# The contribution of timbre attributes to musical tension[a)]

Morwaread M. Farbood[b)]
*Department of Music and Performing Arts Professions, Steinhardt School, New York University, 35 West 4th Street, Suite 1077, New York, New York 10012, USA*

Khen C. Price
*Department of Computer Science, Courant Institute, New York University, USA*

(Received 4 June 2016; revised 25 November 2016; accepted 8 December 2016; published online 20 January 2017)

Timbre is an auditory feature that has received relatively little attention in empirical work examining musical tension. In order to address this gap, an experiment was conducted to explore the contribution of several specific timbre attributes—inharmonicity, roughness, spectral centroid, spectral deviation, and spectral flatness—to the perception of tension. Listeners compared pairs of sounds representing low and high degrees of each attribute and indicated which sound was more tense. Although the response profiles showed that the high states corresponded with increased tension for all attributes, further analysis revealed that some attributes were strongly correlated with others. When qualitative factors, attribute correlations, and listener responses were all taken into account, there was fairly strong evidence that higher degrees of roughness, inharmonicity, and spectral flatness elicited higher tension. On the other hand, evidence that higher spectral centroid and spectral deviation corresponded to increases in tension was ambiguous.
 [http://dx.doi.org/10.1121/1.4973568]

[JFL] Pages: 419–427

## I. INTRODUCTION

Tension is a percept that is an essential aspect of how listeners hear sound and experience music. It has long been regarded by music theorists as a fundamental, emergent phenomenon that is holistic and multidimensional (Nielsen, 1987). Tension as a concept has rarely been defined formally in experimental work, perhaps due to the assumption that it is an intuitive concept to listeners. Generally speaking, a rise in tension has been equated with an increase in excitement or intensity while a decrease in tension has been described as a feeling of relaxation or resolution. Prior work has shown that listeners evaluate tension consistently, as indicated by high within-subject and between-subject agreement in both discrete and continuous tension judgments (Bigand *et al.*, 1996; Farbood, 2012; Farbood and Upham, 2013). Furthermore, tension responses are not influenced by the musical preferences of listeners (Lychner, 1998) or familiarity with the musical stimuli (Fredrickson, 1999). Tension also provides a link between low-level auditory features and high-level psychological and emotional response (Kruhmansl, 1997; Rozin *et al.*, 2004; Eerola and Vuoskoski, 2010; Lehne *et al.*, 2013).

Given the importance of tension in music perception, there have been numerous behavioral and (more recently) neuroimaging experiments that have examined how various auditory and musical features contribute to listeners' experiences of tension. The most frequently examined of these features include loudness (Nielsen 1983; Krumhansl, 1996; Ilie and Thompson, 2006; Granot and Eitan, 2011; Farbood, 2012), melodic contour (Nielsen, 1983; Bigand *et al.*, 1996; Krumhansl, 1996; Granot and Eitan, 2011; Farbood, 2012), and harmony (Nielsen, 1983, 1987; Bigand *et al.*, 1996; Krumhansl, 1996; Bigand and Parncutt, 1999; Lerdahl and Kruhmhansl, 2007; Farbood 2012).

Timbre, on the other hand, has received relatively little attention as a contributing factor to tension perhaps due to the general difficultly in defining its perceptual dimensions. For the purposes of the current work, we employ a working definition of timbre that encompasses aspects of sound that are not accounted for by pitch, loudness, duration, spatial position, or environmental characteristics (McAdams, 2013). The goal of the present study was to systematically examine how different timbre attributes contribute to perceived tension. The experiment described here was designed to feature stimuli with clearly perceptible changes to specific timbre attributes that have been implicated in prior work as being important to timbre perception in general.

Prior studies on timbre and tension fall into two categories: those that explore higher-level features linked to timbre, and those that examine specific spectral features of timbre. High-level features such as orchestration have been shown to affect the perception of tension in musical pieces of varying styles (Paraskeva and McAdams, 1997). Increases in the density of sound arising from complex instrumentation are perceived as increased tension (Nielsen 1983, 1987), while decreasing note density that occurs at the end of major sections in a musical piece corresponds to the greatest drops in perceived tension (Krumhansl, 1996). The only low-level timbre attribute that has been directly linked to tension is

[a)]Portions of this work were presented in "Timbral features contributing to perceived auditory and musical tension," *Proceedings of International Conference on Music Perception and Cognition*, Seoul, Korea, August 2014.
[b)]Electronic mail: mfarbood@nyu.edu

roughness. Bigand *et al.* (1996) found that higher tension is correlated with higher roughness in tonal chord progressions, and Pressnitzer *et al.* (2000) showed that this effect also applies to atonal harmony. Both studies attempted to reproduce experimental results using Hutchinson and Knophoff's (1978) model of dissonance for dyads, which is based on the concept of roughness (or beating) as the source of dissonance.

Related studies have explored specific timbre attributes in the context of affective arousal rather than tension. Although tension and arousal are not necessarily the same phenomenon, they are at the very least closely related; an increase in tension corresponds to an increase in arousal (Krumhansl, 1997; Ilie and Thompson, 2006; Eerola and Vuoskoski, 2010). In the context of work on music and emotion, spectral centroid and spectral flatness are two specific attributes that have been examined. Schubert (2004) found that spectral centroid and texture (the latter defined as the number of instruments playing) did not produce consistent predictions of two-dimensional (valence-arousal) emotion ratings of music. Similarly inconclusive results were obtained by Bailes and Dean (2012), who examined how spectral flatness corresponded to continuous two-dimensional ratings of (primarily) electroacoustic music. They concluded that there was little support for the effect of spectral flatness on arousal.

Aside from perceptual studies, there is also a considerable amount of work in the field of music information retrieval that uses machine learning to classify and predict mood or emotion in music directly from low-level audio features. However, this research primarily focuses on engineering approaches and is beyond the scope of the current work. For a review of literature on machine recognition of emotion in music, see Kim *et al.* (2010) and Yang and Chen (2012).

The work discussed above encompasses widely varying goals and methodologies, and the majority of those studies do not directly focus on the contribution of specific timbre attributes to tension. The present work offers a more in-depth examination of how timbre affects tension perception by systematically examining a number of perceptually relevant spectral features. An experiment was conducted to examine the contribution of five specific timbre attributes on tension perception: inharmonicity, roughness, spectral centroid, spectral deviation, and spectral flatness. These attributes were chosen based on three criteria: (1) they have been reported to a have a strong correlation with listeners' ability to discriminate timbre; (2) they are linked or correlated with musical or tonal tension; and (3) they can be measured and synthesized for the purposes of creating stimuli for an empirical study.

## II. TIMBRE ATTRIBUTES

Spectral centroid is an attribute of timbre that is associated with the perceived brightness of a sound (Schubert and Wolfe, 2006). A substantial body of work has shown that spectral centroid is a prominent factor in timbre perception; these include multidimensional scaling (MDS) studies that analyze dissimilarity ratings for pairs of musical instrument sounds (Grey and Gordon, 1978; Iverson and Krumhansl, 1993; McAdams *et al.*, 1995; Lakatos, 2000; Caclin *et al.*, 2005) as well as music information retrieval applications of timbre attributes (Eronen and Klapuri, 2000; Peeters *et al.*, 2000; Agostini *et al.*, 2003; Zhang and Ras, 2007; Peeters *et al.*, 2011). Spectral centroid is defined as the weighted mean of the energy found in the different frequency bins that are produced by a fast Fourier transform (FFT) or any other applicable transformation between the time and frequency domains. The resulting value is in Hertz:

$$\mu = \frac{\sum_k f_k a[k]}{\sum_k a[k]}, \tag{1}$$

where $a[k]$ is the amplitude corresponding to bin $k$ in a discrete Fourier transform (DFT) spectrum and $f_k$ is the frequency corresponding to bin k.

Spectral standard deviation (also termed spectral spread) is another frequently referenced timbre attribute that is commonly used in the context of music informational retrieval research (Eronen and Klapuri, 2000; Peeters, 2004; Zhang and Ras, 2007; Peeters *et al.*, 2011). It is a perceptually relevant feature as evidenced by its correlation to one of the resulting dimensions of MDS analysis used to differentiate musical instrument timbre (Krumhansl, 1989; McAdams, 2013). Spectral deviation is obtained by calculating the spread of the energy distribution across the spectrum. For example, a pure tone will have no spectral deviation. With the addition of partials, the greater the distance between the frequencies of the complex, the larger the spectral deviation. It is calculated as the square root of the spectral variance:

$$SSD = \left[\frac{\sum_k (f_k - \mu)^2 a[k]}{\sum_k a[k]}\right]^{1/2}, \tag{2}$$

where $a[k]$ is the amplitude of bin $k$ in a DFT, $f_k$ is the frequency corresponding to bin $k$, and $\mu$ is the spectral centroid in Hertz.

The spectral flatness measure of a signal corresponds to how similar its spectrum is to white noise. A spectrum with energy evenly distributed across its range is considered flat; in contrast, a "spiked" spectrum with clearly noticeable peaks is said to have low spectral flatness, and is attributed with more harmonic or pitched sounds. Flatness is another commonly referenced timbre attribute (Eronen and Klapuri, 2000; Peeters, 2004; Peeters *et al.*, 2011) that is included in the MPEG-7 standard for audio and other multimedia content along with spectral centroid and deviation (ISO/IEC, 2002; Zhang and Ras, 2007). The flatness of a signal is the ratio of a geometrical mean to an arithmetic mean:

$$SFM = \frac{\left(\prod_{k=1}^{k} a[k]\right)^{1/K}}{\frac{1}{K}\sum_{k=1}^{k} a[k]}, \tag{3}$$

where $a[k]$ is the amplitude of the $k$th frequency bin, and $K$ is the number of bins.

Inharmonicity is a feature that is based on how partials are offset from integer multiples of the fundamental frequency of a pitch. In addition to being referenced as a timbre attribute in music information retrieval research (Agostini *et al.*, 2003; Peeters, 2004; Peeters *et al.*, 2011; Barbancho *et al.*, 2012), it is an aspect of timbre that has been modeled in the synthesis of various stringed instruments. The inharmonicity of a given signal can be calculated as follows:

$$I = \frac{2}{f_0} \frac{\sum_{n=1}^{N} |f_n - nf_0| \left(A_t^n\right)^2}{\sum_{n=1}^{N} \left(A_t^n\right)^2}, \tag{4}$$

where $f_n$ is the $n$th harmonic of the fundamental frequency $f_0$, and $A_n$ is its corresponding amplitude at time frame $t$.

Roughness is a sensation that occurs when pairs of sinusoids are close enough in frequency such that listeners experience a beating sensation. It is closely associated with sensory dissonance, a term first used by Helmholtz (1885), who proposed that the perception of dissonance corresponded to the beating between partials and fundamental frequencies of two tones. Plomp and Levelt (1965) extended Helmholtz's work, showing that the transition between consonant and dissonant intervals was related to critical bandwidth; their experimental results indicated and that the most dissonant intervals between pure tones occurred when frequency differences were about a quarter of the bandwidth. Hutchinson and Knopoff (1978) then extended the findings of Plomp and Levelt by applying them to multiple simultaneous tones. Roughness is thus a feature that is evident in tonal as well as purely timbral contexts. A widely used approach to measuring roughness has been suggested by Sethares (1998); using this method, the peaks of the spectrum are determined, and then the dissonance between all pairs of spectral peaks are calculated and averaged. The dissonance value for each pair is determined based on the findings of Plomp and Levelt.

Given the formal descriptions of the five spectral features described above, the goal of the present study was to synthesize stimuli with contrasting degrees of those attributes and then evaluate their contributions to perceived tension. Additionally, the stimuli were designed to explore the *directionality* of the features with regard to how they correlated with tension—that is, whether increases in relative levels of an individual feature also corresponded to increases (as opposed to decreases) in tension, or if this relationship was unclear. The working hypothesis was that an increase in each of these attributes would correspond to an increase in perceived tension.

## III. METHOD

### A. Stimuli

The stimuli were designed to have two contrasting states for each attribute: one with a low degree of a particular attribute (state A) and one with a high degree of that attribute (state B). The low and high degrees of each attribute were generated using the formal descriptions of the attributes in conjunction with a perceptual assessment of contrast by the authors. Although it was impossible to avoid covariance between all features when synthesizing the stimuli, special attention was made to keep the spectral centroid constant in cases where it was not the targeted attribute. This was due to the established importance of spectral centroid in timbre discrimination experiments. There were a total of five different categories of stimuli generated, representing the five timbre attributes in question. The subsequent use of the term "category" will refer to the group of stimuli designed to focus on a specific attribute. Likewise, "featured attribute" will refer to the target attribute in a given stimulus category.

Each stimulus consisted of a pair of renderings of either state A or state B separated by 1.2 s of silence. Each A or B state was also 1.2 s in duration, with onset and offset ramps of 20 ms to avoid clicking. For all pairs there were four possible state orderings: AB, BA, AA, and BB. The stimuli were generated at three different pitch registers for each category (labeled low, mid, and high). For example, state A with a low degree of roughness and $f_0 = 220$ Hz would be presented with a corresponding state B with high degree of a roughness and the same $f_0$. Table I lists the fundamental frequencies used to generate the pairs of sounds in each register × attribute category. The $f_0$ values were chosen to represent the same pitch at different octaves. Two different ranges of frequencies were used depending on which set of $f_0$ values resulted in better aural distinction between the pitch registers as well as the general clarity of sound in a given attribute category.

The stimuli in the spectral centroid category were generated by synthesizing a sinusoid representing a fundamental frequency $f_0$ along with several upper partials. Independent amplitude envelopes were applied to each partial to manipulate the spectral centroid, while $f_0$ was not modified. Depending on the frequency of $f_0$, the resulting difference between the A (low) state and B (high) state centroid values ranged from ∼400 to 800 Hz.

Stimuli in the spectral deviation category were first generated to change incrementally over time by synthesizing a sinusoid along with sidebands using frequency modulation (FM) synthesis. Sidebands were added at increasing distances from the sinusoid so that the spectral energy was dispersed while keeping the spectral centroid constant. This was done by increasing the modulation index in the FM synthesis formula,

$$f(t) = f_0 + f_m X \cos(f_m t), \tag{5}$$

TABLE I. Fundamental frequencies used to generate stimuli.

| | Low (Hz) | Mid (Hz) | High (Hz) |
|---|---|---|---|
| Spectral centroid | 110 | 220 | 440 |
| Spectral deviation | 220 | 440 | 880 |
| Spectral flatness | 110 | 220 | 440 |
| Inharmonicity | 110 | 220 | 440 |
| Roughness | 220 | 440 | 880 |

where $f_0$ is the carrier frequency, $f_m$ is the modulating frequency, and $X$ is the modulation index. The A states were represented by the single sinusoid $f_0$, with a deviation of 0, and the B states were represented by the transformed signal, which had a deviation ranging from ~30 to 90 Hz, depending on $f_0$.

To generate stimuli in the spectral flatness category, noise was added to a complex tone with four harmonics, including $f_0$. This was done by filtering pink noise so that the pass band was centered around the tone's spectral centroid. The filters and noise used in the generation of these stimuli were created using the MATLAB DSP toolbox. This method resulted in a uniform timbre rather than two distinct timbres. The ratio between the noisy and tonal parts of the signal determined the degree of spectral flatness. The A states had no noise added while the B states had a significant amount noise added such that the noise component was nearly equivalent in intensity to the pitched portion.

In order to generate timbres with contrasting inharmonicity, a complex harmonic tone was first generated representing a fundamental frequency $f_0$, and partials were added at increasing distances from the harmonic series of $f_0$. More specifically, for $f_0$, partials $P_m$ were added at specific frequencies according to the following equation:

$$P_m = Mf_0(1 \pm \alpha), \quad 0 < \alpha < 1, \quad M \in \mathbb{N}, \tag{6}$$

where $P_m$ is the $M$th partial, and $\alpha$ is the inharmonicity factor, in this case $\alpha = 0.22$. The added partials alternated between frequencies higher and lower than integer multiples of $f_0$. This method allowed the spectral centroid to remain in a fairly restricted range while the degree of inharmonicity could change significantly. The resulting A states had no inharmonicity while the B states had inharmonicity values ranging from ~0.09 to 0.14.

Since closely paired sinusoids generate more beating, or sensory dissonance, the roughness of a stimulus was increased by adding partials of relatively close intervals around existing partials. By increasing the magnitude of these added partials, the perceived beating sensation was correspondingly increased. Neighboring frequencies were added above and below each of the upper partials in order to maintain a consistent spectral centroid. These frequencies were selected by specifying an alpha value, in this case $\alpha = 0.06$. For example, if the first upper partial $f_1$ was 880 Hz, then the following frequencies would be added:

$$f_1(1\text{–}2\alpha) = 774\,\text{Hz},$$

$$f_1(1\text{–}\alpha) = 827\,\text{Hz},$$

$$f_1(1 + \alpha) = 933\,\text{Hz},$$

$$f_1(1 + 2\alpha) = 986\,\text{Hz}.$$

Additionally, an amplitude envelope was applied to the signal so that no increase in intensity was introduced. This resulted in A states that had no apparent roughness and B states with roughness values ranging from ~220 to 3500, depending on $f_0$.

There were 60 stimuli in total (5 features × 3 registers × 4 state orders), all generated in 44.1 kHz, 16-bit mono WAV format. Overall intensity level equalization was done automatically for all stimuli using the Echonest API (Jehan, 2013) so that loudness differences across stimuli were less than 1 dB. The MATLAB Genesis Loudness Toolbox (Genesis, 2009) was used to obtain time-dependent loudness measurements for each synthesized timbre for verification purposes. Amplitude envelopes and changes in intensity were applied to compensate for any changes in loudness between the A and B states, as well as differences in loudness between stimuli.

In addition to the qualitative assessments, objective measures were also used to verify whether the A and B states were sufficiently contrasted. Quantitative measures of all attributes were produced using the MIRtoolbox (Lartillot *et al.*, 2008) in order to provide an independent verification of these differences. The primary MIRtoolbox functions employed were *mircentroid*, *mirspread*, *mirflatness*, *mirinharmonicity*, and *mirroughness*. A frame length and hop factor of 1.1 s were used to cover the duration of each sound not including the initial 100 ms, resulting in a single, nonoverlapping frame of analyzed audio for each sound. In the case of spectral centroid, spread, and flatness, cutoff frequencies were used to exclude frequency bins beyond the range of the generated partials.

### B. Procedure

Forty-six subjects, 21 female and 25 male, mean age 25.54 yr (SD = 8.45), took part in the experiment. In a written questionnaire, all subjects reported having no hearing impairments. A substantial number of participants were undergraduate or graduate music students at New York University, although overall the subject pool encompassed a wide range of musical training. The average years of formal training on a primary musical instrument was 5.50 (SD = 5.36), and the average number of semesters of college-level music theory training was 2.49 (SD = 3.05).

The experiment was presented in a MATLAB graphical interface using the Psychophysics Toolbox Version 3 for audio playback. Participants listened to the stimuli, presented in random order, on Sennheiser HD 650 headphones in a sound-isolated (hemi-anechoic) chamber. For each trial, listeners were asked to judge which of two sounds (first or second) was more tense or if they sounded equally tense. Tension was described generically using the following language: "Less tension corresponds to a feeling of relaxation or resolution, while more tension corresponds to the opposite sensation." This wording was designed to be broad without resorting to circular definitions (e.g., "more tension" described as "feeling more tense") as well as avoiding terminology that evoked specific auditory percepts such as loudness (e.g., "intensity"). In a post-test questionnaire, subjects were asked if they had any problems understanding the experiment, and none reported any difficulties. Subjects were allowed to play each sound only once before responding. After each response, participants rated how confident they were in their judgment (five values ranging from "not confident at all" to "very confident"). In addition to the

auditory stimuli, there was a second block of visual stimuli. This separate experiment exploring visual tension is not reported here (the order of the two blocks were counterbalanced among subjects).

## IV. RESULTS

The response profiles for all of the conditions are shown in Fig. 1. The B states were labeled as more tense by a statistically significant margin in all attribute categories when responses are combined across pitch register categories. Higher spectral centroid, increased spectral deviation, greater spectral flatness, the presence of inharmonicity, and more roughness all corresponded to increased tension. The corresponding chi-square goodness-of-fit statistics are shown in Table II. The results are also consistent when broken down by pitch register: none of the chi-square tests—when employing a $3 \times 3$ (pitch register $\times$ response type) contingency table for each attribute $\times$ order condition—are significant.

Despite the statistical significance of the results, it is evident from the relatively smaller chi-square values as well as visual inspection of the response profiles that perception of the stimuli in the spectral centroid category was not as consistent compared to the other categories. The mean confidence ratings in the spectral centroid category were also the lowest of any category: 4.31 (SD = 0.89) compared to 4.36 (SD = 0.87) for inharmonicity, 4.42 (SD = 0.87) for roughness, 4.48 (SD = 0.82) for spectral flatness, and 4.53 (SD = 0.76) for spectral deviation. Parametric statistics were used to analyze the confidence values due to the manner in which they were presented on the data collection interface: the five confidence options were displayed to subjects as a horizontal row of equally spaced radio buttons with only the ends labeled. In other words, this representation presented the confidence ratings more as interval data rather than ordinal values. A one-way, repeated-measures analysis of variance (ANOVA) showed that the differences between confidence ratings by category were significant, $F(3.45, 155.17) = 5.77$, $p < 0.001$, $\eta_p^2 = 0.11$, $\eta_G^2 = 0.023$ (Greenhouse–Geisser corrected). Tukey–Kramer multiple-comparison tests showed significant differences between the confidence means in the two lowest and two highest-rated categories (between centroid/deviation, centroid/flatness, inharmonicity/deviation, and inharmonicity/flatness).

Despite the statistically significant correspondence between response profiles and featured attributes, it is not possible to vary any one attribute independently such that none of the other attributes would be affected. Thus more than a single attribute may have influenced listeners' responses for a given stimulus. Two correlation matrices shown in Fig. 2 help illuminate the relationship between the attributes and response profiles in each category. Unlike typical correlation matrices, they are not symmetric, as the rows represent correlations corresponding only to the stimuli in the attribute categories labeled on the left-hand side. Figure 2(a) shows the correlations between the objective measures of the attributes for the stimuli in each category, produced by the MIRtoolbox functions, and the proportion of responses corresponding to the A and B states for each stimulus in that category. The rows in Fig. 2(b) show correlations between the objective attribute measures for stimuli in each category and the featured attribute in that category.

Taking spectral centroid as an example, timbre attributes for all of the AB stimuli in the spectral centroid category were produced objectively using the MIRtoolbox functions. The resulting values for each attribute—six values corresponding to the two A and B states for three stimuli (one per pitch register)—were then correlated with the percentage of responses corresponding to each A/B state [Fig. 2(a)] and the spectral centroid values for each state [Fig. 2(b)]. The AA, BB, and BA stimuli were not used in the correlation analysis because the attribute values were redundant. Given the small number of data points, only the very large $r$-values ($>0.95$) were significant at an alpha level of 0.001 (correcting for multiple comparisons). In any case, statistical significance is not particularly useful here—the $r$-values simply provide a helpful measure of how coordinated the various attributes and the respective responses were in each category.

The correlations between featured attributes and response profiles in each category were in general strongly positive. The one exception was the spectral centroid category, in which all attributes except for roughness were strongly correlated with centroid. Both the response profiles and lower confidence ratings were indicative of more ambiguous responses when compared with the other categories. One possible explanation for this could be the opposing influence of roughness; the strong negative correlations between roughness and the other attributes in conjunction with the resulting response ambiguity suggests the possibility that roughness was a confounding factor.

The responses in the spectral deviation category were positively correlated with the featured attribute but more strongly correlated with roughness and inharmonicity. Furthermore, from a qualitative perspective, the attribute that seemed the most salient to the ear for stimuli in this category was roughness. In short, both the objective and subjective measures indicated roughness was again a significant factor in tension judgments. Given the apparent prominence of roughness, as well as no strong correlation between spectral deviation and the response profiles in other categories, it was not possible to conclude that spectral deviation affected tension perception.

In the spectral flatness category, both sets of correlations looked very similar. Flatness, inharmonicity, and roughness all increased when a significant noise component was added to state B. To the ear, this addition of noise was clearly the salient difference between the A and B states. Since the formal definition of spectral flatness is directly linked to noise, we can conclude that an increase in perceived noisiness corresponded to an increase in tension by both objective and qualitative measures.

The results in the inharmonicity category were the clearest due to the lack possible confounding factors. From a qualitative perspective, only changes in inharmonicity were apparent to the ear. From an objective perspective, no other attributes were positively correlated with inharmonicity and the correlations with response profiles showed a very high positive correlation only for inharmonicity; there was

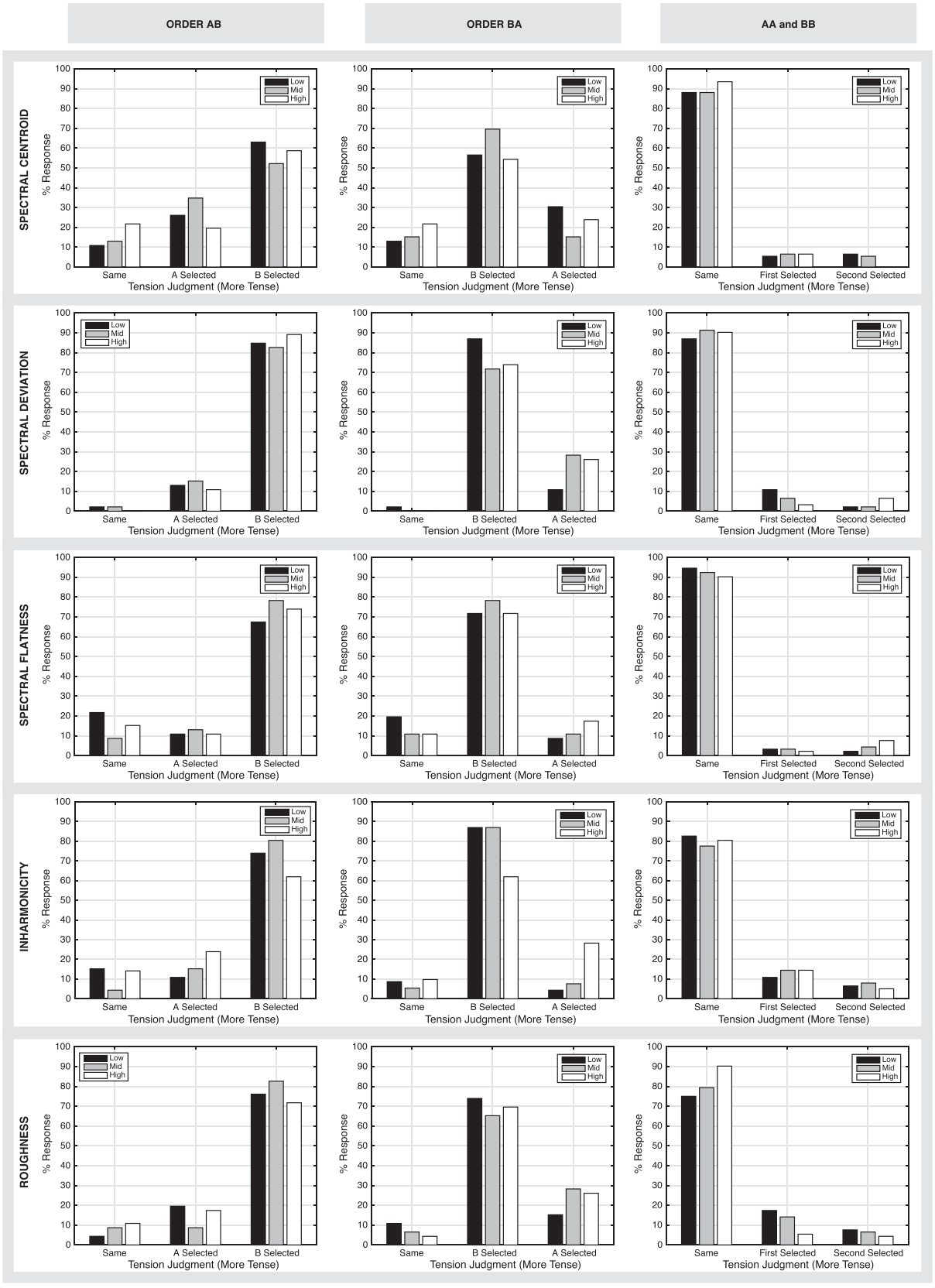



FIG. 1. Tension judgments grouped by attribute category and order type. The values in the legend indicate pitch register categories.

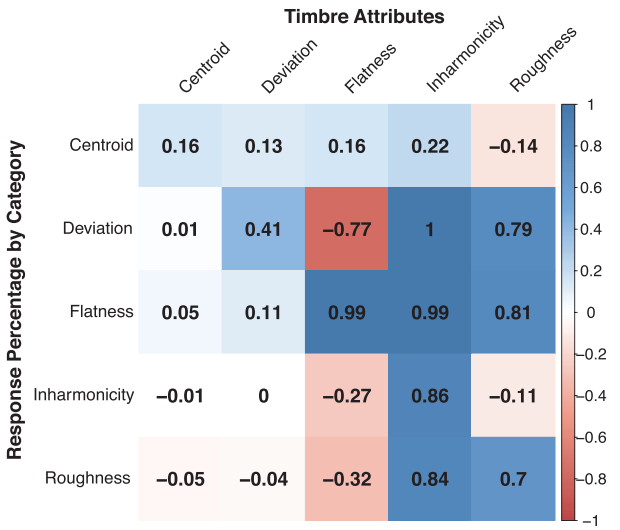
no correlation or a slight negative correlation between the responses and all other attributes.

In the roughness category, both sets of correlations showed that inharmonicity and roughness were both strongly correlated with each other and the response profiles. Using just the objective measures, it was not possible to verify roughness as the primary attribute contributing to tension. However, from a qualitative perspective, the presence or

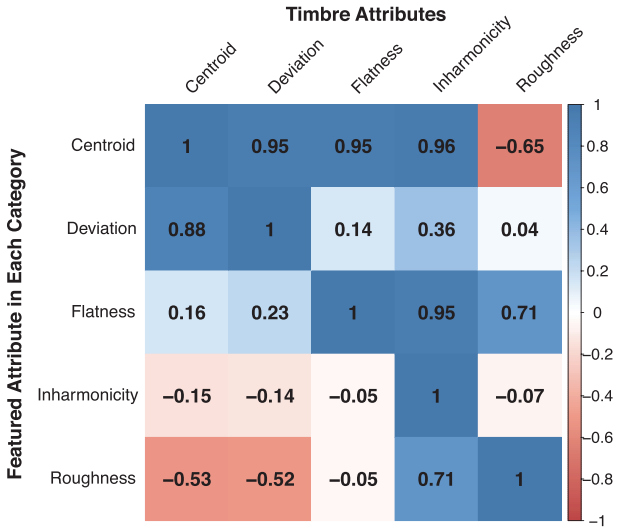
TABLE II. Chi-square goodness-of-fit tests. Note: $df = 2$ in all cases; all values are statistically significant at an alpha level of 0.01 (critical $\chi^2 = 9.21$).

| | AB | | BA | | AA/BB | |
|---|---|---|---|---|---|---|
| Feature | $\chi^2$ | $N$ | $\chi^2$ | $N$ | $\chi^2$ | $N$ |
| Spectral centroid | 40.5 | 138 | 45.5 | 138 | 397.0 | 276 |
| Spectral deviation | 171.8 | 138 | 130.5 | 138 | 392.2 | 276 |
| Spectral flatness | 98.9 | 138 | 102.3 | 138 | 433.3 | 276 |
| Inharmonicity | 106.5 | 138 | 150.8 | 138 | 286.3 | 276 |
| Roughness | 118.5 | 138 | 86.8 | 138 | 290.0 | 276 |

lack of inharmonicity was not evident to the ear; the roughness of the B states of the stimuli was the salient feature. Furthermore, the possible influence of roughness in the other stimuli categories provided additional evidence for its importance in tension perception.




(a)




(b)

FIG. 2. (a) Correlation matrix for response percentages to A and B states, grouped by stimulus category (represented by rows), and attribute values as analyzed directly from the corresponding audio (labeled on top). (b) Correlations between featured attributes and all other attributes in each category. Values shown are pairwise correlation coefficients (Pearson's $r$, two-tailed), $df = 4$.

## V. DISCUSSION

The present study examined the contribution of five timbre attributes—inharmonicity, roughness, spectral centroid, spectral deviation, and spectral flatness—to perceived tension. An experiment was conducted that used stimuli featuring salient changes in each attribute. Subjects were asked to compare two timbre states representing high and low degrees of a particular attribute and then choose which one sounded more tense. The results indicated that increases in all five attributes corresponded to increases in perceived tension. Responses to the stimuli in the spectral centroid category, though statistically significant, were more ambiguous than the others: 59% of subjects chose timbre state B (higher spectral centroid) across all pitch registers, a considerably lower proportion than in the other attribute categories. The other four categories had responses ranging from 73% to 81% for state B.

Correlations between attributes and response profiles in each attribute category provided additional context to these results. In all categories except inharmonicity, multiple attributes strongly correlated with the responses or the featured attribute. Given the correlation results, it is arguable that, with the exception of inharmonicity, definitive conclusions on the direct contribution of individual attributes to perceived tension are not possible. However, these objective measures were unable to quantify the perceptual salience of the featured attributes versus the other attributes in their respective stimuli categories. For example, from a qualitative perspective, states A and B in the spectral flatness category did not distinctly sound harmonic or inharmonic despite the very high correlation between responses and inharmonicity—the salient feature of those states was the presence or lack of noise. Technically speaking, timbres with more noise are also by definition more inharmonic. The point in this case is that from a listener's perspective, the distinctive change in the timbre, and thus what elicits the tension response, is the element best defined by spectral flatness. Similar arguments could be made for the other attribute categories, with the exception of spectral deviation; from a qualitative perspective, change in roughness was evident in parallel with the featured change in spectral deviation. In short, while the correlation results provided another window into how the other attributes corresponded to the response profiles and each other, they did not provide insight into their relative salience.

The correlation results also do not indicate universal relationships between attributes. They are only reflective of the specific timbres created for the current study. Nonetheless, they do not conflict with the results of Peeters *et al.* (2011) who explored intercorrelation between timbre attributes in a large database of musical instrument tones. When Peeters *et al.* grouped attributes by strength of correlation (using interquartile ranges for attribute values since the instrument tones were not static), one cluster included centroid, deviation, and other spectral time-varying attributes; another cluster included inharmonicity, noisiness, and other attributes indicating signal periodicity or lack thereof.

Given the overall results of the analyses presented, there was strong evidence for the contributions of inharmonicity, roughness, and spectral flatness to perceived tension. Increases

in those three features most clearly corresponded to increases in tension, and these positive results support the initial hypotheses. The concept of noisiness as tension-inducing is intuitive; for example, there is a great deal of literature on environmental noise as a source of psychological stress. In the case of inharmonicity and roughness, they are both associated with dissonance (Helmholtz, 1885; Plomp and Levelt, 1965; Hutchinson and Knopoff, 1978; McDermott *et al*., 2010). Roughness in particular, has been previously implicated as a timbre attribute contributing to tension (Bigand *et al*., 1996; Pressnitzer *et al*., 2000). On the other hand, the ambiguous responses to spectral centroid were somewhat surprising given the close association between centroid and brightness. Prior work has indicated that brighter instrument sounds are perceived as more tense (Nikolaidis *et al*., 2012). On the other hand, work on emotion and music has shown no impact of spectral centroid on arousal, suggesting is it not a factor in tension perception as well (Schubert, 2004; Bailes and Dean, 2012). In summary, while there is some evidence—including from the current study—for spectral centroid contributing to perceived tension, it is by no means definitive.

In more ecologically valid contexts, other factors such as loudness, pitch, rhythm, melody, and harmony would have a significant impact on tension perception. Although it is clear from the present results that timbre in isolation influences tension perception, it is not clear how strong this influence is in the presence of other dynamic auditory and musical features. Either in isolation or in parallel with these other features, changes in relevant timbre attributes should align with tension judgments; when in conflict with one or more features, the influence of timbre may be limited. Nonetheless, there are certain genres of music, in particular, electronic or electroacoustic music, where the contribution of timbre is especially important in shaping the tension profiles perceived by listeners.

In conclusion, the results of the present study provide a more detailed account of how timbre contributes to tension perception. From a higher-level perspective, gaining a better understanding of the relative contribution of timbre in comparison with other auditory and musical attributes would shed further light on timbre's general contribution to tension. In order to design experiments that would be effective in exploring the influence of timbre in this broader musical context, determining the individual timbre attributes that clearly affect tension is a crucial step. The results of the current work provide invaluable information toward the design and implementation of this future work.

## ACKNOWLEDGMENTS

The authors would like to thank William Sethares for very helpful suggestions regarding stimuli generation. This study has also benefitted from discussions with Juan Bello, Emilia Gómez, and Stephen McAdams, who offered advice on determining relevant timbre attributes.

Agostini, G., Longari, M., and Pollastri, E. (**2003**). "Musical instrument timbres classification with spectral features," EURASIP J. Appl. Signal Process. **2003**, 5–14.

Bailes, F., and Dean, R. T. (**2012**). "Comparative time series analysis of perceptual responses to electroacoustic music," Music Percept. **29**, 359–375.

Barbancho, I., Tardon, L. J., Sammartino, S., and Barbancho, A. M. (**2012**). "Inharmonicity-based method for the automatic generation of guitar tablature," IEEE Trans. Audio Speech Lang. Process. **20**, 1857–1868.

Bigand, E., and Parncutt, R. (**1999**). "Perceiving musical tension in long chord sequences," Psychol. Res. **62**, 237–254.

Bigand, E., Parncutt, R., and Lerdahl, F. (**1996**). "Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training," Percept. Psychophys. **58**, 125–141.

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**, 471–482.

Eerola, T., and Vuoskoski, J. K. (**2010**). "A comparison of the discrete and dimensional models of emotion in music," Psychol. Music **39**, 18–49.

Eronen, A., and Klapuri, A. (**2000**). "Musical instrument recognition using cepstral coefficients and temporal features," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Plymouth, MA, pp. 753–756.

Farbood, M. M. (**2012**). "A parametric, temporal model of musical tension," Music Percept. **29**, 387–428.

Farbood, M. M., and Upham, F. (**2013**). "Interpreting expressive performance through listener judgments of musical tension," Front. Psychol. **4**, Article 998.

Fredrickson, W. E. (**1999**). "Effect of musical performance on perception of tension in Gustav Holst's First Suite in E-flat," J. Res. Music Educ. **47**, 44–52.

Genesis (**2009**). Loudness Toolbox. Retrieved from http://www.genesis-acoustics.com/en/index.php?page=32 (Last viewed 6/1/16).

Granot, R. Y., and Eitan, Z. (**2011**). "Tension and dynamic auditory parameters," Music Percept. **28**, 219–246.

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

Helmholtz, H. L. F. (**1885**). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 2nd ed. (A. J. Ellis, Trans.) (Longmans, Green, and Co., London).

Hutchinson, W., and Knopoff, L. (**1978**). "The acoustic component of western consonance," Interface **7**, 1–29.

Ilie, G., and Thompson, W. F. (**2006**). "A comparison of acoustic cues in music and speech for three dimensions of affect," Music Percept. **23**, 319–329.

ISO/IEC (**2002**). "MPEG-7: Information Technology—Multimedia Content Description Interface–Part 4: Audio" (No. ISO/IEC FDIS 15938-4:2002) (International Organization for Standardization, Geneva, Switzerland).

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

Jehan, T. (**2013**). Echo Nest Remix API. Retrieved from http://developer.echonest.com/client_libraries.html (Last viewed 6/1/16).

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A., and Turnbull, D. (**2010**). "Music emotion recognition: A state of the art review," in *Proc. Int. Soc. Music Inf. Retr. Conf.*, Utrecht, pp. 255–266.

Krumhansl, C. (**1996**). "A perceptual analysis of Mozart's Piano Sonata K 282: Segmentation, tension, and musical ideas," Music Percept. **13**, 401–432.

Krumhansl, C. L. (**1989**). "Why is timbre so hard to understand?," in *Structure and Perception of Electroacoustic Sound and Music*, edited by S. Nielzen and O. Olsson (Excerpta Medica 846, Elsevier, Amsterdam), pp. 43–53.

Krumhansl, C. L. (**1997**). "An exploratory study of musical emotions and psychophysiology," Can. J. Exp. Psychol. **51**, 336–352.

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

Lartillot, O., Toiviainen, P., and Eerola, T. (**2008**). "A Matlab toolbox for music information retrieval," in *Studies in Classification, Data Analysis, and Knowledge Organization: Data Analysis, Machine Learning and Applications* (Springer, Berlin), pp. 261–268.

Lehne, M., Rohrmeier, M., and Koelsch, S. (**2013**). "Tension-related activity in the orbitofrontal cortex and amygdala: An fMRI study with music," Soc. Cogn. Affect. Neurosci. **9**(10), 1515–1523.

Lerdahl, F., and Krumhansl, C. L. (**2007**). "Modeling Tonal Tension," Music Percept **24**. 329–366.

Lychner, J. A. (**1998**). "An empirical study concerning terminology relating to aesthetic response to music," J. Res. Music Educ. **46**, 303–319.

McAdams, S. (**2013**). "Musical timbre perception," in *Psychology of Music*, 3rd ed., edited by D. Deutsch (Academic Press, New York), pp. 35–67.
McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (**1995**). "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psychol. Res. **58**, 177–192.
McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (**2010**). "Individual differences reveal the basis of consonance," Curr. Biol **20**, 1035–1041.
Nielsen, F. (**1987**). "Musical 'tension' and related concepts," in *The Semiotic Web'86: An International Yearbook*, edited by T. A. Sebeok and J. Umiker-Sebeok (Mouton de Gruyer, Berlin), pp. 491–514.
Nielsen, F. V. (**1983**). *Oplevelse af Misikalsk Spending* (*The Experience of Musical Tension*) (Akademisk Forlag, Copenhagen).
Nikolaidis, R., Walker, B., and Weinberg, G. (**2012**). "Generative musical tension modeling and its application to dynamic sonification," Comput. Music J. **36**, 55–64.
Paraskeva, S., and McAdams, S. (**1997**). "Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of musical tension and relaxation schemas," in *Proc. Int. Comput. Music Conf.*, Ann Arbor, MI, pp. 438–441.
Peeters, G. (**2004**). A large set of audio features for sound description (similarity and classification) in the CUIDADO project, Institut de Recherche et Coordination Acoustique/Musique (IRCAM), Technical Report.
Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., and McAdams, S. (**2011**). "The timbre toolbox: Extracting audio descriptors from musical signals," J. Acoust. Soc. Am. **130**, 2902–2915.
Peeters, G., McAdams, S., and Herrera, P. (**2000**). "Instrument sound description in the context of MPEG-7," in *Proc. Int. Conf. Comput. Music*, Berlin, pp. 166–169.
Plomp, R., and Levelt, W. J. M. (**1965**). "Tonal consonance and critical bandwidth," J. Acoust. Soc. Am. **38**, 548–560.
Pressnitzer, D., McAdams, S., Winsberg, S., and Fineberg, J. (**2000**). "Perception of music tension for nontonal orchestral timbres and its relation to psychoacoustic roughness," Percept. Psychophys. **62**, 66–80.
Rozin, A., Rozin, P., and Goldberg, E. (**2004**). "The feeling of music past: How listeners remember musical affect," Music Percept. **22**, 15–39.
Schubert, E. (**2004**). "Modeling perceived emotion with continuous musical features," Music Percept. **21**, 561–585.
Schubert, E., and Wolfe, J. (**2006**). "Does timbral brightness scale with frequency and spectral centroid?," Acta Acust. Acust. **92**, 820–825.
Sethares, W. A. (**1998**). *Tuning, Timbre, Spectrum, Scale* (Springer-Verlag, London), 345 pp.
Yang, Y.-H., and Chen, H. H. (**2012**). "Machine recognition of music emotion," ACM Trans. Intell. Syst. Technol **3**, 1–30.
Zhang, X., and Ras, Z. W. (**2007**). "Analysis of sound features for music timbre recognition," in *Proc. Int. Conf. on Multimed. Ubiquitous Eng.*, Seoul, pp. 3–8.