

A TIMBRE ATTRIBUTE CALCULATOR AND MUSIC-ANALYTICAL APPLICATIONS

Noah Kahrs

Eastman School of Music
Composition Department
Rochester, NY 14604
nkahrs@u.rochester.edu

ABSTRACT

Although musical tension is well-understood for pitch, it is less clear how it functions for other musical parameters such as timbre. Many timbre attributes can be calculated from a spectrogram, and have documented psychological relations to musical tension in isolation, but not necessarily in musical contexts.

In this paper, I describe a Matlab calculator/toolbox for calculating five timbre attributes (spectral centroid, spectral spread, spectral flatness, roughness, inharmonicity), and suggest a few musical implications. In particular, I compare these quantitative attributes to a more qualitative analysis of a work by Sofia Gubaidulina, to see how such attributes align with music-analytical notions of “consonance” and “dissonance.”

In this particular context, “dissonance” tends to correlate with higher spectral centroid and spectral flatness, and somewhat with spectral spread, but not as much with inharmonicity or roughness. This implies that “consonance” and “dissonance” correspond loosely to pitch and white noise, respectively, and suggests a quantitative approach to theorizing this distinction.

Index Terms— Timbre, Tension, Dissonance

1. INTRODUCTION

Musical tension is a well-understood problem for pitch (melody and harmony), but not so much for other musical parameters such as timbre. However, over the course of the twentieth-century, non-pitch attributes of music became increasingly important to contemporary classical music. In particular, timbre has become much more popular as a compositional parameter, but is not always easily to describe in the usual language of music theory, which is oriented more towards pitch and rhythm.

A number of quantitative timbre attributes are popular with both MIR researchers and Music Perception researchers for both automatic instrument identification and qualitative timbre perception. In this paper, I review existing literature on how such attributes relate to musical tension, introduce a Matlab program that calculates and plots such attributes, and

compare numerical values to a tension-based analysis of a work by Sofia Gubaidulina.

2. TIMBRE ATTRIBUTES AND MUSICAL TENSION

Quantitative timbre attributes, as calculated per frame (in an FFT analysis), generally fall into three categories: (1) statistical properties of the spectrum, (2) attributes calculated from the interpolated peak frequencies, and (3) mel-frequency cepstral coefficients (MFCCs). Although MFCCs are often useful for MIR problems involving instrument identification, they are of limited use for qualitative description and therefore not included in perception-oriented discussions such as that of the McGill Timbre Toolbox [1]; I will not discuss them further in this paper.

Statistical attributes of the spectrum at a given frame are useful for describing basic properties such as “brightness” [2]; although such descriptors do not directly correlate to musical tension as it is usually understood, a number of studies have found that higher-pitched overall spectral envelopes do tend to be considered more tense [3].

My calculator incorporates three such measures calculated directly from the spectrum: spectral centroid, spectral standard deviation, and spectral flatness. These attributes are very standard; formulae are available in [1, 2, 4, 5], and are reproduced below. In all of the following, for a given FFT analysis from which such a measure is calculated, the frequency and amplitude of a given bin k are f_k and a_k , respectively, and K is the number of total bins. Spectral centroid, in Hz, is the weighted mean frequency, as given by:

$$SC = \mu_1 = \frac{\sum_{k=1}^K f_k a_k}{\sum_{k=1}^K a_k}$$

Spectral spread (also known as spectral standard deviation), also in Hz is the corresponding standard deviation per given frame:

$$SS(D) = \mu_2 = \frac{\sqrt{\sum_{k=1}^K (f_k - \mu_1)^2 a_k}}{\sum_{k=1}^K a_k}$$

Spectral flatness (measure) is a unitless quantity calculated as the ratio of geometric mean to arithmetic mean of all bin amplitudes:

$$SFM = \frac{\sqrt[K]{\prod_{k=1}^K a_k}}{\frac{1}{K} \sum_{k=1}^K a_k}$$

Generally, spectral centroid is taken to represent the “brightness” of a signal, ie the overall presence of high frequencies in the spectrum. Spectral spread represents the width of the distribution about the centroid, and would be low for either a pure tone or for a narrow band of noise. Spectral flatness measures similarity to white noise.

Peak-based timbral attributes are somewhat more difficult to calculate directly from the spectrum, but correspond more closely to meaningful descriptors for tension, and align more closely with usual music-theoretical notions of consonance and dissonance. Such attributes, particularly variants of “roughness” and “inharmonic,” are the most common scientific explanations of musical tension and dissonance, and have been ever since Helmholtz’s introduction of roughness and dissonance curves in the nineteenth century [6, 7]. However, because such measures depend on peak extraction, they are not as consistent.

Inharmonicity is given by the following formula, in which f_0 is an extracted fundamental frequency, f_n the corresponding n^{th} harmonic as automatically identified, a_n the corresponding amplitude, and N the number of harmonics; the formula is as in [2, 5, 1]:

$$I = \frac{2}{f_0} \frac{\sum_{n=1}^N |f_n - n f_0| a_n^2}{\sum_{n=1}^N a_n^2}$$

In my calculator, I estimate fundamental frequency f_0 as the median distance between consecutive peaks (this is not the most reliable measure, but given that many sounds are rather inharmonic in the first place, it’s sufficient for my purposes), and calculate f_n as the nearest peak to $n f_0$ for given n .

Roughness the attribute most often tied to tension and dissonance, as discussed in [8, 9, 10, 11]. Roughness has been posited as an explanation of consonance and dissonance in a harmonic context, as with its origin in [7]; timbre analysis via roughness is sort of like considering the peaks of a given timbre to be a chord. Numerous distinct formulae for roughness exist, but are all based on the same general principle: the roughness of a sonority is a weighted sum of the roughnesses between individual frequency components. In my calculator I follow the formula from [10], which has its roots in [12, 13]. In this formula, overall roughness is defined, where $a_{i,j}$ are the amplitudes of given peaks, as

$$\rho = \sum_{j=0}^n \sum_{k=1}^{n-1} \frac{a_j a_k g(f_{cb})}{a_j^2}$$

where f_{cb} is the distance between peaks $f_{i,j}$ in critical band-

widths, namely

$$f_{cb} = \frac{f_i - f_j}{1.72 \left(\frac{f_i + f_j}{2} \right)^{0.65}}$$

and $g(f_{cb})$ is a “standard curve” for dyad roughness in terms of critical bandwidths:

$$g(f_{cb}) = (4e f_{cb} e^{-4f_{cb}})^2$$

A number of music theorists and music perception researchers have explored how these attributes relate to listener perception of “consonance,” “dissonance,” and tension. As discussed above, roughness is most often taken as an indicator of musical tension; however, other attributes are also often correlated to roughness. [2] shows that inharmonicity and spectral flatness also contribute strongly to musical tension, as do spectral centroid and spread to lesser extents. [14] argues from a speculative-theoretical rather than experimental perspective that in addition to roughness and inharmonicity, “brightness” and wide vibrato, corresponding respectively to spectral centroid and spectral spread, should increase musical tension. Thus all five of the timbre attributes described above might bear at least some relation to musical tension. My calculator implements displays for all five of these attributes, as well as for the sheer number of peaks so that one can adjust the peak detection threshold appropriately.

3. MUSIC-ANALYTICAL HYPOTHESIS

To examine how these quantitative timbral attributes intersect with other conceptions of consonance and dissonance, I deployed the calculator on three different recordings of the composition *Meditation on the Bach Chorale “Vor deinen thron tret ich hiermit”*, by Sofia Gubaidulina. I chose this composition because I have for some time been working on a paper discussing how consonance and dissonance govern timbre and form in addition to pitch, not only in this piece but also in Gubaidulina’s music more generally. I will be presenting my research on Gubaidulina this summer at two conferences [15, 16], so this paper was written in conjunction with those efforts.

My analysis of Gubaidulina’s *Meditation*, based on prior music-theoretical work on Gubaidulina’s music, Gubaidulina’s published interviews, and her sketch material for the piece, identifies specific timbres as consonant and dissonant. The piece is scored for five string instruments and harpsichord. When the instruments are played with standard techniques and obtain a full tone, the timbre is “consonant.” When the instruments play with nonstandard techniques such as sul ponticello (placing the bow at the bridge of a string instrument), col legno (hitting the strings with the wood of the bow), or circular bowing (moving the bow in a circle rather than only back and forth), the timbre is “dissonant.”

My hypothesis is that “consonant” sounds, being less tense and more harmonic, would have lower spectral centroid, spectral flatness, inharmonicity, and roughness, although they might have more harmonics and thus higher spectral spread. Correspondingly, as discussed in the review above, “dissonant” sounds should have higher values in all attributes except perhaps spectral spread.

4. METHODOLOGY

To avoid being biased by any one specific recording, I used three different recordings of Gubaidulina’s *Meditation*: one by Thomas Klug et al available on Spotify¹, one by International Contemporary Ensemble available on Vimeo², and one by Viktor Suslin et al available on CD at Sibley Music Library³. The two online recordings were converted to WAV format by playing them in the browser while recording the sound card’s output in Audacity at 44.1kHz with 32-bit quantization; the CD recording was converted to WAV in iTunes at 44.1kHz with 24-bit quantization. Because Spotify and Vimeo are compressed, values are not remotely comparable between recordings; the full CD-quality recording may yield best output relative to the other two.

In Matlab, I implemented a calculator for the five timbre attributes outlined in Section 2. FFT was calculated with the built-in Matlab `fft` function, with Hamming windows and usually with FFT size 1024 samples, hop size 512 samples, and 4x zero-padding. Peaks were identified from local maximum bins whose values were at most 35dB below the global maximum bin, and were determined more precisely via quadratic interpolation. The display allowed for simultaneous displays of the audio file waveform, a spectrogram, and a plot of the user’s choice of timbre attributes.

The calculator was deployed on three audio files, each drawn from a different recording, and each of which contained the same excerpts in the same order. Each file consisted of a series of consonant excerpts, a brief pause, and a series of dissonant excerpts, following my own music-theoretical discussion of consonance and dissonance. Specifically, the first half of each file consisted of the following five consonant sounds: (1) a solo double bass playing with normal technique, (2) the violins playing in counterpoint with normal technique, (3) the harpsichord playing solo chords, (4) the cello and bass playing normally in counterpoint, and (5) all string instruments playing a unison melody, with harpsichord playing higher trills (the harpsichord here is “dissonant,” which throws off the calculations as the sources are grouped together”. After a pause, the second half of each file had the following four dissonant sounds: (1) upper strings playing

with sul pont, (2) upper string playing with tremolo glissando, (3) mid-range strings playing with circular bowing, harpsichord with clusters, and (4) all strings sliding back and forth, changing bow position frequently.

Figure 1 illustrates the calculator interface, showing the full window for the Chojnacka samples. The five consonant and four dissonant samples can easily be identified in all three displays.

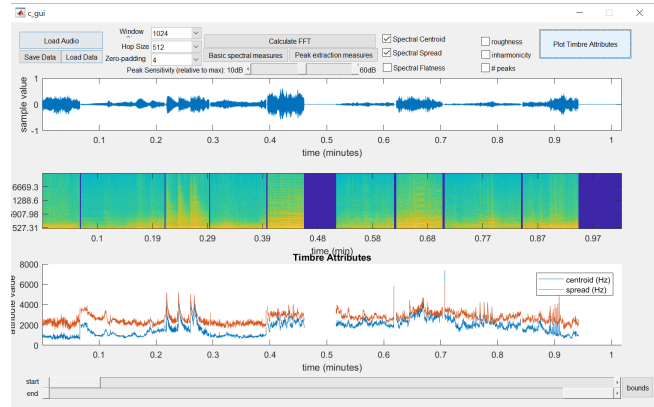


Figure 1: full GUI interface, Chojnacka as input

5. RESULTS

Although all measures had a high degree of variability, even between sounds of a similar ground-truth classification, a few trends could be identified in general. In particular, spectral flatness was consistently higher for dissonant sounds, as were spectral centroid and spectral spread to lesser degrees. Surprisingly, inharmonicity and roughness do not appear to correlate as strongly with either consonance or dissonance—those two measures are particularly volatile, and tend to vary widely even within a specific sound.

Figure 2 highlights the most successful positive results. Each plot indicates which recording it is from in text at the top, and which measures are included in the legend at top right. A horizontal red line indicates a threshold that reasonably segments “consonant” from “dissonant” values for a given timbre attribute. Specifically, spectral centroid is useful for the recordings by Chojnacka and Klug, spectral spread for Klug, and spectral flatness for ICE and Klug. These constant thresholds were established purely by eyeball; in future work, I would like to develop more meaningful statistical measures, which I could also deploy in combination with analysis of more of the recordings.

Some sounds are consistently outside of the thresholds identified in Figure 2. The “consonant” harpsichord triads, identifiable in the plots three large peaks, for instance, exceeds the threshold for nearly all spectral measures, likely because the harpsichord’s attack is particularly “sharp.” Similarly, the last “consonant” excerpt consists of unison strings

¹<https://open.spotify.com/track/5mQsp9wZxGjVhd9bTX902L?si=2EMGgB9rQUaagQVYAn8nBw>
²<https://vimeo.com/106380344>
³<https://catalog.lib.rochester.edu/vwebv/holdingsInfo?bibId=1526402>

but also high harpsichord clusters. The harpsichord likely pushes the spectral measures above the threshold; because the harpsichord is there an element of dissonance in an otherwise consonant sound, this deviation from the threshold is musically meaningful.



Figure 2: positive results, plotted with labels

Figure 3 illustrates a typical negative result, in this case the inharmonicity for the Chojnacka recording. Although the “dissonant” values seem to have an overall slightly higher average inharmonicity, in particular a higher minimum value, there is so much overlap in typical values that it is difficult if not impossible to draw a meaningful threshold.

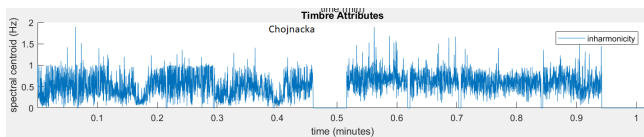


Figure 3: a negative result

6. CONCLUSION

As expected from the literature on timbre perception, some quantitative timbre attributes are meaningful in a music analysis context. Specifically, “dissonant” timbres in Sofia Gubaidulina’s music tend to have higher spectral centroid and/or spectral flatness than “consonant timbres;” surprisingly, roughness and inharmonicity are not as descriptively useful for this particular repertoire. Spectral flatness is particularly analytically suggestive: Gubaidulina’s approach to timbre has often been described as dealing with a “pitch-noise axis” [17], and spectral flatness measures, for a sound, “how similar its spectrum is to white noise” [2], so we might conclude that “dissonance” corresponds to similarity to white

noise. Given that white noise would have very high inharmonicity and roughness compared to a harmonic spectrum, the lack of meaningful correspondence there is somewhat surprising; however, a white noise spectrum also has few if any peaks, so such metrics would not be meaningful and would likely fluctuate, explaining the rapid fluctuations as in Figure 3.

Spectral flatness, unfortunately, does not align as well as roughness with traditional notions of consonance and dissonance. Although the nearest pitch-domain approximation of white noise, a multi-octave cluster chord, would be unambiguously dissonant, spectral flatness would not vary much for different chords in a similar register of a similar instrument. In future work, I would like to explore how some metric might better encompass both the pitch-noise axis and more traditional roughness-oriented notions of tension, as well as how my above analysis might be made more statistically rigorous.

7. REFERENCES

- [1] Geoffroy Peeters, Bruno L. Giordano, Patrick Susini, Nicolas Misdariis, and Stephen McAdams, “The Timbre Toolbox: Extracting audio descriptors from musical signals,” *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2902–2916, Nov. 2011.
- [2] Morwaread M. Farbood and Khen C. Price, “The contribution of timbre attributes to musical tension,” *The Journal of the Acoustical Society of America*, vol. 141, no. 1, pp. 419–427, Jan. 2017.
- [3] Roger A. Kendall and Edward C. Carterette, “Verbal Attributes of Simultaneous Wind Instrument Timbres: I. von Bismarck’s Adjectives,” *Music Perception: An Interdisciplinary Journal*, vol. 10, no. 4, pp. 445–467, July 1993.
- [4] Derry FitzGerald and Jouni Paulus, “Unpitched percussion transcription,” in *Signal Processing Methods for Music Transcription*, pp. 131–162. Springer, 2006.
- [5] Perfecto Herrera, Anssi Klapuri, and Manuel Davy, “Automatic Classification of Pitched Musical Instrument Sounds,” pp. 163–200. Jan. 2006.
- [6] James Tenney, “A history of ‘consonance’ and ‘dissonance,’” 1988.
- [7] Hermann Helmholtz, “On the sensations of tone, 1877,” *Trans. AJ Ellis, Dover, New York*, pp. 390–394, 1954.
- [8] Daniel Pressnitzer, Stephen McAdams, Suzanne Winsberg, and Joshua Fineberg, “Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness,” *Perception & Psychophysics*, vol. 62, no. 1, pp. 66–80, Jan. 2000.
- [9] Philippe Lalitte, “The Theories of Helmholtz in the Work of Varèse,” *Contemporary Music Review*, vol. 30, no. 5, pp. 327–342, Oct. 2011.
- [10] John MacCallum and Aaron Einbond, “Real-Time Analysis of Sensory Dissonance,” in *Computer Music Modeling and Retrieval. Sense of Sounds*. Aug. 2007, Lecture Notes in Computer Science, pp. 203–211, Springer, Berlin, Heidelberg.
- [11] William A Sethares, “Local consonance and the relationship between timbre and scale,” *The Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1218–1228, 1993.
- [12] John MacCallum, Jeremy Hunt, and Aaron Einbond, “Timbre as a psychoacoustic parameter for harmonic analysis and composition,” in *ICMC*, 2005.
- [13] Richard Parncutt, *Harmony: A psychoacoustical approach*, vol. 19, Springer Science & Business Media, 2012.
- [14] Fred Lerdahl, “Timbral hierarchies,” *Contemporary Music Review*, vol. 2, no. 1, pp. 135–160, Jan. 1987.
- [15] Noah Kahrs, “Consonance, Dissonance, and Large-Scale Form in Two Works of Sofia Gubaidulina,” *Music Theory Midwest*, 2018.
- [16] Noah Kahrs, “Consonant and dissonant timbres in two works of Sofia Gubaidulina,” *Timbre is a Many-Splendored Thing*, 2018.
- [17] Rebecca Leydon, “Clean as a whistle: Timbral trajectories and the modern musical sublime,” *Music Theory Online*, vol. 18, no. 2, 2012.