

TWO-DIMENSIONAL FOURIER PROCESSING OF RASTERISED AUDIO

Chris Pike, *

Department of Electronics
Univ. of York, UK
chris.pike@rd.bbc.co.uk

Jeremy J. Wells,

Audio Lab, Dept. of Electronics
Univ. of York, UK
jjw110@ohm.york.ac.uk

ABSTRACT

There is continuous research effort into the expansion and refinement of transform techniques for audio signal processing needs, yet the two-dimensional Fourier transform has seldom been applied to audio. This is probably because audio does not readily allow the application of a 2D transform, unlike images for which its use is common. A signal mapping is first required to obtain a two-dimensional representation. However the 2D Fourier transform opens up potential for new or improved analysis and transformation of audio.

In this paper, raster scanning is used to provide a simple mapping between one- and two-dimensional representations. This allows initial experimentation with the 2D Fourier transform, in which the 2D spectrum can be observed. A straightforward method is used to display the spectral data as a colour image.

The transform gives information on two frequency axes, one in the typical *audible* frequency range and the other in the low frequency *rhythmic* range. This representation can be used to more easily observe rhythmic modulations in the signal. Some novel audio transformations are presented, allowing manipulation of rhythmic frequency content.

The techniques developed using the 2D Fourier transform allow interaction with audio in a new domain, both analytically and creatively. This work shows how two common signal processing mechanisms can be combined to exciting effect for audio applications.

1. INTRODUCTION

The conventional display of an audio signal is a waveform showing amplitude against time. This can be referred to as the 1D time domain representation, since the signal amplitude is shown only against time, in one dimension. Similarly the spectrum of an audio signal can be called 1D frequency domain representation since it displays magnitude and/or phase data against frequency in one dimension.

Another common audio display is the spectrogram which presents amplitude against time and frequency axes. Time-frequency analysis is regularly used in audio processing. We view this as a two dimensional audio representation, just as it is normal to refer to an image as two-dimensional.

For an image there are two spatial dimensions, whereas audio has one temporal dimension. Using an appropriate signal mapping time can be split into two dimensions of different resolution, creating 2D time domain audio. A logical step from this form is to convert to frequency dimensions via the 2D Fourier transform. This 2D frequency domain form yields potential for new insights into

signal analysis and transformation, especially with regard to low-frequency *rhythmic* modulations of the *audible* frequency signal partials.

As a first step in exploring this frequency-frequency representation the 2D discrete Fourier transform (DFT) has been combined with raster scanning. This was chosen because it is a simple method which allows meaningful visual comparisons of the 2D time and frequency representations.

1.1. Matlab Tool

This work was focused on developing a software tool allowing the 2D Fourier analysis and modification of rasterised audio. The outcome is a useful and novel GUI-based application in Matlab, which is intended for use by composers and researchers who are interested in the potential applications of this processing. The Matlab code and additional documentation are available at [1].

1.2. Raster Scanning

Raster scanning is a common technique for producing or recording a display image in a line-by-line manner. It is used in communication and storage of two-dimensional data sets, a common example being video monitors. The scanning path covers the whole image area reading from left to right and progressing downwards as shown in Figure 1.

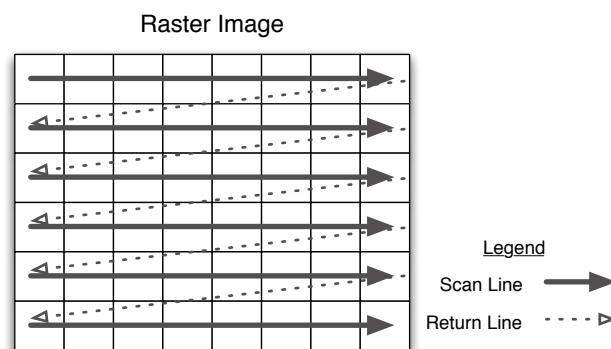


Figure 1: Raster Scanning Path (after [2])

Raster scanning has recently been applied to audio visualisation and image sonification [2]; it allows a simple one-to-one mapping between an audio sample and an image pixel.

Rasterisation is the process by which a 1D audio data array is converted to a 2D matrix using raster scanning. The resulting

* Mr Pike is now working at BBC R&D

representation can be displayed as a grayscale image called a *rastogram*, where the spatial dimensions represent time at two different resolutions and sample amplitude gives the pixel intensity. The temporal resolutions of the horizontal and vertical axes are given by equations (1) and (2) respectively.

$$\Delta n = 1/f_{sa} \quad (1)$$

$$\Delta m = N/f_{sa} \quad (2)$$

N is the chosen image width and f_{sa} is the audio sampling rate, as seen in the horizontal axis of the rastogram, i.e. the reciprocal of equation (1). Thus the sample rate in the vertical axis is given by:

$$f_{sr} = f_{sa}/N \quad (3)$$

The image will have height M where the length of the 1D signal array l is:

$$(M-1)N < l \leq MN \quad (4)$$

If the 1D time domain waveform were displayed as an image without prior rasterisation, it would be a single row of pixels with varying grayscale intensity. The rastogram can be considered as a 2D time domain audio display; it can also be referred to as a time-time representation in comparison to conventional time-frequency analysis.

The inverse process can be used for image sonification or in other words, converting from 2D time domain audio to the 1D waveform. This is known as *derasterisation*. The forward process and its inverse are trivial in digital signal processing, making it useful as part of a 2D analysis-synthesis system.

1.3. Two-dimensional Discrete Fourier Transform

The DFT of a two-dimensional array of $M \times N$ samples can be easily constructed by extending the one-dimensional DFT formula [3]:

$$X[u, v] = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x[m, n] e^{-j2\pi(\frac{um}{M} - \frac{vn}{N})} \quad (5)$$

$$\text{for } \begin{array}{l} u = 0, 1, 2, \dots, (M-1) \\ v = 0, 1, 2, \dots, (N-1) \end{array}$$

As can the inverse DFT:

$$x[m, n] = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} X[u, v] e^{j2\pi(\frac{um}{M} - \frac{vn}{N})} \quad (6)$$

$$\text{for } \begin{array}{l} m = 0, 1, 2, \dots, (M-1) \\ n = 0, 1, 2, \dots, (N-1) \end{array}$$

The equations for the DFT and its inverse have the same relationship in two dimensions as in one i.e. the inverse transform is the complex conjugate of the forward transform, divided by the number of points in the transform. Here u and v are the frequency variables and when considering audio, m and n are time variables but with images they are spatial.

The two frequency domain analysis intervals relate inversely to the size of the matrix. A larger 2D array gives better frequency resolution:

$$\Delta v = \frac{f_{sa}}{N} \quad (7)$$

$$\Delta u = \frac{f_{sr}}{M} \quad (8)$$

The analysis frequencies for each axis can be obtained using equations (9) and (10). It is clear that with rasterised audio, frequencies in the vertical axis f_r are of a lower order than horizontal frequencies f_a since the vertical sample rate f_{sr} is equal to the horizontal resolution Δv . Equation (11) shows that the 2D frequency domain signal is periodic in each axis at the sample frequency of that axis. This is caused by aliasing which occurs in discrete signals.

$$f_a = \frac{vf_{sa}}{N} \quad (9)$$

$$f_r = \frac{uf_{sr}}{M} \quad (10)$$

$$X[u, v] = X[u \pm pf_{sa}, v \pm qf_{sr}] \quad (11)$$

$$\text{for } \begin{array}{l} u = 0, 1, 2, \dots, (N-1) \quad v = 0, 1, 2, \dots, (M-1) \\ p = 0, 1, 2, \dots, \infty \quad q = 0, 1, 2, \dots, \infty \end{array}$$

It is often advantageous to display the 2D spectral data with the DC component (0 Hz in both axes) at the centre of the matrix. The data is divided into quadrants, splitting at the Nyquist frequency $f_s/2$ on each axis; f_s is the sampling frequency for the axis concerned. The data in the analysis frequency range $f_s/2$ to f_s is equal to that in the range $-f_s/2$ to 0 Hz due to the spectral periodicity. Therefore, by shifting the quadrants to reflect this, an analysis range of $-f_s/2$ to $f_s/2$ can be obtained in each axis, placing the DC component in the centre.

This shifting process can also be performed in the time domain, prior to the DFT, multiplying the array by the Nyquist frequency in each axis:

$$x_{shift}[m, n] = x[m, n] (-1)^{m+n} \quad (12)$$

As with the 1D DFT, if the input $x[m, n]$ is real then the Fourier transform is conjugate symmetric so the magnitude spectrum is symmetric, meaning that half of the data is redundant.

$$X[u, v] = X^*[-u, -v] \quad (13)$$

$$|X[u, v]| = |X[-u, -v]| \quad (14)$$

The significance of the 2D spectrum in image processing is well understood [4], each discrete value describes a spatial frequency component. Its precise meaning in an audio context, where it describes temporal frequency, needs to be explored.

1.4. Short-Time Windowed Approach

The 2D DFT equation can be broken into two stages. First the N -point DFT is performed on each of the M rows of the array, so obtaining an intermediate $M \times N$ array. The M -point DFT of each of the N columns of this array is then taken to give the final 2D DFT array. The process can also be done in the opposite order, columns then rows, and the same result will be obtained.

Rasterisation is comparable to the windowing stage of the short-time Fourier transform, where a rectangular window is used with no frame overlap. These frames are then arranged vertically to produce the rastogram.

After the first stage DFT on the rastogram rows, each of the N columns of the intermediate array give temporal variations in magnitude and phase for a particular frequency u , with time interval Δm . Due to the symmetry of the real DFT only the first $N/2$ frequencies are required, the information is duplicated in the second half of the columns.

The second stage is a DFT of the complex data in each of these columns, giving the amount of variation at M different low-frequencies v for each higher frequency u . The complex DFT is not complex conjugate symmetric [4], so there is no duplication and all of the rows are required.

The combined process of rasterisation and the 2D DFT is the simplest implementation of a 2D analysis framework where a DFT is taken for each frequency analysis point across all spectral frames of a STFT. The result is a horizontal dimension with an *audible* frequency range and resolution given by (8) and a vertical axis of sub-sonic *rhythmic* frequency range with resolution given by (7).

This analysis framework was identified in [5], which is a rare example of work describing 2D Fourier analysis of audio. For the rasterised version with no overlap the relationship between Δu and Δv is:

$$\Delta u = \Delta v / N \quad (15)$$

By using the rastrogram as an intermediate step in this analysis rather than a case with overlapping frames, there is a one-to-one sample mapping into the 2D form. Through this method we can more easily observe the relationship between 2D spectral data and the time domain representation for audio.

It was decided to display all four quadrants of the spectrum even though two are redundant. This gives the spectrum the same dimensions as the image to aid direct comparison between them, just as [6] compares the relationship between image features and elements of the 2D spectrum.

2. 2D FOURIER ANALYSIS

An overview of the analysis and processing framework used in this work is shown in Figure 2. The first stage is to perform appropriate rasterisation and 2D Fourier analysis of audio signals and present this information clearly. The raster settings are determined using music information retrieval algorithms or are explicitly set by the user. The signal is then rasterised to create the 2D time domain representation, which can be viewed as a grayscale image. From this rastrogram the 2D frequency domain spectrum is obtained using the 2D FFT and displayed as a colour image.

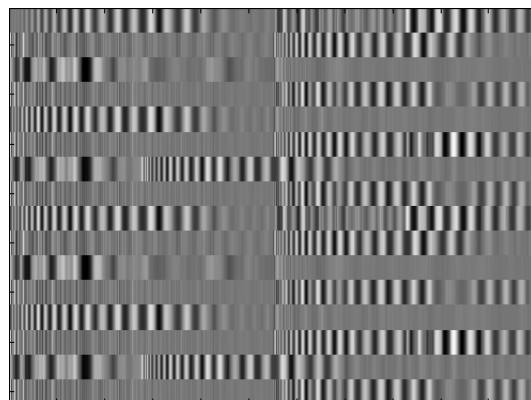
2.1. Rastrogram Width

A rastrogram is often most useful when the image width corresponds to a periodicity within the audio signal. Any slow variation of this periodic element over time can then be observed more easily than with a waveform display. In [2] the fundamental pitch period is used to obtain a rastrogram of audio signals. The period could correspond to any periodic element such as a higher harmonic or it could be larger, like the duration of a quarter-note in a rhythmic audio signal, as shown in Figure 3a.

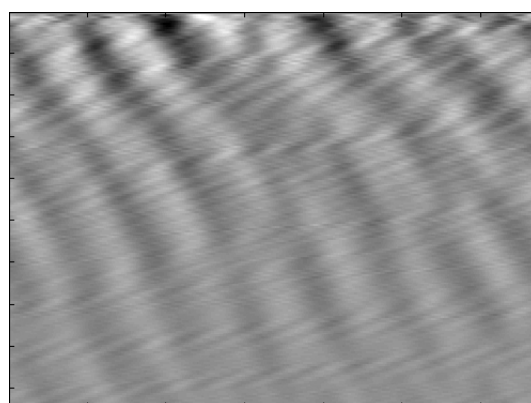
Correct assignment of the rastrogram width is important during 2D Fourier analysis. The Fourier transform analyses the sub-sonic frequency variation between the audible frequency bins of each row. If the image dimensions correspond to a periodic element of the signal then the Fourier data representation will be more informative, see section 2.5.

2.2. Timbral or Rhythmic Analysis

The analysis of audio signals is divided into two categories, timbral and rhythmic, depending on what signal information is used



(a) Drum Pattern at 120 bpm with Width of a Quarter-Note (22,050 pixels)



(b) C2 Piano Note with Width Set to Approximate the Fundamental Pitch Period (672 pixels)

Figure 3: Rastrogram of Different Signal Types

to set the raster width. In timbral analysis mode the width is harmonically related to the pitch whereas in rhythmic analysis mode the width is related to the tempo of the signal.

This implementation uses music information retrieval tools [7, 8] to semi-automate the choice of raster width. These algorithms incorporate techniques like auto-correlation and onset detection to ascertain the pitch and tempo, which can then be used to determine suitable raster widths to divide the audio into rows that contain a common periodic element. These techniques are a vital part of the analysis because a signal periodicity must be synchronised with the raster width in order to obtain a useful 2D spectrum. For timbral mode, note onset information can be used to extract individual notes from a sequence for analysis.

Clearly both modes work on different scales, timbral analysis is best for individual pitched notes and could be useful for observing the internal modulations of partials. Rhythmic analysis is useful on longer signals with rhythmic patterns, where it can show the rhythmic emphasis in the spectrum. In general the range of the rhythmic frequency spectral axis extends higher in timbral analysis mode than rhythmic analysis mode since the image width is lower. Figure 3a was obtained using rhythmic analysis mode while Figure 3b used timbral analysis mode.

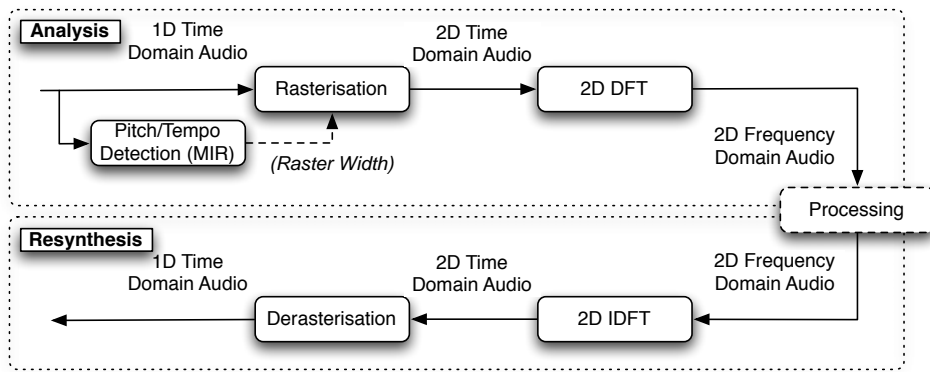


Figure 2: 2D Fourier Analysis-Resynthesis Framework

2.3. Spectrum Display

The 2D frequency domain representation is a large matrix of complex numbers. Some method is needed to present this information clearly. The polar representation of complex data is the most logical when considering audio signals. It is preferable to present magnitude and phase components simultaneously to gain a better understanding of the 2D spectral content.

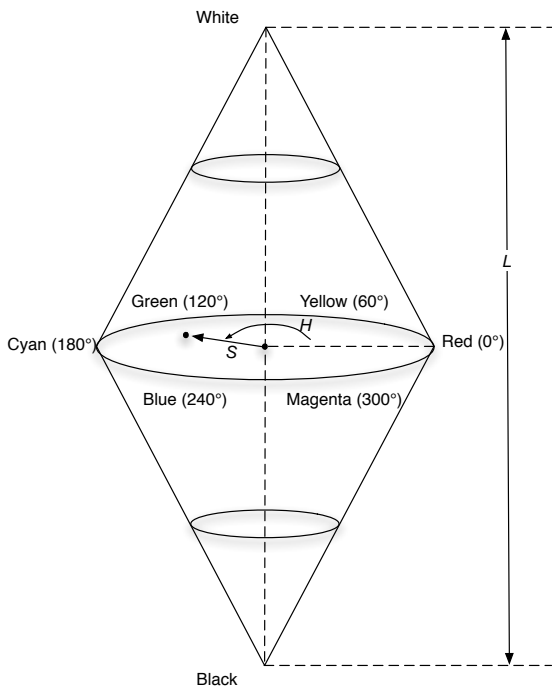


Figure 4: HSL Colour Space

This can be achieved by converting polar data to colour information [9] using the HSL colour space with full saturation, see Figure 4. The following equations show how magnitude is mapped to lightness and (wrapped) phase is mapped to hue:

$$H = \frac{\angle X}{2\pi} \quad (16)$$

$$L = \arctan(|X|) * (2/\pi) \quad (17)$$

The HSL colour space can then be converted to RGB [10] and displayed as a colour image plot. Figure 5 shows an example 2D spectrum. The colour/phase information has been removed and the intensity inverted for printing purposes.

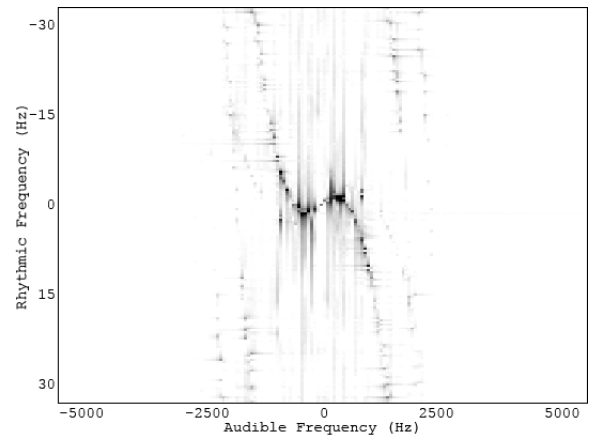


Figure 5: 2D Spectrum Display of a Piano Note (Horizontally Zoomed)

2.4. 2D Spectral Components

The fundamental component of 2D Fourier analysis can be interpreted as an audible sinusoid modulated by a rhythmic sinusoid [5]. However using this rasterisation method it is clear that if component signal periodicities are not synchronised with the analysis window size then the signal will be skewed in the 2D analysis space (Figure 6). The actual frequency of a stationary sinusoidal component is given by:

$$f_{stat} = f_a + f_r \quad (18)$$

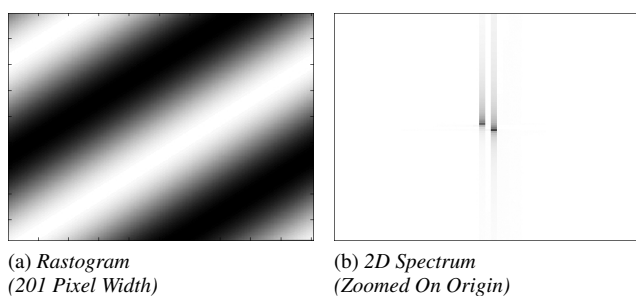


Figure 6: A Non-Synchronised Sinusoid (220 Hz)

The rhythmic frequency (f_r), having a smaller range and finer resolution, essentially shows an error term of the audible frequency (f_a) analysis. These skewed signals can be seen to have non-stationary phase across both dimensions, since in each consecutive row/column the phase angle has changed.

A signal with a phase-stationary frequency in both axes of the rastogram is an amplitude modulated sinusoid. Amplitude modulation is conventionally considered in terms of a carrier frequency and a modulation frequency. 1D Fourier analysis shows that amplitude modulation can be achieved using two sinusoids, f_1 and f_2 , with constant amplitude and frequency, where:

$$f_{carrier} = \frac{f_1 + f_2}{2} \quad (19)$$

$$f_{mod} = |(f_2 - f_1)| \quad (20)$$

If the carrier frequency is synchronised to the width of analysis in the rastogram then the AM signal is represented by symmetrical points in each of the four quadrants of the 2D spectrum (Figure 7). The absolute values of their frequency co-ordinates f_a and f_r will be equal. Due to the conjugate symmetry of the real 2D Fourier transform (13), the 2D spectral data contains duplicate information and two of the quadrants are redundant. The negative quadrants of either axis can be disregarded but conceptually it is easier to ignore the negative audible frequency quadrants.

The remaining two points represent the sinusoids f_1 and f_2 . Both have the same audible frequency which is therefore equal to $f_{carrier}$. They have equal absolute rhythmic frequency but one is negative. The modulation frequency f_{mod} is the difference between positive and negative rhythmic frequencies i.e. $|2f_r|$.

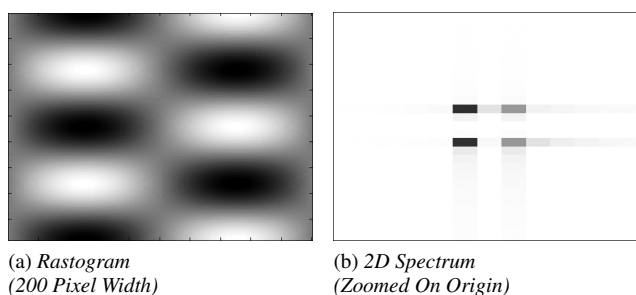


Figure 7: 2D Representations of Bipolar Amplitude Modulation ($f_{carrier} = 220.5$ Hz and $f_{mod} = 2$ Hz)

Figure 8 shows the 2D spectrum of a sequenced drum rhythm at 120 bpm which, when using a sample rate of 44.1kHz, gives an integer period for a crotchet/quarter note. Symmetric points in all quadrants of a 2D spectrum show that an AM "rhythmic" component is present with synchronised carrier frequency. The many symmetric points in this spectrum signify a synchronised rhythmic pattern.

There are pairs of sinusoidal signals skewed in opposing directions across the spectrum. Each sinusoidal component is represented by two points each due to the spectral redundancy, so only the points in quadrants with positive audible frequency are needed. The actual 1D frequency of the sinusoids can be determined by summing the audible and rhythmic frequency co-ordinates of the two points. Combined they create an AM sinusoid, where the mean of their audible frequencies gives $f_{carrier}$ and the difference between their rhythmic frequencies gives f_{mod} .

The rhythmic content of the audio signal is being analysed in terms of a set of low-frequency amplitude modulations. This is the main benefit of the 2D spectrum, it makes rhythmic modulation much easier to detect than in the 1D spectrum.

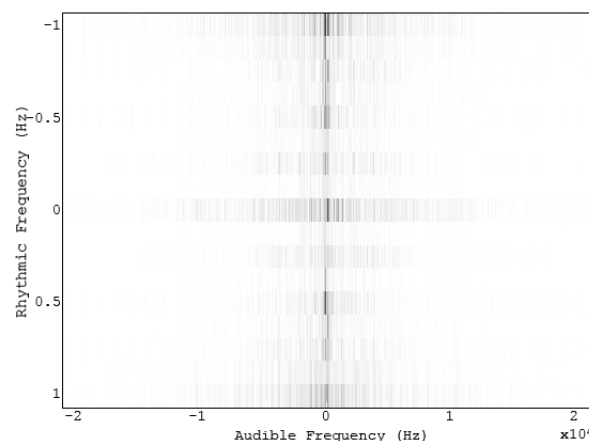


Figure 8: 2D Spectrum Display of a Drum Beat

2.5. Analysis Issues

Spectral analysis in two dimensions is subject to the same issues as conventional 1D methods, such as convolution with the spectrum of the window function and smearing when frequency components are not centred on an analysis point. There are additional limitations for 2D analysis, some as a result of the simple rasterisation approach instead of overlap-add methods. However this method serves as a first step, allowing visual exploration of features to gain understanding. The processing techniques can then be refined and extended in the future.

Rectangular windowing causes particular issues. If a 2D spectral component is not synchronised with the analysis frequencies then spectral energy will leak into adjacent analysis bins due to the high side lobe energy of the window's spectrum [11]. But if a component is synchronised then the spectrum of the window has no effect because there are no discontinuities between the two ends of the window. Unfortunately pitch and tempo related periods very rarely contain an integer number of samples so this spectral smearing is frequent. This can be shown using the previous piano note

examples. In Figure 3b the texture of the rastogram is slanted because the pitch period is between 672 and 673 samples and in the spectrum (Figure 5) there are clear smearing effects as a result. The data could be resampled to fit the period to an integer number of samples, improving the clarity of analysis.

Skewing of non-synchronous signal components is a problem which reduces the intelligibility of the spectral display when combined with the windowing issue. Extending the length of the analysis frames (rastogram rows) will increase the resolution of signal analysis in the the audible axis. Reducing the inter-frame interval to overlap frames will increase the rhythmic axis analysis resolution. The definition of spectral data could be improved further by zero-padding the 2D array before the Fourier transform, increasing the resolution of the display as shown by (7) and (8). The computational complexity of the 2D DFT increases by $O(n^2)$ with the dimensions of the data matrix so these options are limited. Techniques such as time-frequency reassignment [12] could be applied during the analysis to reduce spectral smearing.

The most significant limitation in this analysis framework is that aliasing in the rhythmic axis is intrinsic since the rhythmic sampling rate f_{sr} is much lower than frequencies within the signal. The columns of the rastogram can be viewed as heavily decimated versions of the 1D waveform. The appropriate low-pass filtering to comply with the sampling theorem would remove the required audible frequency content. This is a result of artificially dividing one dimension of time into two with different resolutions.

2.6. Benefits of 2D Spectral Analysis

The outlined 2D spectral analysis provides a useful method for viewing rhythmic range frequency modulations of audible frequencies such as signal harmonics. It could be used for more detailed rhythmic analysis of audio; there is clearly high spectral energy with many symmetric components at sub-sonic frequencies that correspond to rhythmic patterns in the signal. Timbral analysis shows similar 2D spectral envelopes for all notes from an instrument. Understanding timbre in terms of harmonic modulations could have potential for robust pitch and time shifting.

3. AUDIO TRANSFORMATIONS

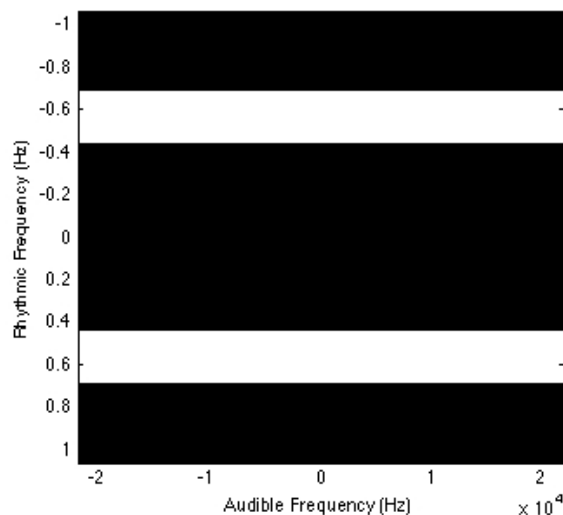
Although there are some issues with the described 2D Fourier analysis method, it clearly shows useful signal information. It is possible to produce interesting audio effects by manipulating the analysis data and resynthesising 1D time domain audio via the 2D IDFT (Figure 2). Some simple transformations of the 2D Fourier data were carried out to investigate the potential for creative and analytical processing in this signal domain.

The effectiveness of the described signal transformations depends heavily upon the signal content being synchronised to the analysis dimensions. With the current analysis process this means that applications are fairly limited.

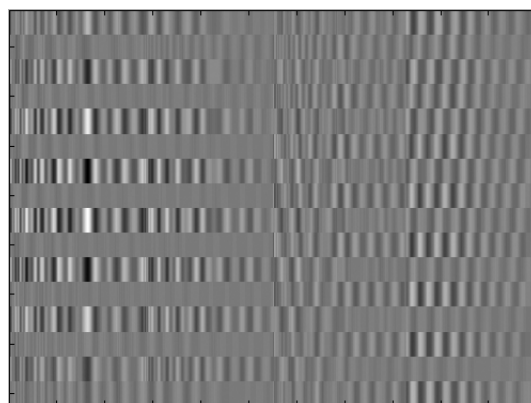
3.1. 2D Frequency domain Filtering

Filtering can be performed in the 2D frequency domain by multiplying the Fourier data matrix of the signal with that of the desired filter. In these initial experiments only simple filters were used such as low-pass, high-pass, band-pass and band-stop in ideal "brick-wall" and Butterworth configurations. The filter operation is performed on either the audible or rhythmic axis.

Filtering in the audible axis yields similar results to conventional 1D filtering provided that the signal components are well synchronised. Filtering on the rhythmic axis produces some interesting effects, altering the rhythmic or timbral structure (section 2.2) whilst maintaining a similar harmonic character. Figure 9 shows an example filter spectrum and the 2D time domain result of multiplying this with the spectrum in Figure 8.



(a) Magnitude Spectrum of an Ideal-Response Band-Pass Filter With 2D Structure. Cutoff Freq. = 0.5 Hz and Bandwidth = 0.25 Hz



(b) Filtered Drum Beat (see Figure 3a for original)

Figure 9: 2D Frequency Domain Rhythmic Filtering

The problem is that due to the inherent aliasing, it is difficult to know what changes have actually been made to the spectral signal content. The results are thus not entirely predictable but still interesting and useful in providing creative permutations of a rhythmic sequence.

The 2D Fourier transform has been used to perform 2D Weiner filtering for noise reduction in speech processing [13], so this method is not unique although its application to rhythmic transformation for creative effect seems novel.

The same effects could be achieved with a time domain filter on the vertical axis of the rastogram, which would be much more efficient computationally. However this is a simple applica-

tion which could be extended to utilise both dimensions and alter the rhythmic structure of a specific audible frequency band or extending it further, perform feature extraction/emphasis.

3.2. Magnitude Thresholding

Thresholding of the spectral magnitude data allows decomposition of the sound into its most or least prominent sinusoidal components. It serves both as a creative effect and an analysis tool, allowing decomposition of the 2D spectral structure of a signal. In the 2D frequency domain thresholding can be performed on rows, columns or spectral points i.e. in both axes.

Thresholding on rhythmic-frequency rows allows the most dominant sub-sonic oscillations of the signal to be separated, an example is shown in Figure 10. Performed along the columns it allows the removal of audible components above or below the threshold, which in timbral analysis mode causes separation of the signal partials. The thresholding for individual points decomposes the signal into the strongest or weakest individual sinusoidal components.



(a) 2D Spectrum - Rhythmic Frequencies <23% of Max. Magnitude Removed (Horizontally Zoomed) (b) Processed Rastogram (see Figure 3b for original)

Figure 10: Rhythmic Frequency Magnitude Thresholding of a Piano Note

The thresholding tends to remove more high-frequency audible energy since audio signals often have larger magnitude in the low frequency components. This aspect could be improved by optionally weighting the audible frequencies according to a perceptual loudness curve [14]. Another issue to consider is that the opposing sinusoids of a rhythmic amplitude modulation do not necessarily have identical magnitude and so may not be retained/removed simultaneously. This would affect the rhythmic structure. It would be easy to prevent this if desired by comparing opposite points within a tolerance range.

3.3. Row/Column Shifts

An obvious and simple manipulation is to move row/column data around in the 2D frequency matrix. In the audible axis column shifting creates a change in the signal partials. In timbral mode this operation never causes a pitch change because the frequency resolution of analysis matches the harmonic spacing of the original signal, although in rhythmic analysis mode it can change the signal pitch.

Row shifting changes the rhythmic frequency of the data so altering the rhythmic/timbral structure. The results are again unpredictable due to aliasing on the rhythmic frequency axis and skewing of non-synchronised sinusoidal components. This effect

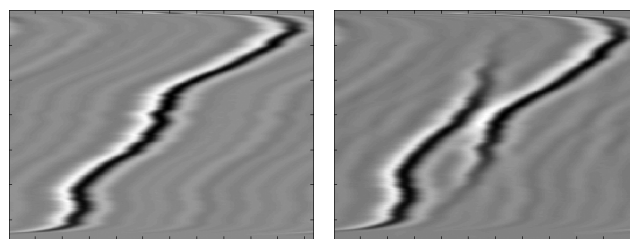
could be improved by incorporating phase difference calculations such as those used in the phase vocoder [15], which would reduce the harmonic distortion caused by rearranging the frames (rows/columns) of spectral analysis.

3.4. Resampling of Data

It was thought that resampling of the 2D spectral data could produce some very useful effects, changing the spectrum size or rescaling the signal within the current analysis range. However this is when the limitations of the analysis cause significant problems.

Pitch-shifting can be implemented by resampling across the audible frequency axis, scaling the data across the spectrum rows by a linear factor to adjust the audible frequencies of signal components. This is functionally similar to pitch shifting methods that use the 1D STFT [16]. Although in the described timbral analysis mode there is a harmonic shift rather than a pitch shift since the resolution is too low in that axis, the analysis frequencies match the original signal harmonics.

By resampling rhythmic frequencies in the same way a change in tempo could be achieved but the rhythmic analysis mode has too low a resolution in this axis. Instead it produces a rhythmic change rather than a tempo change. In timbral analysis mode the tempo is altered without pitch change but unfortunately aliasing occurs because the duration remains the same. If the rhythmic frequency range is halved, the speed of the signal content will halve but since the duration remains the same, the content wraps around and starts again, as shown in Figure 11. When the rhythmic frequency range is extended, the signal speed increases but begins to repeat itself due to the periodic nature of Fourier analysis.



(a) Original Trumpet Note (G3) (b) Rhythmic Frequency Range Reduced to 75%

Figure 11: Attempted Tempo Change By Rhythmic Frequency Range Reduction

Changing the spectrum dimensions will alter the length of the signal on resynthesis. Resizing along one dimension changes the analysis frequency points on that axis. The data can be resampled to keep it at the original frequency.

In rhythmic mode, it was thought that adjusting the width of the spectrum would change the duration of each row of rastogram data whilst maintaining the original pitch. This would essentially alter the tempo of the signal, changing the inter-frame time interval in a similar way to granular time-scale techniques. Each frame/row is either truncated or repeated to achieve the change, rather than overlapping the frames like STFT methods. The rectangular window of rasterisation allows severe distortions using this method.

Changing the height of the spectrum would change the duration of the signal and resampling could be used to maintain the signal tempo by keeping the original rhythmic frequencies. The

results here are very similar to the rhythmic frequency scaling described previously but the duration changes instead of the tempo. The same issues occur when the duration is halved because the signal is at the same tempo, it wraps around to retain all of the signal content, yielding a rastogram with the same form as Figure 11b but twice the height.

The resizing methods also suffer from a lack of spectral resolution. The processing attempted here was basic but it shows that there maybe some potential for future use of these types of technique. If the analysis is refined for precision and transformation rather than visual comparison with the rastogram, then these methods will be much more effective. It will then be easier to determine the appropriate application areas and the transformation techniques themselves can be refined and advanced to make better use of the two frequency axes simultaneously.

One additional factor that needs consideration is that there is often a lot of energy at DC in the rhythmic axis, which requires special handling when resampling is used to avoid strong low-frequency components appearing in the signal.

4. CONCLUSIONS

Combining raster scanning and the 2D Fourier transform allows analysis of audio with two dimensions of frequency. This makes it possible to observe low-frequency rhythmic modulations of audible frequencies in a 2D spectrum display, provided that the width of the raster scan is set to match a periodic component of the signal.

The 2D spectrum can be displayed as a colour image by mapping magnitude to lightness and phase to hue, giving a clear and attractive representation of a large data array. Its horizontal axis shows audible frequency partials in the conventional spectrum range and the vertical axis shows rhythmic frequency in the sub-sonic range. Each 2D spectral point defines a sinusoidal partial which has a frequency of the sum of its audible and rhythmic frequency axis co-ordinates.

The described analysis approach is useful for direct comparison of time domain and frequency domain signal features in two dimensions, however there are limitations when it comes to detailed analysis and processing of the 2D Fourier data. The resolution of analysis could be improved by using overlapping frames with a bell-shaped window function as is common for the STFT. Windowing would reduce distortions and larger Fourier transforms could be utilised to improve the spectral definition.

The initial experiments with signal transformation in the 2D frequency domain show interesting potential for novel effects, with some exciting results, although at the moment these implementations have limited capabilities, partly due to problems with the analysis methods. Future work should look to improve both analysis and processing techniques using the 2D Fourier transform. Similar analysis and processing can be performed using a wavelet transform [17]; a comparison of these methods should also be carried out in future work.

This investigation has established that this is a useful mechanism by which to work with audio and demonstrated some potentially interesting applications both for researchers and sound designers.

5. REFERENCES

- [1] C. W. Pike, "Two-dimensional Fourier processing of rasterised audio," MEng, Department of Electronics, University of York, 2008, http://www.jezwells.org/2D_Fourier/.
- [2] W. S. Yeo and J. Berger, "Application of raster scanning method to image sonification, sound visualization, sound analysis and synthesis," in *Proc. Digital Audio Effects (DAFx-06)*, Montreal, Canada, Sept. 18-20, 2006, pp. 309–14.
- [3] A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing*, chapter 3, pp. 115–20, Prentice-Hall, 1975.
- [4] S. W. Smith, *The Scientist and Engineers' Guide to Digital Signal Processing*, chapter 24, California Technical Pub., 1st edition, 1997.
- [5] C. Penrose, "Chapter 2: Spectral representations," Incomplete thesis available at <http://silvertone.princeton.edu/~penrose/thesis/>, accessed 23rd May 2008.
- [6] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, chapter 4, Prentice Hall, 3rd edition, 2008.
- [7] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.
- [8] O. Lartillot and P. Toiviainen, "A MATLAB toolbox for musical feature extraction from audio," in *Proceedings of the 10th International Conference on Digital Audio Effects*, 2007.
- [9] J. Gallicchio, "2D FFT Java Applet," Available at <http://www.brainflux.org/java/classes/FFT2DApplet.html>, accessed 23rd May 2008.
- [10] K. Fishkin, "A fast HSL-to-RGB transform," in *Graphics Gems*, A. S. Glassner, Ed., pp. 448–9. Academic Press, 1998.
- [11] J. O. Smith III, *Spectral Audio Signal Processing*, chapter 3, CCRMA, Dept. of Music, Stanford University, California, March 2007.
- [12] K. Fitz and L. Haken, "On the use of time-frequency reassignment in additive sound modeling," *J. Audio Eng. Soc.*, vol. 50, no. 11, pp. 879–893, 2001.
- [13] I. Y. Soon and S. N. Koh, "Speech Enhancement Using 2-D Fourier Transform," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 717–724, 2003.
- [14] H. Fletcher and W. A. Munson, "Loudness, its definition, measurement and calculation," *J. Acoust. Soc. Am.*, vol. 5, no. 2, pp. 82–108, 1933.
- [15] J. A. Moorer, "The use of the phase vocoder in computer music applications," *Journal of the Audio Engineering Society*, vol. 26, no. 1/2, pp. 42–45, 1978.
- [16] J. Laroche and M. Dolson, "New phase-vocoder techniques for pitch-shifting, harmonizing and other exotic effects," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 91–94, 1999.
- [17] L. M. Smith and H. Honing, "Time-frequency Representation of Musical Rhythm by Continuous Wavelets," *Journal of Mathematics and Music*, vol. 2, no. 2, pp. 81–97, 2008.