Methods for Real Time Harmonic Excitation of Acoustic Signals

Sean Enderby
Department of Digital Media Technology
Birmingham City University
Birmingham, UK
sean.enderby@mail.bcu.ac.uk

Zlatko Baracskai
Department of Digital Media Technology
Birmingham City University
Birmingham, UK
zlatko.baracskai@bcu.ac.uk

Cham Athwal
Department of Digital Media Technology
Birmingham City University
Birmingham, UK
cham.athwal@bcu.ac.uk

ABSTRACT

In this paper three methods for the introduction of new harmonic content to an acoustic signal are assessed. Each method extracts the amplitude envelope of the fundamental frequency in a signal and applies it to a newly generated harmonic. In one method this is achieved in the frequency domain through use of the short time Fourier transform. The other two methods process audio in the time domain using either instantaneous amplitude and phase measurements or single side band automodulation.

The results from a set of preliminary listening tests are discussed and compared against objective measurements based on psychoacoustic models. It is suggested that frequency domain processing is too inaccurate where low latency is required and a time domain approach is preferential. The two time domain approaches show similar levels of accuracy, however it is considered that extracting the amplitude envelope of harmonics other than the fundamental could increase accuracy. It is noted that the instantaneous amplitude and phase method provides more flexibility in order to achieve this.

1. INTRODUCTION

Harmonic excitation involves the introduction of new harmonic content to an audio signal. This can be used to increase the perceived quality of a piece of audio. They can also be used to restore old recordings where the recording medium may have deteriorated or was not able to capture high frequency signals. [1].

Pitched sounds arise from resonant systems that produce a harmonic spectrum with complex evolution and envelope. In sound synthesis the detailed harmonic spectrum can be produced by synthesising the harmonics separately or by using higher order nonlinearities to process the fundamental. Chebyshev polynomials allow synthesising any proportion of harmonics by applying a transfer function to a full amplitude sine wave [2]. In the context of real-time processing this technique is ideal for its zero latency, however it is not suitable for exciting harmonics of acoustic sounds as the non-unit amplitude produces a varying mixture of the desired and the lower order harmonics as shown in Figure 1.

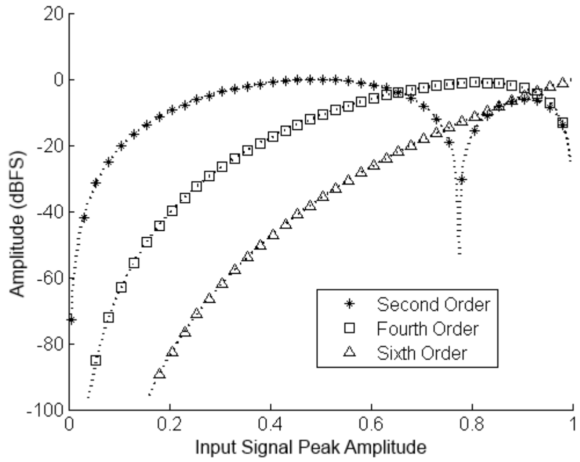


Figure 1: The levels of unwanted harmonics introduced when using Chebyshev polynomials to synthesise the 6^{th} order harmonic of sine waves with various peak amplitudes.

A simple method to introduce new harmonic content to a signal is through the application of a static nonlinear system. The value of each input sample is mapped, using some nonlinear function, to a new output value. Examples of these types of systems are given in [3]. For a sinusoidal input each of these systems will introduce a characteristic set of harmonics. The order and amplitude of these harmonics is defined by the nonlinear function used to process the signal.

The downside of these methods is that more than one harmonic is introduced to the signal. This is not desirable in situations where only a specific harmonic is required. This paper deals with methods by which single harmonics can be introduced to a signal with as little extraneous frequency content as possible.

Three such methods are be described in Section 2. The methods are then assessed on their latency and their ability to introduce specific harmonics into a pitched signal. The later is done through use of perceptual listening tests, as described in Section 3, and a perceptual distortion metric. The results of these test are then compared in order to determine which harmonic excitation method is most suitable for real time applications.

2. METHODS

The process by which individual harmonics will be introduced to a signal can be broken down into four stages.

- Calculate the fundamental frequency of the input signal.
- Extract the amplitude envelope of the fundamental in the input signal.
- Synthesise a new signal with the frequency of the desired harmonic and the amplitude envelope of the fundamental.
- Scale the synthesised harmonic and mix it back into the original input signal.

Three different methods for the synthesis of new harmonics are assessed in this paper. Each has been named according to the mathematical transforms on which they are based. The methods are based on:

- The Short Time Fourier Transform (STFT).
- Instantaneous amplitude and phase measurements (IAP).
- Single side band automodulation (SSB).

2.1 Short Time Fourier Transform

The STFT can be used to analyse the frequency content of sequential time frames of the input signal. This frame based processing introduces an inherent delay into the system. The challenge with generating harmonics using the STFT is to keep the frame length short enough to keep latency from being perceptible whilst still maintaining enough frequency resolution to accurately synthesise the new harmonic [4].

The acceptable levels of latency in live music situations are discussed in [5]. It is suggested that the acceptable level varies from 1.4ms to 42ms depending on the instrument and monitoring system. The frame length used needs to be kept short enough such that the latency of the entire system does not exceed these limits.

The phase vocoder technique can be used to scale the frequency of the fundamental to the desired harmonic. This is achieved by zeroing the bins for frequencies greater than the fundamental in the DFT data for each frame. The new frequency domain data can then be pitch shifted via a phase vocoder to the frequency of the desired harmonic. It was found that when using short frame lengths pitch shifting of several octaves is not achievable due to the poor frequency resolution. This means that higher order harmonics could not be generated.

A simpler method is to calculate the amplitude of the fundamental frequency in each time frame. These values can then be linearly interpolated in order to approximate the amplitude envelope of the fundamental. This amplitude can then be applied during the synthesis of a harmonic with the desired frequency. This allows for a better accuracy for short STFT frame lengths but it does rely on knowing the frequency of the fundamental precicely. The samples used in the listening tests discussed in Section 3 were created using this method.

2.2 Instantaneous Amplitude and Phase

In this technique the fundamental of the input signal is isolated using a low pass filter. The amplitude envelope of the fundamental can then be found using measurements of instantaneous amplitude.

The principles of instantaneous amplitude and phase are discussed in [6]. To take these measurements the filtered signal must be converted to its analytic form. In a true analytic signal the real part will be the original input signal and the imaginary part its Hilbert transform. Calculating a true analytic signal is not possible without introducing delay to the system. The more delay introduced the more accurate the analytic signal will be.

A low latency alternative is to use a pair of all pass filters whose phase responses differ by $\frac{\pi}{2}$ radians across a large proportion of the audible bandwidth. An example of such a pair of filters is given in [7]. Simple calculations can be applied to the output of these filters to produce two new signals. One is a signal which represents the amplitude envelope of the fundamental ($a[t]$) and the other represents the phase of the fundamental ($\phi[t]$). Due to the filters used the phase measurements will not represent the phase of the fundamental in the the original signal. The change in phase measurement with time however, will be consistent with the frequency of the fundamental.

Once the measurements of amplitude and phase have been taken, the new harmonic can be synthesised as done in [8]. Equation 1 shows the calculation for synthesising the n^{th} harmonic ($h[t]$).

$$h[t] = a[t]\cos(n\phi[t]) \tag{1}$$

The accuracy of this method is largely dependant on the order of the low pass filter used to isolate the fundamental. The higher the order of the filter the better the isolation of the fundamental. This leads to less extraneous frequencies being introduced in the synthesis of the new harmonic.

2.3 Single Side Band Automodulation

With this technique, as with the IAP technique, the fundamental is isolated using a low pass filter and then further filtering is applied in order to create an analytic signal. This analytic signal can then be raised to a power in order to scale its pitch to that of the desired harmonic. The underlying principle is that of de Moirve's formula (Equation 2) and single side band modulation.

$$(\cos(x) + i\sin(x))^n = \cos(nx) + i\sin(nx) \tag{2}$$

When signals are multiplied together (or multiplied by them selves in this case) an upper and lower sideband are created. These sidebands are comprised of various intermodulation frequencies which are the sums and differences of the frequencies in the input signals. If the two signals are converted to their analytic representations first only a single sideband will be created. This is the concept of single side band modulation as discussed in [9]. For the generation of harmonics the analytic signal of the fundamental is multiplied with itself rather than a modulator wave. This gives rise to the idea of single side band automodulation.

Where $z[t]$ represents the analytic signal for the fundamental the n^{th} harmonic can be calculated using equation 3.

$$h[t] = \mathrm{Re}(z[t]^n) \tag{3}$$

As with the IAP technique the accuracy of this process relies on the fundamental being well isolated. The better isolated the fundamental the less unwanted intermodulation frequencies will be present in the synthesised harmonic.

An advantage of this technique is that the amplitude envelope of the fundamental does not have to be measured, as such it requires the least computation of the discussed methods. This is beneficial for real time processing but it does cause inaccuracies in the amplitude envelope of the generated harmonic. If the n^{th} harmonic were generated, its amplitude envelope would be the amplitude envelope of the fundamental raised to the power n.

3. LISTENING TESTS

A preliminary series of subjective listening tests were undertaken in order to assess the accuracy of each of the described methods. The assessment criteria for the listening tests were based on the following statement. "If some harmonic content is removed from an audio signal and then reintroduced through harmonic excitation. The newly produced signal should sound the same as the original signal". This is similar to the method by which the quality of perceptual coding algorithms is assessed. This allows a listening test methodology similar to MUSHRA [10] to be used effectively.

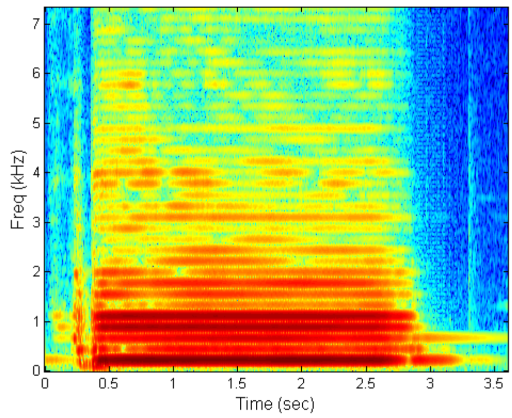
To create the stimuli for the listening test four different audio samples were each processed in nine different ways. The four unprocessed samples were:

- A bowed cello sample.
- A clarinet sample.
- A synthesised harmonic sound.
- A piano sample.

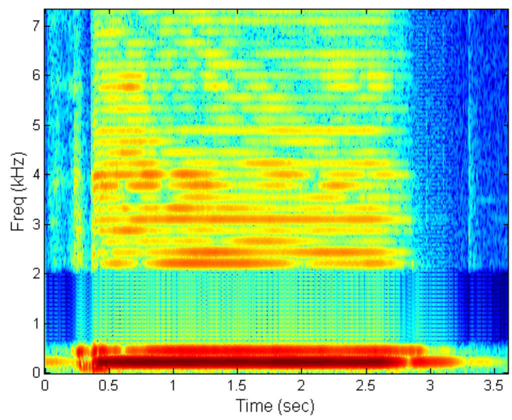
Each of the samples were of the instrument playing a single sustained note. The synthesised sample has very little energy at frequencies that are not its harmonics. This should make it easier to excite harmonics in as the fundamental can be isolated more easily. Owing to the acoustic nature of the other samples they have more energy at these non harmonic frequencies.

In order to reduce the number of variables each of the unprocessed samples was analysed prior to the creation of the test stimuli. The fundamental frequency of each sample was measured along with the amplitudes of the third through ninth harmonics. This information was then used in the reconstruction of the signal. This allowed for any inaccuracies which may be involved with real time calculation of the fundamental frequency or amplitudes of harmonics to be mitigated. Allowing the accuracy of the harmonic generation algorithms to be assessed more thoroughly.

For each sample the third through ninth harmonics were filtered out as shown by the spectrograms in Figure 2. This was in order to cause significant degradation in the quality of the sample such that the difference is plainly audible to the majority of listeners. The second harmonic was left in the signal in order to pose a challenge to the IAP and SSB techniques. As mentioned previously the accuracy of these methods is dependant on how well the fundamental is isolated. Retaining the second harmonic allows for the effects of filter order on the accuracy of the technique to be assessed.



(a) Original Signal

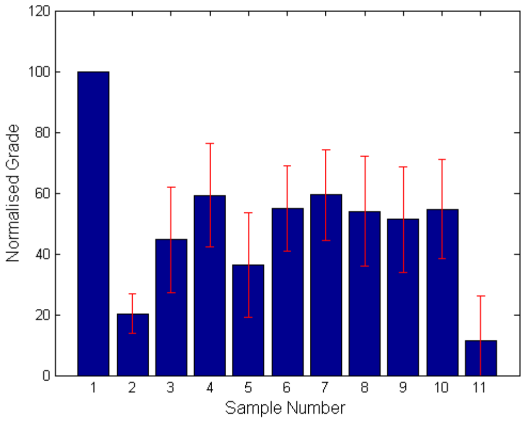


(b) Signal with Harmonics Removed

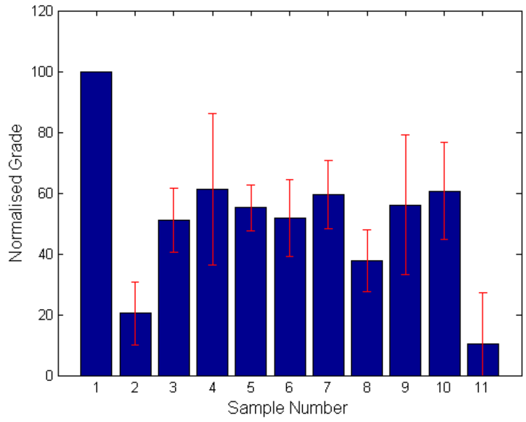
Figure 2: Spectrograms showing the frequency content of the cello sample before and after the harmonics were removed.

The filtered signal was then processed using the techniques discussed in Section 2. Each technique was used to create three stimuli, each with different parameters. For the STFT method, frame lengths of 50, 100 and 500 samples were used. For the IAP and SSB stimuli FIR filters with kernel lengths of 50, 100 and 500 samples were used to isolate the fundamental.

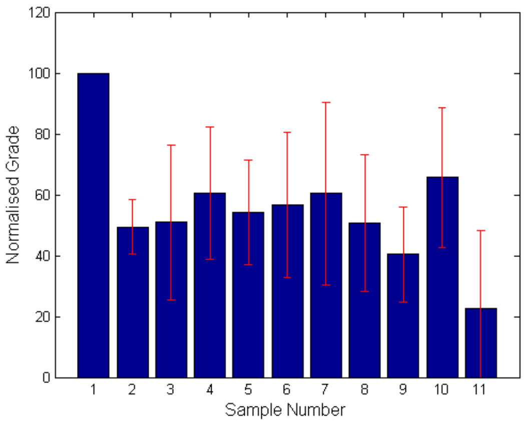
In line with the ITU recommendations [10] test subjects were presented with all the processed versions of a particular sample at once along with a reference sample (the unprocessed sample). Subjects could listen to the samples in any order and as many times as they required. Subjects



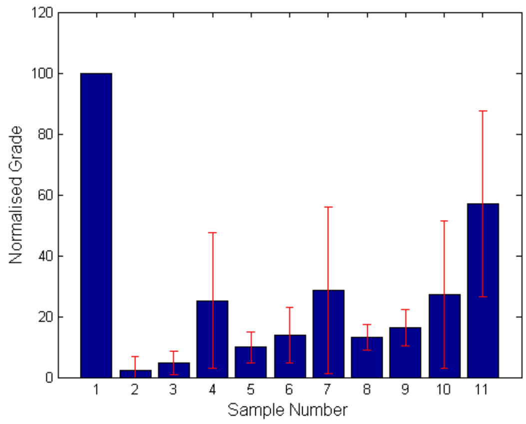
(a) Cello Sample Grades



(b) Clarinet Sample Grades



(c) Synthesised Sample Grades



(d) Piano Sample Grades

Figure 3: Mean grades and confidence intervals for each of the stimuli.

were asked to rate how well each processed samples recreated the reference sample on a scale from 0 to 100. The scale is shown in Figure 4.

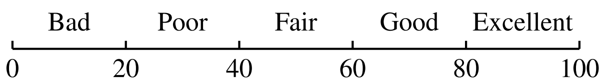


Figure 4: Listening Test Grading Scale

Among the samples to be graded are a hidden reference and anchor. The hidden reference is the same as the reference sample so should be given a score of 100. The anchor is the sample with the third through ninth harmonics removed. As no attempt was made to reintroduce the harmonics, this stimuli should be graded worse that the stimuli which have undergone harmonic excitation.

4. RESULTS

4.1 Listening Test Results

Preliminary testing has been undertaken with six test subjects. While this not a sufficient amount to provide confident assessments for each of the processing algorithms, it was sufficient to find basic patterns in the accuracy achieved by using a different method or changing the parameters of the method.

The grades given by each test subject were normalised to the range of 0 to 100. The mean grade given for each stimulus was then calculated. As suggested in the ITU recommendations [10] a 95% confidence interval was also calculated for each stimulus.

Figure 3 shows the results obtained from this preliminary testing. Each separate graph relates to a particular reference sample. The sample numbers relate to different processing algorithms as follows:

1. The hidden reference sample.
2. STFT reconstructed sample with a frame length of 50 samples.
3. STFT reconstructed sample with a frame length of 100 samples.
4. STFT reconstructed sample with a frame length of 500 samples.
5. SSB reconstructed sample using a filter kernel length of 50 samples.

6. SSB reconstructed sample using a filter kernel length of 100 samples.

7. SSB reconstructed sample using a filter kernel length of 500 samples.

8. IAP reconstructed sample using a filter kernel length of 50 samples.

9. IAP reconstructed sample using a filter kernel length of 100 samples.

10. IAP reconstructed sample using a filter kernel length of 500 samples.

11. The hidden anchor sample.

The error bars on each bar in the graphs show the 95% confidence interval for that stimulus.

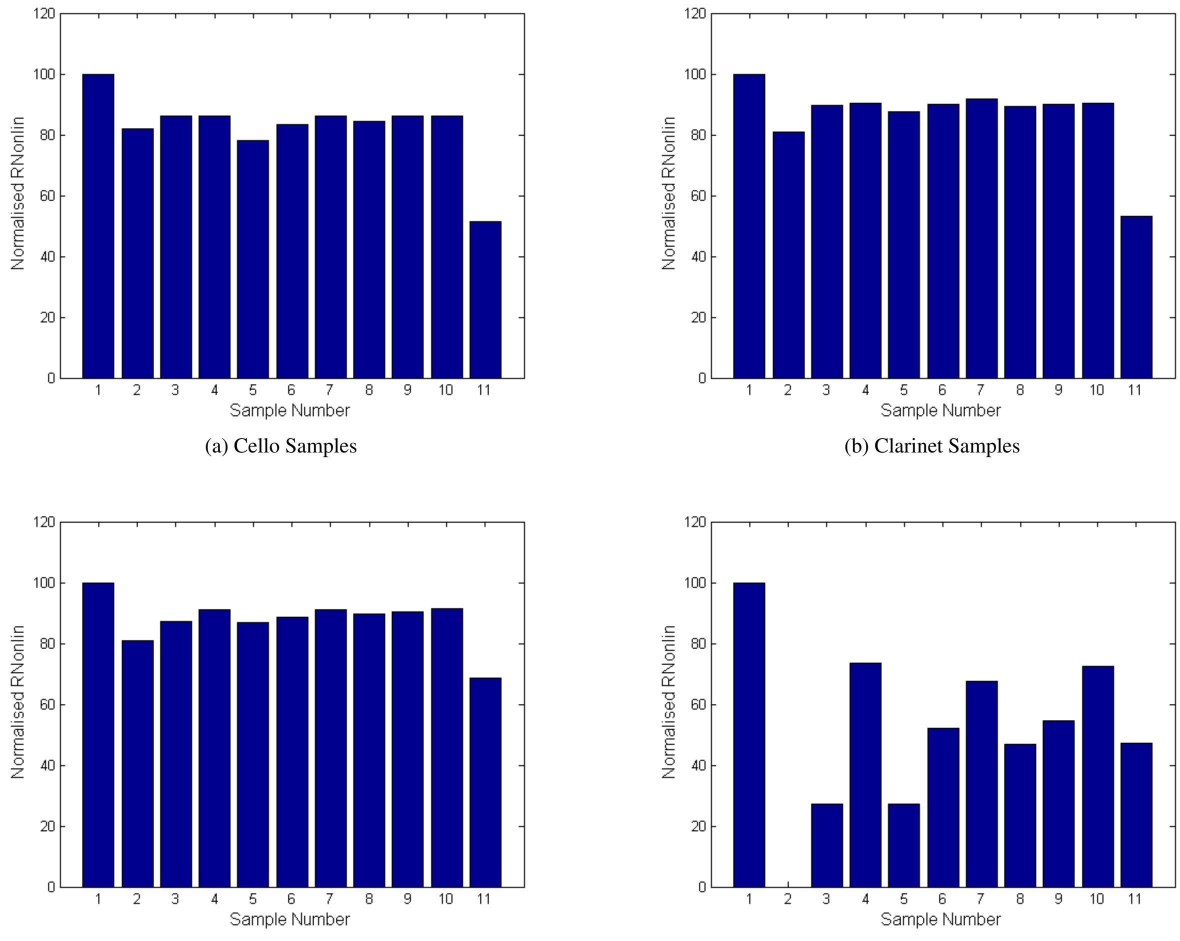
It is immediately apparent that the confidence intervals are fairly large. For most of the stimuli this can be attributed to only having a small cohort of test subjects.

Across the three acoustically recorded samples (Cello, Clarinet and Piano) there is an increase in the perceived accuracy of the algorithms as the frame or filter kernel length is increased as seen in Figures 3a, b and d. The lowest grades in each of these are given to the STFT processing with the shortest window length. The IAP and SSB techniques show greater accuracy while introducing less latency. For the electronically synthesised sample however the different processing algorithms are all given similar grades but with a wider variance in grades between different test subjects.

This could be attributed to the synthetic nature of the sample. There is very little energy in the sample at frequencies that are not harmonics. This makes it easy to isolate the fundamental and generate accurate new harmonics. Because of this even the processed samples which used short filter of frame lengths will be accurate. As all the processed samples sound fairly similar it is then difficult for the test subject to determine where on the scale they should be placed.

In the acoustic signals there is much more energy in frequencies which are not harmonics. This makes it more difficult to generate accurate harmonics so the differences between stimuli with different frame or filter lengths are more perceptible. As the subject is given a larger range of accuracies to assess it is easier for them to place them on the scale in a consistent manner.

The piano sample used is of special interest as its fun-



(a) Cello Samples

(b) Clarinet Samples

(c) Synthesised Samples

(d) Piano Samples

Figure 5: R_{nonlin} measurements for each of the listening test stimuli.

damental frequency is heavily damped. This led to the second harmonic in the signal being the most prominent. As the processing used relies on the amplitude envelope of the fundamental to generate new harmonics this lead to accuracy problems. Figure 3d shows that the reconstruction of the piano sample was not successful. The anchor signal has been graded higher than all of the reconstructed samples. The harmonic excitation has served to make the filtered signal sound less like the original rather than more.

4.2 R_{nonlin} Results

The R_{nonlin} metric was developed for predicting the perceived quality of nonlinearly distorted signals. The process by which it is measured is described in [11]. The metric uses psychoacoustic models and correlation measurements to determine how similar a distorted signal sounds to the undistorted signal. This is very similar to the assessment criteria of the subjective listening tests.

The R_{nonlin} value for each of the listening test stimuli were calculated and are shown in Figure 5. As the R_{nonlin} metric returns a value between 0 and 1 the results have again been normalised to the range of 0 to 100. These values are used to support the results obtained from the listening tests as the cohort of test subjects was not large enough to provide conclusive results.

The results in Figure 5 support some of the correlations noticed in Figure 3. It is shown that increasing the frame or filter kernel length will produce a signal which is objectively more similar to the reference sample. It has also shown that for samples with a prominent fundamental (Figures 5a, b and c) the reconstructed samples are more similar to the original than the anchor sample is.

Figure 5d again highlights the inaccuracies of the discussed methods when the input signal has a severely damped fundamental. Some of the reconstructed samples are less similar to the original than the anchor signal. This was also apparent from the results shown in Figure 3d. Section 5 will suggest methods by which this problem may be overcome.

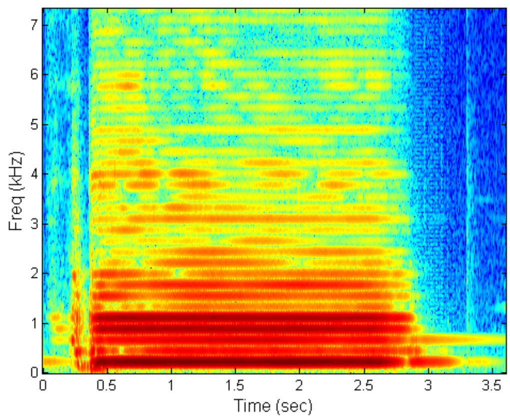
5. FURTHER ISSUES

5.1 Fundamental Amplitude Envelope

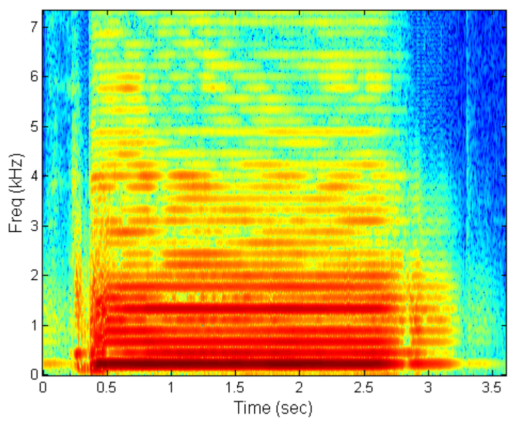
Scaling the amplitude envelope of the fundamental and applying it to the generated harmonic has caused some problems with accuracy. Most obviously in the reconstruction of the Piano sample, where the fundamental was damped and hence its amplitude envelope did not reflect those of the higher order harmonics.

More subtle issues also arose in the reconstructed Cello samples. Figure 6 shows spectrograms of the Cello sample before and after reconstruction using the IAP method.

The decay portion of each of these samples are substantially different from one another. In the original signal (Figure 6a) the first and third harmonics decay in amplitude over a longer period than any of the other harmonics. As the reconstructed harmonics all use the amplitude envelope of the fundamental they also have this extended delay



(a) Original Signal



(b) IAP Reconstruction (50 Sample Filter kernel)

Figure 6: Spectrograms showing the frequency content of the cello sample before and after reconstruction with the IAP method.

time, as shown in Figure 6b. This results in an audible difference in the two samples during the decay phase.

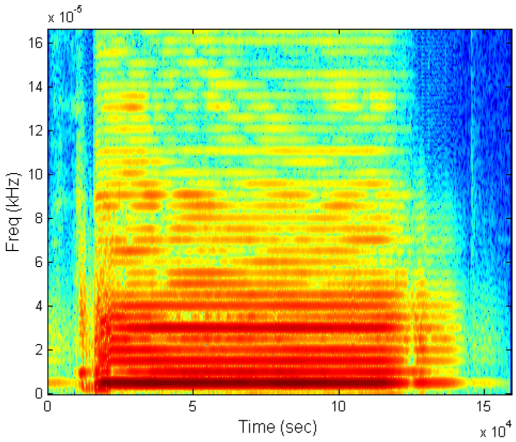
A proposed method to increase accuracy in these cases is to use the amplitude envelope of harmonics closer to the one being generated rather than that of the fundamental. It may also prove more accurate to use the amplitude envelope of a harmonic with the same parity as the one being generated. Note in Figure 6a that the first and third harmonic have an extended decay but not the second.

This is only achievable using the STFT and IAP methods. While the SSB method is fast to compute it does not measure the amplitude envelope of the harmonic used as the input. It merely pitch shifts it by an integer multiple. This means that unless the fundamental is used as the input not every harmonic can be generated.

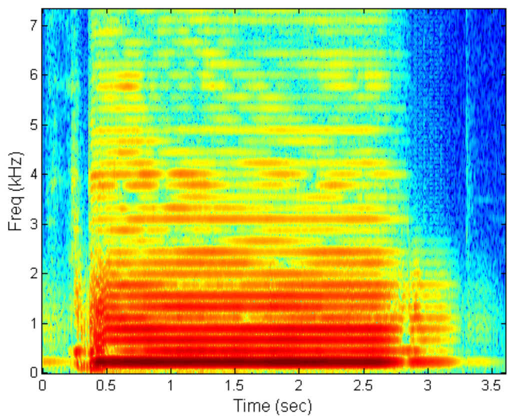
In the samples made for the experiment discussed in Section 3, a large amount of the harmonics were removed prior to harmonic excitation. This meant there was little choice of harmonics to extract an amplitude envelope from. In a more ideal situation the amplitude envelope could be taken from the nearest harmonic with the same parity as the one being generated. Thus an improvement that could be made to the samples created for the experiment would be to use the amplitude envelope of the second harmonic to generate the even order harmonics. Evidently further experimentation is needed to examine what effect this would have on accuracy.

5.2 Post Filtering

Another issue apparent in the samples is the extra high order harmonics that are generated using the SSB and IAP techniques. Any extraneous frequencies in the isolated fundamental will cause these higher order harmonics to be generated along with the desired one. A simple way to overcome this is to apply a band pass filter at the frequency of the desired harmonic after it has been generated. This post filtering process reduces the amount of extraneous high order harmonics in the output signal as evidenced in Figure 7. The extended decay seen at the higher frequencies in Figure 7a are seen to be reduced by the application of post filtering (Figure 7b).



(a) Without Post Filtering



(b) With Post Filtering

Figure 7: Spectrograms showing the frequency content of the SSB recontruction of the Cello sample without and with post filtering.

6. CONCLUSION

It has been shown that resynthesising harmonics using the amplitude envelope of the fundamental gives varying degrees of accuracy depending on the input signal. It is suggested that to increase this accuracy the amplitude envelopes of harmonics closer to those being synthesised could be used. Further experimentation is needed in order to determine the effect this will have on accuracy.

It is also suggested that in order to introduce new harmonics to an audio signal with as little latency as possible, time domain approaches (IAP and SSB) are preferential to a frequency domain approach (STFT). From the preliminary listening tests conducted it is not possible to conclude which of the two time domain approaches is superior.

Were latency the prime concern the SSB method would be appropriate as it requires the least computation. If it can be shown that using the amplitude envelope of harmonics other than the fundamental to synthesise new harmonics would improve accuracy, the more flexible IAP technique would be more appropriate.

7. REFERENCES

[1] J. Chalupper, "Aural exciter and loudness maximizer: What's psychoacoustic about -psychoacoustic processors?-," in *Audio Engineering Society Convention 109*, Sep 2000.

[2] C. Dodge and T. A. Jerse, *Computer Music: Synthesis, Composition and Performance*, 2nd ed. Schirmer: Thomson Learning, 1997.

[3] D. T. Yeh, "Digital implementations of musical distortion circuits by analysis and simulation," Ph.D. dissertation, Stanford University, 2009.

[4] A. De Götzen, N. Bernardini, and D. Arfib, "Traditional (?) implementations of a phase-vocoder: The tricks of the trade," in *DAFx-00*, Dec 2000.

[5] M. Lester and J. Boley, "The effects of latency on live sound monitoring," in *Audio Engineering Society Convention 123*, Oct 2007.

[6] B. Picinbono, "On instantaneous amplitude and phase of signals," *Signal Processing, IEEE Transactions on*, vol. 45, no. 3, pp. 552–560, 1997.

[7] O. Niemitalo. (2003, Jul) Hilbert transform.

[8] M. Puckette, "Patch for guitar," in *Pure Data Convention, Montreal*, Aug 2007.

[9] M. Corinthios, *Signals, Systems, Transforms, and Digital Signal Processing with MATLAB*. Taylor & Francis Group, 2009.

[10] Recommendation ITU-R BS.1534-1, "Methods for the subjective assessment of intermediate quality level of coding systems," 2001.

[11] C. Tan, B. C. J. Moore, N. Zacharov, and M. Ville-Veikko, "Predicting the perceived quality of nonlinearly distorted music and speech signals," *J. Audio Eng. Soc*, vol. 52, no. 7/8, pp. 699–711, 2004.