Automatic Guitar Tablature Transcription from Audio using Inharmonicity Regression and Bayesian Classification

Jonathan J. Michelson[1], Richard M. Stern[2], and Thomas M. Sullivan[2]

[1]*New Sensor Corporation / Electro-Harmonix*
[2]*Department of Electrical and Computer Engineering, and School of Music, Carnegie Mellon University*

Correspondence should be addressed to Jonathan Michelson (`jonathan@ehx.com`)

ABSTRACT

We propose two new methods to classify guitar strings for automated tablature transcription using only monophonic audio. The first method estimates the linear regression of log-inharmonicities of guitar strings with respect to their pitches, and assigns unseen notes to the strings whose means and variances maximize the probability of their measured inharmonicities. The second method, developed as a baseline, characterizes the inharmonicity distribution of each fretboard position as a normal probability density, and then similarly assigns unseen notes to the fretboard positions that maximize the likelihood of their observed inharmonicities. Results from the standard Real World Corpus of guitar recordings show that exploiting regressions generally improves accuracy compared to our baseline, while both achieve adequate performance in guitar-independent test scenarios.

1 Introduction

The guitar is a popular musical instrument whose family comprises a diverse collection of stringed instruments. Despite their variation, the majority of guitars have six strings tuned to $E2$ (82 Hz), $A2$ (110 Hz), $D3$ (147 Hz), $G3$ (196 Hz), $B3$ (247 Hz), and $E4$ (330 Hz). Most guitars also typically have upwards of 20 frets, allowing for versatile musical passage realizations along the fretboard. A consequence of the tuning and extensive fretting locations is a substantial pitch overlap between neighboring strings; multiple fretboard locations can usually be selected to produce a given pitch. This ambiguity renders conventional music scores, which represent passages as notes and chords, undesirable for students trying to learn more skillful placement of scales and arpeggios, or for enthusiasts trying to uncover the fretboard positions used on riffs recorded by

virtuosic players. Figure 1 illustrates this ambiguity.

Tablature is a popular alternative music notation that does not suffer from the one-to-many mapping of scores. The staff, instead of representing pitch as in classical notation, depicts a birds-eye view of the guitar neck from the perspective of the performer. Each of the six horizontal lines represents a corresponding string on the guitar, and numbers on each line specify the fret to be fingered. Reasons for tablature desirability are numerous [1], such as accessibility (legibility does not require formal musical training) and transportability (tablature can be easily represented with ASCII characters).

Accurate tablature transcription is a laborious task. Dependable tablature containing few errors typically require the careful listening of a seasoned musician to an audio recording, and possibly video recordings
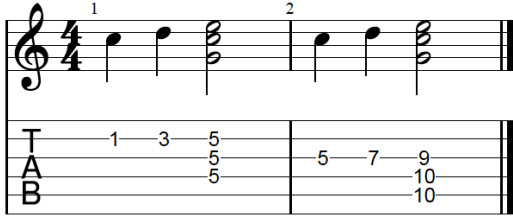
Fig. 1: Two identical bars of a music score (top) and two equivalent realizations in tablature notation (bottom).
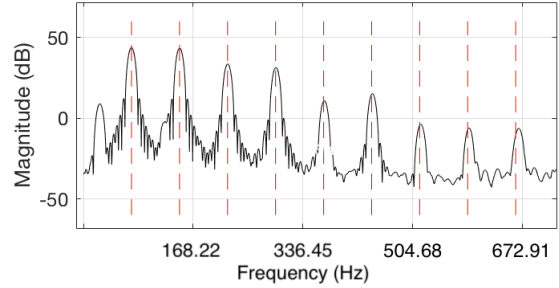

Fig. 2: Power spectrum of note D2 plucked on an electric guitar at string 5 and fret 5. Red dashed lines are drawn at integer multiples of the fundamental. The partial peaks begin visibly skewing rightward after the fifth harmonic.

for additional consultation. Reliable automation of this transcription task would benefit the guitar student community by expediting quality tablature generation and unearthing fretboard positions of guitar riffs in songs for which accurate tablature does not yet exist.

An important feature in systems that transcribe tablature from audio is inharmonicity. When an ideal string fixed at both ends is displaced, the restoring force that causes it to oscillate is its tension. For a real string with stiffness, however, its elasticity contributes to this restoring force [2], causing the frequencies of the modes of vibration to deviate from integer multiples of the fundamental. Instead, the solutions to the equation of motion of a string with stiffness and tension become:

$$f_k = k f_0 \sqrt{1 + \beta k^2} \qquad (1)$$

where f_k is the kth harmonic of fundamental f_0 and β is the inharmonicity of the string, defined by

$$\beta = \frac{\pi^3 Q d^4}{64 T l^2}. \qquad (2)$$

The inharmonicity β of a vibrating string is a dimensionless quantity that depends on the Young's modulus Q, diameter d, tension T, and vibrating length l of the string, and which skews the kth partial of the string upwards in frequency according to Eq. (1). See Fig. 2.

Moreover, inharmonicity in the guitar varies in a deterministic manner with respect to fret positions along a given string [3]. Specifically, the inharmonicity $\beta(s,n)$ of a note produced by string s at fret number n can be expressed in terms of the inharmonicity of the open-string note $\beta(s,0)$ according to:

$$\beta(s,n) = \beta(s,0) 2^{n/6} \qquad (3)$$

Figure 3 illustrates these theoretical trajectories. Barbancho *et al.* [3] have developed a successful transcription system which exploits only these theoretical trajectories. We wondered, though, whether consideration of the *empirical* inharmonicity trajectories could further increase accuracy; knowledge of the inharmonicity of every fret instead of only that of the open-string note could conceivably produce a more accurate picture of the inharmonicity trajectory, which could improve transcription performance.

We therefore propose classifying unknown notes by obtaining comprehensive log-inharmonicities of the notes along the guitar neck, estimating the string-wise linear regressions of these log-inharmonicities with respect to their pitches, and assigning the unknown notes to the strings whose regressions most probably predict the estimated log-inharmonicity of the notes. To contextualize its performance as well as that of the system developed by Barbancho *et al.* [3], we also introduce a baseline Bayesian classifier of fretboard position given an unseen note: the measured inharmonicities of each fretboard position are collected and fit to a normal distribution, and are used to return the string that maximizes the probability of having observed the measured inharmonicity of the unseen note.

2 Previous Work

Empirical estimation of inharmonicity has been investigated for nearly 40 years. The use of pitch extraction techniques to perform estimation was first attempted by Galembo in 1979 and 1986 [4, 5]. In 1994, Galembo *et*
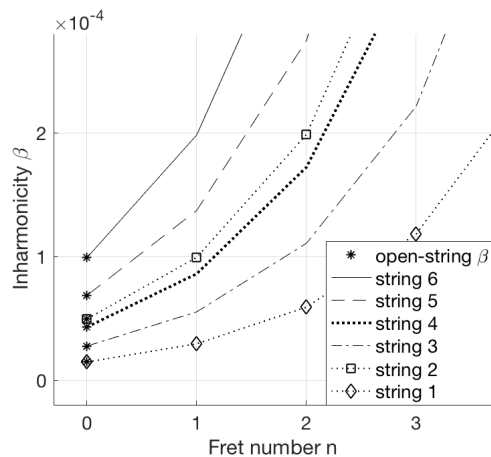
Fig. 3: Example inharmonicity trajectories of acoustic guitar strings, calculated from measured open-string inharmonicities. Only Fret 0 (open-string) through Fret 3 are shown. On this acoustic guitar, the thinner Strings 1 and 2 were unwound while the thicker Strings 3-6 were wound, which helps explain the anomalous inharmonicity-trajectory location of String 2.

Fig. 4: Inharmonicity estimates for Frets 0-12 on Strings 1-6 of RWC Acoustic Guitar AG111, plotted against MIDI fundamental instead of fret number.

al. hypothesized a connection between partials-based fundamental pitch estimates and the degree of inharmonicity in a spectrum [6]. The same authors subsequently introduced another method in 1999 based on inharmonic comb filters [7]. Rauhala introduced an efficient iterative procedure in 2007 [8], and two years later Hodgkinson improved on its efficiency and accuracy with the median-adjustive trajectories (MAT) method [9].

Inharmonicity-based tablature transcription from audio has also been the subject of ample study. In 2012, Barbancho *et al.* introduced a transcription scheme based solely on inharmonicity analysis [3]. We refer to their work as the partial coincidence tally (PCT) method. In the following two years, Abesser [10], Kehling [11], and Dittmar [12] developed machine learning-based transcription systems that used slews of features which included inharmonicity.

This paper builds on the exclusively inharmonicity-based previous work in [3] by considering the inharmonicity of every fret in its estimation of the trajectory instead of only that of the open-string note, and by introducing a baseline likelihood-maximizing method against which to benchmark transcription performance.
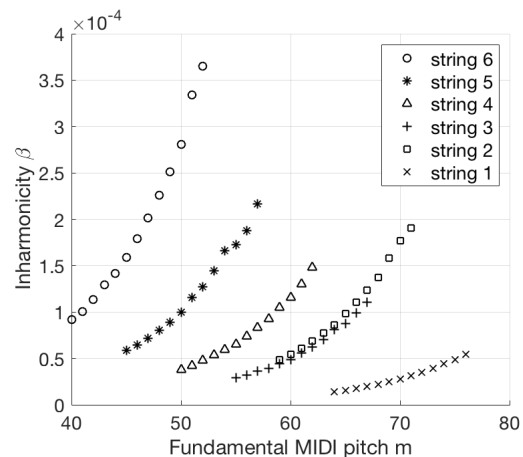
3 Methods

3.1 Inharmonicity Regression

Our approach begins with an estimation of the inharmonicity of an unseen note. The algorithm we employed was the median adjustive trajectories (MAT) method [9] proposed by Hodgkinson in 2009. The MAT algorithm was chosen for its efficiency and accuracy over the algorithms that preceded it [9]. We applied this estimation procedure to five Hann-windowed, 44.1-kHz audio segments of length 2^{14} samples, beginning at the detected onset of a note and at four successive 100-ms intervals (*i.e.* 0 ms, 100 ms, 200 ms, 300 ms, and 400 ms after the note onset). This yielded five inharmonicity estimates that we later compiled into one classification result using the frame-aggregation refinement technique described in [10].

Next, we applied a log transformation to the estimated inharmonicities to make their variation linear with respect to MIDI pitch. Reconsidering Eq. (3) in terms of MIDI pitch number m instead of fret number n, we obtain

$$\beta(s,m) = \beta(s,m_{os})2^{(m-m_{os})/6}, \qquad (4)$$

where m_{os} is the open-string pitch of string s. Figure 4 illustrates these trajectories in terms of MIDI pitches.
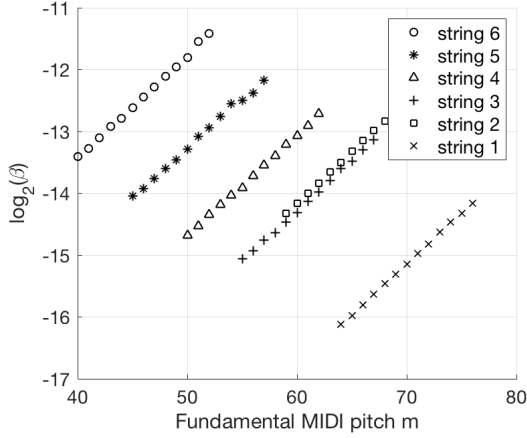
Fig. 5: Log-inharmonicity estimates for the same frets, strings, and guitar as Fig. 4.

Considering the log-inharmonicity $\beta_l(m)$, we see that

$$\beta_l(m) = \log_2 \beta(s,m) = \log_2[\beta(m_{os})2^{(m-m_{os})/6}] \quad (5)$$

$$\beta_l(m) = \log_2 \beta(m_{os}) + (m - m_{os})/6 \quad (6)$$

$$\beta_l(m) = (\log_2 \beta(m_{os}) - m_{os}/6) + (1/6)m, \quad (7)$$

where string s has been dropped from the notation since we are looking only at variation along a fixed string. Substituting w_0 for $\log_2 \beta(m_{os}) - m_{os}/6$ and letting $w_1 = 1/6$, we obtain the standard linear form

$$\beta_l(m) = w_0 + w_1 m, \quad (8)$$

which shows that the log-inharmonicity trajectory of a given string varies linearly with respect to both MIDI pitch and fret number, as seen in Fig. 5.

We next used linear regression to obtain empirical log-inharmonicity trajectories. We collected all notes with common string labels in our training data and performed linear regression of their log-inharmonicities β against their fundamental pitches m (in MIDI note number format), repeating this procedure for each string label.

Let $s \in \{1,2,3,4,5,6\}$ be the string label to which each training note is assigned, and N_s be the number of notes we have belonging to string label s. If we let $\mathbf{x}_s^{(i)} = [1,m_s^{(i)}]^T$ represent the ith note with fundamental pitch $m_s^{(i)}$ belonging to string s , and let $\beta_s^{(i)}$ denote the log-inharmonicity of the ith note belonging to string s, we can solve

$$\mathbf{w}_s = \arg\min_{\mathbf{w}} \sum_{i=1}^{N_s} (\beta_s^{(i)} - \mathbf{w}^T \mathbf{x}_s^{(i)})^2 \quad (9)$$

which yields the weight vector $\mathbf{w}_s$ that minimizes the sum of squared error between the measured and predicted inharmonicities of the notes belonging to string s.

Occasional estimates from the inharmonicity measurement routine were outliers that deviated substantially from the linear trajectory model, so instead of the standard unweighted regression in (9) we performed a bi-square weighted linear regression to approximate the log-inharmonicities. Bi-square weighting is a robust linear regression procedure that minimizes a *weighted* sum of squares which de-emphasizes outliers [13]. This provided us with higher-quality regressions that were more robust to stray inharmonicity measurements.

To classify unknown notes, we imposed Gaussian probability distributions over the inharmonicity measurements of each string centered at their trajectories, and evaluated them at the inharmonicities of unseen notes to determine the most likely string from which it came. The means of the Gaussians are the log-inharmonicity values of points on the line, and the variances of the Gaussians are those of the residuals of the regressions, σ_s^2. In this manner, the probability distribution that characterized the inharmonicity trajectory $\mathbf{w}_s$ of string s at the fundamental pitch $m_0^{(i)}$ of the ith note $\mathbf{x}^{(i)} = [1,m_0^{(i)}]^T$ was defined as the normal distribution $\mathscr{N}(\mu,\sigma^2)$, with mean $\mu = \mathbf{w}_s^T \mathbf{x}^{(i)}$ and variance $\sigma^2 = \sigma_s^2$. Now to classify the nth unseen note $\mathbf{x}^{(n)}$, we measured its log-inharmonicity $\beta^{(n)}$ and obtained the probability that this log-inharmonicity was observed given each string s according to

$$P(\mathbf{x}^{(n)},\beta^{(n)}|s) = \mathscr{N}(\beta^{(n)}|\mathbf{w}_s^T \mathbf{x}^{(n)},\sigma_s^2), \quad (10)$$

and returned the predicted string $\hat{s}^{(n)}$ that maximized this probability,

$$\hat{s}^{(n)} = \arg\max_s P(\mathbf{x}^{(n)},\beta^{(n)}|s). \quad (11)$$

We also attempted classification using a decision function that selected the trajectory which minimized the residual instead of maximizing the probability, but we found that the latter method performed better.

It was straightforward to transform the string classifier output into tablature notation because it was assumed the tuning $\mathbf{t} = [m_1, m_2, m_3, m_4, m_5, m_6]^T$ of the unknown guitar was known, where m_s is the MIDI pitch number of the open note on string s. Since we had an estimate of the MIDI pitch $m^{(n)}$ of the nth unknown note, taking the difference between $m^{(n)}$ and the open pitch m_s of its assigned string yielded the fret on which the note was played.

Finally, we refined the fretboard estimates with plausibility filtering [10], in which we rejected and reassigned string decisions that were implausible given the constraints and assumptions of the data. If our systems assigned to a note a string on which the corresponding fretted position was negative, it would be clear that a mistake was made since frets cannot possibly be lower than 0 (the open string). Similarly, if our system assigned to a note a string on which the corresponding fret is greater than 12, a mistake would have been made because our guitar recordings feature only frets 0-12. When either of these situations arose, we rejected the string classification decision and selected the next most probable string. If the succeeding string was also implausible, the process was iterated until all six strings were considered.

3.2 Baseline Bayesian Classifier

The hypothesis of our regression method was that exploitation of the linear trajectories of log-inharmonicities is helpful for transcription. To objectively evaluate this we developed a baseline classifier, which we simply called Bayesian classification, which *did not* leverage the inharmonicity trajectories and whose results served as a benchmark.

As before, we first measured the inharmonicities of guitar notes using median-adjustive-trajectories (MAT). Frame aggregation was performed using the same specifications discussed in Section 3.1. Next, we estimated the inharmonicity distribution of each fretboard position f. For each of the $F = 78$ positions (13 notes $\times$ 6 strings), we calculated the median $\tilde{\beta}_f$ and variance σ_f^2 of its inharmonicities, and used these to parameterize a Gaussian probability density $\mathcal{N}_f(\tilde{\beta}_f, \sigma_f^2)$. We chose the median to lessen the influence of outliers.

We then classified unseen notes. We measured the inharmonicity $\beta^{(i)}$ of the ith unseen note $\mathbf{x} =$

$[1, m_0]^T$ with fundamental MIDI pitch $m_0^{(i)}$ and evaluated the probability $P_f(\beta^{(i)}|f) = \mathcal{N}_f(\beta^{(i)}|\tilde{\beta}_f, \sigma_f^2)$ of it having been generated by the Gaussian distribution $\mathcal{N}_f$. We again used plausibility filtering to narrow our search space $f \in \{1, 2, 3, ..., F-1, F\}$ of possible fretboard positions by considering just the positions $f \in \{f_{m,1}, f_{m,2}, ...\}$ that agree with the fundamental frequency of the unseen note. For standard tuning, this is a maximum of only three positions. To classify a note, we selected the candidate fretboard position which maximized the likelihood of the measured inharmonicity of the note, just as in equation (11):

$$\hat{f} = \underset{f \in \{f_{m,1}, f_{m,2}...\}}{\arg\max} P(\beta^{(i)}|f). \qquad (12)$$

4 Results

4.1 RWC Transcription

We evaluated our baseline and regression methods using a subset of the Real World Corpus (RWC) database [14]. We transcribed nine RWC guitars (three classical, three acoustic, and three electric) at three levels of context: guitar-specific, guitar-averaged, and guitar-independent. For guitar-specific trials, we trained and tested the system on the same guitar using three-fold cross-validation. For guitar-averaged trials, we trained the system on all recordings of the two guitars in the same class as the test guitar, as well as a two-thirds training portion of the test guitar, and then evaluated on the remaining third of the test guitar recordings. For guitar-independent trials, we trained the system on two of the guitars in the same class as the test guitar, and tested on all recordings of the remaining one. We used error probabilities to quantify the transcription performance for each guitar.

Figure 6 compares the error probabilities and significance ($\alpha = 0.05$, McNemar test [15]) of our two novel methods in all three contexts. Figure 7 compares the error probabilities and significance ($\alpha = 0.05$, t-test) of our methods and the PCT method developed in [3]. Because we did not have access to the trial-by-trial transcription results of PCT, we approximated the standard deviation σ_d of the distribution of the differences by the square root of the average of the variances of the methods $\sigma_d = \sqrt{(\sigma_1^2 + \sigma_2^2)/2}$ according to the Bernoulli approximation, in which $\sigma_x^2 = p_x(1 - p_x)$ where p_x is the probability of a correct classification for a given set of conditions.
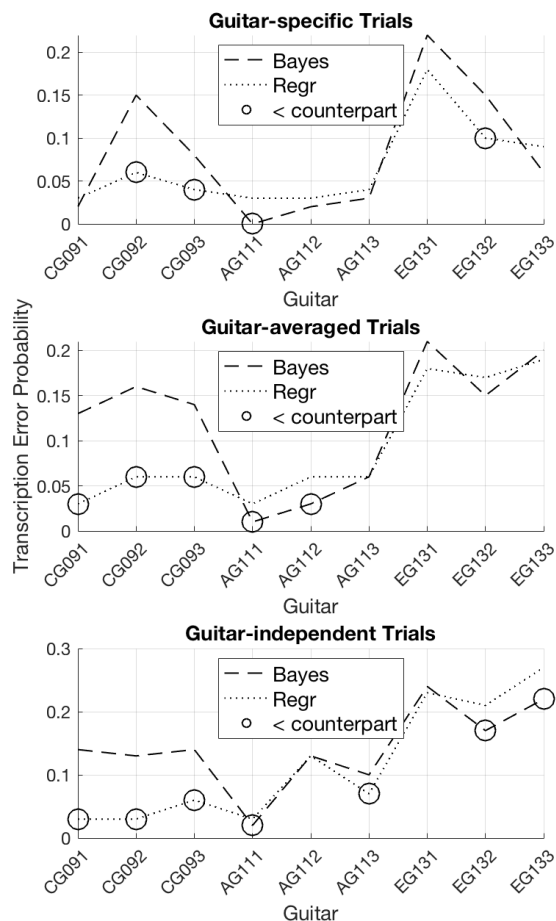
Fig. 6: For each guitar, the method (Bayes or Regression), if either, that performed significantly better than its counterpart is marked with a circle.

5 Discussion

5.1 RWC Transcription

Figure 6 shows that in ten of the twenty-seven RWC classical, acoustic, and electric guitar trials, the regression approach significantly outperformed the Bayesian baseline. The Bayesian baseline provided better performance on six of the audio segments, and there was no significant difference between the two for the remaining eleven recordings. Overall this suggests there is some merit to exploiting the structure of log-inharmonicity trajectories for improved transcription.
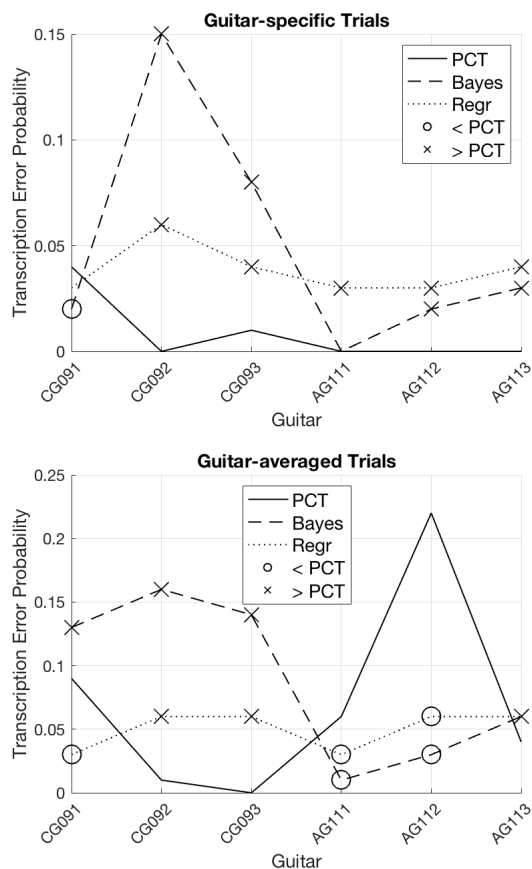

Fig. 7: For each guitar, if our methods performed significantly differently than PCT we mark it with either an 'O' or an 'X'. The former denotes significantly lower error probability while the latter denotes significantly higher error probability.

The "Guitar-independent Trials" plot of Figure 6 shows that the transcription performance of both our regression and baseline methods for unseen guitars remains adequate, which suggests that our systems exhibit an appreciable degree of generalizability. It is unclear, however, whether this degree of generalizability is unique; while we would have liked to compare the performance of PCT and our methods in this trial context, "guitar-independent" results were not reported in [3].

Figure 7 shows that PCT performed at a level that was equal to or better than our two methods in eight

of the twelve RWC classical and acoustic trials. At the time of writing of this paper we were not sure whether the authors of PCT used the same subset of RWC electric guitars, so we omitted them from our analysis. While we had expected that prior consideration of the inharmonicity of every fret would help our classifier obtain a more reliable picture of the inharmonicity trajectory, it appears that this is not the case. The better performance of PCT suggests that the theoretical inharmonicity trajectories are sufficient for transcription, at least for the "guitar-specific" and "guitar-averaged" trial contexts of classical and acoustic guitars.

We believe that the performance of our system could be improved with more reliable inharmonicity estimation. One uncontrolled variable that may have adversely affected performance was guitar string geometry. To understand better the potential impact of guitar string geometry on estimated inharmonicity, we recorded identical note plucks on personal guitars exhibiting a wide variety of combinations of string age, intonation length, and gauge size, and winding. We observed that variations in gauge size and winding affected our log-inharmonicity measurements by orders of magnitude between one-tenth and one, which is a degree of variability that is commensurate with differences between log-inharmonicity estimates for adjacent strings. The RWC database does not annotate string geometry parameters for any of its guitars, so it is possible that variations along these lines may have interfered with the quality of reported transcription results among systems using this database. We will explore the potential impact of this variability on our estimation quality in the future.

6   Summary

Tablature is a popular music notation for guitarists because it uniquely specifies fretboard positions for musical passages. Its manual annotation process, however, is tedious. Automatic transcription systems offer a solution to this difficulty. Inharmonicity is a phenomenon which affects the degree of upward skew of the partials of a note, and it is a key feature used in many systems that attempt automated guitar tablature. We proposed capitalizing on the predictable log-inharmonicity trajectory of a string by characterizing it as a linear function of its pitch, and we developed a baseline Bayesian classifier that exploits this relationship.

Comparisons between our baseline and regression methods suggested that exploiting inharmonicity regressions can be worthwhile, in that the regression-based approach achieved higher accuracy more frequently than the baseline did. The adequate transcription performance of the two methods in guitar-independent scenarios also suggested a degree of generalizability not reported in previous work. Nevertheless, our two approaches on average did not improve transcription accuracy on classical and acoustic guitars compared to the previously-developed PCT method in guitar-specific test scenarios. We also noted that guitar string geometry is an important factor to consider when developing inharmonicity-based methods, and our own results may have been better with more careful control of this variable.

References

[1] Macrae, R. and Dixon, S., "A guitar tablature score follower," in *Proc. Int. Conf. Multimedia Expo*, pp. 725–726, 2010.

[2] Fletcher, H., Blackham, E., and Stratton, R., "Quality of Piano Tones," *Journal of the Acoustical Society of America*, 34(6), pp. 749–761, 1962.

[3] Barbancho, I., Tardón, L., Sammartino, S., and Barbancho, A., "Inharmonicity-based method for the automatic generation of guitar tablature," in *Proceedings of the IEEE Transactions on Audio, Speech, and Language Processing*, pp. 1857–1868, 2012.

[4] Galembo, A., "Method for objective evaluation of the sound quality in piano trebles (in Russian)," in *Proc. of 18th Acoust. Conf. of Czechoslovakia*, 1979.

[5] Galembo, A., *The quality of piano tones (in Russian)*, Legkaya Industria Press, Moscow, 1987.

[6] Galembo, A. and Askenfelt, A., "Measuring inharmonicity through pitch extraction," *STL-QPSR*, 35(1), pp. 135–144, 1994.

[7] Galembo, A. and Askenfelt, A., "Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano," *IEEE Transactions on Speech and Audio Processing*, 7(2), pp. 197–203, 1999.

[8] Rauhala, J., Lehtonen, H. M., and Välimäki, V., "Fast automatic inharmonicity estimation algorithm," *Journal of the Acoustical Society of America*, 121(5), pp. 184–189, 2007.

[9] Hodgkinson, M., Wang, J., Timoney, J., and Lazzarini, V., "Handling Inharmonic series with median-adjustive trajectories," in *Proc. of the 12th Int. Conference on Digital Audio Effects (DAFx-09)*, Como, Italy, 2009.

[10] Abeßer, J., "Automatic String Detection for Bass Guitar and Electric Guitar," in *From Sounds to Music and Emotions - 9th International Symposium, CMMR 2012*, pp. 333–352, London, UK, 2012.

[11] Kehling, C., Abeßer, J., Dittmar, C., and Schuller, G., "Automatic Tablature Transcription of Electric Guitar Recordings by Estimation of Score- and Instrument-related Parameters," in *Proc. of the 17th Int. Conference on Digital Audio Effects (DAFx-14)*, Erlangen, Germany, 2014.

[12] Dittmar, C., Männchen, A., and Abeßer, J., "Real-time Guitar String Detection for Music Education Software," in *Proceedings of the 14th International Workshop on Image and Audio Analysis for Multimedia Interactive Services (WIAMIS)*, Paris, France, 2013.

[13] The MathWorks, "Least-Squares Fitting Documentation," http://www.mathworks.com/help/curvefit/least-squares-fitting.html, 2017.

[14] Goto, M., Hashiguchi, H., Nishimura, T., and Oka, R., "RWC Music Database: Music Genre Database and Musical Instrument Sound Database," in *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pp. 229–230, 2003.

[15] McNemar, Q., "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, 12(2), pp. 153–157, 1947, ISSN 1860-0980, doi:10.1007/BF02295996.