7

The Effect of Dynamic Acoustical Features on Musical Timbre

JOHN M. HAJDA

1 Introduction

Timbre has been an important concept for scientific exploration of music at least since the time of Helmholtz ([1877] 1954). Since Helmholtz's time, a number of studies have defined and investigated acoustical features of musical instrument tones to determine their perceptual importance, or salience (e.g., Grey, 1975, 1977; Kendall, 1986; Kendall et al., 1999; Luce and Clark, 1965; McAdams et al., 1995, 1999; Saldanha and Corso, 1964; Wedin and Goude, 1972). Most of these studies have considered only nonpercussive, or *continuant*, tones of Western orchestral instruments (or emulations thereof). In the past few years, advances in computing power and programming have made possible and affordable the definition and control of new acoustical variables. This chapter gives an overview of past and current research, with a special emphasis on the time-variant aspects of musical timbre. According to common observation, "music is made of tones in time" (Spaeth, 1933). We will also consider the fact that music is made of "time in tones."

The famous music psychologist Carl Seashore recognized that, of the four major perceptual attributes of tone—pitch, loudness, duration, and timbre—timbre is "by far the most important aspect of tone and introduces the largest number of problems and variables" (Seashore, 1938/1967, p. 21). There are many facets to the complexity of timbre, one of these being the dual categorical and continuous nature of timbre as it is used in real-life musical situations. We categorize familiar musical instruments when we hear them: "that's a piano" or "that's a trumpet." Also, even if we do not know the exact name of an instrument, we can often categorize its sound into its correct instrument family, such as bowed string, woodwind, or brass (Clark et al., 1964). However, musical timbres can also be placed along continua, or dimensions, such that one timbre is said to have more or less of a particular perceptual attribute (or simply attributes) than another does. This concept is more elusive than simple categorization but can be easily demonstrated by auditory morphing [e.g., Slaney et al. (1995)]. Finally, we can assign a considerable range of sounds to the same instrument label; consider, for example, the *chalumeau* versus *clarino* registers of the clarinet, or *sul tasto* versus *sul ponticello* playing on

the violin. Sandell (1998) posits that an instrument's characteristic aural signature, or macrotimbre, is learned by exposure to that instrument playing a variety of spectra over different pitches. A single note may be insufficient to satisfactorily code a macrotimbre after a listener has been exposed to different performances, pitches, loudnesses, and durations.

Researchers have used two basic sets of methods for studying the categorical and continuous nature of timbre. The first set of methods falls under the global term classification, which, as its basic operation, is the partitioning of a collection of objects into groups (Estes, 1994). Therefore, categorization, recognition, and identification are all subsets of classification. The second set of methods utilizes what may be called relational measures. Here, an interval or ratio measure allows for comparisons between classes of objects. A measure of similarity, in which a subject hears a pair of sounds and rates them along a scale between "similar" and "not similar," is one such example. Another example is Verbal Attribute Magnitude Estimation (Kendall and Carterette, 1993a), in which a subject rates a sound along a scale that is anchored by a verbal attribute and its negation, such as "nasal" and "not nasal." Although the boundaries between classification and certain relational measures such as similarity become blurred in theories of cognition [e.g., Estes (1994)], from the point of methodological operations the distinction is still useful.

As mentioned above, most previous research has considered only single, isolated, continuant tones. Researchers have investigated the relative salience of both global time-envelope and spectral characteristics of these tones. In general, the global time-envelope constituents are the attack, the steady state, and the decay. The spectral characteristics are more varied, but generally include the relative energy of upper- and lower-frequency components, frequently measured by the spectral centroid; a feature of the spectral envelope shape called spectral irregularity; and various measures of how the individual frequency components change through time, including mean coefficient of variation and spectral flux. The following section will consider each of these parameters.

2 Global Time-Envelope and Spectral Parameters

What we know is largely determined by what we ask and how we ask it (Kendall and Carterette, 1992). In empirical studies of musical timbre, the types of tones researchers choose to investigate and the way in which their parameters are operationally defined can lead to ambiguous—or even conflicting—results. This is illustrated in the ongoing debate regarding the relative perceptual importance of the global envelope constituents of continuant tones.

2.1 Salience of Partitioned Time Segments

In most research, the global time envelope of an isolated continuant tone consists of its attack, steady state, and decay segments. With regard to the attack

and steady-state segments, past and current findings have supported one of the following three hypotheses:

1. The attack is more salient than the steady state.
2. The attack and steady state are equally salient.
3. The steady state is more salient than the attack.

In many studies, tone segments are artificially created by the imposition of a constant time interval from the beginning (for the attack) or from the end (for the decay) of the musical signal. These time intervals are determined *a priori* and sometimes arbitrarily; most researchers choose either a time from onset that is well into the steady-state portion of each stimulus or a time from onset that covers the longest global amplitude rise-time (e.g., time from onset to the first "significant" local maximum) among the stimuli.

Generally, stimuli are presented one at a time to subjects over loudspeakers or headphones, and subjects employ a classification procedure. For a number of studies that date from the 1960s and 1970s, subjects were asked to name—with or without the aid of a word list—the instrument that most likely produced the tone that they heard. In the literature, this procedure is commonly referred to as identification, although, unless the number of choices is equal to the number of stimuli, a more proper term in experimental psychology is name categorization.

In the early identification studies (Berger, 1964; Clark et al., 1963; Elliott, 1975; Saldanha and Corso, 1964; Wedin and Goude, 1972), the durations of attack- and decay-time segments varied from study to study and were usually on the order of a few hundred milliseconds or less. These segments were imposed on every instrument tone, regardless of the type of instrument. These researchers assumed, for the most part, that attack or decay transient segments occurred within these specified segments; the remainder of the signal was considered to be the steady state. Overall, they found that the removal of the attack segments hindered identification, whereas the removal of the decay segments did not affect identification.

Iverson and Krumhansl (1993) examined the role of onsets in similarity-type judgments. Subjects heard consecutive pairs of tones and rated along a scale of "a little" to "a lot" the degree to which they would have to "change the first sound to make it sound like the second sound" (Iverson and Krumhansl, 1993, p. 2597). Three different stimulus contexts were used: the complete tones, onsets only (the segment measured as 80 ms from the beginning of the signal); onsets removed (the complete tone minus the 80-ms onset segment). The authors found that mean subject ratings for all three contexts corresponded highly with one another. They concluded that "the attributes that are salient for timbral similarity judgments are present throughout tones" (Iverson and Krumhansl, 1993, p. 2602). They surmised that the reason their findings did not jibe with those of the earlier identification studies might have been the difference in subject task.

Campbell and Heller (1978, 1979) introduced the influence of melodic context into the onset role issue. Their stimuli were generated from performances of two-note legato phrases (F_4 at 349.2 Hz to A_4 at 440 Hz) played on six different instruments, including piano. The transitional segment between the two notes was

called the legato transient. This transient was operationally defined as a constant time segment before the start of the second steady state, applied uniformly to each instrument recording. The length of the time segment varied from 20 to 110 ms. They also created constant attack-alone and steady-state-alone contexts—generated from the first tone of the sequence. The authors found that the 110-ms legato transients yielded higher identification than either the attacks, steady states, or any of the other shorter legato transients.

Kendall (1986) pursued this issue in two unique ways: (1) He compared the role of transients and steady state across single-note and legato musical phrase contexts, and (2) he included signal characteristics for each stimulus as bases for his operational definitions of transients. Because he also tested for the effect of musical training (musicians vs nonmusicians), Kendall used a non-verbal matching procedure instead of identification. In musical (melodic) contexts, the steady-state-alone contexts—with the attack and legato transients removed—were matched at a mean level (81%) that was statistically equivalent to the unaltered signals (84%). However, in the single-note contexts, both the steady-state-alone (50%) *and* the attack-alone (51%) contexts were matched at the same level as the unaltered single tones (54%). In comparing his results to those of the earlier identification studies, Kendall (1986, p. 210) concluded that "the perceptual importance of transients in defining the characteristic sounds of instruments has been overstated."

The contradictory results given by the myriad of studies that have explored the salience of time-envelope characteristics—with the exception of Kendall (1986)—are most likely directly due to the lack of robust operational definitions based on signal characteristics. The attack is not a duration; it is a transient part of the signal that lasts from onset until a more-or-less stable periodicity and modes of vibration are established. This "steady state" is generally achieved well before the end of the initial rise time, as determined by amplitude. Contemporary with many of the identification studies in the 1960s, Luce (1963) descriptively examined the characteristic attacks and steady states for 14 nonpercussive instruments of the Western orchestra. Notes were recorded across the entire range of each instrument. His associate, William Strong used two methods to calculate the attack durations (Luce, 1963, p. 90):

1. Amplitude transient: the time from onset to the time when the amplitude reached 90% of the amplitude of the steady state.
2. Structure transient: the time from onset to the time when the waveform had essentially the same shape or structural characteristics as the steady state.

For every instrument except the tuba, the structure transient was measured as shorter than the amplitude transient was. In the case of the flute, the structure transient could not be ascertained because "rather large intensity modulations were present" (p. 92). Strong's measurements for 13 instruments (piccolo was excluded) are presented in abbreviated form in Table 7.1.

On average, Strong's structure transients in Table 1 are 53% as long as the amplitude transients. Luce and Clark (1965) modified the amplitude transient definition to the time necessary for the amplitude to reach 50% (−6 dB IL) of the amplitude

TABLE 7.1. Mean Durations of Amplitude and Structure Attack Transients[a]

Instrument	Mean-amplitude transient (ms)	Mean-structure transient (ms)
Violin	218	88
Viola	106	41
Cello	350	124
Double bass	96	84
Oboe	21	16
English horn	52	29
Bassoon	41	30
Clarinet	60	42
Flute	179	not measured
Trumpet	96	24
French horn	34	24
Trombone	51	36
Tuba	73	95

[a]Data adapted from William Strong (Luce, 1963, Table 8.1.1., p. 91). [From Hajda (1999); used by permission.]

at a point 133 ms further into the signal. So if the measured amplitude transient was 30 ms, the amplitude at 30 ms was equal to 50% of the amplitude at 163 ms. In general, this modification brought the new transients into closer concordance with Strong's structure transients. It is likely, therefore, that contemporary researchers who identify the attack as the time from onset to the global or first "significant" local maximum (e.g., McAdams, et al., 1995; Sandell, 1998) have included a sizable segment of the tone in which periodicity (i.e., pitch) and characteristic harmonic relationships (i.e., timbre) are discernable, even though they have based their operational definition on signal characteristics. It is important to note that the effect of using an amplitude transient over a structure transient depends on the subjective tasks and the manner in which the stimuli were constructed.

Ideally, every constituent segment of a musical tone has a structural element in its operational definition; in other words, the evolution of both global amplitude and spectral components should be considered. In addition, the operational definitions of these segments must be perceptually relevant. Hajda et al. (1997) proposed such a model for the signal partitioning of continuant tones. Part of the impetus for this model, called the "amplitude/centroid trajectory" (ACT), was the observation by Beauchamp (1982) that, for certain continuant signals, RMS amplitude and spectral centroid have a monotonic relationship throughout the steady-state portion of a tone.

The ACT model considers the relationship of amplitude and spectral centroid throughout the duration of a tone. Hajda et al. (1997) identified four consecutive contiguous partitions that are evident in the analyses of most continuant musical instrument signals:

1. Attack: that portion of the signal in which the global RMS amplitude is rising and the spectral centroid is falling after an initial maximum.

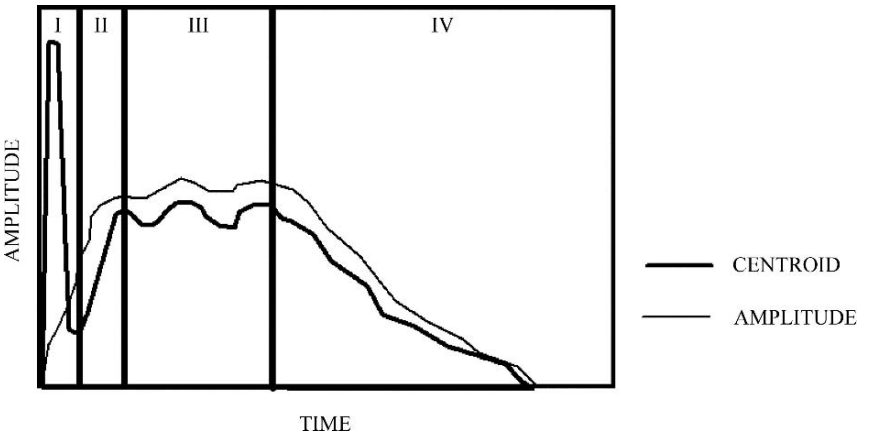

FIGURE 7.1. The RMS-amplitude and spectral-centroid trajectories for a contrived continuant tone. I: Attack; II: Attack/Steady-State Transition; III. Steady State; IV. Decay. [From Hajda (1998), used by permission.]

2. Attack/steady-state transition: the segment from the end of the attack to the first local RMS amplitude maximum.
3. Steady state: the segment during which the amplitude and the centroid both vary around mean values.
4. Decay: the final segment during which the amplitude and centroid both rapidly decrease.

Figure 7.1 illustrates the four ACT segments for a contrived instrument.

Hajda (1996, 1997, 1999) tested the efficacy of this model in a controlled experiment that used single isolated tone stimuli consisting of six "impulse tones" (performed on classical guitar, marimba, piano, pizzicato violin, tubular bell, and xylophone) and six continuant tones (performed on clarinet, flute, oboe, tenor saxophone, trumpet, and bowed violin). The tones were played at concert B_4^b (approximately 466 Hz) in an auditorium and digitally recorded. Two tones from each continuant instrument were used: sustained (about 3.5 s) and staccato (about 600 ms). One tone was recorded from each impulse instrument; because of their different acoustical dampings, the durations of these tones varied. There were 18 unedited tones in all; 12 continuant and 6 impulse. All of the continuant tones used in this study except one manifested characteristics that were consistent with the ACT model. The one exception was a sustained violin tone that was played without an articulated attack.

Continuant tones were partitioned based on three different definitions of attack: (1) fixed attack time from onset to 80 ms into the signal; (2) attack time based on 50% of the average steady-state RMS amplitude, adapted from the operational definition given by Luce and Clark (1965); and (3) the ACT model (Hajda et al., 1997). The partitions for the first two conditions can be described as attack alone

and remainders alone. The partitions for the ACT condition included all possible combinations of the four segments—attack, attack/steady-state transition, steady state, and decay—plus each segment alone. The continuant tones were also subjected to two reverse playback conditions: the entire tone and a 500 ms segment extracted from sustained tones beginning one second after onset.

Nine subjects identified each of the 246 randomly presented stimuli by selecting from a list of the 12 instruments used in the experiment (forced-choice). The probability for a "chance" identification of each stimulus was 8.3%.

The results for continuant tones can be summarized as follows:

1. The unedited signals were correctly identified 93% of the time. The overall results were the same for the unedited sustained and unedited staccato signals, although individual instruments yielded slightly different identifications for different tone durations.
2. For the sustained continuant tones, all three attack-removed conditions yielded a higher percentage of correct identifications than the attack-alone conditions. In addition, the attack-removed conditions yielded results that approached those for the unedited signals. Based on these data, we can conclude that, for these sustained tones, the remainders are more salient than the attacks.
3. For the staccato continuant tones, divergent results were found. For the fixed-80-ms-attack condition, the attacks-alone were identified at a much higher rate than the remainders. In previous studies, the researcher might assume that the removal of the attack adversely affected identification. However, an examination of the raw data showed that the remainders of many of the short signals were confused with impulse instruments (classical guitar, marimba, piano, and pizzicato violin). In fact, removal of the attack was tantamount to imposing an impulse envelope on the staccato tones. In this case, the poor identification results were due to a confounding variable, not experimental control.
4. Therefore, the discussion of the effect of ACT-editing is restricted to the sustained tones. For the sustained ACT conditions, the steady-state-alone edits were identified best. Only the steady-state-alone edits approached the identification rate of the unedited sustained signals (85%–93%). Given all of the above discussion, Hajda (1996, 1997, 1999) concluded that the time-variant steady-state alone is necessary and sufficient for the identification of these isolated sustained continuant tones.
5. For the sustained continuant tones, reverse playback never affected identification.

It seems clear that the process of human identification of an instrument from one of its tones is complex. Listeners can apply a number of strategies, based on the information available. Many of these strategies are determined by the listener's previous knowledge of the instruments' capabilities. Other strategies may stem from basic, seemingly pre-musical distinctions, such as distinguishing an impulse from a continuant envelope. Even in these contrived contexts, it is clear that a single rule will not apply between classes of instruments.

Given the above caveat, it seems that, for sustained continuant tones, the time-variant steady state usually provides sufficient and necessary information for the identification of an instrument. The co-evolution of the amplitude and spectral centroid seems important here, but the direction (i.e., regular vs reverse playback) does not.

The acoustical analyses conducted for this study indicate that when one considers the universe of timbres produced by musical instruments, the issue of attack vs steady state bears little relevance, because impulse instruments cannot be usefully partitioned in such a manner. However, the global RMS amplitude and spectral centroid trajectories and their functional relationship are characteristics of all musical (i.e., time-variant) sounds. Research by Hajda (1998, 1999) focused on the salience of these—and other—trajectories; the results of this preliminary work are reported in the following section.

A caveat should be issued regarding the nature of the attacks of nonpercussive instrument tones. A plethora of measurements made by Luce and his colleagues showed that "the duration of the attack transients depends upon the instrument played, upon the note played on the instrument, and upon the performer, but very little on the dynamic marking at which the instrument is sounded or the duration of the notes played, or whether or not the instrument is played with vibrato" (Luce and Clark, 1965, p. 199). We can add other variables that will probably affect the duration of attack transients, including characteristics of the musical phrase (legato, staccato, etc.), musical style, texture (counterpoint, homophony, heterophony), and other musical contexts.

Finally, although various classification paradigms are simple to operationalize and implement in laboratory experiments, we should question the relevance of classification to the "real world" of musical timbre. To what extent do performers or listeners recognize, categorize, or even identify timbres in the course of their musical experience, Benjamin Britten's *Young Person's Guide to the Orchestra* (1946) notwithstanding? Certainly, orchestration requires a high level of knowledge regarding the timbral characteristics of each instrument of the ensemble. However, more often than not in Western music, timbres are heard in combination. It is not enough to "know" the timbre of a B^b trumpet that is playing "open middle C"; the orchestrator must know how that trumpet tone will sound in the context of a brass quintet, or as part of a jazz ensemble, or part of a marching band. Musical timbre does not operate as a series of unrelated, isolated entities. Every ensemble operates within its own timbral framework, or palette (see Martens, 1985), in a manner analogous to a painter's palette of color. Even a solo instrumentalist manipulates timbre in order to produce "coloristic" effects.[1] More often than

[1] This is particularly true for instruments with multiple degrees of freedom. Consider the classical guitar, on which a given note can be alternatively fingered (stopped) on several strings. Each fingering produces a slightly different timbre due to the physical characteristics (thickness, winding) of the different strings and resonant properties of the instrument. In addition, the right hand can produce a myriad of tonal qualities by plucking the string with different combinations of flesh and nail as well as varying locations relative to the bridge.

not, however, the physical correlates for a palette of timbre are more difficult to determine than those for visual color.

2.2 *Relational Timbre Studies*

Relational measures have been used since the middle part of the 20th century in a variety of experimental contexts. Although this review is by no means exhaustive, it is intended to give the reader an idea about the types of timbre studies that have been conducted as well as a convergence of the findings.

In order to find a palette (or representative geometric structure) for timbre, we must be able to determine its dimensions. Such a determination has been made for pitch. Shepard (1982) has summarized and demonstrated models of Western musical pitch structures that can be expressed in two dimensions (circle of fifths), three dimensions (simple helix), four dimensions (double helix wrapped around a torus), and even five dimensions (double helix wrapped around a helical cylinder)! For a number of reasons, the dimensions for timbre are not nearly so well delineated.

Researchers have used two basic approaches to uncovering the structure of timbre. The first is to directly measure specified attributes of timbre by means of a subject's assignment of a value along a scale of adjectival polar opposites, such as "dullness" and "brightness." This technique, commonly known as the semantic differential (Osgood et al., 1957), is considered a measurement of the meaning of a stimulus and has been used to study other facets of music besides timbre. Lichte (1941) and von Bismarck (1974) used versions of this approach. They constructed steady-state synthetic stimuli with varying spectral characteristics and constant temporal envelopes in order to isolate verbal factors that would identify salient perceptual features. Lichte (1941) found a primary relationship between "brightness" and the midpoint of the energy distribution among frequency partials; von Bismarck (1974) found a similar primary relationship for his stimuli and "sharpness." In their study with dyads produced by recording natural wind instrument performances, Kendall and Carterette (1993a) used English translations of von Bismarck's (1974) semantic differential. They found that these adjectives did not significantly differentiate their stimuli. They replicated the study but replaced the semantic differential adjectives, for example, "dull" and "sharp," with an adjective and its negation, such as "sharp" and "not sharp." This procedure, known as Verbal Attribute Magnitude Estimation (VAME), was used on the same stimuli in a subsequent experiment (Kendall and Carterette, 1993b). This time, the verbal attributes came from a descriptive text on orchestration (Piston, 1955). These final ratings produced the most interpretable results, among them a primary

---

Other instruments, such as the trumpet, maintain timbral control by the prolonged coupling between the energy source (player) and vibrating body. Some of these instruments can also take advantage of additional physical couplings, such as a mute, in order to significantly alter their aural characteristics. From this perspective, instruments such as the piano are impoverished in terms of their degrees of freedom with respect to timbre.

relationship between "nasality" and the relative amount of steady-state energy in the upper frequency partials as compared to the fundamental.

The second approach to determining timbral structures is based on obtaining perceptual qualitative relationships between stimuli, as opposed to directly measuring timbral attributes. Subjects' rating scores are obtained from a direct method of similarity analysis (Ekman, 1965). After hearing a pair of consecutively presented tones, the subject rates how similar those tones sound in relation to the other pairs in the stimuli set. The ratings for every possible paired comparison are then mathematically transformed into distances in a geometrical (usually Euclidean) space. This statistical analysis is commonly referred to as multidimensional scaling, or MDS. There are a number of MDS algorithms, each of which differs slightly in its intricacies.[2] The basic purpose of these procedures is the same: Produce a geometric configuration in which stimuli that are similar appear close together and those that are dissimilar appear far apart. Then, it is up to the researcher to interpret this configuration in terms of the characteristics of the stimuli.

In general, the nature and number of the stimuli limit the number of interpretable dimensions. In a paired-comparisons paradigm, the number of judgments that a subject must make is

$$n = \frac{s(s \pm 1)}{2}, \tag{7.1}$$

where s is the number of stimuli. The quantity $(s + 1)$ is used for experiments that include identities—stimulus x is paired with itself—and $(s - 1)$ without identities.[3] Therefore, a paired-comparison similarity experiment with 25 stimuli requires 325 judgments by a subject with identities, 300 without. If each stimulus is 3 s and a subject requires 5 s for each response, the entire experiment will take about 1 h, not including the time needed for instructions and any practice experiments. In this author's experience, many subjects cannot remain focused for such a duration. In fact, most of the similarity studies conducted for musical timbre have used between 10 and 20 stimuli. The MDS spaces for these experiments have produced interpretable solutions for two or three dimensions. Such is the case with Fig. 7.2, a space generated by the similarity ratings for 11 continuant instruments of the Western orchestra (Kendall et al., 1999).

The interpretation of the dimensions of an MDS space requires a good deal of intuition on the part of the researcher. In general, researchers attempt to find musical and extramusical correlates with each dimension of the solution. The musical correlates might include proximity groupings by instrument family (Wessel, 1973; Grey, 1975), pitch (Miller and Carterette, 1975), or the degree of blend for two simultaneously produced timbres (Kendall and Carterette, 1993c, Sandell, 1995). The extramusical variables are typically verbal attributes (Faure et al., 1996;

[2] For an overview of MDS and related procedures, see Kruskal and Wish (1978) and Arabie et al. (1987).

[3] Although the case of stimulus I paired with itself is obviously trivial, it may be advantageous to include such a pairing in order to identify subjects who produce outlying data.

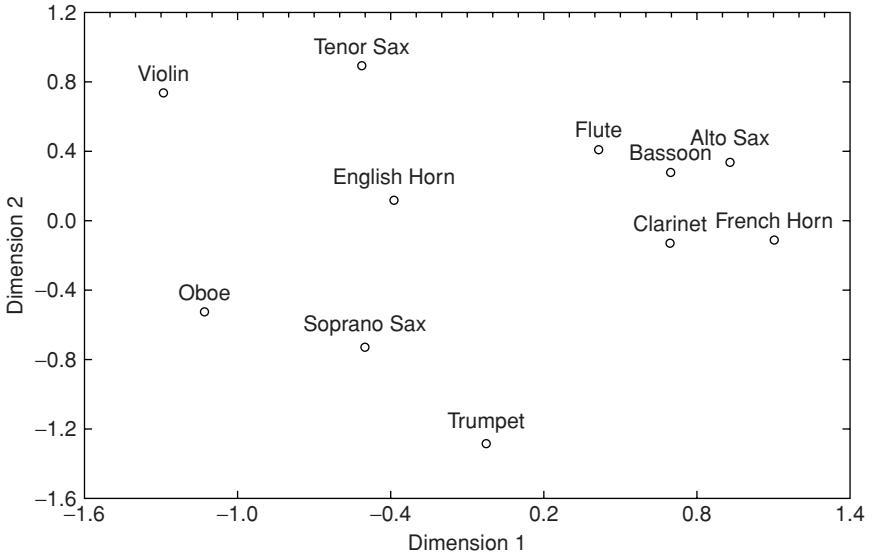

FIGURE 7.2. Two-dimensional MDS solution for similarity ratings of eleven natural instrument tones played at concert B_4^b (ca. 466 Hz). [Reprinted from Kendall et al. (1999). ©1999 by The Regents of the University of California. All rights reserved. Used with permission.]

Kendall et al., 1999) or acoustical parameters. The focus here will be on the correlation of acoustical parameters to the dimensions of MDS solutions.

In general, three acoustical parameters repeatedly appear as correlates to dimensional solutions in timbre studies:

1. Amplitude-vs-time (temporal) envelope, usually expressed in terms of attack or rise times.
2. Spectral energy distribution across frequency components.
3. Spectral variance in terms of the amplitudes of frequency components.

2.2.1 Temporal Envelope

In studies that include both continuant and impulse stimuli, the amplitude-vs-time envelope (aka temporal envelope or amplitude envelope)—in one manifestation or another—is the acoustical correlate to the primary perceptual dimension (Krumhansl, 1989; Iverson and Krumhansl, 1993; McAdams et al., 1995; Kendall et al., 1999). For the most part, researchers have characterized the envelope phenomenon as an issue of attack time; after all, impulse instruments have very brief attacks (less than 10 ms) in comparison to continuant instruments. Therefore, most measures of attack should yield high correlations with a dimension that separates percussive from nonpercussive stimuli. Krimphoff (1993) and McAdams et al. (1995) found precisely such a relationship when they correlated the primary dimension of an MDS space generated by the similarity scaling of impulse and

continuant timbres [from Krumhansl (1989)] with the log-rise-time ("logarithme du temps de montée") of each stimulus. They defined log-rise-time as

$$\text{Log-rise-time} = \log_{10}(t_{\max} - t_{\text{thresh}}), \tag{7.2}$$

where $t_{\max}$ is the time from onset to maximum RMS amplitude and t_{thresh} is the time from onset to a threshold taken as 2% of the amplitude at $t_{\max}$.

2.2.2 Spectral Energy Distribution

Many acousticians have described the steady-state portion of continuant tones in terms of a long-time-average spectrum. The amplitude and frequency components of the two-dimensional spectrum are analogous to a series of weights and distances along a beam. The point at which the sum of moments (weight × distance) equals zero is the fulcrum, or, in the case of the spectrum, the spectral centroid. Such an index for measuring the "quality of a musical instrument" was first described by Knopoff (1963, p. 229).[4] To this author's knowledge, the first correlations of spectral centroid and a perceptual dimension were published by Ehresman and Wessel (1978) and Grey and Gordon (1978). Although the formulas vary in detail, these and later studies use a representative long-time average spectrum such that

$$f_{\text{centroid}} = \frac{\sum_{n=1}^{N} f_n \cdot A_n}{\sum_{n=1}^{N} A_n}, \tag{7.3}$$

where f_n is the frequency and A_n is the amplitude (usually linear) of the nth partial of a spectrum with N frequency components. This equation yields a measure in frequency units, which will suffice in instances where the fundamental frequencies of the stimuli are the same. It is also possible to produce a unitless measure by (1) replacing f_n with the harmonic number or (2) multiplying the denominator by the fundamental frequency.

The Pearson correlation of spectral centroid with Dimension 1 of the two-dimensional MDS space shown in Fig. 2 is 0.9 (Kendall et al., 1999). This result is consistent with other research that has yielded strong correlations between spectral centroid and the primary perceptual dimension of MDS spaces for continuant stimuli [e.g., Ehresman & Wessel (1978); Grey & Gordon (1978)] and secondary perceptual dimension of spaces for mixed impulse and continuant stimuli [e.g., McAdams et al. (1995); Lakatos (2000)].

[4] Knopoff (1963) used the term *center of gravity* (from engineering statics) instead of *spectral centroid*. In fact, his measure involved taking the ratio of (1) a theoretical center of gravity calculated by replacing the amplitude of each frequency partial with the moment of that frequency in the original signal, and (2) the center of gravity from the original signal.

2.2.3 Spectral Time Variance

The individual amplitudes of frequency components for many continuant signals vary significantly throughout the duration of a tone. This dynamic feature has been given a number of labels, among them: Spectral Fluctuation (Grey, 1977); Spectral Variation (Ehresman and Wessel, 1978); Spectral Flux (Krumhansl, 1989); and Time Variance (Kendall and Carterette, 1993b). In spite of the number of phenomenological observations made since Grey (1975), spectral time variance was not quantified until the 1990s.

Kendall and Carterette (1993b) calculated a mean coefficient of variation (MCV):

$$MCV = \frac{\sum_{n=1}^{N=9} \frac{\sigma_n}{\mu_n}}{N}, \tag{7.4}$$

in which σ_n is the standard deviation of the amplitude of frequency component n across time, μ_n is the mean amplitude of component n, and N is the number of frequency components analyzed, in this case $N = 9$. The mean coefficient of variation yielded a moderately strong correlation ($r = 0.7$) with the second dimension of the perceptual space generated by Kendall et al. (1999) shown in Fig. 2.

Krimphoff (1993) examined three different measures of spectral flux ("flux spectral") in order to find the strongest relationship with the third dimension of an MDS space generated by Krumhansl (1989). The first, Spectral Variation ("variation spectrale"), was determined by taking the correlation of respective harmonics of adjacent instantaneous spectra (each corresponding to a single window of analysis of duration $\Delta t = 16$ ms). The absolute values of these correlations were summed and averaged across the entire duration of the tone. The second parameter, Flu*x* ("flux"), was measured as the mean deviation of the spectral centroid of each analysis window with respect to the long-time average measure of spectral centroid. The final parameter, Coherence ("cohérence"), is a measure of the difference in onset times for each harmonic. The term, however, is a bit misleading because a signal in which every harmonic has the same time-to-onset has a coherence value equal to zero; a signal in which harmonics do not have the same time-to-onset has a coherence value greater than zero.

Krimphoff (1993) also examined the relationship of two measures of Fine Spectral Structure ("structure fine du spectrale") to the third dimension of the Krumhansl (1989) MDS space. The first measure was taken from Guyot (1992). It is essentially a ratio with the sum of the energy in the odd-numbered harmonics above the fundamental taken to be the numerator and the sum of the energy in the fundamental plus the energy in the even-numbered harmonics taken to be the denominator. The final parameter, which Krimphoff (1993) called the Spectral Deviation ("déviation"), is the sum of deviations of each harmonic log-amplitude from the mean of three consecutive harmonic log-amplitudes (centered on that harmonic), normalized by a global mean log-amplitude. This parameter, which yielded the highest correlation with Krumhansl's perceptual dimension, has been renamed by Krimphoff et al.

(1994) and McAdams et al. (1995) as Spectral Irregularity and most recently as Spectral Envelope Smoothness by McAdams et al.(1999). Kendall and Carterette (1996) used the following linear version of Krimphoff et al.'s (1994) log-based formula to calculate the linear spectral irregularity (LSI) of static synthetic stimuli:

$$LSI = \frac{\sum_{n=2}^{N-1} \left| A_n - \frac{A_{n+1} + A_n + A_{n-1}}{3} \right|}{\sum_{n=1}^{N} A_n}, \tag{7.5}$$

where A_n is the linear amplitude of the nth harmonic and N is the number of harmonics. A spectral smoothing paradigm used by McAdams et al. (1999) also used linear amplitudes.

In summary, depending on the nature of the stimuli, both long-time average (spectral centroid, spectral irregularity) and time-variant (rise time, mean coefficient of variation) acoustical measures are principal correlates with perceptual spaces generated by relational measures. The experimental control of these acoustical variables has only begun in recent years. Kendall and Carterette (1996) determined difference thresholds for synthetic timbres that varied only in spectral centroid. Jeong and Fricke (1998) found an effect of listening position and reverberation on these difference thresholds. In a separate study, Kendall and Carterette (1996) synthesized timbres with the same centroid but different spectral shapes. These timbres were compared in a separate relational study; as might be expected, spectral irregularity [defined by Eq. (7.5)] correlated very highly with the principal MDS dimension.

3 The Experimental Control of Acoustical Variables

Two recent studies have examined—at least in part—the experimental control of time-variant acoustical variables for tones that were originally produced by acoustical instruments.

McAdams et al. (1999) applied six basic data simplifications and five combinations of these simplifications to seven instrument tones. Five of the instruments were continuant—clarinet, flute, oboe, trumpet, and violin—and two were impulse—harpsichord and marimba. The simplifications are briefly described as follows:

1. Amplitude-Envelope Smoothing: removal of micro time-variations of harmonic amplitudes over the steady-state and decay portions of the tone.
2. Amplitude-Envelope Coherence (spectral envelope fixing): removal of spectral flux while preserving the average spectrum and global RMS envelope over the entire duration of the tone.
3. Spectral-Envelope Smoothness: linear smoothing of the jaggedness or irregularity of a spectral envelope over the entire duration of the tone.

4. Frequency-Envelope Smoothness: removal of micro time-variations of the frequencies of harmonics over the entire duration of the tone.
5. Frequency-Envelope Coherence (harmonic frequency tracking): removal of inharmonicity over the entire duration of the tone.
6. Frequency-Envelope Flatness: removal of frequency variations and inharmonicity over the entire duration of the tone.

Of these six data reduction techniques, numbers 1, 2, 4, and 6 remove a certain amount of time-variance. In all, McAdams et al. (1999) tested the salience of 11 methods of signal simplification: the six methods mentioned above and five combinations of these methods. Listeners were asked to discriminate between (1) sounds that were resynthesized with simplified data and (2) reference sounds that were synthesized versions of the original signal. All analyses and syntheses were conducted with phase-vocoder analysis and oscillator-bank additive synthesis algorithms contained in the SNDAN music sound analysis/synthesis package (Beauchamp, 1993). Overall, the authors found that only amplitude envelope coherence, or the removal of spectral flux, yielded a "very good" proportional mean discrimination (0.91) among the variables that controlled for time-variance. The means of discrimination for other time-variant variables ranged between 0.66 and 0.71; the probability of discrimination due to chance was 0.50. The highest mean discrimination was for spectral envelope smoothing (0.96). In general, edits that combined methods of simplification yielded means of discrimination that were equal to or slightly higher than those for the most salient individual method.

Hajda's pilot study (1998, 1999) investigated the effects of controlling certain time-variant acoustical parameters of continuant tones. The 10 instrument tones used for this research come from the McGill University Master Samples, or MUMS, set of digital recordings (Opolko and Wapnick, 1989): alto flute, cello, clarinet, C trumpet,[5] English horn, French horn, flute, oboe, trombone, and violin. The sustained tones were played at concert B_4^b, or approximately 466 Hz. This pitch is within the normal playing range of all of these instruments although it is toward the high end of the range for some of the instruments.

Three time-variant parameters were controlled in this experiment: global RMS amplitude, spectral amplitude envelope, and frequency deviation for each spectral component. The MUMS signals were trimmed by imposing a 40 dB threshold below the maximum amplitude so that noise floor effects would be minimized when the experimental controls were implemented. Segments of 1.1 s duration were extracted for each of the nine edits beginning 500 ms into each signal. The rationale for this was the finding that relevant timbral information is present in the steady-state portions of sustained continuant tones (Hajda, 1996, 1997, 1999). Linear 50 ms fade-ins and fade-outs were imposed on each edit. The original digital signal was edited in the same fashion for experimental control purposes.

[5] The more common B^b trumpet is not available from the MUMS recordings.

TABLE 7.2. Summary of edits used in Hajda (1998)[a]

Simplification	Frequency deviation	Spectral flux	Global RMS amplitude
SYNTH	Varies	Varies	Varies
FRQ	Controlled	Varies	Varies
SPC	Varies	Controlled	Varies
AMP	Varies	Varies	Controlled
FR/SP	Controlled	Controlled	Varies
FR/AM	Controlled	Varies	Controlled
AM/SP	Varies	Controlled	Controlled
S. S.	Controlled	Controlled	Controlled

[a]SYNTH = full phase-vocoder synthesis; FRQ = remove all frequency deviations; SPC = remove spectral flux; AMP = remove global amplitude variation; FR/SP = combined removal of frequency deviations and spectral flux; FR/AM = combined removal of frequency deviations and global amplitude variation; AM/SP = combined removal of global amplitude and spectral flux; S.S. = true steady state. [From Hajda (1999); used by permission.]

The following spectrotemporal simplifications were made using SNDAN (Beauchamp, 1993, 1998):

1. SYNTH: full (unmodified) phase-vocoder resynthesis.
2. FRQ: replace all frequency deviations by a fixed average frequency for each harmonic.
3. SPC: remove spectral flux by imposing an average spectrum for the duration of the signal during which relative amplitudes of the harmonics are fixed, but the overall RMS amplitude time-variation is preserved.
4. AMP: remove global amplitude variation by imposing a fixed average RMS amplitude on the overall signal while allowing the harmonic relationships to vary relatively as in the original sound.
5. FR/SP: combination of 2 and 3.
6. FR/AM: combination of 2 and 4.
7. AM/SP: combination of 3 and 4.
8. S.S.: combination of 2, 3, and 4 (a steady-state condition).

These simplifications are summarized in Table 7.2.

A relational procedure was employed in which seven subjects rated the dissimilarity of the original digital tone with each of the eight synthesized edits. A zero rating indicated no discriminable difference between the original tone and the synthesized edit. A 100 rating indicated maximum dissimilarity (among all 90 comparisons).

Figures 7.3 and 7.4 show the mean dissimilarity ratings for the alto flute and clarinet edits. For the alto flute (Fig. 7.3), zeroing frequency deviation (FRQ) has no real effect on subject ratings, fixing global RMS amplitude (AMP) has a moderate effect, and removing spectral flux (SPC) has the strongest effect. Multiple controls increase the dissimilarities between the original and edited tones. By comparison, none of the edits for the clarinet tone (Fig. 7.4) has a significant effect on dissimilarity ratings. Informal listening indicated that the alto flute was played with a deep vibrato while the clarinet tone was played without vibrato.

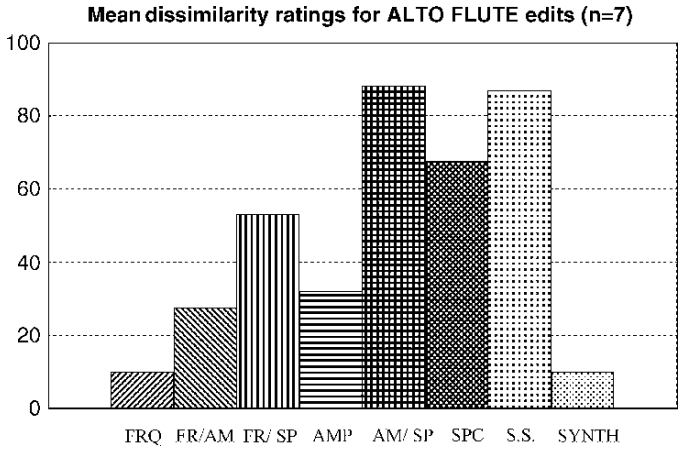

FIGURE 7.3. Mean dissimilarity ratings of seven subjects for the comparison of the original alto flute tone with nine synthetic edits. FRQ = removal of all frequency deviations; FR/AM = combined removal of frequency deviations and global amplitude variation; FR/SP = combined removal of frequency deviations and spectral flux; AMP = removal of global amplitude variation; AM/SP = combined removal of global amplitude and spectral flux; SPC = removal of spectral flux; S.S. = true steady state; SYNTH = full phase-vocoder resynthesis. [From Hajda (1999); used by permission.]

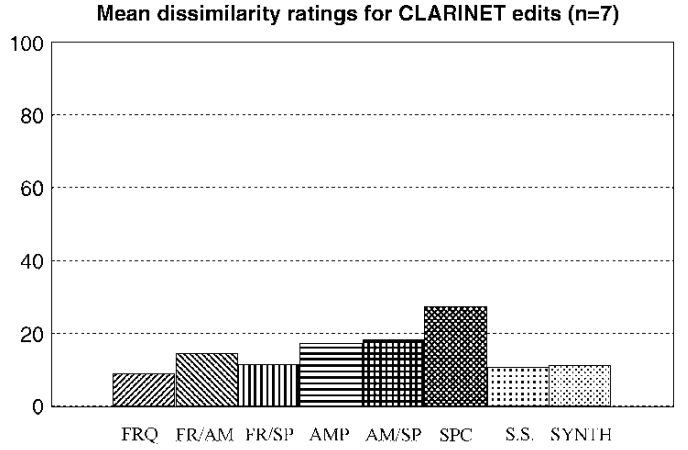

FIGURE 7.4. Mean dissimilarity ratings of seven subjects for the comparison of the original clarinet tone with nine synthetic edits. FRQ = removal of all frequency deviations; FR/AM = combined removal of frequency deviations and global amplitude variation; FR/SP = combined removal of frequency deviations and spectral flux; AMP = removal of global amplitude variation; AM/SP = combined removal of global amplitude and spectral flux; SPC = removal of spectral flux; S.S. = true steady state; SYNTH = full phase-vocoder resynthesis. [From Hajda (1999); used by permission.]

Data analysis indicates the following trends:

1. As one might expect, instruments played with vibrato were affected the most by the acoustical simplifications. However, several instruments played without vibrato—the English horn, oboe, and C trumpet—were affected a moderate amount by the controls. Other instruments played without vibrato—namely, the clarinet, French horn, and trombone—were not affected by the controls.
2. Averaged across all 10 instruments, the mean dissimilarity ratings for zeroing frequency deviations ($\mu = 20.0$) and global amplitude variations ($\mu = 22.5$) are not much different from those of the full resyntheses ($\mu = 16.0$). Removal of spectral flux has a much greater effect on the dissimilarity ratings ($\mu = 42.7$), and, as one might expect, the greatest effect occurs with the steady-state condition ($\mu = 47.5$).

These results are consistent with the findings of McAdams et al. (1999); this is especially interesting given the difference in method (dissimilarity rating versus discrimination).

4 Conclusions and Directions for Future Research

At this point, we can conclude that spectral flux (time variation of the normalized spectrum) is the most salient time-variant parameter of natural continuant tones (Hajda, 1998; Kendall et al., 1999; McAdams et al., 1999). McAdams et al. (1999) found that discrimination of a controlled acoustical variable was strongly correlated to the extent to which it actually varied in the original signal. By a common sense extension, if a parameter varies significantly in a signal, we can hypothesize that a signal resynthesized with the parameter made static will be perceived as significantly different from the original.

In spite of current advances, the salience of time-variant parameters in musical tones is far from fully understood. Part of this is due to the complexity of the musical instrument as a vibrational system, especially in instances in which the performer (driver) maintains a coupling with the generator and resonator. Such is the case with continuant instruments, where the performer controls the time-variant aspect of timbre in an expressive fashion that itself varies from one performance to another.

The next logical extension of this line of research involves musical context. Campbell and Heller (1979) and Kendall (1986) have already conducted work regarding the effect of legato melodic phrases on the classification of timbre. While it is clear that the connection of notes in a melody is important, the manner by which these notes connect has not been investigated in a systematic and controlled fashion. The roles of the time-variant aspects of timbre in a host of other musical contexts, such as expressiveness, dynamics, style, etc., have not been addressed. In addition, orchestral instruments rarely play in an isolated context. The effect of time variance in the presence of vertical combinations of timbres must also be considered.

To this point, the time-variant parameters of impulse signals have not been discussed. This is due to the lack of systematic research on this class of tones. Hajda (1995, 1996, 1997, 1999) found that impulse tones differ from continuant tones in several important ways:

1. Operational definitions of tone segments for continuant signals do not apply to impulse tones, since impulse signals contain no steady state (Hajda, 1996, 1997, 1999).
2. The identification of impulse signals is significantly affected by reverse playback; the identification of continuant signals is not (Hajda, 1996, 1997, 1999).
3. The identification of impulse tones is not affected by any type of partitioning, whether the segment that is presented to listeners is taken from the beginning or middle of a signal; the identification of continuant tones is affected by such signal editing (Hajda, 1997, 1999).
4. The long-time average spectral centroid is the strongest correlate to the primary perceptual dimension of an MDS space generated from the ratings of continuant tones; the *change* in centroid over time is one of several correlates for the primary perceptual dimension of an MDS space generated from the ratings of impulse tones (Hajda, 1995).

The above findings do not jibe entirely with other research (Freed, 1990; Serafini, 1995; Lakatos, 2000). Even if they did, the paucity of research would not warrant generalizations to the entire class of impulse instruments.

As stated by McAdams et al. (1999), two overall goals of research on the time-variant parameters of musical instrument tones are:

1. To facilitate realistic sounding resyntheses with a minimum of control variables.
2. To increase our understanding of the perception of timbre.

As such, musicians of diverse genres—from electronic music composers to orchestrators to music theorists—may benefit from these studies. However, because of the interdisciplinary nature of the research questions, musicians must team with physicists, engineers, and psychologists in order to unravel the mysteries of the "time in tones."

References

Arabie, P., Carroll, J. D., and DeSarbo, W. S. (1987). *Three-Way Scaling and Clustering*. Sage university papers. Quantitative applications in the social sciences; no. 07–065. (Sage Publications, Beverly Hills and London).

Beauchamp, J. W. (1982). "Synthesis by spectral amplitude and 'brightness' matching of analyzed musical instrument tones," J. Audio Eng. Soc. **30**(6), 396–406.

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

Beauchamp, J. W. (1998). "Methods for measurement and manipulation of timbral physical correlates," *Proc. 16th International Congress on Acoustics and 135th Meeting of the*

Acoustical Society of America, 1998, Seattle, Vol. 3, P. K. Kuhl and L. A. Crum, eds. (Acoustical Society of America, Woodbury, NY), pp. 1883–1884.

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

Britten, B. (1946). *Variations and Fugue on a Theme of Henry Purcell (The Young Person's Guide to the Orchestra)* (Boosey & Hawkes, London and New York).

Campbell, W. C. and Heller, J. J. (1978). "The contribution of the legato transient to instrument identification," *Proc. Research Symposium on the Psychology and Acoustics of Music, 1978*, University of Kansas, Lawrence, KS, pp. 30–44.

Campbell, W. and Heller, J. (1979). "Convergence procedures for investigating music listening tasks," Bull. Council for Res. Music Educ. **59**, 18–23.

Clark, M., Jr., Luce, D., Abrams, R., Schlossberg, H., and Rome, J. (1963). "Preliminary experiments on the aural significance of parts of tones of orchestral instruments and on choral tones," J. Audio Eng. Soc. **11**(1), 45–54.

Clark, M., Jr., Robertson, P. T., and Luce, D. (1964). "A preliminary experiment on the perceptual basis for musical instrument families," J. Audio Eng. Soc. **12**(3), 199–203.

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

Ekman, G. (1965). "Two methods for the analysis of perceptual dimensionality," Perceptual and Motor Skills **20**, 557–572.

Elliott, C. A. (1975). "Attacks and releases as factors in instrument identification," J. Res. Music Educ. **23**(1), 35–40.

Estes, W. K. (1994). *Classification and Cognition* (Oxford University Press, New York).

Faure, A., McAdams, S., and Nosulenko, V. (1996). "Verbal correlates of perceptual dimensions of timbre," *Proc. 1996 Int. Conf. on Music Perception and Cognition*, Montreal (Faculty of Music, McGill University, Montreal), pp. 79–84.

Freed, D. J. (1990). "Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events," J. Acoust. Soc. Am. **87**(1), 311–322.

Grey, J. M. (1975). *An Exploration of Musical Timbre* (Report STAN-M-2, CCRMA, Dept. of Music, Stanford University, Stanford, CA).

Grey, J. M. (1977). "Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. **61**(5), 1270–177.

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

Guyot, F. (1992). "Etude de la pertinence de deux critères acoustiques pour caractériser la sonorité des sons à spectre reduits," Unpublished D.E.A. thesis, Université du Maine, Le Mans, France.

Hajda, J. M. (1995). "The relationship between perceptual and acoustical analyses of natural and synthetic impulse signals," masters thesis, University of California, Los Angeles, 1995, *Masters Abstracts International*, **33**(6). (University Microfilms International Publications No. 13–61, 681)

Hajda, J. (1996). "A new model for segmenting the envelope of musical signals: The relative salience of steady state versus attack, revisited," *101st Convention of the Audio Engineering Society*, Los Angeles, Audio Eng. Soc. Preprint 4391.

Hajda, J. M. (1997). "Relevant acoustical cues in the identification of Western orchestral instrument tones" (abstract), J. Acoust. Soc. Am. **102**(5), pt. 2, 3085.

Hajda, J. M., Kendall, R. A., Carterette, E. C., and Harshberger, M. L. (1997). "Methodological issues in timbre research," in *Perception and Cognition of Music,* I. Deliège and J. Sloboda, eds. (Psychology Press, Hove, UK), pp. 253–306.

Hajda, J. M. (1998). "The effect of amplitude and centroid trajectories on the timbre of percussive and nonpercussive orchestral instruments," *Proc. 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America,* Vol. 3, Seattle (Acoustical Society of America, Woodbury, NY), pp. 1887–1888.

Hajda, J. M. (1999)."The Effect of Time-Variant Acoustical Properties on Orchestral Instrument Timbres," doctoral dissertation, University of California, Los Angeles. UMI number 9947018.

Helmholtz, H. L. F. ([1877], 1954). *On the Sensations of Tone as a Psychological Basis for the Theory of Music* (Dover, New York).

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

Jeong, D., and Fricke, F. R. (1998). "The dependence of timbre perception on the acoustics of the listening environment," *Proc. 16th Int. Congress on Acoustics and 135th Meeting of the Acoustical Society of America*, Vol. 3, Seattle (Acoustical Society of America, Woodbury, NY), pp. 2225–2226.

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

Kendall, R. A. and Carterette, E. C. (1992). "Convergent methods in psychomusical research based on integrated, interactive computer control," Behavior Research Methods, Instruments, and Computers **24**(2), 116–131.

Kendall, R. A. and Carterette, E. C. (1993a). "Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck's adjectives," Music Perception **10**(4), 445–468.

Kendall, R. A. and Carterette, E. C. (1993b). "Verbal attributes of simultaneous wind instrument timbres: II. Adjectives induced from Piston's 'Orchestration'," Music Perception **10**, 469–502.

Kendall, R. A. and Carterette, E. C. (1993c). "Identification and blend of timbres as a basis for orchestration," Contemp. Music Rev. **9**(1/2), 51–67.

Kendall, R. A. and Carterette, E. C. (1996). "Difference thresholds for timbre related to spectral centroid," *Proc. 4th Int. Conference on Music Perception and Cognition,* Montreal, Canada, (Faculty of Music, McGill University, Montreal), pp. 91–95.

Kendall, R. A., Carterette, E. C., and Hajda, J. M. (1999). "Perceptual and acoustical features of natural and synthetic orchestral instrument tones," Music Perception **16**(3), 327–363.

Knopoff, L. (1963). "An index for the relative quality among musical instruments," Ethnomusicology **7**(3), 229–233.

Krimphoff, J. (1993). "Analyse acoustique et perception du timbre," unpublished D.E.A. 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 Physique **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: Proceedings of the Marcus Wallenberg Symposium held in Lund, Sweden, on 21–28 August 1988,* S. Nielzen and O. Olsson, eds. (Excerpta Medica, Amsterdam), pp. 43–53.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional Scaling*, Sage university papers, Quantitative applications in the social sciences, no. 07–011 (Sage Publications, Beverly Hills and London).

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

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

Luce, D. A. (1963). *Physical Correlates of Nonpercussive Musical Instrument Tones*, unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Luce, D. and Clark, M. (1965). "Durations of attack transients of nonpercussive orchestral instruments," J. Audio Eng. Soc. **13**(3), 194–199.

Martens, W. L. (1985). "*Palette*: An environment for developing an individualized set of psychophysically scaled timbres," *Proc. 1985 International Computer Music Conference*, Simon Fraser University, Burnaby, British Columbia, (Computer Music Association, San Francisco), pp. 355–365.

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," Psych. Res. **58**(3), 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.

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

Opolko, F. and Wapnick, J. (1989). *McGill University Master Samples User's Manual* (Faculty of Music, McGill University, Montreal).

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning* (University of Illinois Press, Urbana, IL).

Piston, W. (1955). *Orchestration* (W. W. Norton, New York).

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

Sandell, G. J. (1995). "Roles for spectral centroid and other factors in determining 'blended' instrument pairings in orchestration," Music Perception **13**, 209–246.

Sandell, G. J. (1998). "Macrotimbre: Contribution of attack and steady state," *Proc. 16th Int. Congress on Acoustics and 135th Meeting of the Acoustical Society of America*, Vol. 3, Seattle (Acoustical Society of America, Woodbury, NY), pp. 1881–1882.

Seashore, C. E. ([1938], 1967). *The Psychology of Music* (Dover, New York).

Serafini, S. (1995). "Timbre judgments of Javanese gamelan instruments by trained and untrained adults," Psychomusicology **14**, 137–153.

Shepard, R. N. (1982). "Structural representations of musical pitch," in *The Psychology of Music*, D. Deutsch, ed. (Academic Press, New York), pp. 334–390.

Slaney, M., Covell, M., and Lassiter, B. (1995). "Automatic Audio Morphing," *Proc. 1996 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (*ICASSP-96*), Vol. 2 (IEEE, New York), pp. 1001–1004.

Spaeth, S. G. (1933). *The Art of Enjoying Music* (McGraw-Hill, New York).

von Bismarck, G. (1974). "Timbre of steady tones: A factorial investigation of its verbal attributes," Acustica **30**, 146–159.

von Helmholtz, H. L. F. (1877). *Die Lehre von den Tonempfindungen als physiologische Grudlage für die Theorie der Musik*. F. Vieweg and Sohn, Braunschweig. English translation by A. J. Ellis, "On the Sensations of Tone as a Physiological Basis for the Theory of Music (2nd ed., 1885)," reprinted by Dover Publications, New York, 1954.

Wedin, L. and Goude, G. (1972). "Dimension analysis of the perception of instrumental timbre," Scandinavian J. Psych. **13**(3), 228–240.

Wessel, D. L. (1973). "Psychoacoustics and music: A report from Michigan State University," PAGE: Bulletin of the Computers Arts Soc. **30**, 1–2.