Clarifying Spectral And Temporal Dimensions Of Musical Instrument Timbre

Michael D. Hall[1], James W. Beauchamp[2]
[1]Department of Psychology, MSC 7704, James Madison University
Harrisonburg, VA 22807, U.S.A., hallmd@jmu.edu
[2]School of Music and Dept. of Electrical & Computer Engineering, University of Illinois at Urbana-Champaign, Illinois 61801, U.S.A., jwbeauch@uiuc.edu

ABSTRACT

Classic studies based on multi-dimensional scaling of dissimilarity judgments, and on discrimination, for musical instrument sounds have provided converging support for the importance of relatively static, spectral cues to timbre (e.g., energy in the higher harmonics, which has been associated with perceived brightness), as well as dynamic, temporal cues (e.g., rise time, associated with perceived abruptness). Comparatively few studies have evaluated the effects of acoustic attributes on instrument identification, despite the fact that timbre recognition is an important listening goal. To assess the nature, and salience, of these cues to timbre recognition, two experiments were designed to compare discrimination and identification performance for resynthesized tones that systematically varied spectral and temporal parameters between settings for two natural instruments. Stimuli in the first experiment consisted of various combinations of spectral envelopes (manipulating the relative amplitudes of harmonics) and amplitude-vs.-time envelopes (including rise times). Listeners were most sensitive to spectral changes in both discrimination and identification tasks. Only extreme amplitude envelopes impacted performance, suggesting a binary feature based on abruptness of the attack. The second experiment sought to clarify the spectral dimension. Listener sensitivity was compared for a) modifications of spectral envelope shape via variation of formant structure and b) spectral changes that minimally impact envelope shape (using low-pass filters to match the centroids of the formant-varied envelopes). Only differences in formant structure were easily discriminated and contributed strongly to identification. Thus, it appears that listeners primarily identify timbres according to spectral envelope shape. Implications for models of instrument timbre are discussed.

RESUME

Plusieurs études classiques se servant d'une mise en échelle multidimensionnelle pour mesurer des jugements de dissemblance, et de discrimination, pour des sons d'instruments de musique, stipulent un appui concourant à l'importance d'indicateurs de timbre spectrales qui sont relativement statiques (ex., l'énergie dans les harmoniques de haute fréquence, qui a été associée à la brillance perçue), ainsi que d'indicateurs temporels dynamiques (ex., le temps de montée d'une salve , qui est associé à la soudaineté perçue). Comparativement moins d'études ont évaluées les effets de caractéristiques acoustiques sur l'identification d'instruments, malgré le fait que la reconnaissance du timbre est un objectif important de l'écoute. Pour évaluer la nature, et la prédominance, de ces indicateurs de reconnaissance de timbre, deux expériences ont été conçues pour comparer la performance de l'identification et de la discrimination de sons musicaux resynthétisés, dans lesquelles des paramètres spectraux et temporels entre ajustements ont été modifiés systématiquement pour deux instruments naturels. Les stimuli utilisés dans la première expérience étaient composés de plusieurs combinaisons d'enveloppes spectrales (en manipulant les amplitudes relatives des harmoniques) et d'enveloppes d'amplitudes-vs-temps (incluant les temps de montée de slaves). La sensibilité des auditeurs la plus élevée était celle envers les changements spectraux, et ce pour les tâches de discrimination et d'identification. Seules les enveloppes d'amplitude extrêmes ont influencées la performance, ce qui suggère une caractéristique binaire basée sur la soudaineté de l'attaque. La deuxième expérience avait comme objectif de clarifier la dimension spectrale. La sensibilité des auditeurs a été évaluée contre 1) les modifications de forme des enveloppes spectrales par la modification de la structure du formant et 2) les changements spectraux qui ont un effet minime sur la forme de l'enveloppe (en utilisant des filtres passe-bas pour accorder les centroïdes des enveloppes à formants variés). Seules les différences dans la structure du formant étaient facilement discriminées et contribuaient considérablement à l'identification. Ainsi, il paraît que les auditeurs identifient principalement le timbre selon la forme de l'enveloppe spectral. Les implications des résultats pour des modèles de timbre d'instruments sont discutées.

1. INTRODUCTION

Timbre can be thought of as the collection of perceived qualities that help to identify a given sound source. Thus, timbre is what distinguishes a particular person's voice, or what makes a musical instrument, such as a piano, sound like itself. Traditionally, timbre has been "defined" by the exclusion of properties. For example, the American Standards Association (1960) defines timbre as "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar".

Since this definition's inception, considerable gains have been made in research to define timbre by inclusion, that is, by what it is rather than what it is not. Yet, the admittedly inadequate definition by exclusion is still often used today. This is due in large part to the fact that timbre has been repeatedly demonstrated to be multi-dimensional. Much of this evidence has come from research with sounds produced by musical instruments, including multi-dimensional scaling (MDS) of timbre dissimilarity judgments (e.g., Caclin et al. 2005; Grey, 1977; Krumhansl, 1989; Miller and Carterette, 1975; Winsberg et al. 1995), as well as from discrimination tasks in response to manipulations of one or more potentially relevant acoustic dimensions (e.g., Grey and Moorer, 1977; McAdams et al. 1999).

By comparison, relatively little research has been aimed towards the study of timbre identification, even though such a task most closely resembles a typical perceiver goal — to identify the sound source. The current investigation consists of experiments that directly compare data from a timbre identification task with discrimination performance in order to further evaluate the relative contributions of some acoustic dimensions that have been argued to be critical to musical instrument timbre. Prior to summarizing these experiments, a brief overview of some classic methodologies and their corresponding findings will be provided, including the relevant perceptual dimensions that have been identified by that research. Within this overview, potential limitations of traditional methodologies will be discussed, and further justification that timbre identification data is necessary for a more thorough evaluation of the relative importance of timbre dimensions will be given.

1.1. Evidence for critical dimensions of timbre

In typical MDS evaluations of instrument timbre, listeners are presented with all possible instrument pairings from a limited set of stimuli. Their task is to provide estimates of the perceptual distance between each stimuli-pair by rating them on a scale of dissimilarity (usually from *1*-7). All ratings are then submitted to a computer model [e.g., INDSCAL (see Carroll and Chang, 1970), CLASCAL (Winsberg and De Soete, 1993), or CONSCAL (Winsberg and De Soete, 1997)] in order to generate a best-fitting model of perceptual (timbre) space using a minimum number of dimensions. The addition of a given perceptual dimension will significantly increase the proportion of variance in ratings data that is explained by the MDS solution to the degree that the dimension was relied upon by listeners. As a result, MDS is typically argued to reveal the relative strength of contributions from each dimension (e.g., Miller and Carterette, 1975; also see Caclin, et al. 2005).

Probably the most important step in the MDS process comes last, when researchers attempt to correlate the resulting positions of the stimuli along each perceptual axis of the timbre space with changes of a particular physical measure. High correlations between acoustic and perceptual ratings/dimensions are used to infer that a particular source of acoustic variation was likely used by listeners to distinguish between timbres. Using this approach, several acoustic parameters have repeatedly been shown to be closely related to dimensions in timbre scaling solutions. These commonly include not only relatively static characteristics of the spectral envelope, but also more dynamic/temporal attributes, as shown in Krumhansl's (1989) 3-dimensional timbre space.

The primary aspect of the spectral envelope that strongly correlates with dimensions in MDS solutions, as in Ehresman and Wessel (1978), is the spectral centroid, formed by weighting the harmonic frequencies of a musical tone by the amplitudes corresponding to them and by normalizing and adding the resulting values. Thus,

$$f_c = \frac{\sum_{k=1}^{K} f_k A_k}{\sum_{k=1}^{K} A_k} , \qquad [1]$$

where f_c is the spectral centroid (in Hz), k is harmonic number, f_k is harmonic frequency, K is the number of harmonics, and A_k is the amplitude of the k^{th} harmonic. A similar dimension was obtained in the seminal work of Grey (1977), who found a correlation with the distribution of spectral energy as relatively narrow (reflecting a concentration of low-frequency energy) or broad (reflecting contributions from higher harmonics). Insofar as this measure reflects the relative presence of high- or low-frequency energy in the signal, it is frequently argued that the corresponding dimension of a timbre space reflects the perception of brightness (i.e., a tone is perceived as brighter given more high-frequency energy, or less low-frequency energy).

A solution by Krimphoff et al. (1994) also identified a temporal dimension, rise time, which can be defined as the time interval from tone onset to when the most intense portion of the tone is reached. Rise times are typically calculated as the time difference between where a priori criterion values are first reached for very low and high percentages of the signal's maximum amplitude (in the current study, between 10 and 90 percent of peak amplitude). Strong correlations with both the spectral centroid and (log) rise time were later confirmed by Winsberg et al. (1995) using a subset of a stimulus set devised by Krumhansl (1989) and was also used by Krimphoff et al. (1994).

The importance of both spectral and temporal properties has been confirmed by MDS procedures involving stimulus sets that reflect direct manipulations of potentially relevant acoustic characteristics. For instance, timbre space has been shown to be systematically altered when spectral envelopes have been exchanged across instruments, and dimensions from the resulting solutions still correlated with changes in spectral energy distribution (Grey and Gordon, 1978). Artificial spectral manipulations (number of harmonics) and temporal manipulations (rise time conveyed by linear ramps of amplitude) also have been found to correlate with separate dimensions in MDS solutions, thereby presumably indicating separate spectral and temporal contributions to timbre (Samson, Zatorre, and Ramsay, 1997). Furthermore, distortions of these MDS solutions in listeners with right temporal lobe lesions suggest involvement of those brain regions in spectral processing (Samson, Zatorre, and Ramsay, 2002). MDS has been coupled with direct manipulation of stimulus values to confirm the importance of the spectral centroid and attack time as relevant dimensions (Caclin, et al. 2005).

Combined spectrotemporal properties of timbre also have been proposed from MDS solutions. One commonly proposed spectrotemporal dimension (e.g., see Krumhansl, 1989) is spectral flux, which characterizes variation in the shape of the spectral envelope over time. A similar dimension also was described by Grey's (1977) timbre space for attack transients. However, recent MDS evidence has not provided strong support for the dimension of spectral flux, and other alternative dimensions have been proposed. Krimphoff et al. (1994) demonstrated that the dimension originally identified by Krumhansl (1989) as corresponding to spectral flux was better predicted by spectral irregularity, a measure of the relative jaggedness of the spectrum. Caclin, et al. (2005) demonstrated a reduced contribution of spectral flux with increases in the number of concurrently manipulated dimensions. They also suggested that the relative amplitude of even- to odd-numbered harmonics was an important factor in spectral envelope shape. Further evidence for even-odd importance was reported by Beauchamp et al. (2006). Using a set of both sustained and percussion instruments which were normalized with respect to spectral centroid and rise time, Beauchamp and Lakatos (2002) attempted to correlate some of these measures with MDS solutions.

Discrimination performance can additionally indicate what minimal acoustic manipulations are audible, and therefore, which dimensions could potentially constitute a basic feature of timbre. If a given simplification of an instrument tone is easily discriminated from the original or resynthesized version of that tone, then this would indicate that a relevant dimension of timbre was affected. For example, Grey and Moorer (1977) revealed the relevance of attack transients by obtaining evidence of very accurate discrimination of tones from versions of those tones with their attack removed (e.g., Grey and Moorer, 1977). This approach also has been used to identify several potentially relevant, spectrotemporal properties in tones where spectral flux was controlled. These include spectral envelope irregularity (i.e., the relative jaggedness of spectrum, possibly due to a relative emphasis on odd-numbered harmonics), and particularly, amplitude envelope incoherence, the degree to which each harmonic has a unique amplitude envelope shape. For example, McAdams, et al. (1999) showed, using a discrimination method, that both spectral incoherence and spectral irregularity are important for musical instrument tone perception. It should be noted that amplitude envelope incoherence (also known as spectral incoherence) is equivalent to spectral flux (fluctuation in spectral envelope shape) as defined by Krumhansl (1989).

1.2. Need for timbre identification research and the utility of timbre interpolation

While discrimination, and particularly, MDS, procedures have been very informative at revealing several potentially critical dimensions of timbre, the information gathered from any single task in necessarily limited. In this case, data from either of these tasks must be combined with timbre identification data in order to permit strong conclusions about acoustic parameters that listeners utilize for instrument recognition. Although the MDS approach can reveal strong correlations between acoustic parameters and salient dimensions, such correlations do not prove causal relationships. It is possible that other acoustic parameters will better correlate with the MDS dimensions, and that these parameters more closely model what listeners rely on in making their responses. This limitation was indicated effectively by Caclin, et al. (2005), who noted that

> "MDS studies are thus presumed to highlight the most perceptually salient timbre parameters that are likely to be of importance in a variety of situations (voice recognition, music listening). Nevertheless they have a common drawback: given the multiplicity of acoustical parameters that could be proposed to explain perceptual dimensions, one can never be sure that the selected parameters do not merely covary with the true underlying parameters." (p. 472)

Identification performance also need not be directly predicted by discrimination performance, as previously indicated by McAdams (2001):

> "The extent to which an event can be simplified without affecting identification performance is the extent to which the information is not used by the listener in the identification process, even if it is discriminable." (p. 161)

Thus, it is possible that discrimination of variation along one or more acoustic dimensions could be very accurate, and yet this variation might not be sufficiently large for listeners to perceive a change in instrument category. After all, listeners frequently claim to readily perceive differences in timbre between examples of a particular instrument. For example, electric guitarists often discuss their individual preferences for an instrument's signature sound, such as the twang of a *Fender Stratocaster* or the warmth of a large *Gretsch* hollowbody. Furthermore, differences between sets of instruments given supra-threshold levels of stimulus variation could reflect learning of cues to specific timbres. In other words, it is possible that a particular stimulus parameter might prove to be characteristic to a particular instrument (e.g., the presence of inharmonic energy in a piano tone), or instruments, and yet, not particularly informative for others.

Despite the importance of identification data for gaining a more complete understanding of timbre, comparatively few research studies have focused on timbre identification. Most of these studies were early investigations that highlighted the importance of information in the attack to timbre recognition by either eliminating or altering attack transients. For example, Saldanha and Corso (1964) showed that timbre identification performance is reduced for tones whose attack transients have been deleted. A similar reduction/alteration in identification performance was demonstrated in response to swapping of attack transients across instruments (Thayer, 1974).

Many of the existing studies of instrument timbre that rely upon identification tasks (and also, frequently, studies that have used discrimination and MDS procedures) involve a stimulus set that includes some form of synthesized interpolation between natural instrument timbres. Since interpolation requires systematic control of potentially relevant physical parameters, inclusion of interpolated tones should help the researcher in evaluating the relative contribution of those parameters to task performance, and thus, presumably, timbre recognition. Anecdotally, timbre interpolation has had a long history. Any imitation of one instrument or voice by another instrument or voice can be considered timbre interpolation in that aspects of a target instrument are superimposed on a source instrument. Generally this means either using unorthodox manipulation of an instrument's excitation (e.g., vocal folds, reed vibration) or its body resonances (e.g., vocal track or violin body resonances).

To these authors' knowledge, the first instance of systematic timbre interpolation using a computer was accomplished by John Grey as reported in his PhD dissertation (1975, pp. 75-95). In his method the time-varying amplitudes of the individual harmonics were cross-faded between two instruments before resynthesis. The method was tantamount to cross-fading between two signals except for two differences: 1) harmonic phases were aligned and frequencies were flattened; 2) segments before and after amplitude maxima occuring in the endpoint spectra were aligned and interpolated to produce smooth transitions. A series of tones were presented to subjects where the crossfade parameter gradually changed the timbres from a source to a target. Conclusions were that there is a strong hysteresis effect according to the direction of transition but that there is no sharp boundary for identification of which of the two instruments was heard.

Recently, more advanced timbre interpolation has been described by Haken, et al. (2007). In order to align times between eight sounds, a time dilation function is defined for each sound which reveals when prominent points in the sound's structure occur. Then a time envelope is defined which interpolates amongst the eight sounds' time dilations, and this in turn is applied to functions governing amplitudes, frequencies, and noises of each harmonic of each sound which in turn are combined (mixed) to form the additive synthesis control functions prior to final resynthesis.

The current investigation was motivated by an interest in using timbre interpolation to assess the relative contributions of spectral and temporal properties to the ability to identify musical instrument sounds. Such an assessment requires data from a timbre identification task that can be compared with data from procedures that have traditionally been used to evaluate timbre dimensions (e.g., discrimination or MDS). Also required is a direct manipulation of spectral and temporal parameters while excluding any spectrotemporal variation (i.e., spectral flux) that could interact with the perceptual dimensions of interest. Note that these parameters are really vectors, in that their definitions generally require many numerical values.

The current investigation was intended to provide such an assessment. We restricted our focus to manipulating only the most commonly identified timbral parameters — that is, spectral envelope (epitomized by the single-valued spectral centroid) and amplitude-vs.-time envelope (epitomized by the single-valued rise time). (For the sake of brevity, "amplitude-vs.-time envelope" will henceforth be referred to as "amplitude envelope".) Manipulation of each parameter was accomplished by synthesizing a set of hybrid stimuli whose parameters were interpolated between those of two instruments, namely an A_4 violin and an A_4 trombone, whose spectral and temporal properties differed considerably. The nature and relative salience of these two parameters were evaluated in an experiment (Experiment 1) that compared discrimination and timbre identification performance for the hybrid stimuli. A follow-up experiment (Experiment 2) was designed to further clarify whether listeners were using the complex spectral envelope or merely the spectral centroid for instrument recognition.

2. EXPERIMENT 1: SPECTRAL ENVELOPES V. AMPLITUDE ENVELOPES

Experiment 1 was designed to evaluate the relative contribution of a static property, the spectral envelope (epitomized by spectral centroid), and a dynamic property, the amplitude envelope (epitomized by rise time), to timbre identification and discrimination. Sixteen hybrid tones interpolated between two reference tones, a violin and a trombone, both pitched at A_4, were generated for this purpose. Each tone was constructed by amplitude-

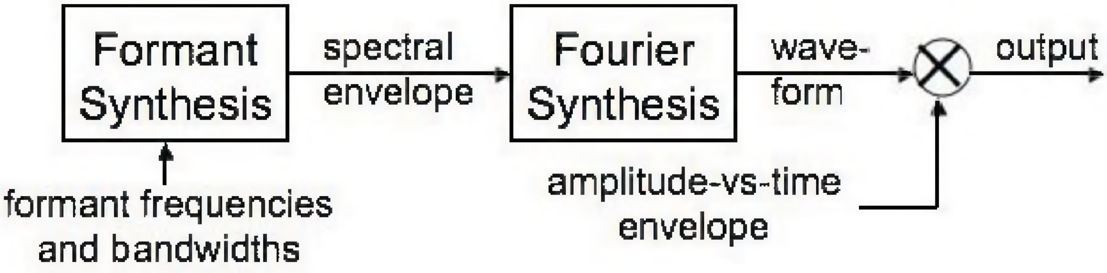


Figure 1. Conceptual diagram of the synthesis procedure in Experiment 1 and their products.

modulating a static waveform formed by interpolation between the reference tone spectra with an amplitude envelope interpolated between those of the reference tones. Based on existing timbre literature (e.g., MDS (Grey, 1977; McAdams et al. 1995; Caclin et al. 2005), discrimination (McAdams et al. 1999)), it was expected that both spectral and temporal parameters would contribute to identification and discrimination. Whether one or the other parameter impacted task performance to a significantly greater degree was left as an open question for the research to address.

2.1. Participants

Seven students from James Madison University participated in exchange for extra credit counted towards a psychology course. All listeners were between 18 and 40 years of age, and none reported having any known hearing deficits.

In neither experiment were participants screened for performance training on a musical instrument. However, some information was collected about extent and type of training via questionnaire. The participants in Experiment 1 could generally be characterized as having some musical experience. Participants had a mean of 3.2 years of performance training on a musical instrument (with a standard error of 0.9 years) and a range of 0 to 6 years. Two listeners had no musical training at all; the remaining 5 listeners had some formal musical training. No listener had previous performance experience with one of the target instruments from the experiment. Only 1 participant was occasionally playing an instrument at the time of testing; no other participant had played an instrument in the preceding 4 years.

2.2. Stimuli

Sixteen stimuli of 500 ms duration were generated that orthogonally combined 4 levels of spectral envelope with 4 levels of amplitude envelope. This was accomplished using a source-filter synthesis method that involved a series of operations. Figure 1 shows a conceptual equivalent block diagram for this process. In reality the Praat program was used to process a sawtooth waveform by a bank of band-pass filters to yield a waveform whose harmonic amplitudes form a static spectral envelope. The block "Formant Synthesis" in Figure 1 represents this process. An important aspect of the spectral envelope is the formants that correspond to the filter resonances. Another important aspect is the general negative slope of the spectral envelope that corresponds to the inverse frequency characteristic of the sawtooth waveform. (Similar kinds of formant-based estimations of spectral envelopes have been previously applied successfully to musical instrument tones; for a detailed summary of the general utility of such estimations, including a detailed description of several related synthesis methods, see Rodet and Schwarz (2007).) The waveform which corresponded to the spectral envelope was then multiplied by an amplitude contour to impose an instrument-specific (or hybrid) amplitude envelope on the waveform, and thereby obtain the stimulus (labeled "output"). This approach to synthesis, in which a time-varying amplitude envelope was supplied separately from a static spectral envelope, eliminated all inharmonicity, fundamental frequency (F_0) deviations, and spectral flux, thus allowing a direct assessment of the relative perceptual contributions of the spectral and temporal parameters.

To aid in understanding the synthesis model for this experiment, let the resynthesized violin tone (henceforth called *Vn*) be characterized by its amplitude envelope $A_V(t)$ and spectral envelope $S_V(f)$. Likewise, let the resynthesized trombone tone (henceforth called *Tr*) be characterized by its amplitude envelope $A_T(t)$ and spectral envelope function $S_T(f)$. Then the interpolated signal is given by

$$s(t) = \{\alpha A_V(t) + (1-\alpha)A_T(t)\} \times \sum_{k=1}^{K} A(\beta, S_V(kf_0), S_T(kf_0)) \cos(2\pi k f_0 t + \theta_k), \qquad [2]$$

where t = time, k = harmonic number, K = number of harmonics, α = interpolation value for amplitude, β = interpolation value for the spectral envelope, f_0 = fundamental frequency, and θ_k = phase of harmonic k. Note that the time-varying amplitude in front of the summation sign does not depend on frequency. Likewise, the $A()$ function weights in front of the cos functions (that give the harmonic sinusoidal variations) do not depend on time, but rather only on the frequency of the corresponding harmonic.

The original violin and trombone tones, both pitched at A_4 (440 Hz), were taken from McGill University Master Samples (MUMS) library (Opolko and Wapnick, 1987). These instruments were selected to share a similar total number and distribution of harmonics, while differing significantly with respect to both spectral centroid and rise

	Vn	Vn Hybrid	Tr Hybrid	Tr
Rise Time (ms)	404	398	335	236
Centroid (Hz)	1082	984	953	922
F1	492	600	715	839
F2	2159	1899	1660	1441
F3	3651	3205	2802	2438
F4	4457	4767	5094	5440
F5	7006	7177	7352	7531
F6	8465	8833	9216	9613
B1	277	311	347	383
B2	193	400	643	928
B3	222	507	857	1290
B4	953	1694	2717	4128
B5	744	823	906	993
B6	955	944	932	921

Table 1. Formant center frequencies (F) for each stimulus in Experiment 1 (Vn = violin; Tr = trombone) and their bandwidths (B). Also provided are corresponding measures of spectral centroid for each spectral envelope, as well as measured rise time values for each amplitude envelope.

time. The open string production of the violin tone was selected to eliminate vibrato that was otherwise present throughout the chromatic series of recordings. Measured envelope values for the violin tone and trombone tone represented endpoints along the temporal and spectral dimensions.

The computer program Praat was used to determine the average spectra of the natural violin and trombone tones, as well as to construct all stimuli/spectra for our experiment (Boersma and Weenink, 2007). The source-filter synthesis model in Praat that was used is similar to that described for speech by Klatt and Klatt (1990); a vibrational source (sawtooth wave) was submitted to a bank of band-pass filters specifying six formants. Formants were derived from an LPC analysis (extended over an 11,025 Hz range, with a .05 s analysis window length, and based upon a 30 dB range for each measured formant). Artificial spectra for *Vn* and *Tr* were determined by combining the formants whose mean center frequencies and corresponding average bandwidths were measured over the initial 500 ms of each tone's steady-state (i.e., immediately after the tone's peak amplitude, reflecting completion of its attack).

Formant center frequencies and bandwidths for *Vn* and *Tr* (as well as for hybrid stimuli) are provided in Table 1. As Table 1 reveals, extremely wide bandwidths were measured and synthesized for the mid-frequency range of the trombone tone. These very broad spectral peaks reflect the smooth regions that naturally occur in instrument spectra. Mean measures of spectral centroid for each stimulus also are included in Table 1. Previous MDS studies (e.g., Krumhansl, 1989; McAdams et al. 1995) have established that the range of spectral centroid variation across these instruments is moderate relative to other pairs of instruments.

Formant frequencies and bandwidths for two hybrid spectral envelopes were chosen to be in-between the natural instrument (endpoint) values and were spaced for equal perceptual distance (with respect to frequency discrimination) from each endpoint stimulus. The frequency spacings were obtained by a log/Mel scale transformation of the formant center frequencies and their corresponding bandwidths according to the approximation proposed by Fant (1973),

$$M = 1000 \log(1 + f/1000)/\log(2), \qquad [3]$$

where M is mels and f equals frequency in *Hz*. Hybrid values along each dimension that were closest to the natural violin and trombone values will henceforth be referred to as *Vn Hybrid and Tr Hybrid*, respectively (see Table 1 for the center frequencies and bandwidths of their formants). Spectral envelopes for the four stimuli, *Vn*, Vn *Hybrid*, *Tr Hybrid*, and *Tr*, are shown in Figure 2.

Filtered stimuli representing each level of spectral envelope were multiplied separately by each level of amplitude envelope in Praat to complete synthesis of the stimulus set. Amplitude envelope estimates for the attack transients of the violin and trombone tones were determined from measurements in Praat of waveform maxima (in relative dB) taken every 2 ms over the first 500 ms of the original tone production. Hybrid values were obtained by interpolation to form equal steps between violin and trombone amplitudes. Additionally, amplitude was down-ramped linearly over the final 20 ms of each tone to avoid the perception of abrupt offsets. A depiction of each amplitude envelope is provided in Figure 3. Measures of rise time, the time interval from tone onset during which the signal moved from 10 to 90 percent of its peak amplitude, also are included in Table 1. Rise times of the endpoint stimuli reveal that the stimulus set reflected a reasonably broad range (236 – 404 ms) along this characteristic.[1]

Several aspects of the stimuli and stimulus presentation were shared with Experiment 2. All stimuli were 500 ms in duration, and were synthesized with a 44.1 kHz (16-bit) sampling rate. Furthermore, all tones were presented through a low-pass (Butterworth) anti-aliasing filter with a cutoff frequency of 11 kHz, and the peak sound level (to the nearest dB) of the presented stimuli was 80 dB[A]. Both experiments were conducted in a quiet room, and all stimuli were delivered over Sennheiser HD 280 earphones.

2.3. Procedure

Listeners completed two tasks—a timbre identification task and a tone discrimination task. These tasks were counterbalanced across participants, and participants were afforded a brief rest break in between the two tasks. A few procedures were shared across tasks, as well as with Experiment 2. First, stimulus delivery and the collection of responses across tasks were controlled by Music Experiment Development System software (v. 2002-B-1; Kendall, 2002). Also, within each task, a 500 ms inter-trial interval followed any given response prior to stimulus delivery for the next trial.

Discrimination task

In the discrimination task listeners were instructed to respond whether the two tones presented on each trial were either physically identical (i.e., “same”) or different. On any given trial either the *Vn* tone or the *Tr* tone was used as a “standard” stimulus ($p = 0.5$ for each instrument). The other stimulus on each trial was either identical to the standard ($p = 0.5$ across trials), or alternatively, a different stimulus. As in the discrimination task of the subsequent experiment, a 250 ms inter-stimulus interval separated the pair of tones on each discrimination trial.

To enable an evaluation of the relative contribution of the spectral and amplitude envelope parameters to timbre, stimulus comparisons on different trials could involve a manipulation of either parameter in isolation, or alternatively, both parameters together. Discrimination would be expected to improve with increasing distance along a given perceptual dimension. Therefore, step size was manipulated across trials. Included were *1-step comparisons* (i.e., *Vn⇔VnHybrid*, *Tr⇔TrHybrid*), *2-step* comparisons (i.e., *Vn⇔TrHybrid*, *Tr⇔VnHybrid*), and *3-step* comparisons (*Vn⇔Tr*) along each dimension, or combination of dimensions. There were 17 such “different” pairs of stimuli.

Participants used a laptop keyboard to indicate their response on each trial. Listeners were instructed to press the ***1*** key if the two tone stimuli on a trial were perceived to be identical or the ***3*** key if the tone stimuli on a trial were perceived to be different.

A brief familiarization period preceded the discrimination task. During this familiarization, listeners were presented with the 14 tone stimuli (once in random order) that they would subsequently hear during discrimination trials so that they would have a clear sense of the range of stimulus differences that they would be exposed to during the task. As with each subsequently described familiarization procedure, a 1-sec inter-trial-interval separated each tone. No responses were made during familiarization. Listeners could request to repeat the familiarization sequence as needed to feel comfortable with the differences between stimuli before proceeding with the task; no such request was made.

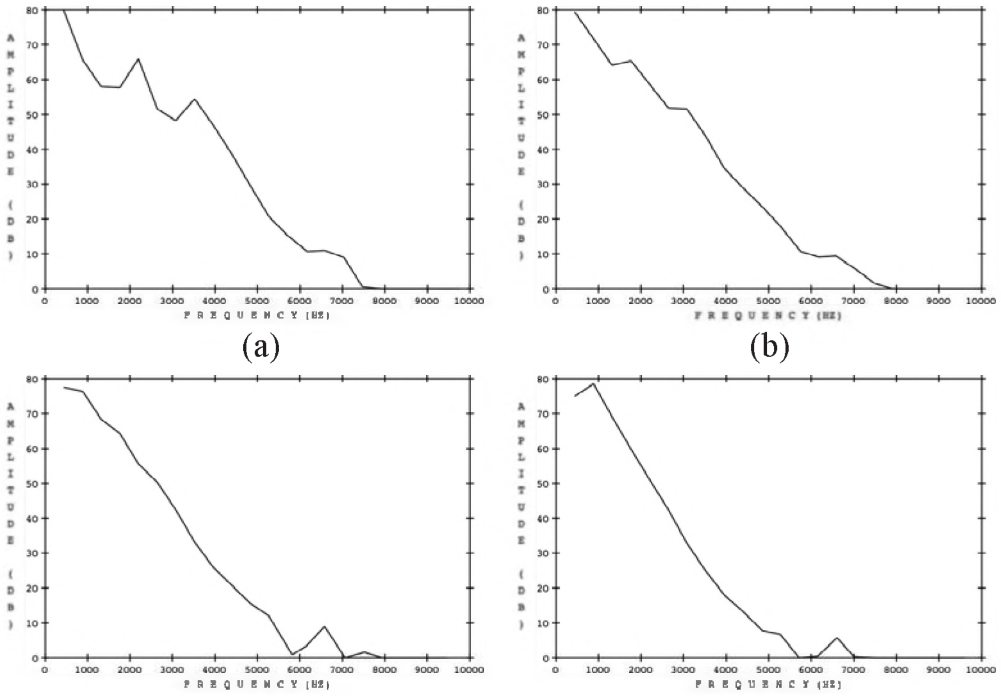
The familiarization period was immediately followed by two blocks of 272 randomized discrimination trials (i.e., 544 total trials). Each block of trials consisted of 8 repetitions for each of the 17 pairs of different stimuli (4 repetitions for each ordering of the standard and comparison tones). For the “same” trials, which were half of the trials within each block, *Vn* was presented on an equivalent number of trials as the *Tr* stimulus.

Timbre identification

In the timbre identification task listeners were instructed to indicate whether the tone they heard on a given trial corresponded to a violin, a tenor trombone, or the timbre of a different instrument. The participants also were informed that the tones that they would be hearing had been resynthesized based on naturally occurring parameters for the violin and tenor trombone, and that they may sound quite artificial as a result of being simplified in several ways. They were told that some trials would contain a tone that was a simplified version of a natural instrument tone (either *Vn* or *Tr*), and that other trials would contain a hybrid tone that resulted from a combination of attributes that differed from those of a natural instrument. Participants were asked simply to categorize the timbre of each instrument tone to the best of their ability.

Responses on each trial were made by using the computer’s mouse to click on a small bitmap image on the laptop screen corresponding to the perceived instrument. Each image consisted of a verbal label for the instrument in white lettering against a blue rectangular background (e.g., *violin* or *trombone*). The background of each image was lighter on the edges so that the collection of images appeared as a series of buttons from a button box. In addition, the response category of *other* was included so that listeners could indicate if any of the stimuli (particularly, the hybrid tones) sounded like they were derived from an instrument other than the violin or tenor trombone.

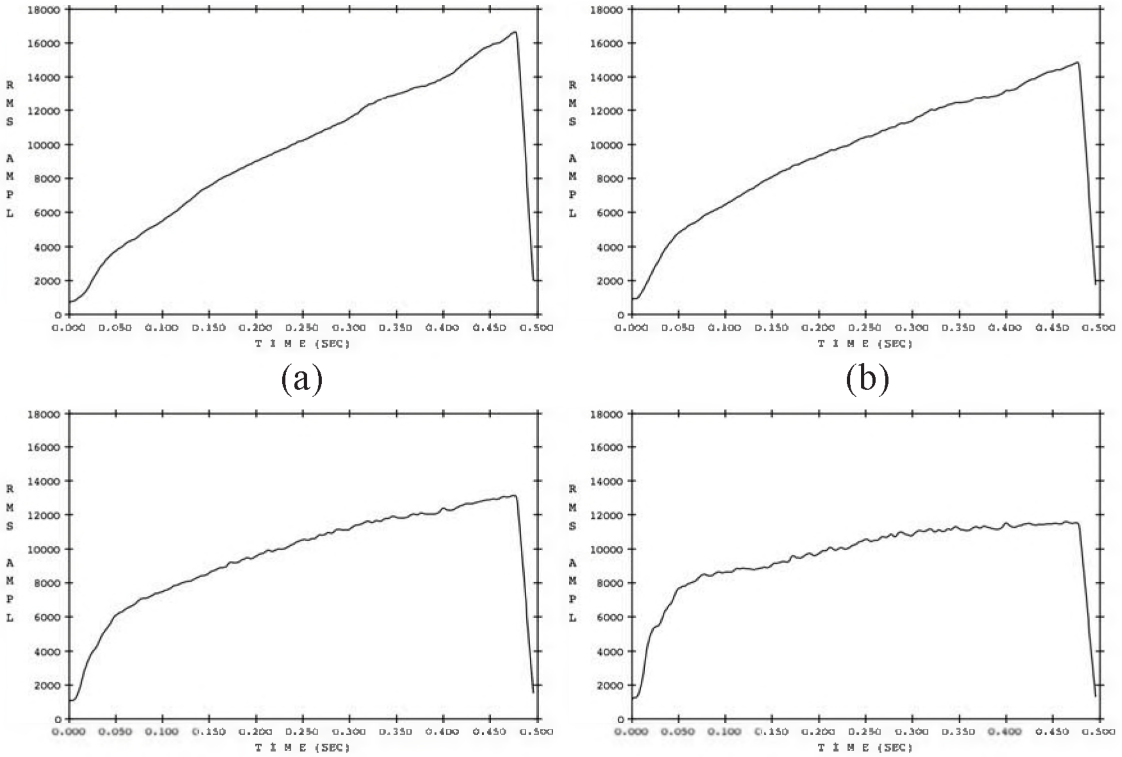
Before proceeding with the task, listeners first were familiarized with the tones that most closely approximated the original violin and trombone tones. Ten tones were presented in non-random order, comprising five repetitions of the *Vn* tone followed by the *Tr* tone. No responses were made during this familiarization period. Rather, participants just closely listened to each sound to get a better sense of the violin-like and trombone-like sounds in the stimulus set. Listeners were permitted to repeat the familiarization procedure again if they felt that they had any trouble recognizing basic timbre differences between the target timbres, but, in fact, no listener requested to repeat the procedure. This familiarization period was immediately followed by a block of 320 randomized experimental trials consisting of 20 repetitions of each tone stimulus.



(a) (b)

(c) (d)

Figure 2. Spectral envelopes used in Experiment 1 to synthesize a violin tone, *Vn* (panel a), a tenor trombone, *Tr* (panel d), and two hybrid tones, *Vn Hybrid* (b) and *Tr Hybrid* (c), based on resonances equally spaced in *Mels* between *Vn*, *Tr*, and each other.



(a) (b)

(c) (d)

Figure 3. Average (RMS) amplitude-v.-time (in seconds, 0 to 0.5 s) displays corresponding to each amplitude envelope in Experiment 1. Envelopes ranged from the onset of a violin tone, *Vn* (panel a), to that of the tenor trombone, *Tr* (panel d), including hybrid envelopes with instantaneous amplitudes that were equally spaced from these natural instrument values and each other, *VnHybrid* (panel b) and *TrHybrid* (panel c).

2.4. Results and Discussion

Discrimination task

Discrimination performance was assessed using d′, a theoretically bias-free measure of sensitivity from Signal Detection Theory. Sensitivity (d′) was calculated according to an Independent Observations Model, which assumes an individual assessment of each stimulus on a trial (see Macmillan and Creelman, 2005). The probability of false alarms was obtained from performance on same trials for the given standard tone (i.e., violin or trombone). The resulting measures of sensitivity were submitted to a 3x2x3 3-way repeated measures ANOVA with timbre dimension(s) (spectral, temporal, both), standard instrument (*Vn*, *Tr*), and step size (1, 2, or 3) as factors. Corresponding mean calculations of d′ (along with standard error bars) are provided in Figure 4 for each timbre dimension (displayed as differently shaded bars), standard stimulus (displayed to the left for *Vn*, and to the right for *Tr*), and step size (varying along the horizontal axis).

As can be seen in Figure 4, listeners were readily able to discriminate changes in spectral envelope, but not amplitude envelope. This corresponded to a main effect of dimension, $F(2,12) = 207.285$, $p = .001$. Post-hoc pair-wise comparisons of means via Tukey HSD tests further revealed that discrimination of different spectral envelopes was significantly better than for amplitude envelopes ($p < .05$), and that no further performance gains were obtained when both dimensions varied across tone stimuli. In fact, ceiling levels of discrimination performance were approached for discrimination of spectral envelope differences. In contrast, d′ never exceeded criterion levels of performance (d′ = 1) for the amplitude envelope dimension.

As expected, discrimination performance also generally improved with increasing distance along perceptual dimensions, as indicated by a significant main effect of step size, $F(2,12)= 9.305$, $p < .01$. The fact that sensitivity was essentially at ceiling across step sizes for comparisons involving the *Vn* standard (see Figure 4, left) also contributed to a standard instrument x step size interaction, $F(2,12) = 8.142$, $p < .01$. No other effects approached significance ($p > .10$).

Identification task

For each listener the probabilities of each type of timbre identification response (i.e., *violin*, *trombone*, and *other*) were calculated as a function of each combination of spectral envelope and amplitude envelope. In order to assess the relative contribution of each parameter to timbre identification, the probabilities of each type of response were submitted to a separate 4x4 2-way repeated measures ANOVA with spectral envelope level and amplitude envelope level as factors.

Mean response probabilities (and standard error bars) across participants are displayed in Figure 5 for the *violin* (displayed to the left) and *trombone* responses (right), with spectral envelopes displayed along the horizontal axis and amplitude envelopes as differently shaded bars. Corresponding means for *other* responses are not shown because they were rarely used, occurring on less than 1.5 percent of trials. Anecdotal reports from several participants indicated that they only felt it necessary to use the *other* response category in instances when they were indecisive about the instrument that produced the perceived timbre. No significant effects were obtained from ANOVA and post-hoc analyses involving *other* responses (all *F*'s < 1).

As can be seen in Figure 5, spectral envelopes strongly affected timbre identification, whereas amplitude envelopes did not. Specifically, the incidence of *violin* responses decreased, and *trombone* responses increased, as spectral envelopes were systematically altered from *Vn* to *Tr* (i.e., from left to right for either side of Figure 5). This trend resulted in a significant main effect of spectral envelope for both types of responses [$F(3,18) = 54.378$, $p < 0.0001$ and $F(3,18) = 59.546$, $p < 0.0001$ for *violin* and *trombon*e responses, respectively]. Post-hoc pair-wise comparisons of means (Tukey HSD tests) further revealed that response probabilities for tones with the *TrHybrid* and *Tr* spectral envelopes did not significantly differ. However, a significant increase in *violin* responses (plus a corresponding decrease in *trombone* responses), was obtained for the *VnHybrid* spectral envelope, and another such increase (or decrease for *trombone* responses) was obtained for the *Vn* spectral envelope ($p < .05$). In contrast, the probabilities of neither *violin* responses nor *trombone* responses significantly changed as a function of amplitude envelope, as indicated by the absence of a main effect of amplitude envelope [$F(3,18) = 1.213$, $p > 0.33$ and $F(3,18) = 1.161$, $p > 0.35$, for *violin* and *trombone* responses, respectively].

Cross-task comparisons

It is clear that listeners in Experiment 1 relied more heavily on the static spectral envelope manipulation than the dynamic amplitude envelope manipulation in both timbre identification and tone discrimination tasks. The fact that amplitude envelope did not contribute significantly to timbre identification appears to be attributable to the fact that the range of variation along that dimension was not really discriminable in the context of roving spectral envelopes across trials. Insofar as a quite broad (70%), naturally occurring range of rise times was used in Experiment 1, it is unlikely that this difference in performance across dimensions is due to reliance on a truncated range of amplitude envelopes. Furthermore, insofar as spectral centroids varied by only 160 Hz (17%), it also seems unlikely that listeners' reliance on spectral envelopes for making judgments was simply due to an exaggerated range of spectral centroid variation. We will elaborate on this argument in the general discussion section.

Based on the findings from Experiment 1, it could potentially be claimed that common spectral aspects of timbre, including perceived brightness due to shifts in the spectral centroid, could frequently be more salient than a

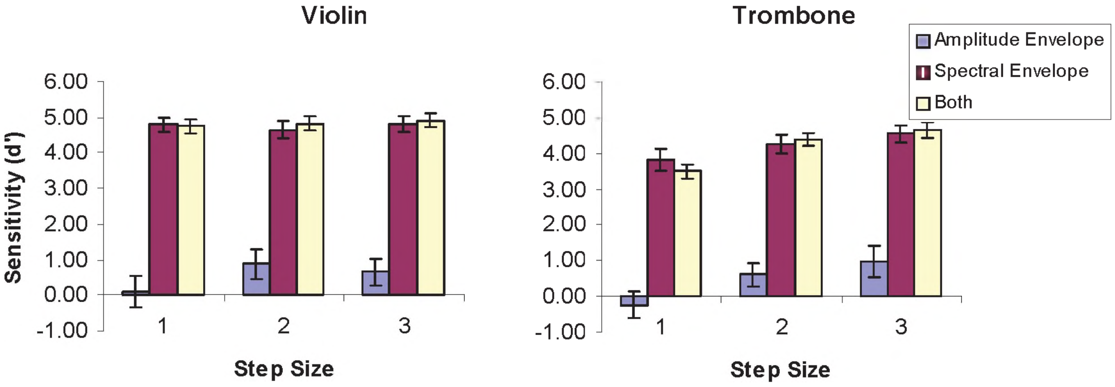


Figure 4. Mean sensitivity and corresponding standard errors for tone discrimination in Experiment 1 as a function of acoustic dimension (temporal, spectral, or both dimensions), standard instrument [violin (left) or trombone (right)] and physical distance (step size along the stimulus continuum).

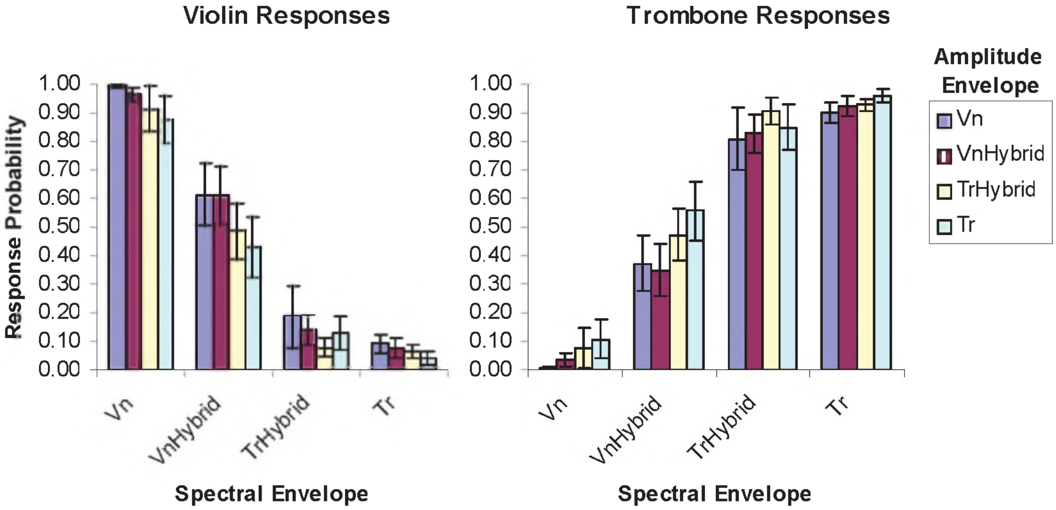


Figure 5. Mean probabilities of *violin* responses (left) and *trombone* responses (right), along with corresponding standard errors, for timbre identification in Experiment 1. Results are displayed for each combination of spectral envelope (varying along the horizontal axis) and amplitude envelope (distinguished by differently shaded bars).

commonly identified dynamic aspect of timbre, rise time. After all, the entirety of the original attack functions for the violin and trombone were retained within the amplitude envelopes used in this study.

3. EXPERIMENT 2: FORMANT STRUCTURE V. SPECTRAL CENTROID

While it is clear that listeners in Experiment 1 relied on static spectral information more than dynamic amplitude information, the nature of that spectral cue required additional clarification. Even though traditional interpretations of MDS results have suggested that listeners likely respond to differences in perceived brightness, as indicated by the spectral centroid, there is another possibility: detailed formant structure.

Manipulation of the spectral envelope in Experiment 1 was accomplished by shifting spectral peaks between naturally occurring values, thereby creating hybrid envelopes. As a result, shifts in the spectral centroid were confounded with a corresponding shift in formants. Spectral centroids gradually decreased from the *Vn* value (1082) to the *Tr* value (922). Furthermore, the center frequencies for second and third formant also decreased systematically from the *Vn* to the *Tr* spectrum (although it is obvious from Figure 2 that the formants virtually disappear in *Tr*'s spectral envelope). It is therefore possible that listeners responded to the spectral envelope shape, rather than changes in spectral centroid (brightness), in both discrimination and timbre identification tasks.

Distinguishing between the potential contribution of spectral centroid and spectral envelope structure to instrument timbre is complicated by the fact that they are very closely related properties. Raising (or lowering) the center frequency of a formant should increase (or decrease) the spectral centroid, particularly when that formant is high in amplitude to begin with. This close relationship is also seen in the timbre literature, where the same property could be argued to reflect either envelope structure or centroid. For example, one dimension from an MDS solution for a set of FM-synthesized tones was labeled by Krumhansl (1989) as indicating “spectral envelope”. However, the same dimension was later found to be more strongly correlated with spectral centroid. This was demonstrated through subsequent acoustic analyses (Krimphoff, 1993; Krimphoff et al. 1994), as well as through additional MDS data involving a large subset of Krumhansl’s (1989) stimuli (McAdams et al. 1995; for a summary of findings, see Donnadieu, 2007).

Despite this strong correspondence between variables, there is evidence that general information about the shape of the spectral envelope and the spectral centroid correspond to distinct perceptual properties. For example, discrimination of harmonic series according to differences in spectral slope, a characteristic of natural sources of vibration, is relatively unaffected by changes in the number of spectral peaks, a filter characteristic (see Li and Pastore, 1995). This finding is particularly relevant to the current study insofar as spectral slope manipulations should impact the perceived brightness of tones, whereas peaks in the spectral envelope determine formant structure.

Teasing apart which spectral cue listeners relied upon more heavily in Experiment 1—spectral centroid or spectral envelope/formant structure—constituted the goal of Experiment 2. Addressing this issue required manipulation of the spectral centroid in a manner that was largely independent of formant structure (and thus, minimized its impact on the basic shape of the spectral envelope). This was accomplished by using low-pass filtering to match the spectral centroids of three Experiment 1 tones that had a different formant structure from *Vn*.

There are indications from the literature on timbre recognition by machine that a cepstral coefficient measure (which accounts for formant structure) consistently leads to significantly more accurate classification of orchestral timbres than reliance solely on the spectral centroid (e.g., Brown, Houix and McAdams, 2001). Thus, there were reasons to anticipate that listeners in Experiment 2 would rely more heavily on the shape of spectral envelopes in making their judgments. It was therefore hypothesized that these listeners would exhibit greater sensitivity in timbre discrimination based upon differences in spectral envelope detailed structure rather than simply differences in spectral centroids, and that as a result, timbre identification also would be primarily affected by manipulations of this structure.

3.1. Participants

Eighteen students from introductory psychology courses at James Madison University participated in partial fulfillment of course requirements. All listeners were between 18 and 40 years of age, and none reported having any known hearing deficits.

As with Experiment 1, participants in Experiment 2 typically had not received much prior training on a musical instrument. Participants had a mean of 3.1 years of training (with a standard error of 0.6 years), ranging in experience from 0 to 8 years. Two listeners had no musical training. Of the remaining listeners, 14 had received some formal musical training, with 11 of them receiving 3 years or less. One listener was a former violinist who had not played the instrument for the preceding 6 years. None of the participants had continued practicing their instrument(s) at the time of testing, and only 4 had actively practiced in the past 4 years.

3.2. Stimuli

Seven tone stimuli were used in Experiment 2. All of them shared the violin's amplitude envelope. One of these (stimulus 1) was the *Vn* tone of Experiment 1, from which all of the remaining stimuli were derived. Note that *Vn* combined both the average spectrum of a violin tone and its amplitude envelope during its attack. It had no spectral or frequency variations.

Three other tones (stimuli 2 - 4) that were taken from Experiment 1 provided manipulations of the spectral envelope's formant structure (i.e., *VnHybrid*, *TrHybrid*, and *Tr*). The remaining three tones (stimuli 5 - 7) were produced by submitting *Vn* to a first order (i.e., shallow slope) low-pass filter to match the spectral centroids of stimuli 2 – 4; this was accomplished by setting the filter’s cutoff frequencies to 3300, 2640, and 2165 Hz, respectively. In this way, the general shapes of the spectral envelopes for stimuli 5 - 7 were minimally impacted (relative to *Vn*) by the changes in centroid. Henceforth in this experiment, the labels *VnHybrid*, *TrHybrid*, and *Tr* will be used to refer to stimuli with either the corresponding manipulation of spectral centroid via filtering or a particular spectral envelope/formant structure.

3.3. Procedure

Both a timbre identification task and a tone discrimination task were used to assess the relative contribution of each manipulated dimension (formant structure v. spectral centroid alone) to timbre. These tasks were closely modeled after the corresponding tasks in Experiment 1. Unless otherwise noted below, remaining aspects of the procedure, including the timing of stimulus presentation, as well as the means of making responses and collecting data, were as described for that task in Experiment 1.

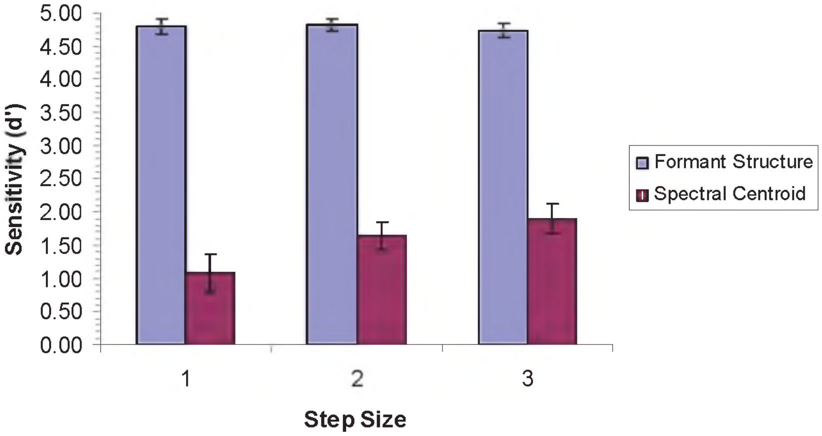


Figure 6. Mean sensitivity and corresponding standard errors for tone discrimination performance in Experiment 2 as a function of acoustic dimension (formant structure v. spectral centroid alone) and physical distance (step size along the stimulus continuum) from the standard stimulus.

Discrimination task

The discrimination task of Experiment 2 was very similar to that of Experiment 1. Listeners again were instructed to respond whether the two tones presented on each trial were either physically identical (i.e., “same”) or were different. One fundamental difference in discrimination procedures was that there was only a single standard stimulus in Experiment 2. This standard was the *Vn* tone, which was presented on every trial. The remaining stimulus on any given trial was either a second presentation of the standard (p = 0.5) or one of the six alternative tone stimuli. In other words each “different” trial included a manipulation of either formant structure or spectral centroid. There were three step-sizes for both the formant structure and the spectral centroid conditions. Comparison stimuli for either the formant structure or spectral centroid conditions were as follows: *1-step* comparisons *(VnHybrid)*, *2-step* comparisons *(TrHybrid)*, and *3-step* comparisons (*Tr*).

Before proceeding with experimental trials, listeners first were familiarized with the stimuli that would be presented in the task. This was accomplished by randomly presenting each of the seven tones once. The listeners could request repetition of the familiarization sequence until they felt comfortable with the range of perceived differences in timbre. No listener requested that the tones be repeated.

A single block of 240 randomized experimental trials followed the familiarization procedure. Within this block of trials there were 120 “same” trials, in which the standard constituted both stimuli on a trial. The remaining 120 trials were “different” trials, which paired the standard with a different comparison tone. Each of the six comparison tones were provided on twenty trials. On 10 of these 20 trials the standard was presented first; the standard was the second stimulus on the remaining 10 trials.

Identification task

As in Experiment 1, listeners in the timbre identification task of Experiment 2 were instructed to indicate whether the tone they heard on a given trial was that of a violin, a trombone, or a different instrument. Before proceeding with this task, listeners first were familiarized with the tones that most closely approximated the natural violin and trombone timbres. During this familiarization procedure, ten tones were presented in non-random order, including five repetitions of *Vn* followed by *Tr*. No listener requested that the familiarization procedure be repeated before proceeding immediately to experimental trials. A single block of 140 randomized experimental trials was given, including 20 repetitions of each individual of the 7 tone stimuli.

3.4. Results and Discussion

Discrimination task

Sensitivity (d′) for each stimulus condition in the discrimination task was calculated in the manner described for Experiment 1. The resulting measures of sensitivity were submitted to a 2x3 2-way repeated measures ANOVA with manipulated dimension (i.e., formant structure v. spectral centroid alone) and step size (1, 2, or 3) as factors. Corresponding mean calculations of d′ (along with standard error bars) are provided in Figure 6 for each timbre dimension (displayed as differently shaded bars) and step size (varying along the horizontal axis).

As can be seen in Figure 6, discrimination was much easier for any manipulation of spectral envelope shape compared to corresponding manipulations of spectral centroid alone via low-pass filtering. In fact, discrimination of spectral envelope shape approached ceiling levels for each step size, whereas mean d′ calculations for the spectral centroid manipulation did not exceed 2.0. This difference contributed to a main effect of timbre dimension, $F(1,17)$ = 182.522, p < .001. Pair-wise comparisons of means (using adjusted Bonferroni values) further confirmed that sensitivity to changes in spectral envelopes/formant

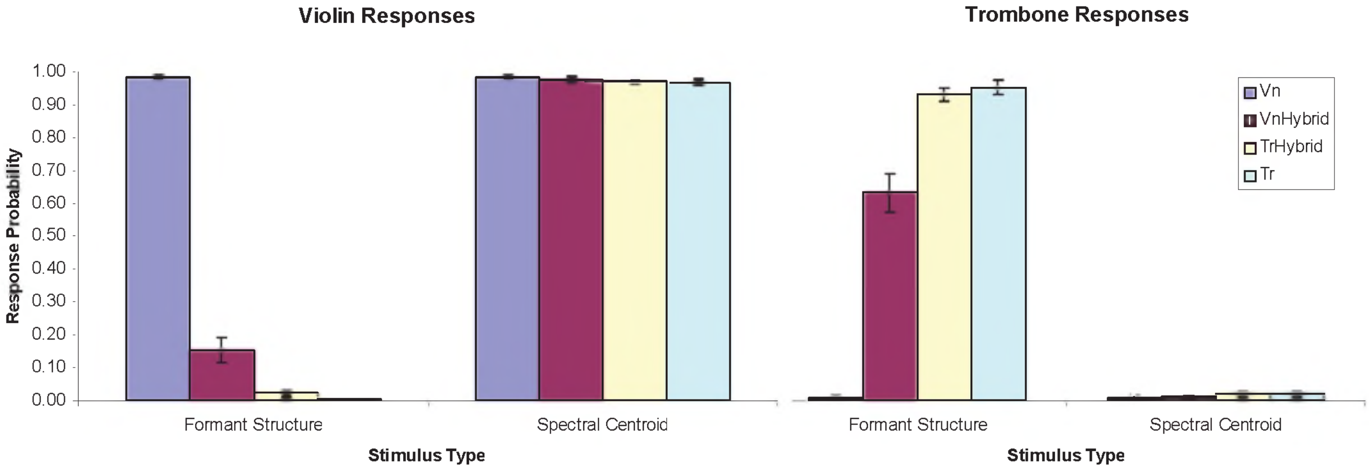


Figure 7. Mean probabilities of *violin* responses (left) and *trombone* responses (right), along with corresponding standard errors, for timbre identification in Experiment 2. For each type of response, results are displayed individually for each manipulation of spectral envelope shape (formant structure, shown to the left) and filtering (to have a corresponding effect on the spectral centroid, shown to the right), progressing gradually from spectral characteristics of a natural violin (*Vn*) to that of a trombone (*Tr*).

structures was significantly higher than for corresponding changes in spectral centroid at each step size ($p < .001$).

Discrimination performance also improved with increasing acoustic distance between stimuli, as revealed by a main effect of step size, $F(2,34) = 4.593$, $p < .05$. However, examination of the means across stimulus conditions in Figure 6 reveals that this effect of step size was driven solely by improvements in performance across spectral centroid conditions. Consistent with this interpretation, pair-wise comparisons of means revealed that the average sensitivity to 1-step changes in spectral centroid alone was significantly lower than for 2- and 3-step comparisons ($p < .05$), whereas sensitivity to changes in spectral envelopes did not significantly differ across the different step sizes. This difference in the effect of step size across (spectral envelope v. centroid) manipulations also contributed to a significant dimension x step size interaction, $F(2,34) = 7.577$, $p < .01$. No other effects approached significance ($p > .10$).

Identification task

For every listener the probabilities of each type of timbre identification response (i.e., *violin*, *trombone*, and *other*) were determined individually for the seven timbre stimuli. The resulting probabilities for each response category were submitted to a separate 1-way repeated measures ANOVA with the 7 levels of stimuli as the sole factor.[2] The decision to collapse these analyses to single-factor ANOVAs permitted the inclusion of timbre identification data for the *Vn* tone without violating assumptions of the statistical test. Pair-wise comparisons of means (according to Bonferroni adjustments) additionally were used to assess the relative contribution of spectral envelope shape and spectral centroid to timbre identification.

Figure 7 displays mean probabilities (and standard error bars) across listeners for the *violin* responses (displayed to the left), as well as the *trombone* responses (right). Within the graph for each response category, average responses to manipulations of formant structure are depicted as the left set of bars, whereas responses to corresponding changes to the spectral centroid alone are depicted to the right. Mean identification data for the *Vn* tone is displayed as the leftmost bar within both formant structure and spectral centroid displays. This duplication of data is intended to simplify visual comparisons with the mean response probabilities that were obtained from each dimension.

The pattern of results displayed in Figure 7 shows that timbre identification performance was strongly impacted by changes in spectral envelope shape (formant structure), but not by changes solely in the spectral centroid. The mean probabilities of *violin* responses decreased, and *trombone* responses increased, as spectral envelopes varied from that of *Vn* to *Tr* (i.e., from left to right for either side of the figure). In fact, *violin* responses approached a mean probability of 1.0 when listeners were given the spectral envelope of *Vn*, and *trombone* responses approached a similar maximum when listeners were presented with the spectral envelope of *Tr*. This trend resulted in very robust main effects of formant structure for *violin* responses [$F(6,102) = 1027.102$, $p < .001$] as well as for *trombone* responses [$F(6,102) = 378.098$, $p < .001$]. Pair-wise comparisons of means further revealed that a significantly greater probability of a *violin* response (and a corresponding reduced probability of a *trombone* response) was obtained for the *Vn* spectral envelope relative to each of the alternative formant structures ($p < .001$). Additionally, the mean probability of a *violin* response also was greater for the stimulus with the *VnHybrid* formant structure than for either the *TrHybrid* or the *Tr* formant structure, ($p < .05$). A corresponding reduction in the probability of a *trombone* response was likewise obtained for the stimulus reflecting the *VnHybrid* spectral envelope relative to the alternative formant structures ($p < .001$).

In contrast, changes to the spectral centroid through low-pass filtering had a negligible effect on timbre identification. The probabilities of neither *violin* responses nor *trombone* responses significantly changed as a function of low-pass filtering; the probabilities of *violin* responses remained near maximum across the different filter settings, and the probabilities of *trombone* responses remained near minimum values (see the mean probabilities displayed above the *Spectral Centroid* label in Figure 7). This was apparent within pair-wise comparisons of means, which revealed a distinct lack of variation in the obtained probabilities for either response category as centroid alone was varied relative to the *Vn* tone ($p > .87$).

The response of *other* was utilized more frequently in Experiment 2 than in Experiment 1. The incidence of *other* responses differed across the stimulus set, as reflected by a significant main effect of stimulus, $F(6,102) = 10.680$, $p < .001$. An examination of the mean probabilities across stimuli reveals that *other* responses were generally reserved for the tone with the *VnHybrid* formant structure ($M = 0.22$, with a standard error of 0.06). *Other* responses were rarely used to identify the remaining stimuli, occurring on 5 percent of trials involving tones with alternative formant structures (*TrHybrid* and *Tr*; standard errors = 0.02) and on only 1 percent of the trials for each of the other 4 stimuli (standard errors $\leq$ 0.01). Pair-wise comparisons of means confirmed that the mean probability of *other* responses for the *VnHybrid* formant structure was significantly greater than for the corresponding manipulation of spectral centroid alone ($p < .05$), and was marginally greater than the mean probability obtained for any other stimulus ($p < .10$). Anecdotal reports from participants additionally indicated that they used the *other* response to indicate when a particular stimulus was perceived ambiguously with respect to the violin and trombone categories. Clearly, the tone with the *VnHybrid* formant structure was often perceived as such a stimulus. The perceived ambiguity of this particular timbre stimulus also was likely heightened relative to the corresponding stimulus in Experiment 1 due to the reduced number of stimuli (7) in Experiment 2. As a result, listeners in Experiment 2 probably were able to store some information about each stimulus in working memory for the duration of the identification task.

Cross-task comparisons

The major findings from both tasks in Experiment 2 provide supporting evidence for the general hypothesis that listeners rely more heavily on information about detailed spectral envelope shape rather than perceived brightness in making timbre judgments. Not only were listeners able to maximally discriminate any change in formant structure, but such changes in the spectral envelope strongly affected instrument identification as well. In contrast, while isolated changes to the spectral centroid through low-pass filtering still produced moderate levels of discrimination performance, such changes had almost no impact on timbre identification.

It also may be inappropriate to attribute the moderate levels of discrimination performance at the larger step size in the centroid conditions (d′ approaching 2.0) to changes in spectral centroid. The manipulation of centroid using a first-order low-pass filter was done to closely match the spectral centroid of formant-manipulated stimuli (to the nearest Hz). At the larger step size this necessarily required that the cutoff frequency be moved much lower than for the other comparisons, substantially reducing the intensities of higher-frequency components within the original waveform. This raises the possibility that the filter's cutoff frequency could have been sufficiently low that filtering also might have strongly impacted higher resonances, and thus, begun to affect the general shape of the spectral envelope. This was confirmed by follow-up acoustic analyses. Figure 8 displays spectral envelopes for the *Vn* tone both before (panel a) and after (panel b) low-pass filtering to match the *Tr* spectral centroid. This side-by-side spectral comparison reveals that the lower-amplitude resonances (F5 and F6) that are present in the original tone are virtually absent after filtering. It thus appears that listeners were really only sensitive to changes in spectral centroid when such changes also began to impact the shape of the spectral envelope.

When taken collectively, the results of Experiment 2 indicate that, by itself, brightness was not a particularly salient attribute of instrument timbre. Some important caveats to this conclusion are necessary. For example, in the current investigation judgments about centroid manipulations were made in the context of other information about the shape of the spectral envelope. Thus, it is possible, even likely, that the perceived effect of filtering could have been greater in the absence of other salient spectral envelope structure (e.g., including obvious formants). Spectral centroid also was varied across only two instruments in the current investigation. While these instruments were selected to reflect a reasonably wide difference in centroids, it also is acknowledged that more support for the role of spectral centroid might have been obtained using a wider array of instruments, and therefore, greater variation in spectral slopes. However, it is noteworthy that such increased variation also would further increase the difficulty in isolating changes in spectral centroid from the shape of the spectral envelope.

4. GENERAL DISCUSSION

4.1. Assessment of amplitude envelope contributions to timbre

In Experiment 1 it was expected that both spectral envelope and amplitude envelope manipulations would contribute significantly to timbre identification and discrimination performance. Thus, it was somewhat surprising that the various amplitude envelopes were not only less salient than our spectral envelope manipulation insofar as they contributed minimally to timbre

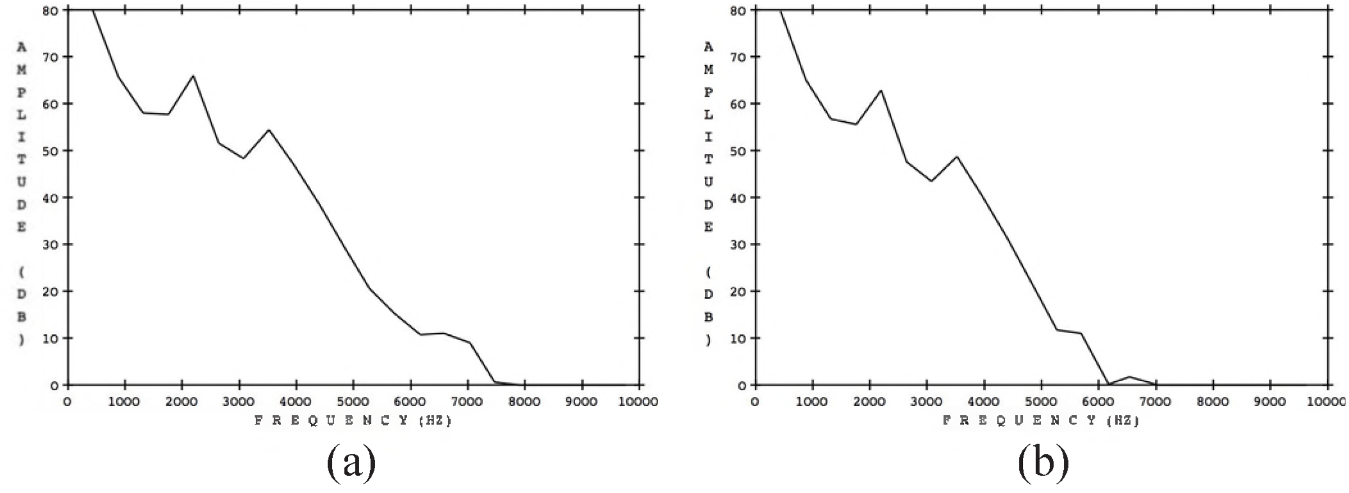


Figure 8. Spectral envelopes [amplitudes (in dB) v. frequency (in Hz)] for the resynthesized violin tone before (panel a) and after (panel b) low-pass filtering to match the spectral centroid of a tenor trombone tone.

identification, but also were not very discriminable from each other. This suggests that timbre identification performance was probably a reflection of basic psychoacoustic limits that were demonstrated by performance within the discrimination task.

Given that the amplitude envelopes in Experiment 1 were distinguished by rise time, these findings seem initially contradictory to several MDS studies (e.g., see Grey and Gordon, 1978; Iverson and Krumhansl, 1993; Krimphoff, et al. 1994; McAdams, et al. 1995) that have concurred that rise time is a primary dimension of instrument timbre. What factor or factors are likely responsible for this apparent discrepancy in results across studies?

A few potential explanations can probably be considered less likely. For example, it could be argued that the amplitude envelope dimension was weaker in the current study because entire amplitude envelopes were not included; natural decay portions of the tones were replaced with 10 ms linear ramps to tone offset. Furthermore, part of the steady-state portion of the original trombone tone was missing because it was truncated to 500 ms. Steady-states were almost absent from the tones derived from violin since the attack constituted the majority of the tone. It is therefore possible that these restrictions of the amplitude envelope limited its contribution to timbre identification and discrimination performance. However, it is the amplitude envelope of the *attack* that has typically been argued to be a critical timbre dimension, and this part of the envelope (epitomized by rise time) was retained. Furthermore, some MDS results have indicated a greater reliance on basic spectral information (number of harmonics) when it was manipulated along with certain amplitude envelope types (horn, string, or trapezoidal), although it is noteworthy that attack length was fixed (Miller and Carterette, 1975).

It also is acknowledged that only a single fundamental frequency was used throughout the current investigation. It therefore could be argued that different results might have been obtained had different pitches, from different registers, been included. For example, it has been demonstrated that identification of some timbres can change depending on pitch register, raising the possibility of “characteristic pitches” for an instrument, as well as a dependence upon training and exposure to a wide array of pitches to truly understand the overall timbre of any instrument (e.g., see Sandell and Chronopoulos, 1997). However, it should be noted that the A_4 pitch was selected to be well within the range that is typically produced by both instruments. Thus, there does not seem to be much reason to expect different outcomes with different pitches unless samples were selected to include atypically high or low pitches for one or both instruments.

It also could be suggested that the impact of amplitude envelope variation might have been different had the tones been presented within melodic sequences rather than in isolation. While this possibility is acknowledged, available evidence suggests that the contribution of amplitude envelope to timbre would actually be expected to be further reduced in such sequences. For example, Grey (1978) demonstrated that simplifications to the attacks of trumpet and clarinet sounds were more poorly discriminated in musical contexts, whereas the discrimination of simplifications to the spectral envelopes of bassoon sounds was unaffected by context. Likewise, Kendall (1986) found that the presence/absence of attack transients did not aid instrument recognition across pairs of short melodic sequences.

Additionally, it is acknowledged that the current investigation does not take into account known visual influences on instrument timbre. Specifically, there have been demonstrated shifts in timbre (i.e., the report of plucked vs. bowed strings) depending upon the presence of congruent/incongruent visual information (the synchronous movie of a musician plucking or bowing the instrument; see Saldaña & Rosenblum, 1993). While corresponding shifts would be likely to occur given corresponding visual productions for the tones in the experiments reported here, there is no reason to expect either categorically different responding or increased/decreased contributions from a given acoustic dimension in the presence of incongruent visual information.

Another possible explanation was raised at the end of Experiment 1, where we indicated that poor discrimination performance for the amplitude envelope parameter could be argued to be due to a truncated range of amplitude envelopes relative to the spectral envelope dimension. We pointed out that the instruments were selected to reflect a fairly wide distribution of values across both spectral and temporal dimensions, including rise times (see Table 1). Thus, it is not likely that our results are simply due to a lack of physical variation along the temporal dimension.

Further support for this interpretation comes from the results of pilot experiments for the current investigation. We had collected identification and discrimination data for tones derived from a larger set of instruments (piano, vibraphone, clarinet, and violin). Although the amplitude envelopes for these preliminary tones included linear onset ramps, and thus were not as natural as those used in the experiments reported here, they also essentially had maximally different rise time values across the set (40 ms for piano v. 458 ms for violin; the remaining rise times were 67 and 111 ms for the vibraphone and clarinet, respectively). In that pilot study listeners were reasonably sensitive (with mean values of d′ between 1 and 2) to differences in amplitude envelopes only for conditions involving the longest rise time, and mean values of d′ approached 2 only for the largest difference in rise times (an inordinately long 418 ms difference). In light of these results, it is unlikely that discrimination performance in the current investigation would have improved much unless extreme differences in rise times were used, likely needing some of the largest rise time differences that occur in natural instruments. Thus, it appears that, despite our reliance on a reasonably broad range of rise times within the included amplitude envelopes, listeners in Experiment 1 were not able to perceive much variation along that dimension.

A more reasonable explanation for the relatively poor discrimination performance with respect to amplitude envelope, as well as for the minimal contribution of this parameter to timbre identification, may be the possibility that rise time acts like a binary feature characterized by the presence or absence of a very abrupt attack. In other words, listeners would categorize amplitude envelopes during the attack as either abrupt or not. This argument does not require that rise times be categorically perceived, although there have been debates about whether or not categorical perception occurs along rise time continua (e.g., see Cutting, 1982; Donnadieu, McAdams, and Winsberg, 1996; Rosen and Howell, 1981, 1983). Rosen and Howell (1983) provided evidence that discrimination performance was at a maximum at the short rise time end of their continuum, and thereafter linearly decreased with increasing rise times. Thus, distinct performance differences along a rise time dimension are possible in the absence of a fixed category boundary between instruments with abrupt and gradual onsets.[3] This finding also is consistent with the notion that an abrupt stimulus with a particularly short rise time could act as a type of perceptual anchor against which all other stimuli are evaluated. If so, poor discrimination performance could be obtained despite large physical differences in rise time when the distribution of rise times does not include very short values. This was indeed the case in our Experiment 1, where the shortest rise time exceeded 150 ms.

A closer look at some classic MDS results that report rise time as a critical timbre dimension also lends further support for regarding rise time as a binary feature. One such study comes from McAdams, et al. (1995), who collected timbre dissimilarity ratings for pairs of tones taken from a larger stimulus set (that was previously developed by Wessel, Bristow and Settel, 1987 and used in a frequently cited MDS study by Krumhansl, 1989). Perceptual dimension 1 of their scaling solution (in their Figure 1), which is strongly correlated with rise time, shows a very large gap in the middle along with a clustering of instruments to either side of the dimension, particularly for instruments lacking abrupt onsets. Thus, despite a broad range of physical differences in rise time, listeners grouped instruments into perceptually abrupt (e.g., for vibraphone, guitar, piano, harp, and harpsichord) versus other values (e.g., bowed string, bassoon, English horn), and all non-abrupt rise times were perceived as quite similar to each other.

A similar conclusion can be reached upon an examination of MDS results from Iverson and Krumhansl (1993), which, like the current investigation, were based upon samples from the MUMS database, including the violin and trombone. They found similar MDS solutions based upon pair-wise ratings obtained for complete tones, their onsets only, or the remainder of the tones, leading to the conclusion that attributes used in making similarity judgments were present throughout the entire tone rather than being confined to the attack. Furthermore, in the horizontal dimension of their scaling solution for tone onsets (their Figure 6), which was labeled as relating to dynamic attributes of timbre, there was a perceptual clustering of all instruments except those with abrupt onsets (tubular bells, piano, vibraphone, and cello). Thus, very few instrument tones were judged as very dissimilar along that perceptual dimension, and those that were distinctly perceived were tones with an abrupt attack.

When these classic findings from MDS and categorical perception are taken together with the lack of a strong contribution from rise time to either discrimination or identification performance in the current investigation, they collectively suggest that rise time is only likely to permit reliable instrument identification when very short values along the dimension are contrasted with longer values. This suggestion should not be viewed as contradictory to seminal timbre research that reveals that timbre identification is negatively impacted when the attack is excised (e.g., see Saldanha and Corso, 1964). After all, removal of the attack effectively alters all amplitude envelopes to have abrupt onsets. Such a transformation should be easily perceived if rise time is perceptually evaluated as the presence or absence of abruptness. In fact, such results should be expected because relatively few sustained tone instruments' attacks approximate immediate onsets.

The idea of heightened salience for abrupt attacks also is consistent with other findings. For example, temporal order judgments are more accurate for tones with short rise times (e.g., see Bregman, Ahad and Kim, 1994; also see Pastore, Harris and Kaplan, 1981). This suggests that listeners may have difficulty detecting, or attending to, the temporal locations of intense portions of tones that have more gradual onsets.

Finally, while spectrotemporal variation was necessarily eliminated from the current investigation in order to focus on the respective contributions of spectral envelope and amplitude envelope to timbre, it is quite possible that one or more spectrotemporal dimensions could correlate highly with rise time as well, and thus enhance perception of attack transients in tones produced by natural instruments. As alluded to in the introduction, several important spectrotemporal parameters related to instrument timbre have been defined, including spectral centroid variation, spectral incoherence, and spectral irregularity (Beauchamp and Lakatos, 2002). Traditionally, in MDS studies naturally occurring spectrotemporal variations have been retained in the (attacks of) tones used to evaluate timbre. Under such stimulus conditions it is conceivable that spectral variation could contribute to, or even explain, findings for a perceptual dimension based on rise times that are overall measures of complex spectrotemporal phenomena (e.g., spectral flux or other spectrotemporal variables that are functions of rise time). This would be consistent with spectrotemporal changes that occur in a relatively systematic way when moving from instruments with abrupt to more gradual onsets. While beyond the scope of the current investigation, the relative salience of the spectral envelope and weakness of the amplitude envelope information in Experiment 1 leaves open the possibility that spectral variation could contribute in cases where a strong perceptual relation to rise time is found and spectrotemporal variation is not controlled. This possibility warrants future investigation to permit a more complete assessment of the contribution of amplitude envelope to timbre.

4.2. Clarifying the role of the spectral envelope

The advantage observed in Experiment 1 for the spectral envelope dimension relative to the amplitude envelope dimension does not appear to be due to a reliance on brightness perception. Experiment 2 was designed to directly evaluate this possibility for discrimination and timbre identification tasks. Spectral centroid, which is generally regarded as an acoustic measure related to the perceptual dimension of brightness, was equivalently manipulated in Experiment 2 by either altering the center frequencies of formants or by low-pass filtering tones. The latter manipulation preserved most of the original shape of the spectral envelope; only moderate discrimination performance was obtained in response to such variation, and virtually no effect on timbre identification was observed. Had spectral centroid been the primary cue that listeners used to evaluate timbre, then performance in the filtering conditions should have instead approached the near-ceiling levels of discrimination and sharp changes in timbre identification that were observed for the spectral envelope conditions.

The conclusion that listeners in both our experiments relied upon perceptual information about the shape of the spectral envelope, rather than brightness, should not be considered surprising. After all, it is the entire spectral envelope that reflects the natural resonances of the instrument body. In contrast, brightness reflects much less information for the listener, indicating the relative contribution of components within the spectral envelope having higher or lower frequencies.

This conclusion also is consistent with several findings from the timbre literature. This includes existing evidence that machine recognition of instrument timbres is significantly improved when supplying information about formant structure (via cepstral coefficients; see Brown, et al. 2001) rather than simply supplying data about the spectral centroid. The major findings from Experiment 2 also could be regarded as further evidence for the perceptual separability of spectral slope, which directly impacts the spectral centroid, and the shape of the spectral envelope (see Li and Pastore, 1995). This suggestion comes from the fact that our filtering manipulation, which was essentially a manipulation of spectral slope, resulted in much poorer discrimination and timbre identification performance than our formant-based manipulation of the spectral envelope. Finally, the conclusion for listeners' greater reliance on spectral envelope information also is consistent with the initial interpretations of spectral dimensions in some early MDS solutions (e.g., see Krumhansl, 1989).

So what should be made of the results from the current investigation in light of several classic MDS studies that demonstrate that the spectral centroid is the primary spectral measure that strongly correlates with an obtained perceptual dimension (e.g., Ehresman and Wessel, 1978; Krimphoff, 1993; Krimphoff, et al. 1994; McAdams, et al. 1995)? Given the lack of another purely spectral dimension within the MDS solutions from these studies, it is likely that the researchers initially attributed perceptual effects of spectral envelope shape to brightness. The latter perceptual dimension presumably reveals corresponding changes in the physical signal that could be summarized by the spectral centroid.

Then, how can we account for the very high correlation coefficients that have been obtained between spectral centroid measures and perceptual dimensions in these MDS solutions? It is important to remember that the spectral centroid is just an acoustic measure, one that reflects the center of the distribution of energy across the spectrum. The spectral centroid should be expected to be highly correlated with uniform shifts in peaks within the spectral envelope. For example, a typical vibrational sound source produces a prominent spectral peak near the sound's fundamental frequency, and the amplitudes of other harmonics are eventually reduced with increasing frequency (indicating a negative spectral slope or rolloff). Harmonics are also typically grouped within spectral envelopes or resonances. As these peaks, or formants, are distributed more widely, the spectral centroid also should increase. Likewise, displacement of any formant higher in frequency should also increase the spectral centroid. Both of these cases should result in a brighter timbre. Insofar as this description reflects natural acoustic consequences of different resonance patterns within musical instruments, there should be a systematic relationship between the spectral centroid and the shape of the spectral envelope. Perceived brightness based on changes in the spectral centroid should still be expected to contribute heavily to timbre in instances where there is a particularly strong or weak contribution from higher frequency components. However, the latter stimulus conditions are likely to also drastically reduce energy across the spectrum, and thus, minimally specify the spectral envelope.

It should not be too difficult to disentangle these properties. For the sake of simplicity, let us assume that we are presented with a signal consisting of just two formants. The same spectral centroid should result regardless of whether the center frequencies of both formants are compressed to the middle of the spectrum or are carefully adjusted in opposing frequency directions. However, clearly these two sounds have drastically different spectral envelopes, and therefore, they should be perceived as drastically different. After all, this difference between formant structures is the very distinction that exists between vowels based on changes in tongue height and position; modal productions of low back vowels like /a/ tend to have a compact spectrum based on F1 and F2 center frequencies, whereas high front vowels like /i/ tend to have more diffuse center frequencies for F1 and F2 (e.g., see Klatt and Klatt, 1990).

Insofar as brightness can be distinguished from other timbre information, it also is possible that it could contribute to tone perception in unexpected ways. For example, there are numerous demonstrations of pitch judgments being impacted by changes in timbre, particularly in musically untrained listeners. Large individual differences in the weighting of tone height and chroma have been attributed to brightness (e.g., Demany and Semal, 1993). Consistent with this possibility, shifts in the position of a spectral envelope with a fixed (bell) shape spectral have been found to impact pitch judgments in the *tritone paradox* (Repp, 1997; also see Deutsch, 1987). Recently, it has been demonstrated that musically untrained listeners will often perceive a tone's pitch to change upon removal of its fundamental frequency (Seither-Preisler, Johnson, Seither and Lütkenhöner, 2007). Also, Pitt (1994) used speeded classification tasks to demonstrate that pitch and timbre are perceptually integral properties, and that, furthermore, non-musicians are frequently likely to confuse timbre variation for changes in pitch. The reported size of pitch intervals by listeners with little musical training has even been shown to depend upon differences in the relative weighting of amplitudes in the synthesized tones' (upper and lower) harmonics (Russo and Thompson, 2005; Warrier and Zatorre, 2002).

It is possible that these various demonstrations share a common basis. A possible explanation for these phenomena is that some (particularly musically untrained) listeners frequently attribute changes in brightness, presumably in response to changes in the spectral centroid, to changes in pitch. According to this view, pitch would frequently be perceived to increase with increases in the relative weighting of higher-frequency components.

Ongoing efforts in our laboratories are therefore comparing performance across these traditional demonstrations, coupled with our manipulations of spectral centroid, in order to determine the potential impact of brightness on pitch judgments. Then the extent to which purely spectral dimensions can contribute to tone perception will hopefully be better understood. For now, the results of the current investigation provide further indications of the relative salience of some important dimensions of timbre. These results also suggest that a more thorough understanding of the nature of critical timbre dimensions is likely to be attained by comparing performance across an array of tasks that includes timbre identification.

5. ACKNOWLEDGEMENTS

The authors would like to thank the reviewers for their helpful comments, as well as the following research assistants for all their help with data collection: Charlotte Claflin, Rachel Ledebuhr, and Benjamin Rodes.

6. REFERENCES

American Standards Association (1960). *American Standard Acoustical Terminology*. New York.

Beauchamp, J. W. and Lakatos, S. (2002). New spectrotemporal measures of musical instrument sounds used for a study of timbral similarity or risetime-and centroid-normalized musical sounds. *Proc. 7th Int. Conf. on Music Perception & Cognition*, Univ. of New South Wales, Sydney, Australia, 592-595.

Boersma, P. and Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.6.01**)** [Computer program]. Retrieved May, 2007, from http://www.praat.org/.

Bregman, A. S., Ahad, P. and Kim, J. (1994). Resetting the pitch analysis system 2: Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones. *J. Acoust. Soc. Am.*, ***96***, 2694-2703.

Brown, J. C., Houix, O. and McAdams, S. (2001). Feature dependence in the automatic identification of musical woodwind instruments. *J. Acoust. Soc. Am.*, ***109***(3), 1064-1072.

Caclin, A., McAdams, S., Smith, B. K. and Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *J. Acoust. Soc. Am.*, ***118***(1), 471-482.

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.

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

Demany, L. and Semal, C. (1993). Pitch versus brightness of timbre: Detecting combined shifts in fundamental and formant frequency. *Music Perception*, ***11***(1), 1-13.

Deutsch D. (1987). The tritone paradox: Effects of spectral variables. *Perception & Psychophysics*, ***41***, 563-575.

Donnadieu, S. (2007). Mental representation of the timbre of complex sounds. In J. W. Beauchamp (Ed.), *Analysis, Synthesis, and Perception of Musical Sounds* (pp. 272-319). New York: Springer.

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

Ehresman, D. and Wessel, D. (1978). Perception of timbral analogies. *Rapports IRCAM*, ***13***, Paris: IRCAM.

Fant, G. (1973). *Speech Sounds and Features*. Cambridge, Massachusetts: MIT Press.

Grey, J. M. (1975). *An exploration of musical timbre*. Doctoral dissertation, Dept. of Music, Report No. STAN-M-2. Stanford Univ., Stanford, CA.

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

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

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

Haken, L., Fitz, K. and Christensen, P. (2007). Beyond traditional sampling synthesis: Real-time timbre morphing using additive synthesis. In J. W. Beauchamp (Ed.), *Analysis, Synthesis, and Perception of Musical Sounds* (pp. 122-144). New York: Springer.

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

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

Kendall. R. A. (2002). Music Experiment Development System (Version 2002-B) [Computer software]. Los Angeles, CA: Author,

www.ethnomusic.ucla.edu/systematic/Faculty/Kendall/meds.htm.

Klatt, D. H. and Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.*, ***87***(2), 820-857.

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. *Journal de Physique*, ***4***(C5), 625-628.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielsen and O. Olsson (Eds.), *Structure and Perception of Electroacoustic Sound and Music* (pp. 43-53). Amsterdam: Elsevier.

Li, X. and Pastore, R. E. (1995). Perceptual constancy of a global spectral property: Spectral slope discrimination. *Journal of the Acoustical Society of America*, ***98***(4), 1956-1968.

Macmillan, N. A. and Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

McAdams, S. (2001). Recognition of sound sources and events. In S. McAdams and E. Bigand (Eds)., *Thinking in Sound: The Cognitive Psychology of Human Audition* (pp. 146-198). New York: Oxford University Press.

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), pt. 1, 882-897.

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. *Psychological Research*, ***58***, 177-192.

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

Opolko, F., and Wapnick, J. (1987). McGill University Master Samples [CD-ROM]. Montreal, Quebec, Canada: jwapnick@music.mcgill.ca.

Pastore, R. E. (1990). Categorical perception: Some psychophysical models. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 29-52). New York: Cambridge University Press.

Pastore, R. E., Harris, L. B., and Kaplan, J. K. (1981). Temporal order identification: Some parameter dependencies. *J. Acoust. Soc. Am.*, ***71***, 430-436.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *J. Exp. Psych.: Human Perception and Performance*, ***20***(5), 976-986.

Repp, B. H. (1997). Spectral envelope and context effects in the tritone paradox. *Perception*, ***26***, 645-665.

Rodet, X. and Schwarz, D. (2007). Spectral envelopes and additive + residual analysis/synthesis. In J. W. Beauchamp (Ed.), *Analysis, Synthesis, and Perception of Musical Sounds* (pp. 175-227). New York: Springer.

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

Rosen, S. and Howell, P. (1983). Sinusoidal plucks and bows are not categorically perceived, either. *Perception & Psychophysics*, ***34***(3), 233-236.

Russo, F. A. and Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics*, ***67***(4), 559-568.

Saldaña, H. M. and Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & 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.

Samson, S., Zatorre, R. J. and Ramsay, J. O. (1997). Multidimensional scaling of synthetic music timbre: Perception of spectral and temporal characteristics. *Canadian J. Exp. Psych.*, ***51***(4), 307-315.

Samson, S., Zatorre, R. J. and Ramsay, J. O. (2002). Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain*, ***125***, 511-523.

Sandell, G. J. and Chronopoulos, M. (1997). Perceptual constancy of musical instrument timbres; generalizing timbre knowledge across registers. P*roc. Third Triennial Conf for the European Society for the Cognitive Sciences of Music*, Uppsala, Sweden, 222-227.

Seither-Preisler, A., Johnson, L., Seither, S. and Lütkenhöner, B. (2007). Ambiguous tone sequences are heard differently by musicians and non-musicians [A]. *Program for the 8th Conf. of The Society for Music Perception and Cognition (SMPC)*, 7.

Thayer, R. C. (1974). The effect of the attack transient on aural recognition of instrumental timbres. *Psychology of Music*, ***2***(1), 39-52.

Warrier, C. M. and Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics*, ***64***(2), 198-207.

Wessel, D.L., Bristow, D. and Settel, Z. (1987). Control of phrasing and articulation in synthesis. *Proc. 1987 Int. Computer Music Conf* (ICMC), pp. 108-116. Computer Music Association, San Francisco.

Winsberg, S. and De Soete, G. (1993). A latent-class approach to fitting the weighted Euclidean model, CLASCAL. Psychometrika, ***58***, 315-330.

Winsberg, S. and De Soete, G. (1997). Multidimensional scaling with constrained dimensions: CONSCAL. *British J. of. Mathematical and Statistical Psychology*, ***50***(1), 55-72.

7. NOTES

[1] Envelopes from both instruments had a very shallow increase in amplitude immediately prior to reaching (90 percent) criterion for the measurement of rise time, This was particularly true for the trombone tone. Adjusting the end measurement for rise time down by 1 dB, a decrease that should not be audible over an extended time period, produced rise time approximations of 345 ms and 156 ms for violin and trombone, respectively. The difference between these measures is 21 ms more than for our initial measures. Thus, it is quite possible that functional differences in rise times for the listeners in Experiment 1 were actually slightly greater than indicated by the values in Table 1.

[2] Although the assumption of sphericity appeared to have been violated bv the ANOVA for each identification response, all reported effects continued to be significant when relying instead upon the Greenhouse-Geisser correction procedure. Thus, critical findings remained the same regardless of which statistical procedure was used.

[3] The observed changes in instrument timbre identification might lead some readers to question whether or not the instrument timbres in the current investigation were categorically perceived. It should be noted that the experiments reported here were not designed to directly assess categorical perception. There were very broad acoustic differences between adjacent steps along continua composed of very few stimuli. This complicates any determination of whether true categorical boundaries were perceived (i.e., discrete changes in timbre given a relatively small physical change in the middle of an acoustic dimension). Furthermore, in contrast to what is typically done in studies of categorical perception, in the discrimination tasks not all adjacent stimuli were compared along a given dimension. Specifically, the two hybrid stimulus values were not directly compared.

Despite this apparent limitation, there are indications that the violin and trombone timbres were not categorically perceived. For example, discrimination performance for single-step comparisons was nearly perfect for within-category comparisons (based on timbre identification data) along the spectral envelope/formant structure dimension in both experiments, which contrasts with the expected troughs of within-category discrimination performance that is characteristic of categorically perceived events (see Figure 4). Furthermore, the changes in instrument identification that were observed along the formant structure dimension were clearly gradual in Experiment 1 (see Figure 5), which is inconsistent with a clear demonstration of a category boundary. Evidence from the amplitude envelope dimension also is inconsistent with arguments for categorical perception. Highly accurate levels of discrimination performance were never obtained for the amplitude envelope dimension (even when larger step sizes were used). Additionally, timbre identification in Experiment 1 (see Figure 5) was generally unaffected by this dimension. [For a review of criteria and stimulus conditions needed to effectively demonstrate categorical perception, see Pastore (1990)].