Paul Rudmik

Richard Mann

CS 489: Computational Audio

April 8, 2019

Computing Clarity in Audio Samples of Guitars

0. ABSTRACT

This paper describes and evaluates three algorithmic approaches of modelling human perceptions of clarity of guitar music: Inter-Band Relationship score (IBR), Harmonic Bandwidth Measurement (HBM), and Harmonic-to-Signal Ratio (HSR). After comparing the algorithmic output to human perceptions of clarity within a small dataset of guitar samples, HBM was shown to most accurately model perceptions of clarity.

1. INTRODUCTION

Although guitarists widely recognize the importance of good tone in the production of music, there currently exist few software options for gaining this form of feedback. Popular educational offerings to aspiring guitarists, such as Yousician and Guitar Tricks, often focus more on detecting the timing of a user's performance [1], or on providing video instructions [2], than they do providing feedback on tone and style of playing. Such tools could aid guitarists gain feedback on their technique while playing, which is often only accessible around other peers skilled enough to discern good tone from bad. The software application market, then, lacks a solution for analyzing and providing feedback on user's quality of musical performance, including with regard to clarity.

This paper suggests and evaluates three algorithms for samples of guitar tones with regard to clarity, the quality of a guitar tone that sharply, distinctively, or expressively expresses a particular note on the guitar. The algorithms tested include one measure previously proposed by Steven Fenton and Jonathan Wakefield, the Inter-Band Relationship score (IBR), in addition to two newly proposed algorithms, Harmonic Bandwidth Measurement (HBM) and Harmonic-to-Signal Ratio (HSR). These algorithms are evaluated with the aim of producing an approach to computing clarity that accurately models the human perception of clarity.

2. METHODOLOGY

Each algorithm was tested against a small dataset of guitar tones to determine how accurately it may model the human response to each sample. In total, twenty-three trials were run to determine the most accurate algorithm under a variety of settings. The dataset was also distributed to four human subjects to measure the human perception of clarity in the samples. This feedback was used as the benchmark to evaluate how successfully the algorithms measured clarity.

2.1 Dataset

Of the three data samples, two were recorded by the study author with the aid of the course instructor. The third was downloaded from an online sound repository. All three samples were perceived to have a distinct amount of clarity from one another. So as to not influence human subjects' perceptions of the samples, each file was assigned a number at random, which is what is used to refer to them in this paper.

2.2 Algorithmic Approaches

Three algorithms were designed and implemented in Python for the purposes of the study. Each has a variable that can be altered. All code used in the project can be publicly accessed at https://github.com/PaulMHR/analyzing-clarity.

2.2.1 Inter-Band Relationship Score

Steven Fenton and Jonathan Wakefield developed the Inter-Band Relationship score as an attempt to measure clarity in completed musical productions that overlaid several instruments together. Where earlier measures of clarity within the same field attempted to analyze where the onset and offset of individual notes within the song were, the author's approach was relatively simple: in essence, divide the song into three bands, and measure the relationship between the Dynamic Ranges between the band. The study ultimately provided some evidence to suggest that should there be less "dynamic content between the frequency bands the listeners will grade the music as lacking clarity" [3].

Algorithmically, IBR first filters the sample into the three frequency bass, mid, and treble bands using an FIR:

Filter	Low Freq (Hz)	High Freq (Hz)
Band 1: Bass	-	947
Band 2: Mid	947	3186
Band 3: Treble	3186	-

For each of the three bands with n samples, the Dynamic Range of the band is found as:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S_{RMS} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$S_{peak} = \max(X_{1 \ldots n})$$

$$DR = 20\log(\frac{S_{peak}}{S_{RMS}})$$

For each of the Dynamic Range calculations across m bands, the IBR is calculated as:

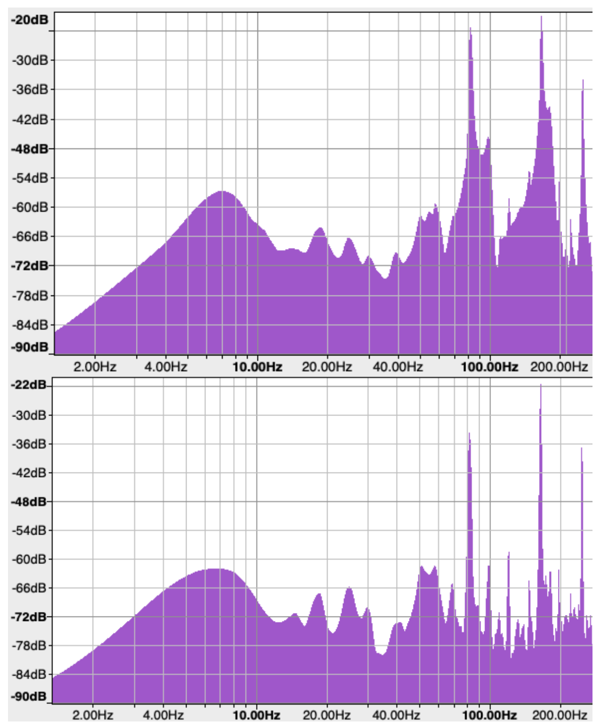
$$\overline{DR} = \frac{1}{m}\sum_{i=1}^{m}(DR_i)$$

$$IBR = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(DR_i - \overline{DR})}$$

This IBR measurement is taken as the measure of whether or not the sample is "clear" or not. Generally speaking, an IBR above a threshold of 4 is considered clear.

2.2.2 Harmonic Bandwidth Measurement

In general, samples that are regarded as clearer tend to have thinner peaks in their Fourier Transformations. For instance, the FFT graphs of sample2.wav, which was considered to be less clear than sample1.wav from human response data, has thicker peaks in its FFT:



The Harmonic Bandwidth Measurement, then, measures the width of these harmonic peaks directly. The pseudocode for this approach is itself fairly simple:

1. let diff_sum = 0
2. for the first n harmonic peaks in the Fourier Transformation:
   A. let peak_height be the maximum of the harmonic

B. let left_side, right_side be the frequencies on the harmonic peak that are d dB below peak_height
C. diff_sum = diff_sum + (right_side - left_side)
3. output diff_sum / n

In the Python implementation, HBM was set to iterate over the first 12 harmonics. Smaller outputs of HBM generally indicate a clearer sample.

2.2.3 Harmonic-to-Signal Ratio

The Harmonic-to-Signal Ratio, similarly to the HBM, provides a measurement of how large the harmonic peaks are within the signal by taking the power between two frequencies surrounding the harmonic. These frequencies are determined to be of b Hz apart, with the frequencies centred around the tallest peak of the harmonic. These powers are calculated using the formula:

$$Power(f_1, f_2) = \frac{1}{N}\sum_{i=k_1}^{k_2} 2 * |X_i|^2$$

The pseudocode for this algorithm follows:
1. let power_sum = 0
2. for the first n harmonic peaks in the Fourier Transformation:
   A. let peak_freq be the frequency of the maximum of the harmonic
   B. power_sum = power_sum + Power(peak_freq - b/2, peak_freq + b/2)
3. output power_sum / n

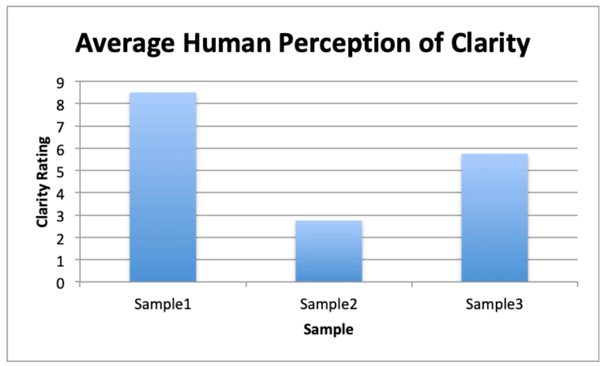
Within the Python implementation, HSR was set to iterate over the first 12 harmonics.

2.3 Human Evaluation

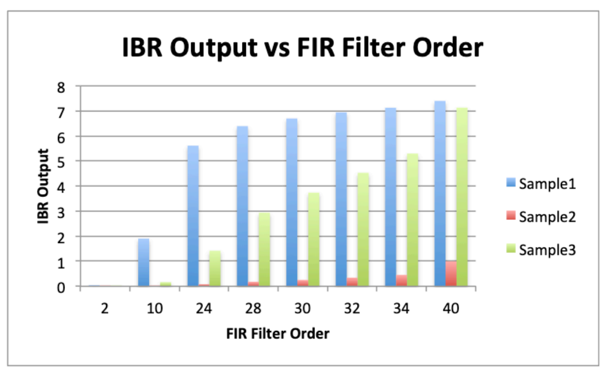
Four human subjects were sent copies of the three samples and were asked to rate each sample on a scale of 1 to 10, and rank the samples in order of least to most clear. This data was used to be an objective measure of how clear these samples were, and were the primary method by which the algorithms were evaluated.

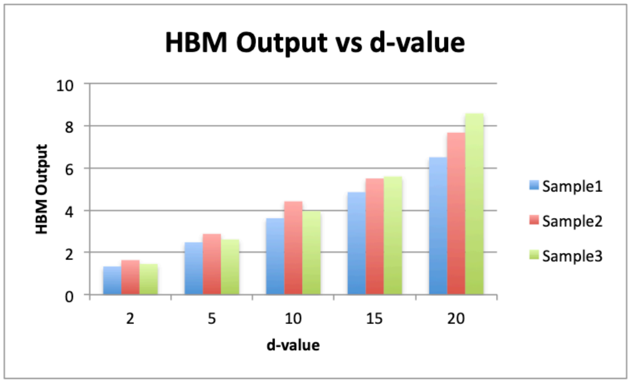
3 EVALUATION

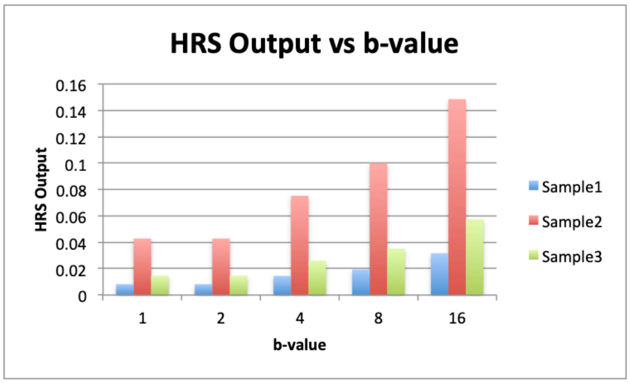
Human responses to the samples were as follows, with Sample1 rated as most clear, and Sample 2 rated as least clear:



Responses from algorithmic trials follow. Each is listed with the set of values changeable to the algorithm:

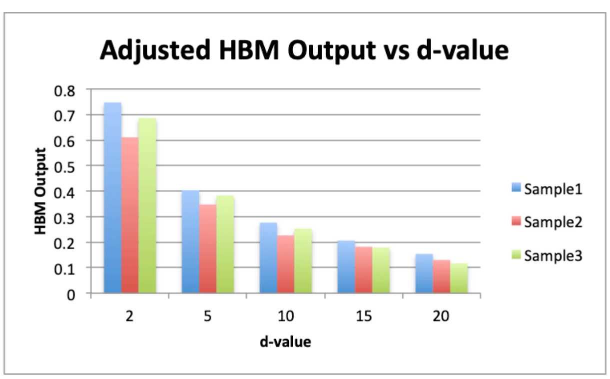


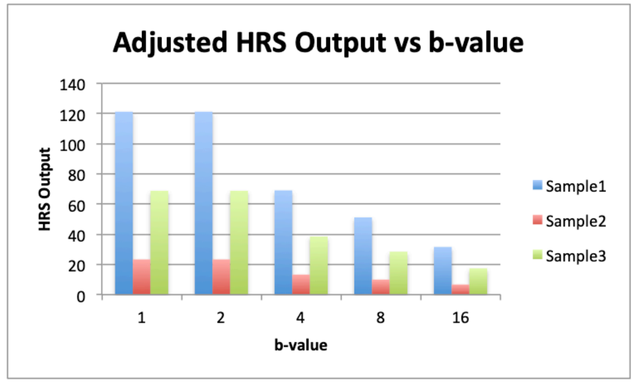




3.1 Adjusting HBM and HSR

One shortfall of the design of HBM and HSR is that they output values in opposite order to the human response (i.e. greater clarity is represented with a larger number), so some transformation has to be done on the values to produce the order needed to fit the human response pattern. Each value from the algorithms were inverted (i.e. applied to a power of -1). The adjusted values are as such:





4.2 Deciding Between Algorithms

Between these various algorithms and settings, the HBM-1 is the most accurate model of the human response data, as it produces the ratio of clarity values most similar to the human response data (1:1.12:1.22).

All algorithms tended to correctly reproduce the human ordering of clarity among the samples for low values. Notably, IBR decidedly resulted in the worst output between the three algorithms, with its most accurate version of IBR-40 producing the ratio 1:7.16:7.43. This result entails that IBR generally “squishes” together values of higher clarity, which may make distinguishing between higher-clarity samples difficult in future applications.

5 CONCLUSION

Each of the three proposed algorithmic approaches to evaluating clarity have been shown to accurately model, to some degree, the human perception of clarity in guitar music, with greater effectiveness shown in HBM. This is certainly positive indication that clarity can be computationally determined — it isn’t simply a subjective human impression of music, but a mathematical property of sound systems. Although the scope of this project is limited by further access to data and human subjects, it is indicative that this is certainly an area which can be easily further researched within the domain of computational audio, and even extended to include other aspects of guitar tone, such as articulation and brightness, among others. By doing so, a new range of software options for aspiring musicians might be openly explored.

Works Cited

[1] https://yousician.com/guitar

[2] https://www.guitartricks.com

[3] http://www.aes.org/e-lib/browse.cfm?elib=1626