THE TEMPERAMENT POLICE: THE TRUTH, THE GROUND TRUTH, AND NOTHING BUT THE TRUTH

Simon Dixon[1], **Dan Tidhar**[2], **and Emmanouil Benetos**[1]
[1]Centre for Digital Music, Queen Mary University of London
[2]AHRC Research Centre for Musical Performance as Creative Practice, King's College London
[1]{simond,emmanouilb}@eecs.qmul.ac.uk, [2] dan.tidhar@kcl.ac.uk

ABSTRACT

The tuning system of a keyboard instrument is chosen so that frequently used musical intervals sound as consonant as possible. Temperament refers to the compromise arising from the fact that not all intervals can be maximally consonant simultaneously. Recent work showed that it is possible to estimate temperament from audio recordings with no prior knowledge of the musical score, using a conservative (high precision, low recall) automatic transcription algorithm followed by frequency estimation using quadratic interpolation and bias correction from the log magnitude spectrum. In this paper we develop a harpsichord-specific transcription system to analyse over 500 recordings of solo harpsichord music for which the temperament is specified on the CD sleeve notes. We compare the measured temperaments with the annotations and discuss the differences between temperament as a theoretical construct and as a practical issue for professional performers and tuners. The implications are that ground truth is not always scientific truth, and that content-based analysis has an important role in the study of historical performance practice.

1. INTRODUCTION

Recent years have seen a renewed interest in keyboard temperament both in scholarly work [14] and in more popular literature [9]. The modern tuning literature is abundant with detailed specifications of hundreds of different keyboard temperaments; some are directly taken from historical manuscripts and some are based on reconstruction or speculation [3,7]. A prescriptive approach taken by some scholars and performers regards adherence to specific temperaments as a desirable aim, and moreover, promotes the notion that for particular styles or even particular pieces there exists the "right" temperament [14]. An alternative approach, not less common amongst tuners and performers, is based on the view that since temperament is by definition a compromise, it is primarily a practical matter, and allows room for deviations from the underlying theoretical constructs. Rather than mistakes, such deviations are considered creative solutions to constraints arising from different instrument characteristics, inharmonicity, stylistic preferences, and the combinations of keys (tonalities) played in a concert programme.

Not all harpsichord CD sleeve notes specify the temperament, but when they do, there appears to be a tendency toward the former, prescriptive, approach. It is therefore intriguing to analyse such recordings and explore their adherence to the advertised temperaments. In this work, we analyse a dataset of over 500 harpsichord recordings for which temperament information is specified on the CD sleeve notes, aiming to shed some light on the relation between tuning theory and tuning practice, and more generally, on the nature of human "ground truth" annotations. We extend recent work demonstrating the feasibility of temperament estimation from solo harpsichord recordings [8, 18]. The proposed system uses a conservative NMF-based automatic transcription algorithm followed by frequency estimation using quadratic interpolation and bias correction. Multiple pitch estimates for each pitch class are combined with a median weighted by the pitch salience output of the transcription system. Results show significant gaps between advertised and actual temperaments, which can be interpreted as evidence for the more pragmatic approach to tuning.

2. BACKGROUND

2.1 Temperament

For the last two centuries, the scales used in Western music have been built predominantly upon equal temperament. This situation has been changing since the second half of the twentieth century, as part of the revival of interest in historical performance practice of early music on period instruments, resulting in increased attention to historical, unequal temperaments. We give a brief introduction to temperament, referring the reader to thorough treatments elsewhere [3, 7].

Explanations of musical consonance are based on the fact

E. Benetos is funded by a Westfield Trust research studentship (Queen Mary University of London).

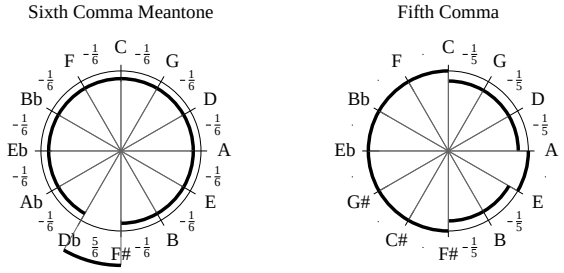

Figure 1. Circle of fifths representations for 2 temperaments used in this paper. The deviation of each fifth from a pure fifth (the lighter cicle) is represented by the positions of the darker segments. The fractions specify the distribution of the comma between the fifths (if omitted the fifth is pure).

that listeners prefer sounds with harmonic spectra and without beats [17]. For combinations of harmonic tones, the sensation of consonance correlates to small integer frequency ratios between fundamental frequencies, and particularly ratios of the form $\frac{n+1}{n}$ where $n \leq 5$ (corresponding to the following pure intervals for successive values of n: octave, perfect fifth, perfect fourth, major third and minor third).

The two most consonant intervals, the octave (ratio $\frac{2}{1}$) and perfect fifth (ratio $\frac{3}{2}$) correspond to intervals of 12 and 7 semitones respectively in Western music. From a given starting note, either a succession of 7 octave steps or a succession of 12 perfect fifth steps will lead to the same note. However, $(\frac{3}{2})^{12} \neq 2^7$, so it is not possible for all of these intervals to be pure simultaneously. Temperament refers to the various methods of adjusting some or all of the fifth intervals (octaves are always kept pure) with the aim of reducing the dissonance in the most commonly used intervals in a piece or programme of music. One way of representing temperament is by the distribution of the "Pythagorean comma" (the ratio $(\frac{3}{2})^{12} : 2^7 \approx 1.0136$) around the cycle of fifths (see Figure 1). For example, equal temperament diminishes all fifths by $\frac{1}{12}$ of a comma relative to the pure ratio 3:2. The other common way to represent temperament is by the frequency differences of each pitch class from their equal tempered counterparts, which is the representation we use in our results and analysis.

Theoretical models of temperament ignore the fact that stringed instruments are slightly inharmonic. This means that a pure fifth, maximally consonant when the 3rd partial of the lower tone coincides with the 2nd partial of the upper tone, will not correspond to a fundamental frequency ratio of $\frac{3}{2}$, as the partials are not precisely at integer multiples of the fundamental. We have shown [8] that this effect is of the order of a fraction of a cent for the harpsichord, which is negligible. The modelling of inharmonicity in the frequency estimation step is however important, and this is addressed in section 5.

2.2 Precise Frequency Estimation

Despite the vast literature on frequency and pitch detection (reviewed in [5, 12]), there is no general purpose method suitable for all signals and applications. Many systems assume monophonicity, stationarity and/or harmonicity, none of which hold for polyphonic harpsichord music, and only few papers address high-precision frequency estimation to a resolution of cents, which we require for the present work. The highest precision is obtained using the FFT with quadratic interpolation and correction of the bias due to the window function [1], which outperforms instantaneous frequency estimation using phase information [18]. Given a local peak a_p in the log magnitude spectrum $\log|X(n,p)|$ at frame n, that is, $a_{p-1} < a_p$ and $a_p > a_{p+1}$, then the three points $(-1, a_{p-1})$, $(0, a_p)$, and $(1, a_{p+1})$ uniquely define a parabola with maximum at:

$$\delta = \frac{a_{p-1} - a_{p+1}}{2(a_{p-1} - 2a_p + a_{p+1})} \tag{1}$$

where $-0.5 \leq \delta \leq 0.5$ is the fractional offset from the integer bin location p. This estimate is further refined using the following formula for bias correction, based on the window shape and zero padding factor [1, equations 1 and 3]:

$$\delta' = \delta + \xi_z \delta(\delta - 0.5)(\delta + 0.5) \tag{2}$$

where δ' is the bias-corrected offset in bin location, z is the zero-padding factor, $\xi_z = c_0 z^{-2} + c_1 z^{-4}$ is the bias correction factor and the constants $c_0 = 0.124188$ and $c_1 = 0.013752$ were determined empirically for the Blackman-Harris window [1, table 1].

3. DATA

The dataset used for this study consists of 526 tracks from 22 CDs and the 48 tracks from [18] [1] . Generally, the CDs present a rather balanced sample of recorded harpsichord music, including famous and less famous players, and a range of composers including J. S. Bach, D. Scarlatti, F. Couperin, M. Locke, and J. P. Sweelinck. The CDs provide details of the temperament used for the recordings. A few provide details of the reference frequency as well (e.g. $\mathrm{A} = 415$ Hz), but this is mostly not specified. In some cases the temperament information is precise and unambiguous, as in "Werckmeister III" or "Sixth comma meantone with the wolf between B and $G\flat$". In other cases it is underspecified, such as with "Neidhardt 1724", for which different versions exist both in the original manuscripts and in the secondary literature, or with "Quarter comma meantone" where the wolf interval (i.e. the widened fifth) is not specified. Some underspecification can be resolved by convention: although "meantone" can refer to several different temperaments – e.g. quarter comma or fifth comma meantone – the normal use of "meantone" without any qualification refers to quarter comma meantone.

[1] For details, see http://www.eecs.qmul.ac.uk/~simond/ismir11

4. TRANSCRIPTION

Our pitch estimation algorithm in Section 5 assumes that the existence and timing of each note is known. Therefore a transcription system for solo harpsichord was developed, using pre-extracted harpsichord templates, NMF with beta-divergence [13] for multiple-F0 estimation, and hidden Markov models (HMMs) [16] for note tracking. NMF with beta-divergence is a computationally inexpensive multiple-F0 estimation method which has been used for piano transcription [6]. It has been shown to produce reliable results for instrument-specific transcription, being highly ranked in the MIREX 2010 piano-only note tracking task.

4.1 Extracting Pitch Templates

Firstly, spectral templates were extracted from three different harpsichords, from the RWC musical instrument sounds database [11]. For extracting the note templates, the constant-Q transform (CQT) was computed with spectral resolution of 120 bins per octave. The standard NMF algorithm [15] with one component was employed for template extraction: $\mathbf{V} \approx \mathbf{wh}$, where $\mathbf{V} \in \mathbb{R}^{f \times n}$ is the input CQT spectrum, $\mathbf{w} \in \mathbb{R}^{f \times 1}$ is the extracted spectral template, and $\mathbf{h} \in \mathbb{R}^{1 \times n}$ is the component gain (since only one component was set, it corresponds to the frame energy).

For template extraction, the complete harpsichord note range was used (F1 to F6). Thus, three spectral template matrices were extracted, $\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \mathbf{W}^{(3)} \in \mathbb{R}^{f \times 61}$, corresponding to each harpsichord model.

4.2 Multiple-F0 estimation

For the multiple-F0 estimation step, we used the NMF algorithm with beta-divergence [13]. The basic model is the same as in the standard NMF algorithm: $\mathbf{V} \approx \mathbf{WH}$, where $\mathbf{W} \in \mathbb{R}^{f \times r}$, $\mathbf{H} \in \mathbb{R}^{r \times n}$, and r is the number of components. The beta-divergences (or β-divergences) are a parametric family of distortion functions which can be used in the NMF cost function to influence the NMF update rules for $\mathbf{W}$ and $\mathbf{H}$. Since in our case the spectral template matrix is fixed, only the gains $\mathbf{H}$ are updated as:

$$\mathbf{h} \leftarrow \mathbf{h} \otimes \frac{\mathbf{W}^T((\mathbf{Wh})^{\beta-2} \otimes \mathbf{v})}{\mathbf{W}^T(\mathbf{Wh})^{\beta-1}} \quad (3)$$

where $\mathbf{v} \in \mathbb{R}^{f \times 1}$ is a single frame from the test signal and $\beta \in \mathbb{R}$ the divergence parameter, set to 0.5 for this work, as in [6]. Although the update rule (Equation 3) does not ensure convergence, non-negativity is ensured [6].

For the harpsichord transcription case, the spectral template matrix was created by concatenating the spectral templates from all instrument models:

$$\mathbf{W} = [\mathbf{W}^{(1)} \ \ \mathbf{W}^{(2)} \ \ \mathbf{W}^{(3)}] \quad (4)$$

thus, $\mathbf{W} \in \mathbb{R}^{f \times 183}$. After the NMF update rule was applied to the input log-spectrum $\mathbf{V}$, the pitch activation matrix was created by summing the component vectors from $\mathbf{H}$ that correspond to the same pitch p:

$$\mathbf{H'}_{p,n} = \mathbf{H}_{p,n} + \mathbf{H}_{p+61,n} + \mathbf{H}_{p+122,n} \quad (5)$$

4.3 Note tracking

Instead of simply thresholding the pitch activation $\mathbf{H'}$ as was done in [6], additional postprocessing is applied in order to perform note smoothing and tracking. Here, the approach used in [4] was employed, where each pitch p is modeled by a two-state HMM, denoting pitch activity/inactivity.

The hidden state sequence for each pitch is given by $Q_p = \{q_p[t]\}$. MIDI files from the RWC database [11] from the classic and jazz subgenres were employed in order to estimate the state priors $P(q_p[1])$ and the state transition matrix $P(q_p[t]|q_p[t-1])$ for each pitch p. For each pitch, the most likely state sequence is given by:

$$\hat{Q}_p = \arg\max_{q_p[t]} \prod_t P(q_p[t]|q_p[t-1]) P(o_p[t]|q_p[t]) \quad (6)$$

which can be computed using the Viterbi algorithm [16]. For estimating the observation probability for each active pitch $P(o_p[t]|q_p[t]=1)$, we use a sigmoid curve which has as input the pitch activation $\mathbf{h}_p = \mathbf{H'}_{p,n}$ from the output of the transcription model:

$$P(o_p[t]|q_p[t]=1) = \frac{1}{1+e^{-(\mathbf{h}_p - \lambda)}} \quad (7)$$

where λ is a parameter that controls the smoothing (a high value will discard pitch candidates with low energy). The result of the HMM postprocessing step is a binary piano-roll transcription which can be used for evaluation.

For setting the parameter λ for the harpsichord transcription experiments, we employed a training dataset consisting of the 7 harpsichord recordings present in the RWC classical music database [11]. As a ground truth for the recordings, the syncRWC MIDI files were used [2]. Since for the present system a conservative transcription with high precision is favorable, λ was set to 0.25, which results in a false alarm rate of 5.33% with a missed detection rate of 46.49% (see [4] for metric definitions). An example harpsichord transcription is shown in Figure 2, where the piano-roll transcription of recording RWC MDB-C-2001 No. 24b is seen along with its respective MIDI ground truth.

5. PRECISE F0 ESTIMATION

Based on the transcription results, we search for spectral peaks corresponding to the partials of each identified note. For identification of the correct peaks, the tuning reference frequency and inharmonicity of the tone also need to be estimated. For Baroque music, the tuning reference frequency (expressed as the fundamental frequency of the note A4) is

[2] http://staff.aist.go.jp/m.goto/RWC-MDB/AIST-Annotation/SyncRWC

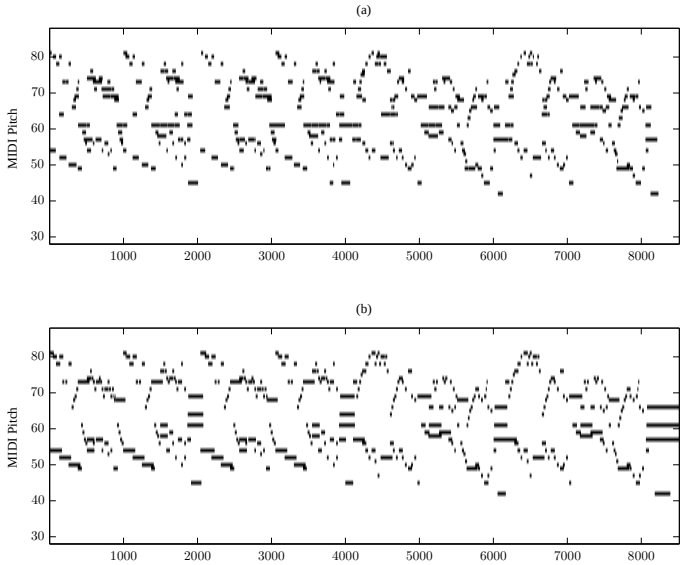

Figure 2. (a) The piano-roll transcription of J.S. Bach's *Menuet in G minor* (RWC MDB-C-2001 No. 24b). (b) The pitch ground truth of the same recording. Units on the abscissa correspond to 10ms.

usually lower than the modern standard of 440 Hz. For our data set, the CD sleeve notes mention reference frequencies of 405, 415 and 440 Hz, with the majority of CDs not giving any value. This introduces a problem: without knowing the score (or at least the key) of a piece of music, it is not possible to determine the reference frequency unambiguously, since, for example, a note with F0 around 415 Hz could be A4 (reference 415 Hz) or G♯4 (reference 440 Hz).

The tuning frequency is ascertained by the following iterative process: 40 frames are selected (equally spaced throughout the piece) and the fundamental frequency estimation stage described below is computed, using an initial value of 440 Hz for the tuning frequency and taking the inharmonicity estimates from measurements of other harpsichords [8]. The frequencies are divided by their nominal values (given the reference frequency and assuming equal temperament), and a weighted average of the deviations is computed. The reference frequency is updated by the result and the process is repeated for 5 iterations, or until it converges (the update is less than one cent) if sooner.

The inharmonicity of each note is estimated jointly with its fundamental frequency. For a string with (ideal) fundamental frequency f_0 and inharmonicity constant B, the frequency f_k of the kth partial is given by [10]:

$$f_k = k f_0 \sqrt{1 + Bk^2} \tag{8}$$

where the constants f_0 and B depend on the physical properties of the string. Given any two partials of a note, it is possible to solve for f_0 and B, assuming the partial numbers are known. We compute these two parameters for each pair of partials estimated below, and use a robust statistic, the median over all frames and partial pairs, to estimate the true values, using the inter-quartile range as an inverse measure of confidence in the estimates.

The fundamental frequency and inharmonicity of each transcribed note are computed as follows:
1) Compute the STFT using the following parameters: $f_s = 44100$ Hz, Blackman-Harris window with support size of 4096 samples (93 ms), zero padding factor $z = 4$ ($N = 16384$), and hop size of 1024 samples.
2) For each note w given by the transcription, compute an initial estimate of the frequency f_k^w of partial $k = 1...40$ with equation 8, using the reference frequency computed above, the inharmonicity estimate from [8], and assuming equal temperament for the fundamental.
3) For each partial frequency, a local spectral peak in a window of ± 30 cents around f_k^w is sought, and if found the frequency estimate is refined as described in subsection 2.2.
4) Using the transcription, any overlapping partials are identified and deleted from the estimate, as they are likely to give unreliable values. Partials are deemed to overlap if their frequency separation is less than $3.03 f_s z/N$ [2].
5) For each pair of partials remaining, solve for F0 and B using equation 8.
6) For each pitch class k, convert each frequency estimate to cents deviation from equal temperament and return the weighted median $\hat{c}_k$ as the overall tuning value for the pitch class, where the weights are given by the pitch activation $\mathbf{H}'_{p,n}$ (Equation 5). This gives a 12-dimensional temperament vector, which can be compared with the profiles of known theoretical temperaments. For simplicity we represent the pitch class k by an integer from 0 (C) to 11 (B), corresponding to the MIDI pitch number modulo 12.

6. TEMPERAMENT ESTIMATION

Our temperament classifier recognises the following temperaments: equal, fifth comma, Vallotti, quarter comma meantone (QCMT), fifth comma meantone (FCMT), sixth comma meantone (SCMT), Kellner, Werckmeister III, Lehman, Neidhardt (1,2 and 3), Kirnberger (2 and 3) and just intonation. We also recognise rotations of these temperaments, although this is not a typical tuning practice for all temperaments, as illustrated by the example of the Young II temperament, a rotation of the Vallotti temperament, which is considered a different temperament in its own right. Rotations are specified via the wolf interval where applicable (e.g. SCMT-FD has wolf interval F♯-D♭, as in Figure 1), otherwise by the number of semitones rotated (e.g. Vall+7).

Given the estimate $\hat{c} = (\hat{c}_0, \ldots, \hat{c}_{11})$ and a temperament profile $c^i = (c_0^i, \ldots, c_{11}^i)$ for temperament i, we calculate the divergence between estimate and profile, $d(\hat{c}, c^i)$:

$$d(\hat{c}, c^i) = \sum_{k=0}^{11} \frac{u_k (\hat{c}_k - c_k^i - r)^2}{\sum_{j=0}^{11} u_j} \tag{9}$$

where $u_k = \sum_n \sum_{p \equiv k \mod 12} \mathbf{H}'_{p,n}$ is the weight for pitch

class k, and $r = \sum_{j=0}^{11} u_i(\hat{c}_j - c_j^i) / \sum_{j=0}^{11} u_j$ is the offset in cents which minimises the divergence and thus compensates for deviations in the reference tuning frequency (pitch A4) from the reference computed above in previous calculations. A piece is classified as having the temperament i whose profile c^i gives the least divergence $d(\hat{c}, c^i)$. We also consider rotations of temperaments, $c^{i,r}$, given by $c_k^{i,r} = c_m^i$, where $m \equiv (k + r) \mod 12$, in order to deal with different positions of the wolf interval in meantone temperaments, as well as the tuning ambiguity discussed in section 5.

7. SUMMARY OF RESULTS

The results are summarised in Table 1 [3]. Column 1 is our CD index, where letters are used to distinguish groups of tracks with different temperament metadata. Column 2 shows the annotated reference tuning, while the mean and standard deviation of the estimated reference tuning are given in columns 3 and 4 respectively. Columns 5 to 8 give the annotated temperament, the average divergence $d(\hat{c}, c^i)$ from this temperament, the most frequent highest ranked temperament according to $d(\hat{c}, c^i)$, and the average difference in divergence between the annotated temperament and the best ranked temperament.

The results for tuning show agreement with the ground truth values where they were available, with the exception of CD 21, which had only 2 tracks at 440 Hz. The CDs generally show tuning consistency across all tracks, with high standard deviations ($>$ 2 Hz) being due to a bimodal distribution of tuning frequency (CD 18) and 5 outlier tracks (CDs 2,7,19). Summarising by CD assumes fixed tuning for all tracks, which is clearly not always the case.

The temperament results vary from close agreement to the metadata (CDs 4,5,8,9,16,21,22) to moderate agreement (e.g. CDs 15, 18) to disagreement (e.g. CDs 12,13, 17). An example is shown in Figure 3. For a number of tracks it was not possible to find a single "best fit", as some temperaments are only distinguished by a pitch class which does not appear (or is not detected) within the piece. The large divergences of CDs 2 and 19 are explained by the tuning frequency being at the half-way point between two semitones relative to the 440 Hz reference assumed by the transcription algorithm, making the transcriptions unreliable.

On CD 17 and some other tracks specifying QCMT, the temperament was often closer to FCMT. This is an interesting tendency, as two are fairly similar, with FCMT being milder (slightly larger major thirds and a smaller wolf interval). It seems plausible that QCMT was intended but then tempered to bring it (inadvertantly) closer to the less extreme FCMT. However, the opposite tendency appears on CD 3a. Werckmeister 3 is specified on five CDs, but only fulfils the claim on two. The reason may be that Werckmeister 3 is popular as a starting point for tuners while they experiment and develop their own temperaments, or that it is very close to other temperaments such as Kellner (note the low value of ΔDiv in each case).

[3] It is not possible to fit all results into this paper. For more details, please see: http://www.eecs.qmul.ac.uk/~simond/ismir11

	Tuning			Temperament			
CD	Not.	Est.	StD	Notated	Div.	Estimated	ΔDiv.
1		417.6	0.2	Ordinaire		Neid2	
2	405	405.7	3.2	FCMT	21.8	Various	16.4
3a		416.8	0.2	SCMT-BG	3.3	FCMT-BG	2.5
3b		413.9	0.2	Kellner*	8.5	Various	1.2
3c		414.2	0.2	Kellner	3.3	Kellner	0.0
4b		416.9	0.3	FCMT-FD	1.1	FCMT-FD	0.0
5	415	417.1	0.9	QCMT	1.4	QCMT-GE	0.0
6		413.8	0.7	Late17		Vall+7	
7		432.6	4.8	FCMT	7.6	Various	4.1
8b		416.8	0.4	QCMT	1.2	QCMT-GE	0.0
9	415	415.3	0.3	Neid	1.1	Neid1/2	0.0
10	415	416.5	0.4	Werck3	3.4	Various	1.7
11	415	416.6	0.6	Werck3	3.0	Various	0.9
12	415	415.3	0.2	Kirn3	11.1	Neid1	9.4
13	415	415.1	0.3	Kirn3	7.3	Neid1	5.9
14a	(415)	412.7	0.3	QCMT	10.0	Various	7.0
14c	(415)	435.2	0.2	QCMT	2.7	QCMT-GE	0.0
15		415.7	1.3	Werck3	3.4	Werck3	0.5
16		416.1	1.1	Werck3	0.0	Werck3	0.9
17		413.9	1.2	QCMT	6.0	FCMT	2.2
18		440.5	2.4	QCMT	5.0	QCMT-GE	2.7
19	440	447.6	5.6	QCMT	19.5	FCMT	15.2
20		412.9	0.6	Werck3	2.6	Various	0.8
21		414.5	1.6	FCMT	1.0	FCMT-GE	0.0
22		408.7	0.3	Lehman	1.1	Lehman	0.1
RH	415	415.5	0.8	Various	7.1	Various	0.3
PT	415	415.6	0.7	Various	0.1	All correct	0.0

Table 1. Summary of results, with columns for CD number, notated reference tuning, estimated reference tuning, standard deviation across tracks of CD, notated temperament, highest ranked temperament (Eqn 9), and average difference in divergence $d(\hat{c}, c^i)$ between notated and highest ranked temperaments. The last two rows refer to the data from [18].

Since we are claiming that CD sleeve notes are a questionable source of "ground truth", we need an independent means of ascertaining the reliability of our system. The bottom row of Table 1 shows the results for 4 pieces recorded with six different temperaments using the physical modelling synthesiser Pianoteq [18]. Using the current approach, these tracks were all classified correctly from the set of 180 possible temperaments (15 temperaments by 12 rotations). Confidence in classification results can also be gained by considering the divergence value and consistency of results (i.e. if a number of related tracks are classified with the same label and low divergence from the given temperament).

8. CONCLUSION

We have presented a method for analysing harpsichord temperament directly from audio recordings, using an NMF-based transcription system, followed by bias-corrected quadratically interpolated short-time spectral analysis to estimate partial frequencies, estimation of inharmonicity, deletion of overlapping partials, and robust statistics weighted by the pitch salience given by the transcription system. We anal-

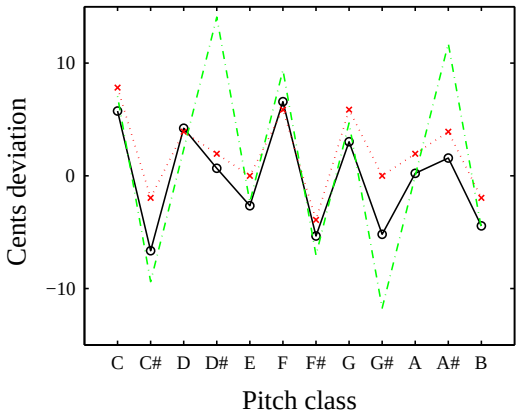

Figure 3. Estimated temperament profile (solid line, circles) compared with the temperament specified on the CD (dot-dash) and that with least divergence from the estimate (dotted line, crosses). In this case the data matches the Vallotti profile (d = 2.2) more closely than the specified Fifth Comma Meantone (d = 17.1).

ysed a collection of CDs which provide metadata about the tuning system, and found that while this information is mostly correct, there were several cases in which another temperament matches the data more closely than the advertised one. This is perhaps more surprising to a music theorist than to a practising tuner or performer, reflecting the dichotomy between those who see temperament as a mathematical system and those who have to retune their instrument during the interval of a concert. This also raises an interesting issue about the nature of human annotations and their use as "ground truth". The metadata provided with the CD is intended to give an indication of the tuning system rather than scientifically accurate documentation, and we need to be discerning in the use of metadata that has been collected for a purpose other than scientific analysis or evaluation.

9. REFERENCES

[1] M. Abe and J. Smith. CQIFFT: Correcting bias in a sinusoidal parameter estimator based on quadratic interpolation of FFT magnitude peaks. Technical Report STAN-M-117, CCRMA, Dept of Music, Stanford University, 2004.

[2] M. Abe and J. Smith. Design criteria for the quadratically interpolated FFT method (II): Bias due to interfering components. Technical Report STAN-M-115, CCRMA, Dept of Music, Stanford University, 2004.

[3] J.M. Barbour. *Tuning and Temperament, A Historical Survey*. Dover, Mineola, NY, 2004/1951.

[4] E. Benetos and S. Dixon. Polyphonic music transcription using note onset and offset detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011. 37–40.

[5] A. de Cheveigné. Multiple f0 estimation. In D.L. Wang and G.J. Brown, editors, *Computational Auditory Scene Analysis: Principles, Algorithms and Applications*, pages 45–79. IEEE Press/Wiley, Piscataway, NJ, 2006.

[6] A. Dessein, A. Cont, and G. Lemaitre. Real-time polyphonic music transcription with non-negative matrix factorization and beta-divergence. In *11th International Society for Music Information Retrieval Conference*, pages 489–494, 2010.

[7] C. Di Veroli. *Unequal Temperaments: Theory, History, and Practice*. Bray Baroque, Bray, Ireland, 2009.

[8] S. Dixon, M. Mauch, and D. Tidhar. Estimation of harpsichord inharmonicity and temperament from musical recordings. *Journal of the Acoustical Society of America*, 2011. To appear.

[9] R. E. Duffin. *How equal temperament ruined harmony (and why you should care)*. W. W. Norton, 2007.

[10] H. Fletcher. Normal vibration frequencies of a stiff piano string. *Journal of the Acoustical Society of America*, 36(1):203–209, 1964.

[11] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. RWC music database: Music genre database and musical instrument sound database. In *4th International Conference on Music Information Retrieval*, pages 229–230, 2003.

[12] A. Klapuri and M. Davy, editors. *Signal Processing Methods for Music Transcription*. Springer, New York, NY, 2006.

[13] R. Kompass. A generalized divergence measure for nonnegative matrix factorization. *Neural Computation*, 19(3):780–791, 2007.

[14] B. Lehman. Bach's extraordinary temperament: our rosetta stone. *Early Music*, 33(1):3–23, 2005.

[15] D. D. Li and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, October 1999.

[16] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[17] E. Terhardt. The two-component theory of musical consonance. In E. Evans and J. Wilson, editors, *Psychophysics and Physiology of Hearing*, pages 381–390. Academic, London, 1977.

[18] D. Tidhar, M. Mauch, and S. Dixon. High precision frequency estimation for harpsichord tuning classification. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 61–64, 2010.