceptual transitions based on a few physical dimensions are possible. In the same way, studies on the role played by timbre in auditory organization (Bey and McAdams, 2003; Gregory, 1994; Hartmann and Johnson, 1991; Iverson, 1993; McAdams and Bregman, 1979; Wessel, 1979) indicate that the auditory-stream-segregation process is based on the same perceptual attributes as those used by listeners when they were asked to do dissimilarity judgments between different timbres. These results suggest that MDS timbre spaces can account for the similarity relations between different timbres. Fusion and segregation processes could be based on the metric distance separating timbres in a geometric space, and the perceptual dimensions of timbres may be related to the representation upon which such processes operate. Finally, the development of verbal attributes of timbre (Faure et al., 1996; Samoylenko et al., 1996) allow us to complete our knowledge concerning timbre space and to establish the relation between perceptual representations and semantic representations.

The second approach is related to ecological considerations (Gibson, 1966) and the notion of perceptual categories of timbre. According to this approach, timbre perception is a direct function of the physical properties of the sound source. In this case, the aim of various studies (e.g., Donnadieu, 1997; Lakatos, 2000) has been to describe perceptually relevant physical properties of sound objects and their relative roles in the perception of musical instrument sounds, in addition to the major roles played by perceptual attributes such as pitch salience, spectral envelope, and roughness. The idea is that the auditory system can code the timbre of complex sounds in terms of the details of physical source sound production. Indeed, this research suggests that the relation between timbre and physical causality could be a fundamental aspect of our perception and of the categorical organization of the perceptual structure of timbre. However, studies that investigated whether the attack quality of complex sounds is categorically perceived gave contradictory results. In summary, these results suggest that attack qualities can be continuously perceived and support a model of perceptually continuous timbre space, but they do not exclude the possibility that higher-level classification organizations could be present and that timbre categories could be organized in such a timbre space.

References

American National Standards Institute (1973). *Psychoacoustical Terminology, S3.20-1973* (American National Standards Institute, New York).

Barthélemy, J.-P. and Guénoche, A. (1988). *Arbres et les représentation des proximités* [Trees and proximity representations]. (Masson, Paris).

Beauchamp, J. W. (1993). "Unix workstation software for analysis, graphics, modifications, and synthesis of musical sounds," *94th Convention of the Audio Engineering Society*, Berlin, (Audio Eng. Soc., New York), Audio Eng. Soc. Preprint 3479.

Berger, K. W. (1964). "Some factors in the recognition of timbre," J. Acoust. Soc. Am. **36**(10), 1888–1891.

Bey, C. and McAdams, S. (2003). "Postrecognition of interleaved melodies as an indirect measure of auditory stream formation," J. Exp. Psychol.: Human Percept. Perform. **29**, 267–279.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).

Bregman, A. S., Liao, C., and Levitan, R. (1990). "Auditory grouping based on fundamental frequency and formant peak frequency," Canadian J. Psychol. **44**, 400–413.

Cadoz, C. (1991). "Timbre et causalité," in *Le timbre: Métaphore pour la composition*, J.-B. Barriere, ed. (Christian Bourgois, Paris), pp. 17–46.

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

Charbonneau, G. R. (1981). "Timbre and the perceptual effects of three types of data reduction," Computer Music J. **5**(2), 10–19.

Cutting, J. E. and Rosner, B. S. (1974). "Categories and boundaries in speech and music," Perception and Psychophysics **16**(3), 564–570.

Cutting, J. E. and Rosner, B. S. (1976). "Discrimination functions predicted from categories in speech and music," Perception and Psychophysics **20**, 87–88.

Cutting, J. E., Rosner, B. S., and Foard, C. F. (1976). "Perceptual categories for musiclike sounds: Implications for theories of speech perception," Quarterly J. Exp. Psychol. **28**, 361–378.

Cutting, J. E. (1982). "Plucks and bows are categorically perceived, sometimes," Perception and Psychophysics **31**, 462–476.

De Brujin, A. (1978). "Timbre classification of complex tones," Acustica **40**, 108–114.

Dictionnaire de l'Academie Française, 1835.

Diehl, R. (1976). "Feature analyzers for the phonetic dimension stop vs. continuant," Perception and Psychophysics **19**, 267–272.

Donnadieu, S. and McAdams, S. (1996). "Effect of context change on dissimilarity, discrimination and categorization task on timbre perception," in *Proc. 12th Annual Meeting of the Int. Society for Psychophysics*, Padua, Italy, S. Masin, ed. (Univ. of Padua, Padua, Italy), pp. 239–244.

Donnadieu, S., McAdams, S., and Winsberg, S. (1996). "Categorization, discrimination and context effects in the perception of natural and interpolated timbres," in *Proc. 4th Int. Conf. on Music Perception and Cognition* (ICMPC4), Montréal, Canada, B. Pennycook and E. Costa-Giomi, eds. (McGill University, Montréal), pp. 73–78.

Donnadieu, S. (1997). "Représentation mental du timbre des sons complexes et effets de contexte [Mental representation of timbre of complex sounds and the effects of context]," unpublished doctoral dissertation, Université Paris V.

Ehresman, D. and Wessel, D. (1978). *Perception of Timbre Analogies*, IRCAM Technical Report 13/78 (Centre Georges Pompidou, Paris).

Eimas, P. D. (1975). "Auditory and linguistic processing of cues for place of articulation by infants," Perception and Psychophysics **16**, 513–521.

Faure, A., McAdams, S., and Nosulenko, V. (1996). "Verbal correlates of perceptual dimensions of timbre," in *Proc. 4th Int. Conf. on Music Perception and Cognition* (ICMPC4), B. Pennycook and E. Costa-Giomi, eds., McGill University, Montreal, Canada, pp. 79–84.

Faure, A. (2000). "Des sons aux mots: Comment parle-t-on du timbre musical [From Sounds to Words: How Does One Speak of Musical Timbre?]", unpublished doctoral dissertation, Ecoles des Hautes Etudes en Sciences Sociales, Paris.

George, W. H. (1954). "A sound reversal technique applied to the study of tone quality," Acustica **4**, 224–225.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems* (Houghton-Mifflin, Boston).

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception* (Houghton-Mifflin, Boston).

Granier-Deferre, C. and Busnel, M-C. (1981). "L'audition prénatale [Prenatal Hearing]", in *L'aube des sens,* Cahiers du Nouveau-né [*The dawn of the senses*, Newborn Journal], E. Herbinet and M-C. Busnel, eds. (Stock, Paris), pp. 147–175.

Granier-Deferre, C. and Lecanuet, J-P. (1987). "Influence de stimulations auditives précoces sur la maturation anatomique et fonctionnel du système auditif [Influence of early auditory stimulation on anatomical and functional maturation of the auditory system]," Progrès en Néonatalogie **7**, 236–249.

Gregory, A. H. (1994). "Timbre and auditory streaming," Music Perception **12**(2), 161–174.

Grey, J. M. (1975). "An Exploration of Musical Timbre," unpublished doctoral dissertation, Stanford University, Stanford, CA. Also available as Stanford Dept. of Music Report STAN-M-2.

Grey, J. M. (1977). "Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. **61**(5), 1270–1277.

Grey, J. M. and Moorer, J. A. (1977). "Perceptual evaluations of synthesized musical instrument tones," J. Acoust. Soc. Am. **62**(2), 454–462.

Grey, J. M. and Gordon, J. W. (1978). "Perceptual effects of spectral modifications on musical timbres," J. Acoust. Soc. Am. **63**(5), 1493–1500.

Grey, J. M. (1978). "Timbre discrimination in musical patterns," J. Acoust. Soc. Am. **64**(2), 467–472.

Guyot, F. (1992). "Etude de la pertinence de deux critères acoustiques pour caractériser la sonorité des sons à spectre réduit [Study of the relevance of two acoustic criteria for characterizing the sonorities of simplified sounds]," unpublished DEA thesis, Université du Maine, France.

Hartmann, W. M. and Johnson, D. (1991). "Stream segregation and peripheral channeling," Music Perception **9**(2), 155–183.

Hary, J. M. and Massaro, D. W. (1982). "Categorical results do not imply categorical perception," Perception and Psychophysics **32**(5), 409–418.

Iverson, P. and Krumhansl, C. L. (1991). "Measuring similarity of musical timbres," J. Acoust. Soc. Am. **89**(4), Pt. 2, 1988 (abstract).

Iverson, P. (1993). "Auditory segregation by musical timbre," doctoral dissertation, Cornell University, Ithaca, NY. *Dissertation Abstracts International*, 54 (4-B), 2249.

Iverson, P. and Krumhansl, C. L. (1993). "Isolating the dynamic attributes of musical timbre," J. Acoust. Soc. Am. **94**(5), 2595–2603.

Johnson, S. C. (1967). "Hierarchical clustering schemes," Psychometrika **32**, 241–254.

Jusczyk, P. W., Rosner, B. S., Cutting, J., Foard, C. F., and Smith, L. B. (1977). "Categorical perception of nonspeech sounds by 2-month-old infants," Perception and Psychophysics **21**(1), 50–54.

Kat, D. and Samuel, A. G. (1984). "More adaptation of speech by nonspeech," J. Exp. Psych: Human Percept. Perform. **10**, 512–525.

Kendall, R. A. (1986). "The role of acoustic signal partitions in listener categorization of musical phrases," Music Perception **4**, 185–214.

Krimphoff, J. (1993). "Analyse acoustique et perception du timbre," unpublished DEA thesis, Université du Maine, Le Mans, France.

Krimphoff, J., McAdams, S., and Winsberg, S. (1994). "Caractérisation du timbre des sons complexes. II : Analyses acoustiques et quantification psychophysique. [Characterization of the timbre of complex sounds. 2. Acoustic analysis and psychophysical quantification]," J. de Phys. **4**(C5), 625–628.

Krumhansl, C. L. (1989). "Why is musical timbre so hard to understand?" in *Structure and Perception of Electroacoustic Sound and Music: Proc. Marcus Wallenberg Symposium*, Lund, Sweden, August, 1988, S. Nielzén and O. Olsson, eds. (Excerpta Medica, Amsterdam), pp. 43–53.

Kuhl, P. K. and Meltzoff, A. N. (1982). "The bimodal perception of speech in infancy," Science **218**, 1138–1144.

Lakatos, S. (2000). "A common perceptual space for harmonic and percussive timbres," Perception and Psychophysics **62**(7), 1426–1439.

Lecanuet, J-P., Granier-Deferre, C., and Busnel, M-C. (1988). "Fetal cardiac and motor responses to octave-band noises as a function of central frequency, intensity and heart rate variability," Early Human Development **18**, 81–93.

Lecanuet, J-P., Granier-Deferre, C., Jacquet, A-Y., and Busnel, M-C. (1992). "Decelerative cardiac responsiveness to acoustical stimulation in the near-term fetus," Quarterly J. Exp. Psychol. **44B**, 279–303.

Liberman, A. M. (1957). "Some results of research on speech perception," J. Acoust. Soc. Am. **29**, 117–123.

Lichte, W. H. (1941). "Attributes of complex tones," J. Exp. Psychol. **28**, 455–480.

Lindsay, P. H. and Norman, D. A. (1977). *Human Information Processing: An Introduction to Psychology*, 2nd ed. (Academic Press, New York).

Locke, S. and Kellar, L. (1973). "Categorical perception in a nonlinguistic mode," Cortex **9**(4), 355–369.

MacDonald, J. and McGurk, H. (1978). "Visual influences on speech perception processes," Perception and Psychophysics **24**, 253–257.

Macmillan, N. A. (1979). "Categorical perception of musical sounds: The psychophysics of plucks and bows," Bull. Psychonomic Soc. **14**, 241 (abstract).

Manoury, P. (1991). "Les limites de la notion de 'timbre'," in *Le timbre: Métaphore pour la composition*, J.-B. Barriere, ed. (Christian Bourgois, Paris), pp. 293–299.

Mathews, M. V., Miller, J. E., Pierce, J. R., and Tenney, J. (1965). "Computer study of violin tones," J. Acoust. Soc. Am. **38**, p. 912 (abstract).

McAdams, S. and Bregman, A. (1979). "Hearing musical streams," Computer Music J. **3**(4), 26–43.

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

McAdams, S. (1993). "Recognition of sound sources and events," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, S. McAdams and E. Bigand, eds. (Oxford University Press, Oxford), pp.146–198.

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.

McAdams, S., Beauchamp, J. W., and Meneguzzi, S. (1999). "Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters," J. Acoust. Soc. Am. **105**(2), 882–897.

McGurk, H. and MacDonald, J. (1976). "Hearing lips and seeing voices," Nature **264**, 746–748.

Miller, G. A. and Heise, G. A. (1950). "The trill threshold," J. Acoust. Soc. Am. **22**, 637–638.

Miller, J. R. and Carterette, E. C. (1975). "Perceptual space for musical structures," J. Acoust. Soc. Am. **58**(3), 711–720.

Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., and Dooling, R. J. (1976). "Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception," J. Acoust. Soc. Am. **60**, 410–417.

Opolko, F. and Wapnick, J. (1987). *McGill University master samples* [CD-ROM] (McGill University, Montreal).

Pastore, R. E. (1976). "Categorical perception: A critical re-evaluation," in *Hearing and Davis: Essays Honoring Hallowell Davis (contributed by present and former colleagues on the occasion of his 80th birthday)*, S. K. Hirsh, D. H. Eldredge, I. J. Hirsh, and S. R. Silverman, eds. (Washington University Press, St. Louis), pp. 253–264.

Pisoni, D. B. (1977). "Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops," J. Acoust. Soc. Am. **61**, 1352–1361.

Pitt, M. A. (1995). "Evidence for a central representation of instrument timbre," Perception and Psychophysics **57**(1), 43–55.

Plomp R. (1970). "Timbre as a multidimensional attribute of complex tones," in *Frequency Analysis and Periodicity Detection in Hearing*, R. R. Plomp and G. F. Smoorenburg, eds. (Sijthoff, Leiden), pp. 397–414.

Plomp, R. (1976). "Timbre of complex tones," in *Aspects of Tone Sensation: A Psychophysical Study*, R. Plomp, ed. (Academic Press, London), pp. 85–110.

Preis, A. (1984). "An attempt to describe the parameters determining the timbre of steady-state harmonic complex tones," Acustica **55**(1), 1–13.

Remez, R. E. (1978). "An hypothesis of event-sensitivity in the perception of speech and bass violins." *Dissertation Abstracts International,* **39** (11-B), 5618-B (University Microfilms No. 7911404).

Remez, R. E., Cutting, J. E., and Studdert-Kennedy, M. (1980). "Cross-series adaptation using song and string," Perception and Psychophysics **27**, 524–530.

Risset, J.-C. and Mathews, M. V. (1969). "Analysis of musical-instrument tones," Physics Today **22**(2), 23–30.

Risset, J-C. and Wessel, D. (1982). "Exploration of timbre by analysis and synthesis," in *The Psychology of Music*, D. Deutsch, ed. (Academic Press, New York), pp. 25–58.

Rosch, E. H. (1973a). "Natural categories," Cognitive Psychology **4**, 328–350.

Rosch, E. H. (1973b). "On the internal structure of perceptual and semantic categories," in *Cognitive Development and the Acquisition of Language*, T. E. Moore, ed. (Academic Press, New York), pp. 111–144.

Rosen, S. M. and Howell, P. (1981). "Plucks and bows are not categorically perceived," Perception and Psychophysics **30**(2), 156–168.

Rosenblum, L. D. and Fowler, C. A. (1991). "Audiovisual investigation of the loudness-effort effect for speech and nonspeech events," J. Exp. Psychol.: Human Percept. Perform. **17**, 976–985.

Rumelhart, D. E. and Abrahamson, A. A. (1973). "A model for analogical reasoning," Cognitive Psych, **5**, 1–28.

Saldana, H. M. and Rosenblum, L. D. (1993). "Visual influences on auditory pluck and bow judgments," Perception and Psychophysics **54**(3), 406–416.

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

Samoylenko, E., McAdams, S., and Nosulenko, V. (1996). "Systematic analysis of verbalizations produced in comparing musical timbres," Intern. J. Psychol. **31**, 255–278.

Samson, S., Zatorre, R. J., and Ramsay, J. O. (1996). "Multidimensional scaling of synthetic musical timbre: Perception of spectral and temporal characteristics," Canadian J. Psychol. **51**, 307–315.

Samuel, A. G. and Newport, E. L. (1979). "Adaptation of speech by nonspeech: Evidence for complex acoustic cue detectors," J. Exp. Psychol.: Human Perception Perform. **5**, 563–578.

Samuel, A. G. (1988). "Central and peripheral representation of whispered and voiced speech," J. Exp. Psychol.: Human Percept. Perform. **14**, 379–388.

Schaeffer, P. (1966). *Traité des objets musicaux* [Treatise on musical objects] (Seuil, Paris).

Serafini, S. (1993). "Timbre Perception of Cultural Insiders: A Case Study with Javanese Gamelan Instruments," unpublished masters thesis, University of British Columbia, Vancouver, Canada.

Schoenberg, A. (1911). *Harmonielehre* [Harmony] (Universal, Leipzig/Vienna) [French translation (1983), Lattes, Paris].

Shepard, R. N. and Arabie, P. (1979). "Additive clustering: Representation of similarity as combinations of discrete overlapping properties," Psychol. Rev. **86**, 87–123.

Shepard, R. N. (1982). "Structural representations of musical pitch," in *The Psychology of Music*, D. Deutsch, ed. (Academic Press, New York), pp. 343–390.

Siegel, J. A. and Siegel, W. (1977). "Categorical perception of tonal intervals: Musicians can't tell sharp from flat," Perception and Psychophysics **21**, 399–407.

Singh, P. G. (1987). "Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre?" J. Acoust. Soc. Am. **82**(3), 886–899.

Smurzynski, J. (1985). "Noncategorical identification of rise time," Perception and Psychophysics **38**(6), 540–542.

Solomon, L. N. (1959). "Search for physical correlates to psychological dimensions of sounds," J. Acoust. Soc. Am. **31**, 492–497.

Strong, W. and Clark, M. (1967a). "Synthesis of wind-instrument tones," J. Acoust. Soc. Am. **41**, 39–52.

Strong, W. and Clark, M. (1967b). "Perturbations of synthetic orchestral wind-instrument tones," J. Acoust. Soc. Am. **41**, 277–85.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., and Cooper, F. S. (1970). "Motor theory of speech perception: A reply to Lane's critical review," Psychol. Rev. **77**, 234–249.

Terhardt, E. (1974). "On the perception of periodic sound fluctuations (roughness)," Acustica **30**, 201–213.

Tversky, A. (1977). "Features of similarity," Psychol. Rev. **84**, 327–352.

Van Heuven, V. J. J. P. and van den Broecke, J. P. R. (1979). "Auditory discrimination of rise and decay time in tone and noise bursts," J. Acoust. Soc. Am. **66**, 1308–1315.

Van Noorden, L. P. A. S. (1975). "Temporal Coherence in the Perception of Tone Sequences," unpublished doctoral dissertation, Eindhoven Univ. of Technology, Eindhoven, Pays-Bas, Germany.

Vogel, A. (1974). "Roughness and its relation to the time-pattern of psychoacoustical excitation," in *Facts and Models in Hearing*, E. Zwicker and E. Terhardt, eds. (Springer-Verlag, Berlin), pp. 241–250.

von Bismarck, G. (1974). "Sharpness as an attribute of the timbre of steady sounds," Acustica **30**, 159–172.

Wedin, L. and Goude, G. (1972). "Dimension analysis of the perception of instrumental timbre," Scandinavian J. Psychol. **13**, 228–240.

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

Wessel, D. L. (1983). *Le concept de recherche en musique*, IRCAM, Paris, Communication.

Wessel, D., Bristow, D., and Settel, Z. (1987). "Control of phrasing and articulation in synthesis," in *Proc. 1987 Int. Computer Music Conf.*, Urbana, IL (Computer Music Assoc., San Francisco), pp. 108–116.

Winsberg, S. and Carroll, J. D. (1989). "A quasi-nonmetric method for multidimensional scaling of multiway data via a restricted case of an extended INDSCAL model," in *Multiway Data Analysis*, R. Coppi and S. Bolasco, eds. (North-Holland, Amsterdam), pp. 405–414.