Accurate temporal resolution of harmonic content in both amplitude and phase

Damien Holloway[1]

(Received 16 December 2013; revised 13 June 2014)

Abstract

Frequency content of a periodic signal is easily computed using fast Fourier transforms (FFT). While the magnitude is well predicted, phase information is usually meaningless and temporal changes are difficult to resolve accurately over short time periods, even with techniques such as discrete time Fourier transforms or wavelets. These problems arise in the analysis of musical sounds and should be solvable since a well trained human ear can detect subtle and rapid changes to timbre and pitch that occur with expressiveness, techniques such as 'vibrato', and lack of tone control exhibited by music students. Starting with the premise that a musical sound is truly periodic, we obtain much more accurate information from a Fourier *series* than from a Fourier *transform*. In this article the Fourier series of a periodic signal is evaluated using a least

squares fit, as was done before the popularisation of the FFT algorithm, but the difference is that the frequency is precisely defined before fitting the coefficients, which succeeds with as few as three or four cycles. The proposed technique achieves the above objectives and opens up the possibility of exploring the role of phase in the quantification of musical sound, a critical component that is traditionally ignored.

Contents

1 Introduction

The objectives of this study are to accurately quantify musical timbre in terms of both amplitude and phase of the harmonic content; to detect subtle but rapid changes in loudness, pitch and timbre that may arise from the use of techniques such as 'vibrato', expressiveness or simply lack of technical control of the sound; and to resolve and quantify rapid sequences of notes.

This problem originated from the idea of quantifying an 'ideal' sound produced by professional musicians for use as a pedagogical tool. Gaskell [3] studied the sound produced by flautists, with subjects divided into three categories

(professional or semi-professional, advanced student and beginner student) with the aim of identifying features that consistently differ between the sounds produced by professional musicians and those produced by students at various levels. In this study the signals were processed using fast Fourier transforms (FFT) and timbre was quantified purely in terms of energy content in the different harmonics.

Gaskell succeeded in identifying clear difference between his subject groups, but several of the objectives above were not achievable using conventional tools such as FFT, wavelets or discrete-time Fourier transforms (DTFT) for reasons described in Section 2. A method proposed in Section 3 overcomes these limitations. We see in Section 4 that harmonic amplitudes can vary as much or more due to player, pitch and loudness as they can from one musical instrument to another. However, a trained musician is easily capable of identifying the instrument being played, as well as the quality of timbre, pitch, vibrato, etc. Therefore I hypothesise that the phase of the harmonic content of a musical note is at least as important as amplitude, and that analysis techniques exist that can quantify it accurately.

2   The problem (an example)

Consider the signal $y(t)$ in Figure 1(a) (which we revisit in Section 3), containing 4096 samples of the signal at the standard audio sampling rate of 44.1 kHz[1]. The pitch is C4 ('middle C'), nominally 262 Hz (based on equal tempered A440 tuning). The total duration is approximately 93 ms and contains 24 highly regular 'cycles' (as would be expected for a musical note from a defined pitch instrument). It is typical of the signals to be analysed (although significantly longer than often desired).

---

[1]This is effectively the instantaneous sound pressure at the microphone. Units are arbitrary since the signal $y(t)$ (which is in the range 0–1) depends on the microphone sensitivity and amplifier gain.
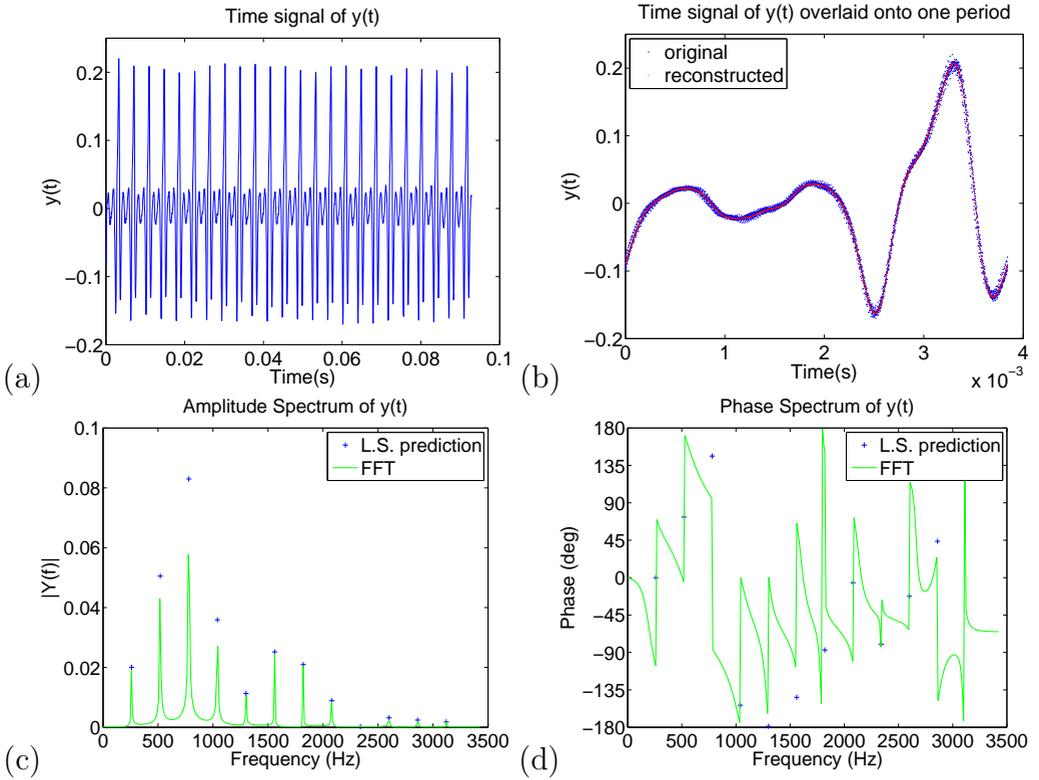
Figure 1: (a) C4 (nom. 262 Hz) played *mezzo forte* by professional flautist; (b) 24 cycles of (a) overlaid onto one period (blue) with a perfectly periodic 'reconstructed' signal (red) added; (c) amplitude spectrum of reconstructed signal in (b), the continuous line is the Fourier transform of $y(t)$ and the points are the coefficients of the Fourier series; (d) phase spectrum of reconstructed signal in (b).

Figure 1(b) shows an equivalent signal reconstructed from Figure 1(a) using the procedure that is outlined in Section 3 (hence it is perfectly periodic). The Fourier *series* coefficients for this signal are the blue + in Figures 1(c) and (d), while the Fourier *transform* (specifically the FFT) evaluated over the 4096 samples are the green lines.

The Fourier series and Fourier transform should of course be identical if the sample is of infinite length (since the reconstructed signal is perfectly periodic), or *exactly* an integer number of periods. However, in practice this will not be the case because

- real signals are always of finite length (and usually quite short);

- the sample period does not in general divide evenly into the signal period; and

- real signals will change with time, so even if they are long we must analyse short segments to obtain the desired information.

The FFT gives a reasonable approximation to the amplitude spectrum for the higher harmonics; however

- the frequencies computed by the FFT are limited to discrete values that will not in general coincide with the true frequency, in this example the resolution is about $\pm 2\%$, at least an order of magnitude short of the accuracy useful for the musical applications mentioned above;

- the amplitudes of the first four harmonics (the most critical ones) are in error by up to 30%;

- there is no meaningful information at all in the phase spectrum (and a real signal is much worse), in particular the are abrupt phase shifts at the component frequencies and the true phase value could be any point on a near-vertical line.

Similar problems exist with other traditional techniques. They always fall well short of achieving the objectives desired by this study.

3   Procedure

Historically, Fourier series were often computed by fitting the sine and cosine coefficients by least squares [1] prior to the rapid growth in popularity of the FFT algorithm in the decades immediately following publication of Cooley and Tukey's seminal paper [2]. A return to the older least squares method proves to be the key to exploiting the unique properties of the present problem to achieve the objectives stated in Section 1. However, I emphasise that the modern techniques will still be preferred for most other applications, especially in identifying weak signals in noisy or random data, and I do not wish to undervalue their importance.

In the present work the least squares approach described by Chatfield [1] is improved upon by exploiting the knowledge that musical sounds from a definite pitch sustaining instrument very closely approximate being truly periodic over short periods of time. Therefore, the method is extended by accurately defining the fundamental period before fitting the coefficients, and by not constraining the period to bear any relation to the sample rate. The accurate determination of the fundamental frequency is absolutely critical to the success of the method, otherwise results are little, if any, better than FFT. Accurate determination of the fundamental frequency also means that the coefficients need only be fitted for arbitrarily few (or many) integer multiples of the fundamental frequency without any loss of accuracy, so only a small number of coefficients generally need to be computed to produce useful, accurate and meaningful results.

To accurately define the frequency the signal is smoothed, zero crossings are identified (sometimes a trigger other than zero may be more effective) and a linear fit is applied to the time versus crossing count relation. Using only every second zero crossing has two advantages: the slope of the linear fit line is the period (or an integer multiple), and it avoids inaccuracy where the upward and downward zero crossings are not equally spaced and an odd number of points is identified. The computed period is then compared with

any good estimate of the expected period (which usually need only be $\pm 25\%$) to correct for the case where the predicted frequency is an integer multiple of the true fundamental frequency (i.e., where there are strong harmonics, hence more than two zero crossings per period—Figure 1(b) is such a case).

The accurate period identification method may fail if the signal is so noisy that the number of zero crossings differs from one period to another, or if the sound is unmusical (e.g., not of definite pitch). For example, Figure 5, discussed more fully below, shows some deviations from the expected regular stepwise monotonically increasing frequencies of a typical musical scale. This is most likely due to 'key slap' or other noise associated with the flute's mechanism. Such frequency misdiagnosis is minimised by experimenting with the amount and type of smoothing, the trigger threshold, and the length of the analysis window. For musical sounds the frequency misdiagnosis will be confined to individual data points, which will usually be identifiable as outliers.

Once the period T is accurately known, coefficients C_i are fitted in the least squares sense to

$$y_j = \sum_{k=1}^{M} \left[C_k \cos(k\omega t_j) + C_{M+k} \sin(k\omega t_j) \right], \qquad (1)$$

where y_j is the jth sample taken at corresponding time t_j and $\omega = 2\pi/T$ is the fundamental angular frequency. The number of computed harmonics M is chosen arbitrarily (subject to having sufficient samples, which in practice will never be problematic since typically only the first few harmonics are desired) and does not affect the values of the fitted coefficients. The coefficient fitting is performed almost trivially using the Matlab code

```
k = (1:M);
C = lsqlin([cos(2*pi*k*t/T)' sin(2*pi*k*t/T)'],y);
```

where t and y contain all the t_j and y_j values. The code puts C_k and C_{M+k} end to end into a single vector $C(1:2*M)$.

To extract the phase, rewrite (1) as

$$y_j = \sum_{k=1}^{M} A_k \cos\left[k(\omega t_j - \phi_k) - \phi_0\right],$$ (2)

where ϕ_0 is chosen such that $\phi_1 = 0$, thus the phase ϕ_k of the harmonics are expressed relative to the phase of the fundamental ϕ_0. The following Matlab code evaluates the phase of each harmonic:

```
Z     = C(k) - i*C(M+k);
phi0  = imag(log(Z(1)));
Z0    = Z.*exp(-i*phi0*k);
phase = imag(log(Z0));
```

4   Applications

Figures 2, 3 and 4 show the technique described in Section 3 applied to four different versions of the musical note G4 (nominally $392\,\text{Hz}$) using an analysis window of $15\,\text{ms}$ (about 4.9 cycles). The top two examples in each of these figures illustrate the difference between a professional player and beginner student. In particular, they show the consistency and degree of control with which a professional can sustain a sound over a duration of $2\,\text{s}$.

Figure 3 clearly shows the professional flautist's sound to have relatively much stronger, higher harmonics. The instrument is possibly slightly 'overblown', as evident from the high second harmonic sounding an octave above the fundamental. However, the ear still picks up the fundamental frequency in spite of its low energy but perceives it as having a richer and more carrying sound. This much could have been deduced using FFTs, but it is in Figures 2 and 4 where only the present method is able to show accurately the rapid variation of the signals in time, and the lack of control in the student's sound compared to the professional flautist's is clearly evident. Figures 2 shows how
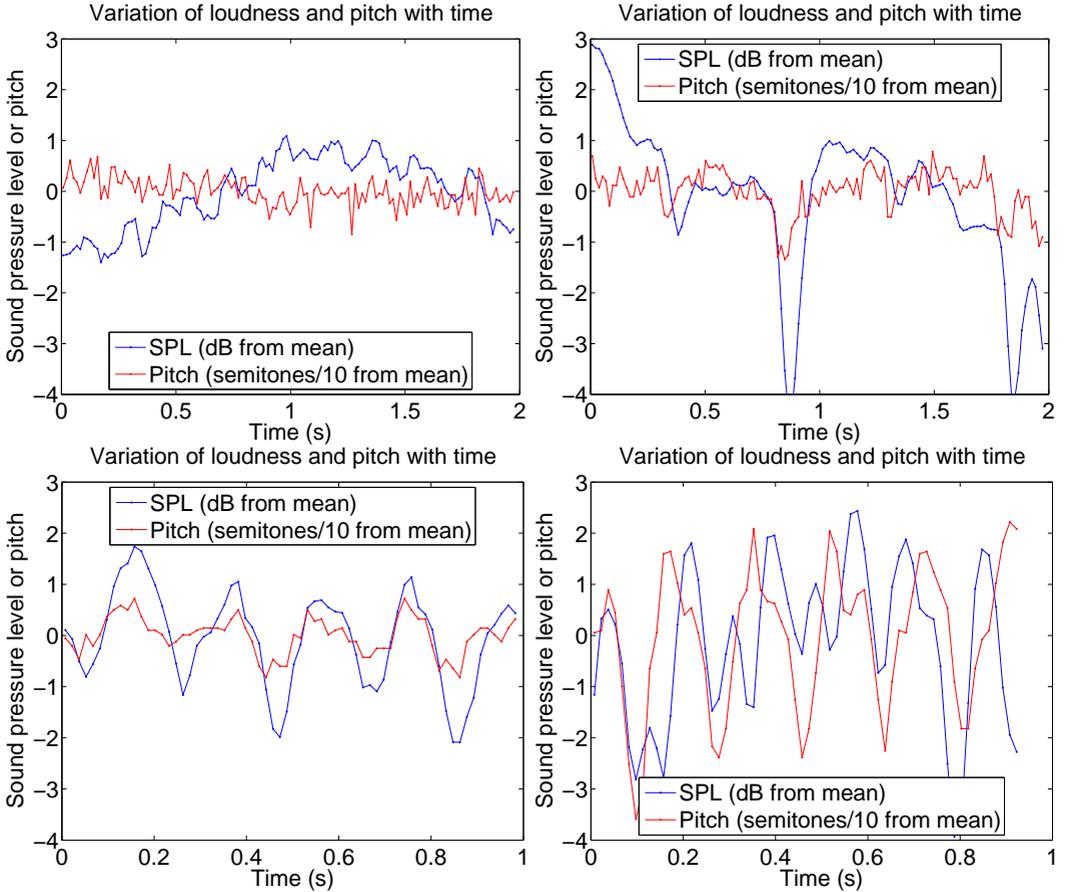
Figure 2: Time variation of amplitude and frequency for the note G4 (392 Hz) (analysis window 15 ms). Top left: professional flautist, no vibrato; right: beginner flute student, no vibrato. Bottom left: semi-professional flautist with vibrato; right: semi-professional violist with vibrato.
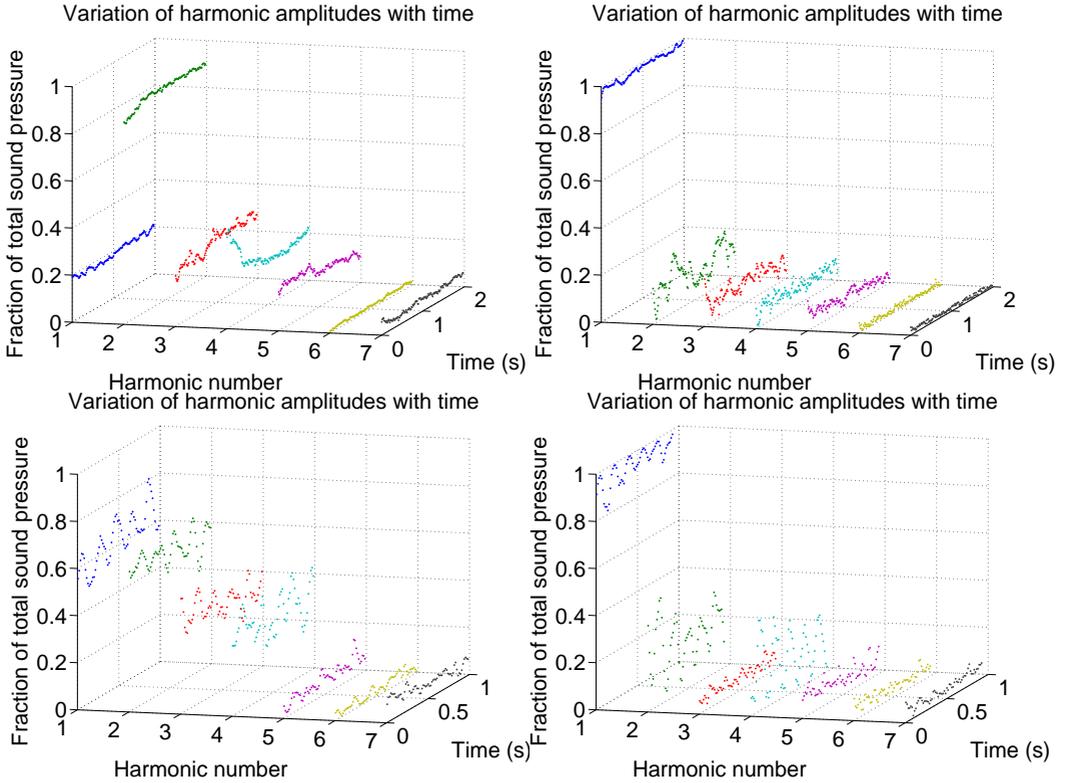
Figure 3: Time variation of relative amplitude of harmonics (normalised on the sum of squares of harmonic components) for the note G4 ($392\,\text{Hz}$) (analysis window $15\,\text{ms}$). Top left: professional flautist, no vibrato; right: beginner flute student, no vibrato. Bottom left: semi-professional flautist with vibrato; right: semi-professional violist with vibrato.
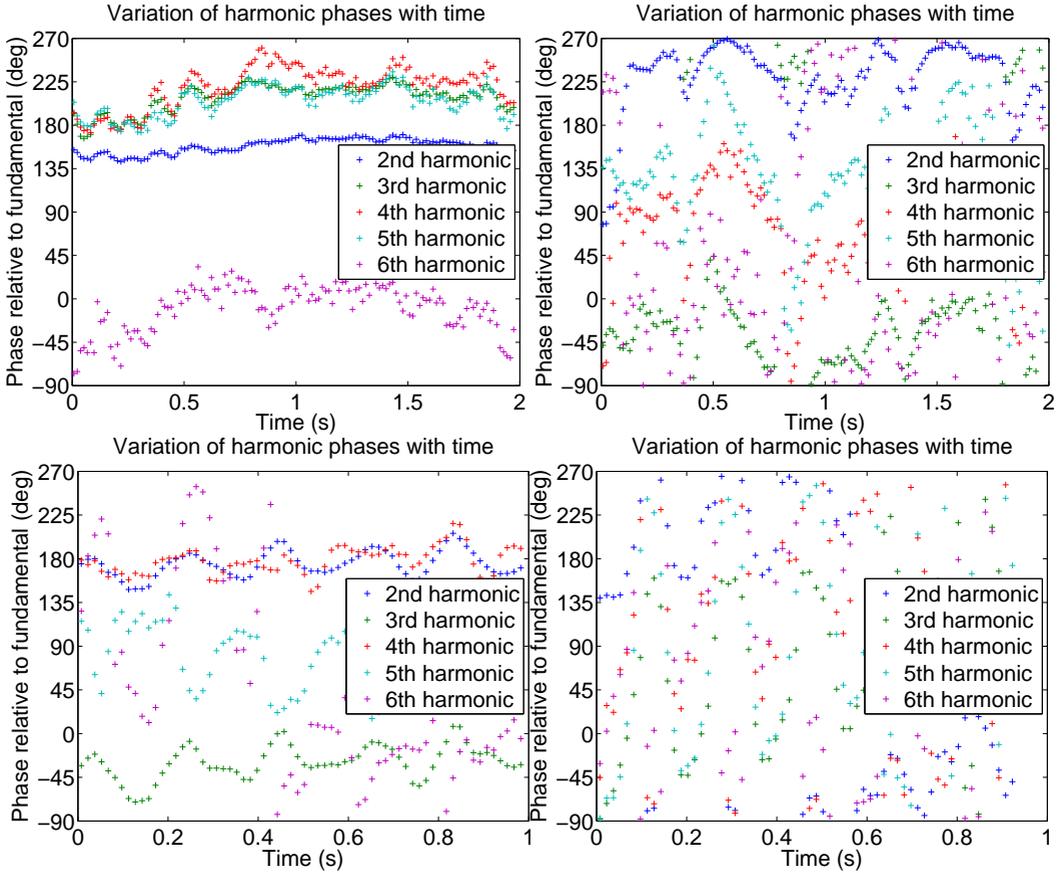
Figure 4: Time variation of relative phase of harmonics for the note G4 (392 Hz) (analysis window 15 ms). Top left: professional flautist, no vibrato; right: beginner flute student, no vibrato. Bottom left: semi-professional flautist with vibrato; right: semi-professional violist with vibrato.

the professional player is able to maintain a constant pitch as the sound gets louder, while student's pitch is clearly correlated with the loudness, showing how it is affected by poor breath control. Figure 4 shows how the student's poor sound control is illustrated by substantial random variation of the phase spectrum with time, which is not nearly so clearly evident in the amplitude spectrum (Figure 3).

The third and fourth examples in each of Figures 2, 3 and 4 illustrate the use of an expressive technique called 'vibrato', in which the musician modulates the amplitude and/or frequency of the note at about 5–6 Hz, producing a 'wobbly' sound—typically the frequency (pitch) modulation is more obvious to the listener than the amplitude modulation. These figures compare the vibrato on two different instruments, a flute and viola. Again, FFT would not be able to resolve the frequency accurately enough to meaningfully quantify the vibrato given the short analysis windows required by the closeness of the modulation and carrier frequencies (less than two orders of magnitude difference).

Figure 2 shows some interesting differences between the two instruments. Apart from the slightly different modulation frequencies (about 5 Hz for the flute and 6 Hz for the viola) we see that the viola has relatively more pitch variation and the flute relatively more amplitude modulation. This is not surprising considering the method of vibrato production—breath control for the flute, and left (fingering) hand oscillations for the viola. We also see sharper troughs for the flute and sharper peaks for the viola.

A peripheral observation from Figure 3 is that there is nothing at all about the amplitude spectrum for the viola that distinguishes it from that of the three flute samples, primarily because of the significant variations between the latter. I hypothesise that there is key information in the phase spectrum that can be used to define a characteristic sound for, or to distinguish between, various musical instruments. The phase spectra (Figure 4) shows the flute to be considerably more consistent than the viola, although careful observation shows that the second, third and fourth harmonics for the viola are better
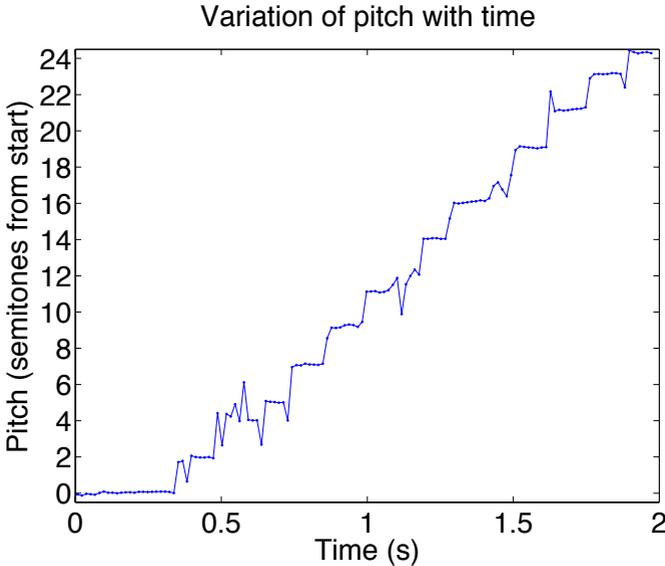
Figure 5: E♭ major scale played on flute, semiquavers at approximately 132 beats per minute (analysis window 15 ms).

correlated with each other than with the fundamental—this could perhaps indicate that the overtones of the viola's sound are not perfectly harmonic, and could be one such characteristic point of difference between these instruments.

As a final illustration, Figure 5 shows how pitch is detected for a very fast scale played on the flute. The frequency range is 311–1245 Hz, and there are approximately nine changes of note per second. At the lower frequencies of this example the analysis window contains only 4.7 vibration cycles. Some challenges with pitch detection are evident (possibly noise from key slap) but mostly the pitches are clearly identified with sufficient accuracy to give helpful feedback to the player about tuning etc.

5   Conclusions

The proposed method

- accurately defines both phase and amplitude of harmonics that make up musical notes;

- works on very short analysis windows containing as few as four periodic cycles of data, and of arbitrary length (e.g., not constrained to 2^N samples), so quantifies variations of timbre and pitch with sufficient accuracy to study vibrato, consistency of tone production, etc;

- is robust to noise, provided the correct fundamental frequency is identified; and

- is performed rapidly with no loss of accuracy if arbitrarily few coefficients are computed.

References

[1] C. Chatfield. *The Analysis of Time Series: Theory and Practice.* Chapman and Hall, 1975. http://www.crcpress.com/product/isbn/9781584883173 C240

[2] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.*, 19:297–301, 1965. doi:10.1090/S0025-5718-1965-0178586-1 C240

[3] A. Gaskell. *Timbre of the Flute.* Honours thesis, University of Tasmania, 2012. C236

Author address

1. **Damien Holloway**, School of Engineering, University of Tasmania, Hobart, Australia.
   mailto:Damien.Holloway@utas.edu.au