HANDLING INHARMONIC SERIES WITH MEDIAN-ADJUSTIVE TRAJECTORIES

Matthieu Hodgkinson, Jian Wang, Joseph Timoney and Victor Lazzarini
Digital Sound and Music Technology Group
National University of Ireland, Maynooth
matthew.hodgkinson@nuim.ie, Jwang@cs.nuim.ie,
jtimoney@cs.nuim.ie, victor.lazzarini@nuim.ie

ABSTRACT

A new method for the analysis of inharmonic instrumental tones is presented. The method exploits an equation derived from the well-know inharmonic series equation, where the inharmonicity coefficient is balanced with the frequencies and numbers of any two partials extracted from a pseudo-harmonic series. A serial search for increasingly deviating spectral peaks is aided with the integrated refinement of increasingly reliable inharmonicity coefficient and fundamental frequency estimates. This firsthand approach to the problem of evaluating inharmonic spectra brings about an unprecedented level of simplicity, efficiency and accuracy.

1. INTRODUCTION

As opposed to the solutions of the wave equation, the frequencies of the modes of vibration of actual instrumental strings are not exact integer multiples of the fundamental frequency. In fact, material stiffness imparts an additional, curvature-related restoring force, which raises higher frequency components above their "ideal" harmonic position.

In mathematical terms, the relation between the k^{th} partial f_k and the fundamental frequency f_0 of a spectrum featuring such inharmonicity can reliably be expressed as

$$f_k = k f_0 \sqrt{1 + \beta k^2}\,, \tag{1}$$

where β is the so-called *Inharmonicity Coefficient* (IC). The latter can itself be related to physical parameters of a plain string (i.e. not wounded) after the equation

$$\beta = \frac{\pi^3}{64}\frac{Qd^4}{Tl^2}, \tag{2}$$

where Q is the material's elasticity modulus, d the diameter, T the tension and l the length of the string [1].

While inaudible in most plucked- and hit-string instruments, this acoustical effect is perceptually noticeable on such instruments featuring sufficiently thick strings in their lower end. In the case of the piano, for instance, the effect is so conspicuous it cannot be ignored during the tuning process. Also, in sound synthesis, the proper emulation of instruments featuring audible inharmonicity necessarily requires the comprehension of the phenomenon. In any case, it is desirable to possess the means of measuring the inharmonicity of tones.

2. MEDIAN-ADJUSTIVE TRAJECTORIES

To help the reader situate the proposed method of Median-Adjustive Trajectories (MAT), the next section opens with a chronological overview of the main existing methods for automatically estimating the analytical values of β and f_0. The proposed method is thereafter explained in detail.

2.1. Previous methods

Ever since it has been considered, the problem of automatically estimating the inharmonicity coefficient has been approached indirectly. In 1994, Galembo and Askenfelt pointed out that the one peak produced in the cepstra and Harmonic Product Spectra of inharmonic tones features distinctive width, by the interpretation of which it is possible to derive an estimate for β [2]. Thence was cleverly suggested to take account of the "partial stretch" caused by inharmonicity in the frequency-axis rescaling process, leading to the expression for the *Inharmonic Product Spectrum* (IHPS)

$$\pi_{\text{IHPS}}(f_0, \beta) = \prod_{p=1}^{P} \left| \mathrm{F}\left[p 2\pi \frac{f_0}{f_s} \sqrt{1 + \beta p^2} \right] \right|^2, \tag{3}$$

where f_s is the audio sampling rate, p is the spectrum compression factor, P is the order of the IHPS, and $\mathrm{F}(\omega)$, the Fourier transform of the signal at the angular frequency ω.

The aim there was to obtain as high and narrow a peak as possible, implying accordingly accurate estimates of β and f_0. Practically, the IHPS can be implemented by the means of a non-uniform Fourier transform, evaluating

$$\pi_{\text{IHPS}}(f_0, \beta) = \prod_{p=1}^{P} \left| \sum_{n=0}^{N-1} x[n] w[n] e^{-jpn 2\pi \frac{f_0}{f_s}\sqrt{1+\beta p^2}} \right|^2 \tag{4}$$

for a number of f_0 and β values across their respective range of possibilities. In (4), $x[n]$ and $w[n]$ are the n^{th} samples of the discrete analysed signal x and window function w, respectively.

An alternative method, the Inharmonic Comb Filter (ICF), was presented in 1999 by the same authors [3]. Here, (1) is used to make the notches of a frequency-domain comb filter coincide with the peaks of an inharmonic spectrum, as illustrated in Fig. 1.

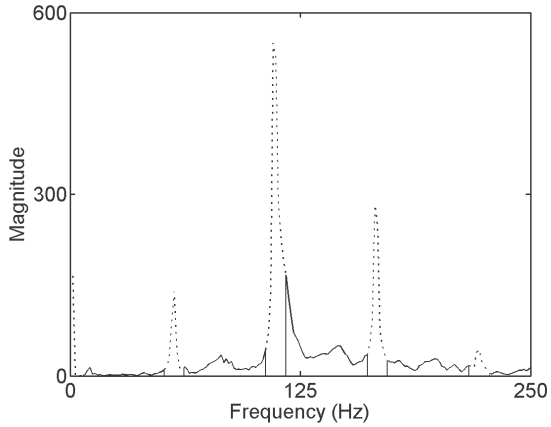

Figure 1: *Inharmonic comb-filtered spectrum*

As in the case of the IHPS, f_0 and β are swept across their respective, user-estimated range of realistic values. For each combination, the sum of the values of the accordingly comb-filtered power spectrum is evaluated. The f_0 and β couple yielding the least sum is finally taken as the best estimate.

The major inconvenient of both above-presented methods is the necessity of evaluating the IHPS/ICF for a number of f_0 and β combinations. In [3], for instance, β is given an initial value of $300 \cdot 10^{-6}$, while f_0 is varied logarithmically between 20 and 200Hz in 200 steps to get a rough fundamental frequency estimate. Secondly, f_0 is linearly varied in 30 steps across +/−10% the previously obtained estimate, and at each step, β is varied between 0 and 0.001 in 200 logarithmic steps. Thirdly, f_0 and β are both varied in 30 linear steps around +/−5% the estimates obtained in the previous run. Altogether, this brings the power spectrum to be comb-filtered and summed 7,100 times.

Rauhala, Lehtonen and Välimäki addressed in 2006 the need for a far more efficient algorithm [4]. The method uses a peak detection where the peak-detection frequency bands are centred on estimates for the partial frequencies f_k. The latter are obtained from the substitution into (1) of estimates for β and f_1,

$$\hat{f}_k = k\hat{f}_1\sqrt{\frac{1+\hat{\beta}k^2}{1+\hat{\beta}}}\,. \tag{5}$$

The trend of the frequency differences between the estimated and measured inharmonic series is considered: a positive trend suggests too great a β estimates, and vice-versa. The estimate is accordingly adjusted in an iterative peak-detection process, and thus made to converge towards a precise estimate.

2.2. Median-Adjustive Trajectories (MAT)

As suggested by Alexander Galembo's *Inharmonic Calculator* [5], the IC can be expressed in terms of the frequency and number of any two partials of a series. This is done by expressing f_0 in terms of the m^{th} partial f_m as

$$f_0 = \frac{f_m}{m\sqrt{1+\beta m^2}}\,, \tag{6}$$

and then substituting (6) into (1), yielding

$$f_k = k\frac{f_m}{m\sqrt{1+\beta m^2}}\sqrt{1+\beta k^2}\,. \tag{7}$$

Now we solve for β,

$$\beta = \frac{\left(f_k\frac{m}{k}\right)^2 - {f_m}^2}{k^2{f_m}^2 - m^2\left(f_k\frac{m}{k}\right)^2}\,. \tag{8}$$

Thus, any couple of partials of a spectrum known to belong to the series stemming from the fundamental frequency, and whose numbers are known, can provide with an estimate of β. The difficulty resides in finding partials when β remains unknown, especially at high partial indexes, as the k^{th} partial of an inharmonic series deviates by $k((1+\beta k^2)^{0.5}\text{-}1)$ times the fundamental frequency. Fig. 2 illustrates such deviation in the example of a bass guitar open E with fundamental frequency $f_0 \approx 20.6356$Hz and IC $\beta \approx 3.9602 \cdot 10^{-4}$. For values of β that substantial, this frequency stretch renders unreliable the detection of pseudo-harmonics using frequency bands centred on integer multiples of the fundamental frequency.

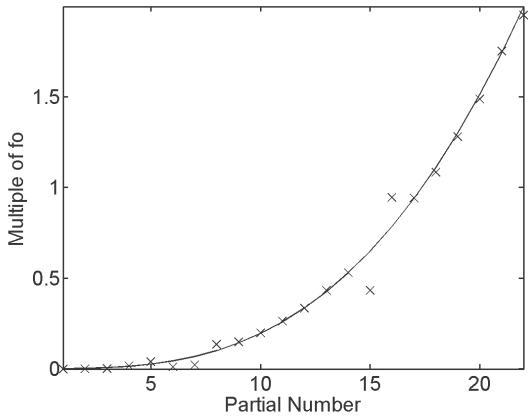

Figure 2: *A Bass Guitar example of partial frequency deviation as multiple of the fundamental frequency.*

On the other hand, the deviation caused by realistic inharmonicity values on the first two partials is trivial. In the previous example, for instance, the deviation of those does not exceed one thousandth of the fundamental frequency.

The present method takes advantage of this fact. Here, we estimate the frequencies of the two prominent peaks within narrow peak-detection frequency band, respectively centered around $f_{0,\text{ET}}$ and $2f_{0,\text{ET}}$ ($f_{0,\text{ET}}$ is the user-input, Equal-Temperament fundamental frequency). Thereafter, those values and corresponding peak numbers are substituted into (8) for the calculation of the first

entry of an array **B** of β estimates. This estimate is used in (6), in turn with each of the above-mentioned partial frequencies, to calculate the first two entries of another array of estimates **Fo**. The arrays' medians are finally substituted into (1) in place of β and f_0 to obtain a frequency estimate for the next partial. This estimate is used as the centre frequency of the peak-detection frequency band of the next "partial step".

Fig. 3 pictures the first three steps of the method in detail, and can be regarded as a pictorial aid to the implementation of the algorithm, to be used with equations (1), (6) and (8) at hand.

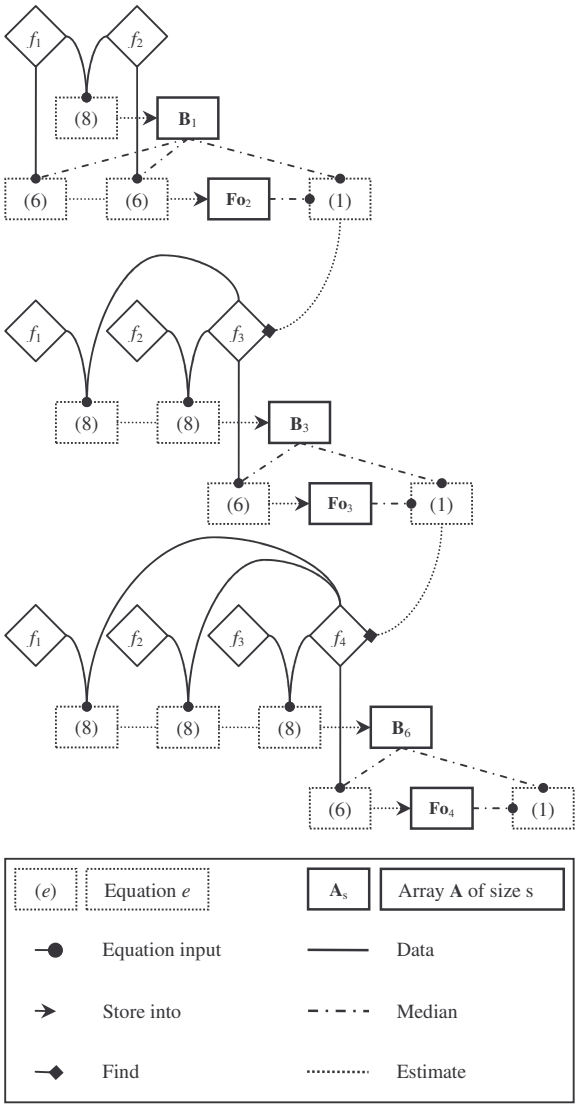

Figure 3: *Median-Adjustive Trajectory method*

As the figure suggests, there is a potential triangle number

$$E = \frac{K^2 - K}{2} \tag{9}$$

of β estimates for K measured partial frequencies, along with $K f_0$ estimates.

Another way of interpreting the procedure is to visualise a trajectory of the form (1) for a continuous k, on either sides of which are picked up the closest peaks. The frequency information related to those peaks feeds back into (1) to adjust the trajectory and progress further into the spectrum. Fig. 4 shows such a trajectory at three different stages of its adjustment, during the analysis of a C#3 of a hammered dulcimer (with a 69.0789Hz fundamental frequency). As it is more convenient for visualization, the ordinate axis here represents the derivative of the frequency series for a continuous k,

$$\frac{df_k}{dk} = f_0 \frac{1 + 2\beta k^2}{\sqrt{1 + \beta k^2}}. \tag{10}$$

As for the measured frequencies (circles), they were substituted in (10) using (6).

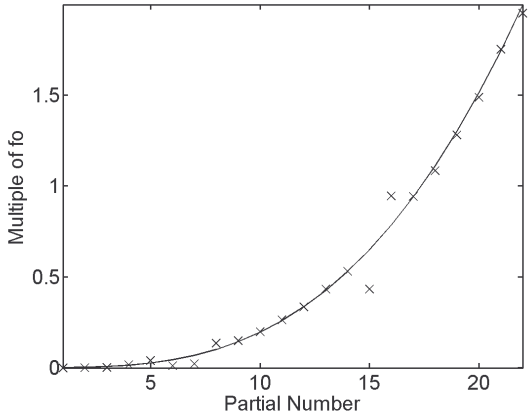

Figure 4: *Three states of a median-adjustive trajectory*

The peak-detection process can be carried as far as the series features sufficient energy to raise peaks above the threshold of noise. It was empirically determined most convenient to use the average of the magnitude spectrum as the value against which to compare the magnitude of each newly detected peak. Fig. 5 gives an idea of the level of such a threshold within the spectrum of a mandolin's G3. This magnitude threshold check, introduced in the peak detection loop, allows the latter to stop when no more significant partials are found.

2.3. Refinement of the partial frequency estimates

The accuracy of the β estimate is necessarily dependent on the accuracy of the spectral analysis. It is therefore desirable to use long window sizes, at least of 2^{14} (16,384) samples for a CD-quality sample rate (44,100Hz). Also, when dealing with instruments such as the piano, which feature courses of strings (i.e. two or more strings are coupled to increase the volume of each note), windows of 2^{15}, or even 2^{16} samples enhance the chances of making the distinction between parallel series.

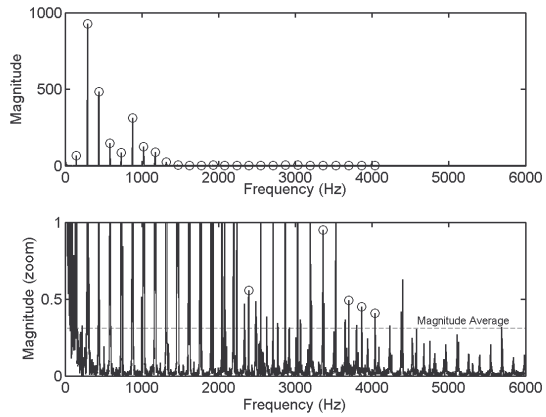

Figure 5: *Magnitude spectrum average used as peak detection threshold*

Meanwhile, regardless of the size of the window, it is possible to refine the DFT-quantised frequency of a peak standing on a discrete-frequency axis. For example in [4], a continuous peak is imagined between the detected discrete-frequency maximum and its immediate neighbours. Assuming the shape of a second-order polynomial, the frequency value at which the corresponding first-order gradient is nil is taken as a better estimate as to where the frequency component actually stands.

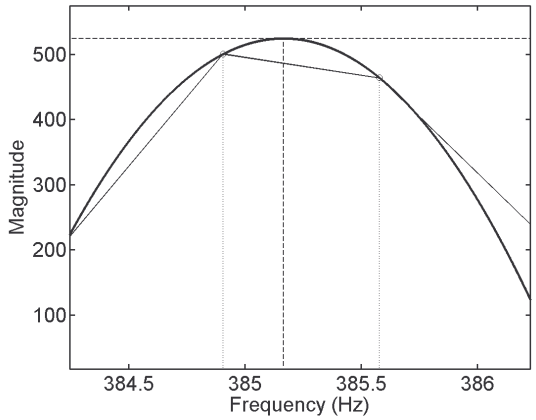

Figure 6: *An example of partial frequency estimation with quadratic fit.*

Alternatively, the use of Complex Spectral Phase Evolution (CSPE) [6] offers a significant frequency refinement for all components detected in the Fourier analysis, and can be integrated in the algorithm without so much as a handful of extra lines before the peak detection loop. Here, the Fourier analysis is performed twice, the second time upon a one-sample shift of the same discrete signal. The product of the time-shifted spectrum with the complex conjugate of the initial spectrum conveniently isolates the actual frequencies of the detected components.

Let us express a segment of N samples of a complex signal $\boldsymbol{x}[n]$ as the sum of N frequency components, each of complex amplitude $\boldsymbol{a}_k$ and angular frequency ω_k,

$$\boldsymbol{x}[n] = \sum_{k=0}^{N-1} \boldsymbol{a}_k e^{j\omega_k n}, n = 0,1,..., N-1 \cdot \tag{11}$$

The signal's one-sample shifted version, $\boldsymbol{x}'[n]$, can likewise be expressed as

$$\boldsymbol{x}'[n] = \sum_{k=0}^{N-1} \boldsymbol{a}_k e^{j\omega_k (n-1)} = \sum_{k=0}^{N-1} \boldsymbol{a}_k e^{j\omega_k n} e^{-j\omega_k} \cdot \tag{12}$$

Now we express the Fourier series $\boldsymbol{X}[p]$ of $\boldsymbol{x}[n]$,

$$\boldsymbol{X}[p] = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} \boldsymbol{a}_k e^{j\omega_k n} e^{-j\frac{2\pi}{N}pn}, p = 0,1,..., N-1 \cdot \tag{13}$$

We take the component frequencies ω_k to deviate from being harmonics of the analysis' fundamental frequency by ε_k radians per sample, i.e. $\omega_k = 2\pi k/N + \varepsilon_k$. Thus we can reformulate (13) as

$$\boldsymbol{X}[p] = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} \boldsymbol{a}_k e^{j\frac{2\pi}{N}(k-p)n} e^{j\varepsilon_k n} \cdot \tag{14}$$

Meanwhile,

$$\sum_{n=0}^{N-1} e^{j\frac{2\pi}{N}(k-p)n} = 0, k \neq p \cdot \tag{15}$$

Thus, on the condition that the components are exact multiples of the analysis' fundamental frequency (implying $\varepsilon = 0$), we can write

$$\boldsymbol{X}[p] = \boldsymbol{a}_p, \tag{16}$$

and likewise,

$$\boldsymbol{X}'[p] = \frac{1}{N} \sum_{n=0}^{N-1} \boldsymbol{a}_p e^{-j\omega_p} = \boldsymbol{a}_p e^{-j\omega_p} = \boldsymbol{X}[p] e^{-j\omega_p} \cdot \tag{17}$$

By taking the product of the time-shifted spectrum $\boldsymbol{X}'[p]$ with the conjugate of the original spectrum, $\boldsymbol{X}^*[p]$, we obtain

$$\boldsymbol{X}'[p] \cdot \boldsymbol{X}^*[p] = |\boldsymbol{X}[p]|^2 e^{-j\omega_p}, \tag{18}$$

and therefore can extract the desired frequency taking

$$\omega_p = \angle\left(\boldsymbol{X}'[p] \cdot \boldsymbol{X}^*[p]\right). \tag{19}$$

Equations (16) and (17) strictly hold only when the fundamental frequency of the analysed signal is an integer multiple of the nat-

ural frequency of the analysis, in which case, ironically, any frequency refinement is unnecessary. However, the method allows room for error, as will verify the following tests.

The estimates of the quadratic fit were compared with those of the CSPE. A sequence (f_c) of twenty centre frequencies, equally spaced between the excluded limits of 0Hz and the Nyquist frequency, was created:

$$(f_c) = c\frac{f_s}{41}, c = 1,...,20 \tag{20}$$

For each c, 100 random, f_c-centered frequency values were generated:

$$f_{r,c} = f_c + \text{rand}\left(-\frac{f_s}{N}, \frac{f_s}{N}\right), r = 1,...,100 \tag{21}$$

i.e. the frequencies $f_{r,c}$ vary randomly around f_c within one frequency bin of the Fourier analysis of size N to come. Those frequencies were used to generate equally many discrete pure tones as follows:

$$Y_{r,c} = \sin\left[2\pi\frac{f_{r,c}}{f_s}n + \text{rand}(0,2\pi)\right], n = 0,..., N \tag{22}$$

The Quadratic Fit (QF) and CSPE techniques were used to get frequency estimates for each of those pure tones. For each centre frequency, the RMS of the error of the 100 estimates was computed as

$$RMS_{c,\text{QF}} = \text{RMS}\left[f_{r,c} - \text{QF}(Y_{r,c})\right], \tag{23}$$

$$RMS_{c,\text{CSPE}} = \text{RMS}\left[f_{r,c} - \text{CSPE}(Y_{r,c})\right]. \tag{24}$$

Those error RMS values are presented graphically in Fig. 7.

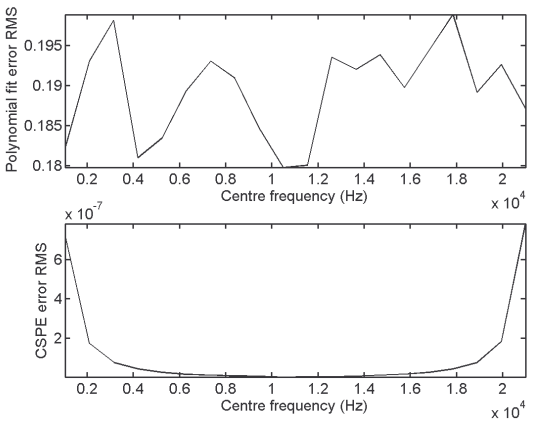

Figure 7: *Accuracy comparison of quadratic fit (top) and CSPE (bottom) frequency refinements*

The error RMS of the quadratic fit estimate exhibits little, irrelevant accuracy fluctuation across the frequency domain. On the other hand, it appears that the accuracy of the CSPE is highly dependent on frequency. At a quarter of the sampling frequency, the CSPE error RMS curve exhibits a minimum of 10^{-9}, which is 10^8 lesser than for the quadratic fit refinement. The CSPE exactitude falls off dramatically on either side of that frequency. Yet it was found to remain almost a hundred times lesser than the quadratic fit error RMS at the lower musical extreme of 20Hz.

2.4. Peak detection faithfulness

A major advantage of the proposed method, stemming from the firsthand accuracy of its partial frequency estimates, is its potential to work with extremely narrow peak-detection frequency bands. In [4], large such bands opposed little protection against the intrusion into the measurements of peaks belonging to other series such as parallel series, or series of longitudinal vibrations.

Fig. 8 shows a good example of how such partials can overshadow sought-for peaks, as well as the extent to which narrow peak-detection frequency bands reduce this risk. The analysed tone is a Steinway F2, with a fundamental frequency of 43.46Hz and $1.18\cdot10^{-4}$ inharmonicity value. Here, the frequency bands span 4 analysis bins across, which represents here one 16th of the fundamental frequency, and one 12th of the bandwidth used in the PFD method [4].

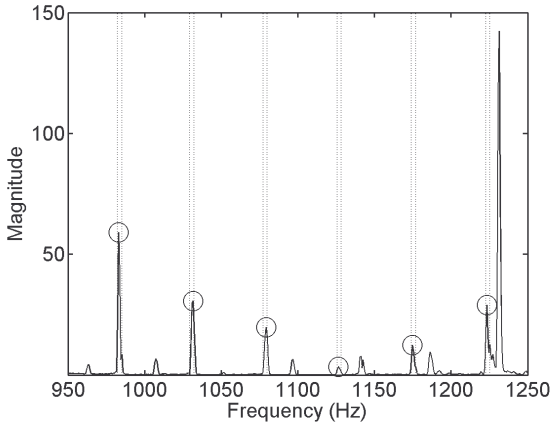

Figure 8: *Peak detection faithfulness of the MAT method*

3. COMPARISON

Comparison tests were run which confronted the MAT method with the Partial Frequency Deviation (PFD) method proposed by Rauhala *et al.* in [4].

Sixteen fortissimo tones of a Steinway grand piano, ranging from C#3 to C#7 (of respective equal temperament fundamental frequency 34.6 and 1,108.7 Hz) in steps of major third were obtained from [7]. Prior to analysis, the sound files were formatted

to mono-channel, 44.1 kHz sample rate. Then, in both methods' cases:

- The first 0.3 seconds of all tones were discarded, to minimise the presence of transient-related noise in the spectra.
- A window size of 2^{14} samples was adopted.
- The algorithms were arranged to use the above-described CSPE method for refinement of the partial frequency estimates.
- The algorithms' respective performances were individually timed from the time-domain windowing of the signals until the final estimates were obtained.
- To improve the relevance of the statistics, each tone was analysed four times, bringing the number of performances up to 64.

As mentioned earlier, the PFD method requires the number of partials to account for in the analysis to be specified. All the while, the number of existing partials varies widely across the range of subject tones. Fig. 9 exhibits the spectra of the two extreme tones C#3 and C#7.

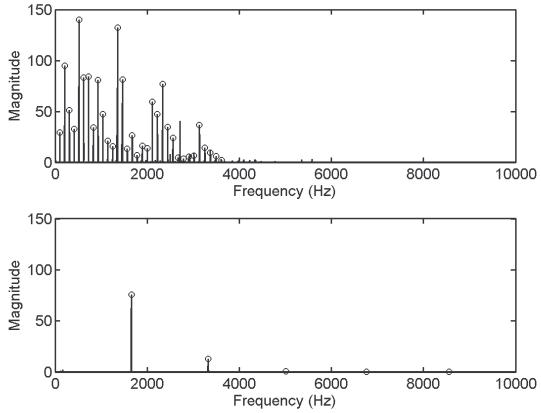

Figure 9: *Peak series of a Steinway's C#3 (top) and C#7 (bottom).*

The MAT method code was arranged to return the number of detected partials, to be used in the PFD in this regard. The β estimates for the 16 tones, as issued by either method, are sequenced in Fig. 10.

To the exception of the first and second last entries, the two sets are consistent throughout, which denotes the validity of either method. Also, the dramatic inharmonicity increase towards the upper end of the piano keyboard corroborates the measurements shown in [8].

To measure with more precision and certainty the accuracy of each method, the above-presented β estimates were used in turn to create inharmonic test tones using additive synthesis. The initial phase of each sinusoid was given a random radian value in the interval [0 2π]. Also, because the PFD method in its first step detects twice as many peaks as there are relevant partials, noise was added to the time-domain signal to provide the spectra with dummy peaks.

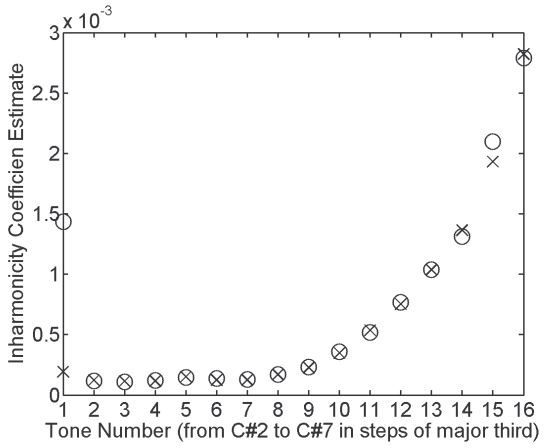

Figure 10: *Compared* β *estimates from the PFD (circles) and MAT (crosses) methods.*

The accuracy of the peak detection proposed by the MAT method is such that it was deemed unnecessary to compare the algorithmic estimates with estimates obtained by the means of visual identification of the relevant serial peaks. Until now, it was common for such means to be resorted to, as in [4] and [9].

The comparison tests were implemented in Matlab, and run on a laptop equipped with an Intel Core 2 Duo T8100 2.1GHz processor and 2GB RAM. The total runtime and RMS error of each method, for the 64 analyses, are presented in Table 1. Put in perspective with similar tests produced in [5], where the PFD is confronted with Galembo and Askenfelt's ICF method, the Median-Adjustive Trajectory method stands as the most efficient and accurate to date.

Table 1: *Comparison of PFD and MAT methods*

	PFD	MAT	PFD/MAT
Runtime (*s*)	2.48	0.719	3.4492
RMS_{error}	0.001067	0.00049268	2.1657

4. CONCLUSION

The leading methods for the automatic estimation of the IC and, in some cases, theoretical fundamental frequency of string instrumental tones were presented and discussed in chronological order. The proposed Median-Adjustive Trajectory (MAT) method, featuring a firsthand peak detection approach to the estimation of those theoretical values, was thereafter described and schematised. The Complex Spectral Phase Evolution [6] was also introduced as a beneficial tool for partial frequency estimation refinement. Comparison of the MAT method with the PFD exhibits unprecedented accuracy and computational efficiency.

In the future, the proposed method could be extended to account for several pseudo-harmonic series. In the tones produced by finely tuned courses of strings, the frequency difference between partials of corresponding indexes cannot be seen early in the series. When eventually the parallel series break apart, the method has presently no command as to which series the median-adjustive trajectory will follow, and will return IC and fundamental frequency estimates for one of the strings only. A significant improvement would therefore be for the MAT method to be able to return as many β and f_0 couples of estimates as there are strings involved in the production of the analysed tone.

5. REFERENCES

[1] H. Fletcher, E. D. Blackham and R. Stratton, "Quality of Piano Tones", Journal of the Acoustical Society of America, vol. 36, no. 6, pp. 749-761, 1962.

[2] A. S. Galembo and A. Askenfelt, "Measuring inharmonicity through pitch extraction", STL-QPSR, Vol. 35, No. 1, pp. 135-144, 1994.

[3] A. S. Galembo and A. Askenfelt, "Signal Representation and Estimation of the Spectral Parameters by Inharmonic Comb Filters with Application to the Piano", IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 2, pp. 197-203, March 1999

[4] J. Rauhala, H.-M. Lehtonen and V. Välimäki, "Fast automatic inharmonicity estimation algorithm", Journal of the Acoustical Society of America, vol. 121, no. 5, pp. EL184-EL189, 2007.

[5] http://www.geocities.com/CapeCanaveral/Lab/8779/chc5.html (Latest date visited: 07-04-2009)

[6] K. M. Short and R. A. Garcia, "Signal Analysis using the Complex Spectral Phase Evolution (CSPE) Method", Audio Engineering Society 120th Convention, May 2006, Paris, France. Paper Number: 6645.

[7] http://theremin.music.uiowa.edu/MIS.piano.html (Latest date visited: 07-04-2009).

[8] A. Askenfelt and A. S. Galembo, "Study of the Spectral Inharmonicity of Musical Sound by the Algorithms of Pitch Extraction", Acoustical Physics, Vol. 46, No. 2, pp. 121-132, 2000.

[9] J. Lattard, "Influence of inharmonicity on the tuning of a piano – Measurements and mathematical simulation", J. Acoust. Soc. Am. 94 (1), pp. 46-53, July 1993.