# AN INVESTIGATION INTO MANIPULATION OF SPECTRAL DATA AS AN EDUCATIONAL TOOL

AUTHOR: ASHLEY L. NOEL-HIRST

STUDENT NUMBER: 1602254

DATE: APRIL 15TH 2020

# Table of Contents

# INTRODUCTION

This project concerns the efficacy of the implementation of spectral processing tools for a multi-sensory (in this case audio-visual) education/exploration of timbre. Timbre was stratified into tone-colour and temporal properties, and as a means of simplifying the project, we focused on tone-colour. The team created software, a presentation and demonstrations all exploring the titular brief, and as the project group operated, so our reports will be divided. This report will primarily be concerned with the application of spectral processing tools using Fourier transformation algorithms in real time. I will abstract away from the particular tools/languages used, and focus instead on the processes and their interactions, as a means of making this research as transferable as possible. The specificities of implementation will likely be covered by my colleague Sam Sutton. For discussion of the educational aspects of our project, please see the work of my colleague, Imogen Kelly, who has investigated this at length.



# BACKGROUND

As is discussed by Ian Johnston (Johnston 2009:28), a sound may be identified by a number of qualities including: pitch, loudness, and duration. These are pretty straightforward to perceive in a note, but the fourth quality, timbre, is the most elusive of the properties, appearing to be comprised of different combinations of the other three (Siedenburg et al. 2019:2).

If we consider the volume envelope (using the standard “Attack, Decay, Sustain, Release” descriptors[i]), of an acoustic double bass in comparison to an electric double bass[ii] (both belonging to the violin family, and therefore subject to its models and modes) for example, one might argue that the acoustic had more “fullness” or “body”. We would be justified in the hypothesis that this is a result of longer decay time for the low end of the instrument, occurring due to resonance with a larger body. This is because the lowness of frequencies excited, as well as the amount of excitement, are proportional to the size of the body, among other factors (Välimäki and Holm 2000:2,3). That is to say that as the body gets smaller, the decay time decreases, and the pitch of frequencies amplified becomes higher and higher. With the quality of modern recording technology, as well as processing such as pitch-shifting and time-stretching (Heller Murray et al. 2019:2272), one can see how this definition of timbre can be influenced by all other qualities of a sound, making it difficult to comprehend. This troublesome three-dimensionality is demonstrated by the movement of the frequency analysis which audacity can perform[iii].

…But what if we wanted to talk about the relation of JUST amplitude and frequency, whether that is with regards to one of those slices, or the overall sound produced by the instrument? What quality would we then be discussing? Perhaps part of untying this definition, particularly in relation to this research project, lies in the reappreciation of a term often synonymized with timbre – “tone-colour”, in fact my separation of these two terms flies in the face of “measured tones” (Johnston 2009:28), but I believe that this is rightfully so. How I wish to distinguish between the two as a means of unpacking timbre is purely on the basis of their immediate lexical semantics. “Tone-colour” appears to refer to the palate of a sound – its overall tonal qualities – completely freeing us from the envelopes which timbre seems so married to[iv]. This distinction/subsetting of definitions is not an arbitrary/manufactured one, as is demonstrated by the foundational work of Helmholtz (1886; Siedenburg et al. 2019:6), the framework of which was[v] instrumental for the construction of this project. From an educational standpoint, we must first foster an understanding of this tone-colour of a sound, since it is only with this that we can start to describe how it moves to affect perceived timbre. This sub-sectioning is fundamental to our project, due to the limited nature of our timeframe/resources, we were only able to focus on the first section. In line with Helmholtz’s work, for the rest of this report discussion of tone in relation to timbre “shall disregard these peculiarities of beginning and ending, and confine our attention to the peculiarities of the musical tone which continues uniformly" (H. L. F. Helmholtz and Ellis 1886:67).

Operationalising tone-colour:

But what are these peculiarities of the musical tone which continues uniformly? All sound is comprised of oscillations in a medium, and the number of oscillations per second is called the frequency, measured in Hertz (Hz). For the majority of humans, all perceivable sound falls in the continuous range of 20hz to 20,000hz, and is discretized into pitch classes by temperament systems[vi]. The tone-colour of a sound could thus be considered as the space which it occupies on this frequency spectrum[vii].

For a single tone, the most immediate demonstrations of such spectral content are harmonic or inharmonic. The harmonic series denotes the mathematical relation between a pure fundamental and the overtones which occur above it[viii]. Content from this can occur at varying amplitudes depending on the placement and amplitude of excitation, medium of wave (for example stringed vs wind instruments) and more. For example, bowing a violin harder and closer to the bridge excites its upper harmonics more than if one were to bow softly and nearer the neck (Charlotte Desvages 2016: 1:50 ; Johnston 2009:127).

Inharmonic content can arise in the following situations.

- The introduction of a noise source to increase the footprint of the signal in the frequency spectrum, such as a guitarist in a rock band using distortion to hype up yet another average solo, or the use of subtle tape saturation in mixing and mastering to make signals appear to occupy more sonic space[ix].
- Anything with wave-equation of dimension n>1, such as drums (n = 2) or bells (n=3) (Haraux 2018: 9-13).

- Inharmonic synthesis, often modelled by displacement “adding a constant frequency value k to all harmonic frequencies . . . for example, 230, 430, 630, 830, ...,1630 Hz (with the original f1 = 200 Hz, and k = 30 Hz)” (Schneider and Frieler 2009:23), or intermodulation “. . . no fundamental is present (e.g.,1230, 1430, 1630, 1830…Hz)”(de Boer 1976; Schneider and Frieler 2009:23).

Since we wanted to allow for combinations of either of these scenarios, our program had to be free from any dependence on detection of a fundamental, since it was possible there simply wouldn’t be one.

Since the cause for the differentiation of tone-colour from timbre was the removal of time-domain qualities introduced by envelopes, we should say that discussion of the tone-colour of a sound pertains instead to the frequency-domain, within which the frequency range of the signal can be manipulated by a method known as spectral processing (Rao, Kim, and Hwang 2010:236).

## Designing the Processing Tools

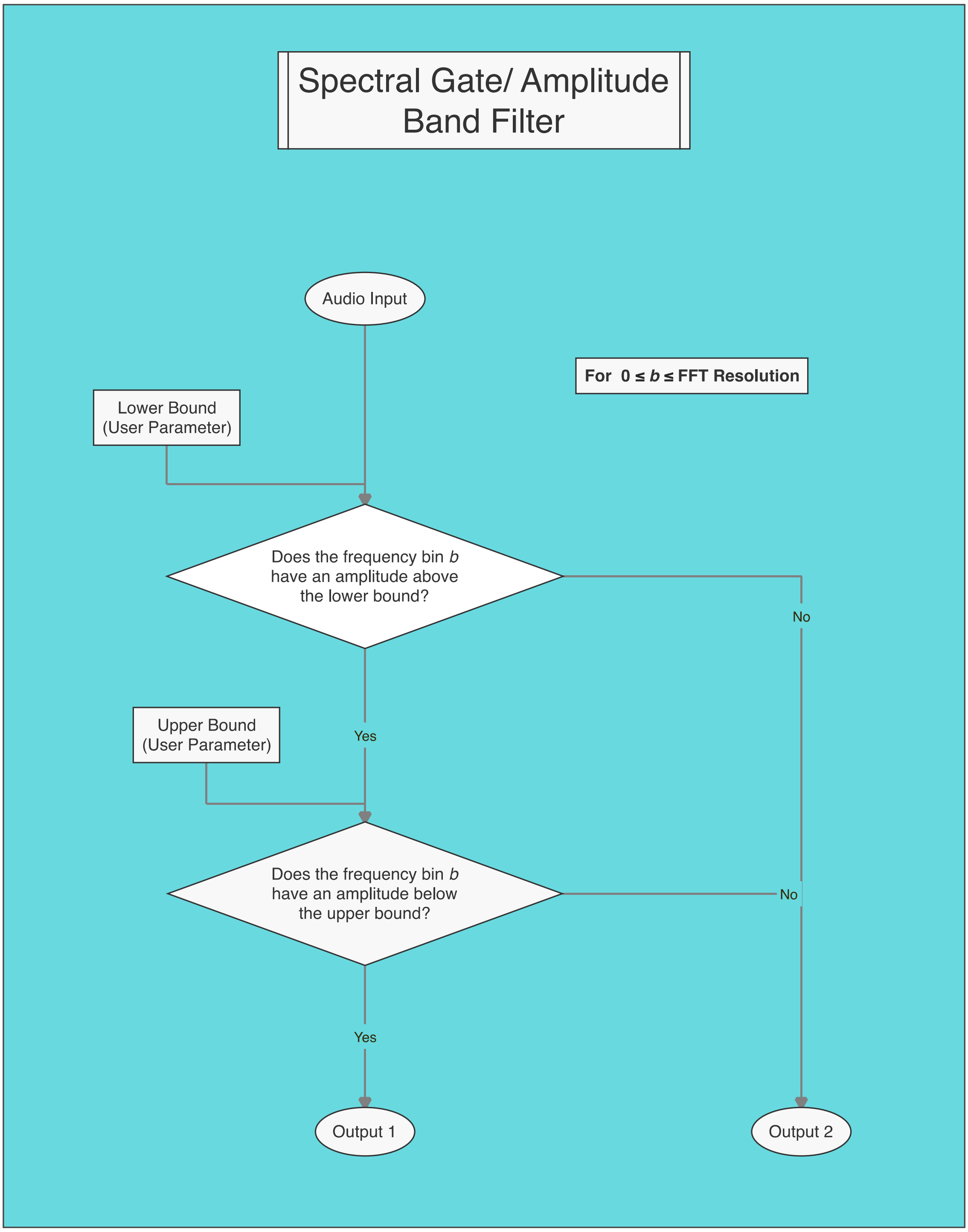
This spectral processing is where our research project really begins. So how do we do it, and how does one even access the frequency domain? Such a problem was addressed by Mathematician Jean-Baptiste-Joseph Fourier, who devised a method to turn periodic waveforms into variables for a series of sine waves (Lewis and Welch, n.d.:1675). His method was further developed by Cooley and Turkey into a Discrete Fourier Transform (Cooley and Tukey 1965), which “maps a sequence x(n) into the frequency domain” (Rao, Kim, and Hwang 2010:1), so is suitable for digital signal processing, which is always discrete.

*A brief aside:*

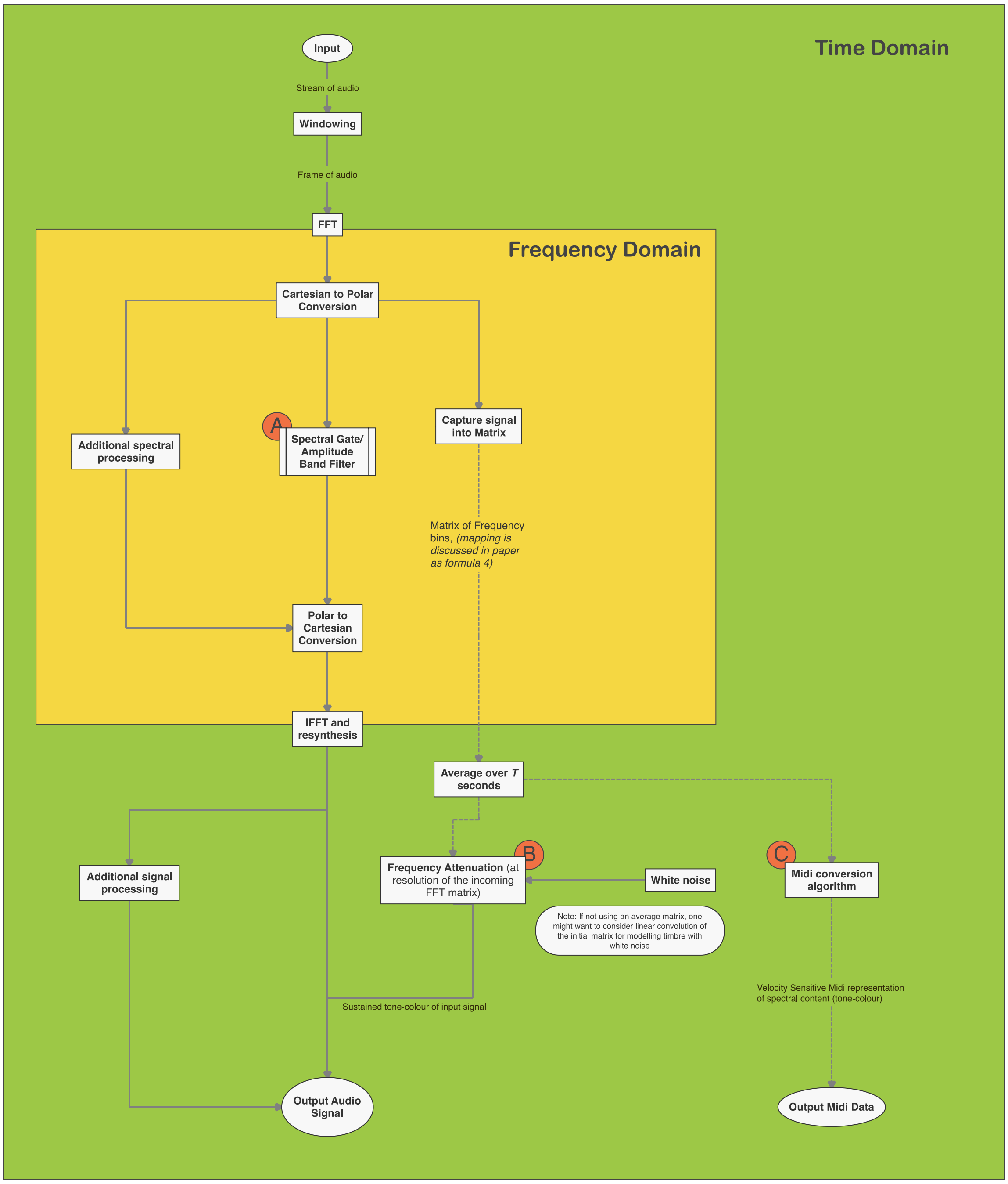
*I will do my best to keep this report process driven rather than number driven, but for the mathematically inclined among you, implementation and further discussion of these techniques is in the appendix* [x] *(I would advise that one reads this report with two copies open, so that both the discussion and diagrams are simultaneously viewable). Though I implemented this method in Max (https://maxforlive.com) – due to its accessibility, interactivity, and Ableton live integration – this could be done in any DSP tool.*

The numbers coming out of the FFT matrix represent ***N = Frame size*** divisions of the frequency range from 0 to the sampling rate. The first number represents the energy at 0 Hz, the second number represents the energy at (1/FS) * the sampling rate, the third number represents the energy at (2/FS) * the sampling rate, and so on.

It is at this point where our first spectral manipulator, the Frequency amplitude filter/gate, comes into play:




I designed a piece of code to filter out background noise from our data. In the time-domain, such a device is commonplace, the most rudimentary of which being the audio-gate, which requires a signal to have an amplitude within a certain range for it to be made audible. This frequency gate did exactly the same thing, but in the frequency-domain and with upper, as well as lower bounds. The amplitude of each frequency was considered, and any frequencies which had amplitudes outside of the user-defined range were removed from the signal. Despite its simplicity, this worked remarkably well[xi], and served as a proof of concept for the re-appropriation of time-domain processes to the frequency domain. With time-domain variables, such as attack and delay times, the results could have been smoother, but this was sufficient for our needs, such devices have already been developed to far greater sophistication that we hoped to achieve, and this had shown us how accessible spectral interaction could be, so we moved onto our primary task.

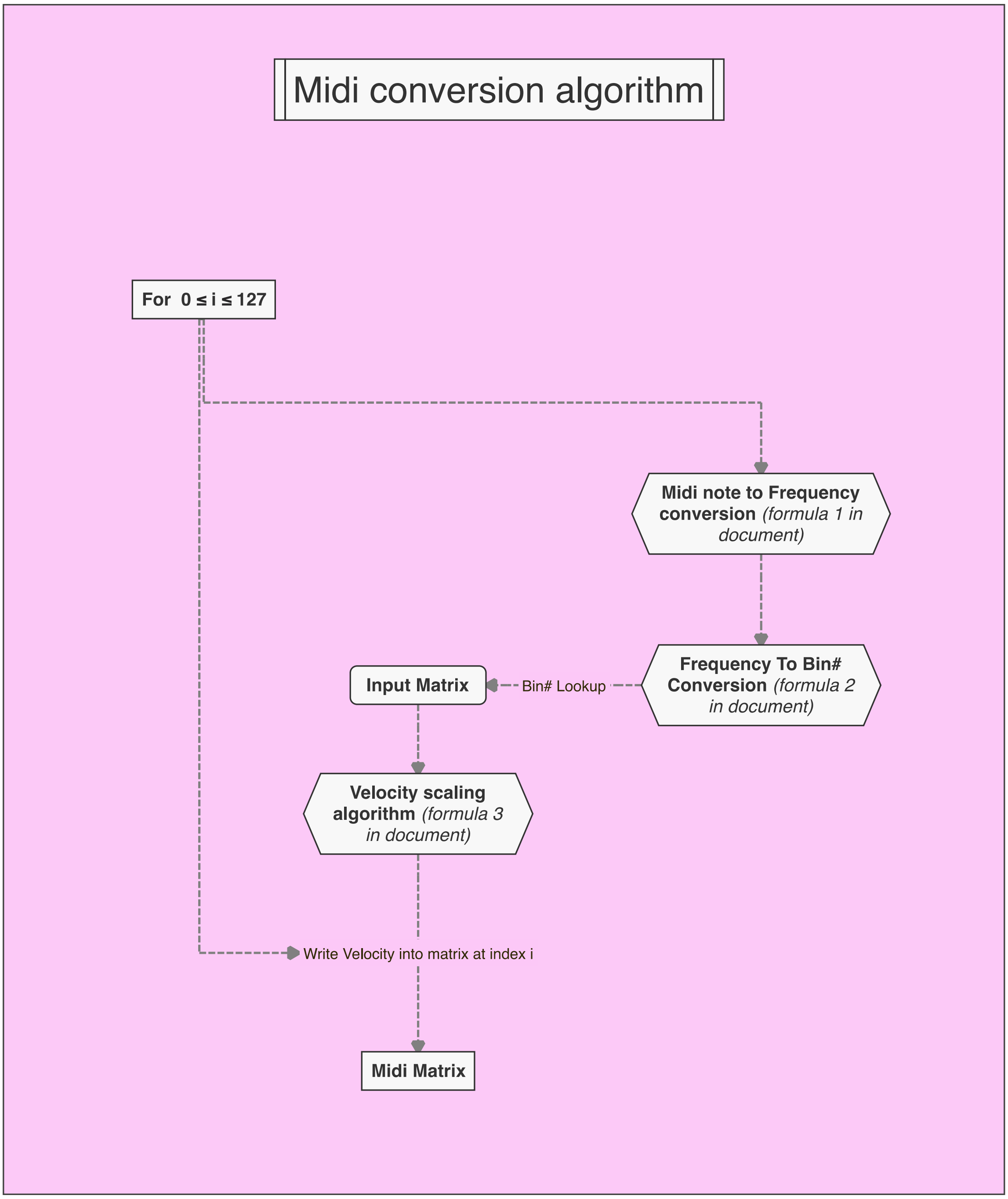



1 - Note: to remove noise from our input signal before processing, task A would have to be placed in series and before processes B and C, they are merely placed in parallel to demonstrate their behavior in a mutually independent situation

Now that we have signals denoting the amplitudes and phases of each frequency (expand on we want to create some imprint of the tone-colour of the sound, in particular, removing the change of this over time. This desire resulted in processes B and C. Before processing, we mapped our DFT into matrices, chopping the frequency spectrum into bins and inscribing their amplitudes and phases into a 2-plane matrix within the frequency-domain [xii]. Now, for each bin of frequencies we have the average amplitude and phase displacement of the waves within. If we access this matrix in the time domain, we can perform matrix operations on our waveform in real time[xiii]. The first thing I did was to average the matrix over a time *T*, *so* instead of a matrix which updates at the sampling rate of the processor, it is a more stable/indicative matrix, displaying the average of the values for (in the case of my tool) the past 5 seconds. This matrix contains the data which we would say corresponds to our earlier definition of tone-colour but – short of bombarding students with values for amplitude and phase – this data is still inaccessible. In order to remedy this, we set about not only visualizing it, but making it audible. Inspired by the findings of Al-Ahrani et al (2010) from the field of mathematics education, we hope that this will aid the reinforcement of understanding of tone-colour.

The first task, audiation, corresponds to process (B) on the flowchart. We use the average matrix as a buffer matrix (imagine an equalizer but with a volume slider for each frequency bin – that's ***N = matrix size*** frequencies being independently controlled!). This buffer is then used to attenuate frequencies of a white noise signal in correspondence to the average matrix[xiv]. This subtractive synthesis process results in an indefinite sound which bears the same overall tone-colour as the source signal[xv].

We can achieve the visual task, as well as work around any equaliser errors of the auditory[xvi] by using additive resynthesis. We can use midi data to drive both lighting and sine waves with specific frequencies and amplitudes as given by the FFT (we used a lightbox which was created last year for visualization, the implementation of which will surely be covered by my colleague Sam Sutton). This midi conversion initially appeared to simply be a task of matrix transformation and resampling[xvii] (it was not, but this will be discussed to greater length in the results and conclusion sections of the paper) – making our original matrix into a midi compatible one, which could generate corresponding midi notes.




These midi notes could be sent to a polyphonic sine wave synthesizer to recreate this tonal-imprint, and we affectionately named these recreations “tone-chords”– a play on the jazz-theory emphasis on chord-tones/colour-tones, and an acknowledgement that our midi output was chordal, though generated from the tone-colour of a single note. This midi data could also be sent to a visualizer, such as the lightbox created by the previous cohort of the course, though the mappings from midi to light vary depending on the constraints of the lightbox. In the case of our lightbox, it was limited to one octave, and without velocity sensitive/ dimmable LED's, so we arpeggiated our output, so that we could cycle through the octaves.

Limitations of these methods:

Since this process is now taking audio into the digital domain and back, the limitations of Digital Signal Processing (DSP) must be acknowledged. Even if we disregard latency and consider that I am running a system at 24bits with a 48Hz sampling rate, digital artefacts must still be acknowledged. As one approaches the Nyquist frequency (half of the sample rate), the number of samples per waves becomes such that the negative part of the oscillation is lost, and so the wave can be reconstructed as being of a much lower frequency. The frequency perceived by the computer reflects back down from this point and can start to cause interference/inter-modulation (known as aliasing) before/during processing[xviii]. For this reason, it is important to use low pass filters at this frequency throughout an implementation, or to: Filter -> upsample during processing -> filter -> down-sample before output. However, there are still issues of intermodulation which may occur even before processing, and below our Nyquist frequency[xix] .

## RESULTS AND REFLECTIONS

The spectral shaping aspect of our project worked extremely well, as demonstrated by the use of the effects (demonstrated in our presentation as well as in this appendix). This also has uses for the sound design world, as the methods used to recreate a tonal imprint with white noise could be used to subtractively superimpose the tone-colour of a sound onto any other audio source.

However, our midi efforts were plagued by false positives, rounding errors, aliasing, as well as artefacts introduced by windowing, all of which made the midi implementation too unreliable for use. However, promising work from Lagrange, Marchand, and Rault (2007) describes methods of analysis which could be combined with our own to create more reliable polyphonic data, while the work of Tian Wenbo et al (2012) demonstrates how the implementation of alternative algorithms for both windowing and FFT can lead to more suitable frequency analysis. We could also have tried to eliminate nearby false positives (as can be found near the root in ***appendix vi***), with discrete methods, or using differential equations to estimate the turning points (and thus maximum frequency bins) for a region. If I were to do this project again, I would stay away from midi for these reasons, and instead pursue sonography as a means of visualisation. The visualisation of spectral data has long history with Jitter – a visual variant embedded in Max for live, and so extremely suitable for integration into the analysis we had already conducted at this point. Another route could be non-real-time processing, which affords us the opportunity to compare different window sizes, reducing the errors and false positives which we came across. This could also be developed to include the machine learning techniques explored by the other groups.

Our project's mapping was based on an equal tempered scale, but we know that overtones don't work exactly like this. And what about out of tune notes? If we used additional processing such as machine learning to estimate a fundamental frequency of the tone, then we could construct a just-intonated overtone scale above it and create a more corelated midi output by employing midi pitch bends protocols.

In the case that there is a complex tone which generates a high number of midi notes, we noticed that lights/notes would randomly disappear as the polyphony limit of the instrument/lightbox was reached. To avoid this in the future, one could implement a system to sort midi messages by velocity, so that the output message is ordered by priority, and the important partials don't disappear.

## CONCLUSIONS

As is demonstrated by the educational literature on multi-sensory education, it is a highly effective field which will surely begin to find its way into modern educational spaces. As a result, it is very important that an inherently novel, sensory field of study – such as music – is not left behind. Though the accuracy of some of our results was not as concrete as we would have hoped, the findings were still extremely exciting, as they proved that audio-visual spectral interaction is feasible for educational purposes. The interactivity of Max-For-Live helped us to prototype quickly, however for robustness I would suggest the development of lower level tools using languages such as JavaScript and C++ – both of which can operate within the max for live domain, should educators wish to imbue their programs with some interactivity. It is also worth noting that I am not a mathematician or computer scientist by trade, and so I would recommend that the primary take away from this project is a proof of concept for the synthesis of these processes, and their potential to create a powerful whole.

Word Count: 3020

# **BIBLIOGRAPHY**


Agarwal, Amit, S.n. Sur, Arun Kumar Singh, Hemanth Gurung, Abhishek Kumar Gupta, and R. Bera. 2012. 'Performance Analysis of Linear and Non-Linear Equalizer in Rician Channel'. *Procedia Technology* 4 (January): 687–91. https://doi.org/10.1016/j.protcy.2012.05.111.

Al-Zahrani, F. A., H. M. Mustafa, and A. Al-Hamadi. 2010. 'On Analysis and Evaluation of Multi-Sensory Cognitive Learning of a Mathematical Topic Using Artificial Neural Networks'.

Boer, E. de. 1976. 'On the "Residue" and Auditory Pitch Perception'. In *Auditory System: Clinical and Special Topics*, edited by E. de Boer, W. K. Connor, H. Davis, J. J. Eggermont, R. Galambos, C. D. Geisler, G. M. Gerken, et al., 479–583. Handbook of Sensory Physiology. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-66082-5_13.

Charlotte Desvages. n.d. *The NESS Project: Bowed String Synthesis*. Edinburgh, UNITED KINGDOM. Accessed 14 April 2020. https://www.youtube.com/watch?v=fQMJm-YMXuQ&feature=youtu.be.

Cooley, James W., and John W. Tukey. 1965. 'An Algorithm for the Machine Calculation of Complex Fourier Series'. *Mathematics of Computation* 19 (90): 297–297. https://doi.org/10.1090/S0025-5718-1965-0178586-1.

H. L. F. Helmholtz, and Alexander J. Ellis. 1886. 'On the Sensations of Tone, as a Physiological Basis for the Theory of Music'. *The Musical Times and Singing Class Circular* 27 (522): 481. https://doi.org/10.2307/3363877.

Haraux, Alain. 2018. *Nonlinear Vibrations and the Wave Equation*. SpringerBriefs in Mathematics. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-78515-8.

Heller Murray, Elizabeth S., Ashling A. Lupiani, Katharine R. Kolin, Roxanne K. Segina, and Cara E. Stepp. 2019. 'Pitch Shifting With the Commercially Available Eventide Eclipse: Intended and Unintended Changes to the Speech Signal'. *Journal of Speech, Language, and Hearing Research* 62 (7): 2270–79. https://doi.org/10.1044/2019_JSLHR-S-18-0408.

Johnston, Ian. 2009. *Measured Tones: The Interplay of Physics and Music, Third Edition*. Bosa Roca, UNITED STATES: CRC Press LLC. http://ebookcentral.proquest.com/lib/warw/detail.action?docID=1648337.

Lagrange, M., S. Marchand, and J.-B. Rault. 2007. 'Enhancing the Tracking of Partials for the Sinusoidal Modeling of Polyphonic Sounds'. *IEEE Transactions on Audio, Speech, and Language Processing, Audio, Speech, and Language Processing, IEEE Transactions on, IEEE Trans. Audio Speech Lang. Process.* 15 (5): 1625–34. https://doi.org/10.1109/TASL.2007.896654.

Lewis, W, and Peter D Welch. n.d. 'Historical Notes on the Fast Fourier Transform', 3.

Rao, K.R., D.N. Kim, and J.-J. Hwang. 2010. *Fast Fourier Transform - Algorithms and Applications*. Signals and Communication Technology. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-6629-0.

Sansaloni, T., A. Perez-Pascual, V. Torres, V. Almenar, J.f. Toledo, and J. Valls. 2007. 'FFT Spectrum Analyzer Project for Teaching Digital Signal Processing With FPGA Devices'. *IEEE Transactions on Education, Education, IEEE Transactions on, IEEE Trans. Educ.* 50 (3): 229–35. https://doi.org/10.1109/TE.2007.900025.

Schneider, Albrecht, and Klaus Frieler. 2009. 'Perception of Harmonic and Inharmonic Sounds: Results from Ear Models'. In *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music*, edited by Sølvi Ystad, Richard Kronland-Martinet, and Kristoffer Jensen, 5493:18–44. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-02518-1_2.

Siedenburg, Kai, Charalampos Saitis, Stephen McAdams, Arthur N. Popper, and Richard R. Fay, eds. 2019. *Timbre: Acoustics, Perception, and Cognition*. Vol. 69. Springer Handbook of Auditory Research. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-14832-4.

Tian Wenbo, Yu Jianming, Ma Xiaojin, and Li Ji. 2012. 'Power System Harmonic Detection Based on Bartlett-Hann Windowed FFT Interpolation'. *2012 Asia-Pacific Power and Energy Engineering Conference, Power and Energy Engineering Conference (APPEEC), 2012 Asia-Pacific*, March, 1–3. https://doi.org/10.1109/APPEEC.2012.6307426.

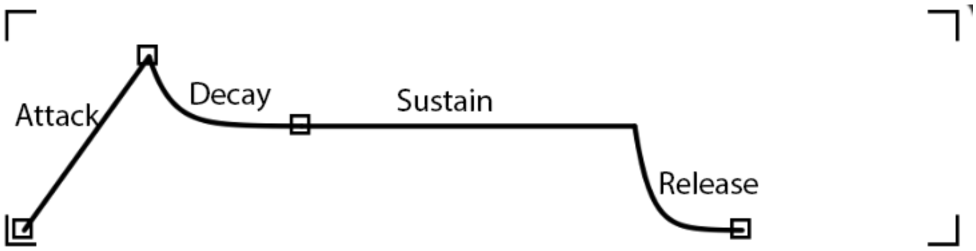
Välimäki, Vesa, and Jan-Markus Holm. 2000. 'Modeling and Modification of Violin Body Modes for Sound Synthesis'. In , Volume: vol. 4:2229–32. Tampere, Finland.

Worrall, Dan. 2013. *EQ: Linear Phase vs Minimum Phase*. https://www.youtube.com/watch?v=efKabAQQsPQ.

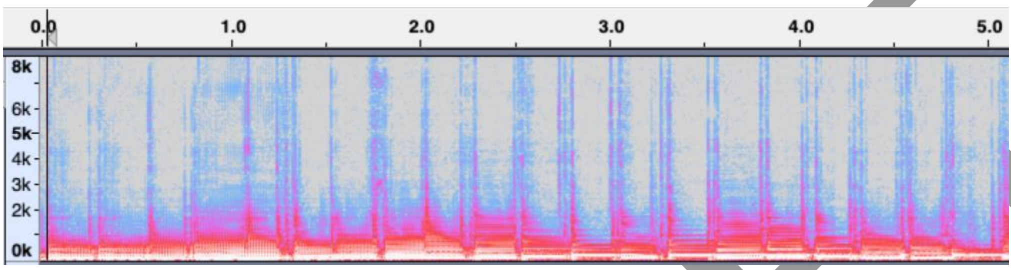
———. 2020. *Samplerates: The Higher the Better, Right?* https://www.youtube.com/watch?v=-jCwIsT0X8M&t=704s.

## APPENDICES

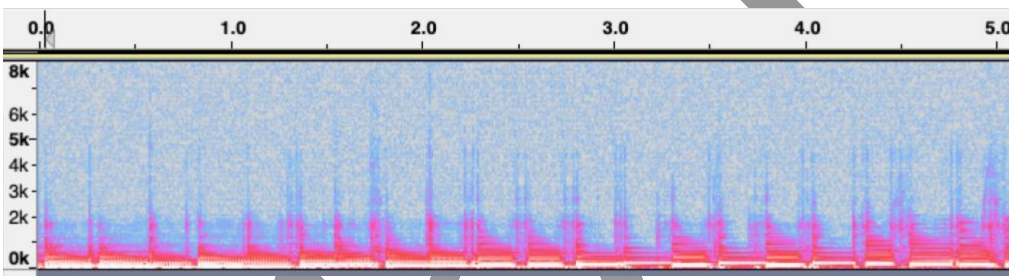
[i] ADSR Diagram:



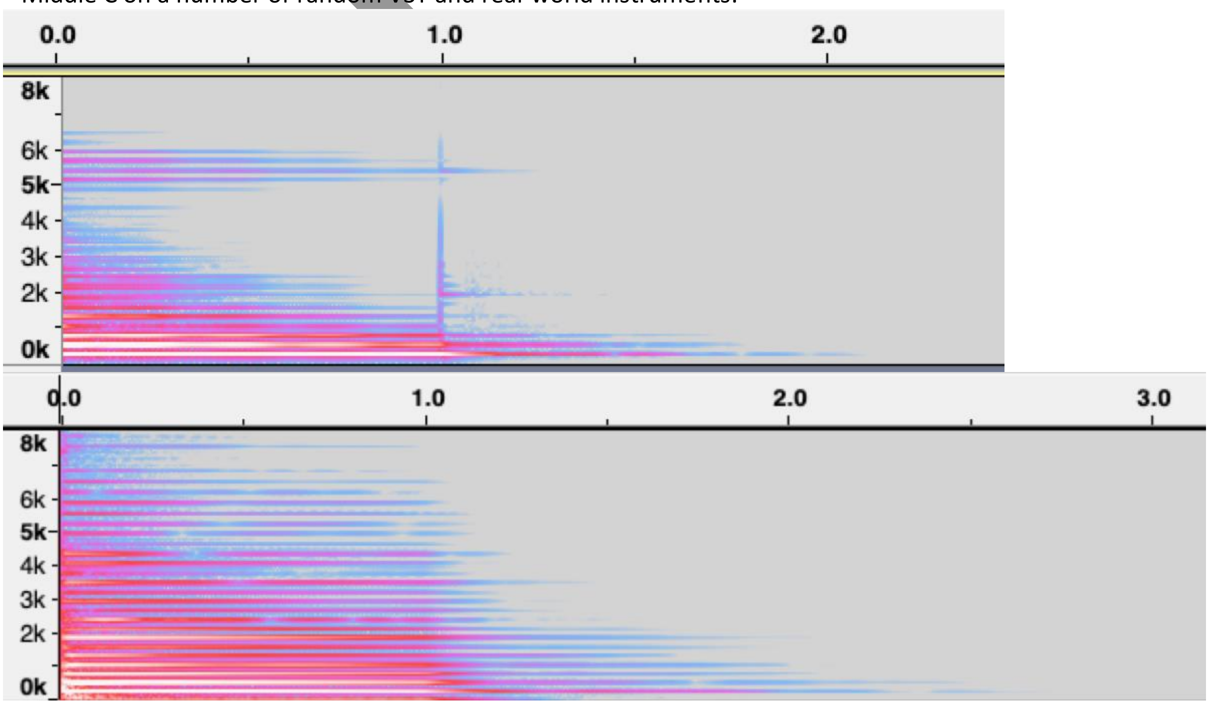

[ii] Spectrogram (frequency against time, heat as amplitude) for Acoustic Double Bass, generated in audacity:

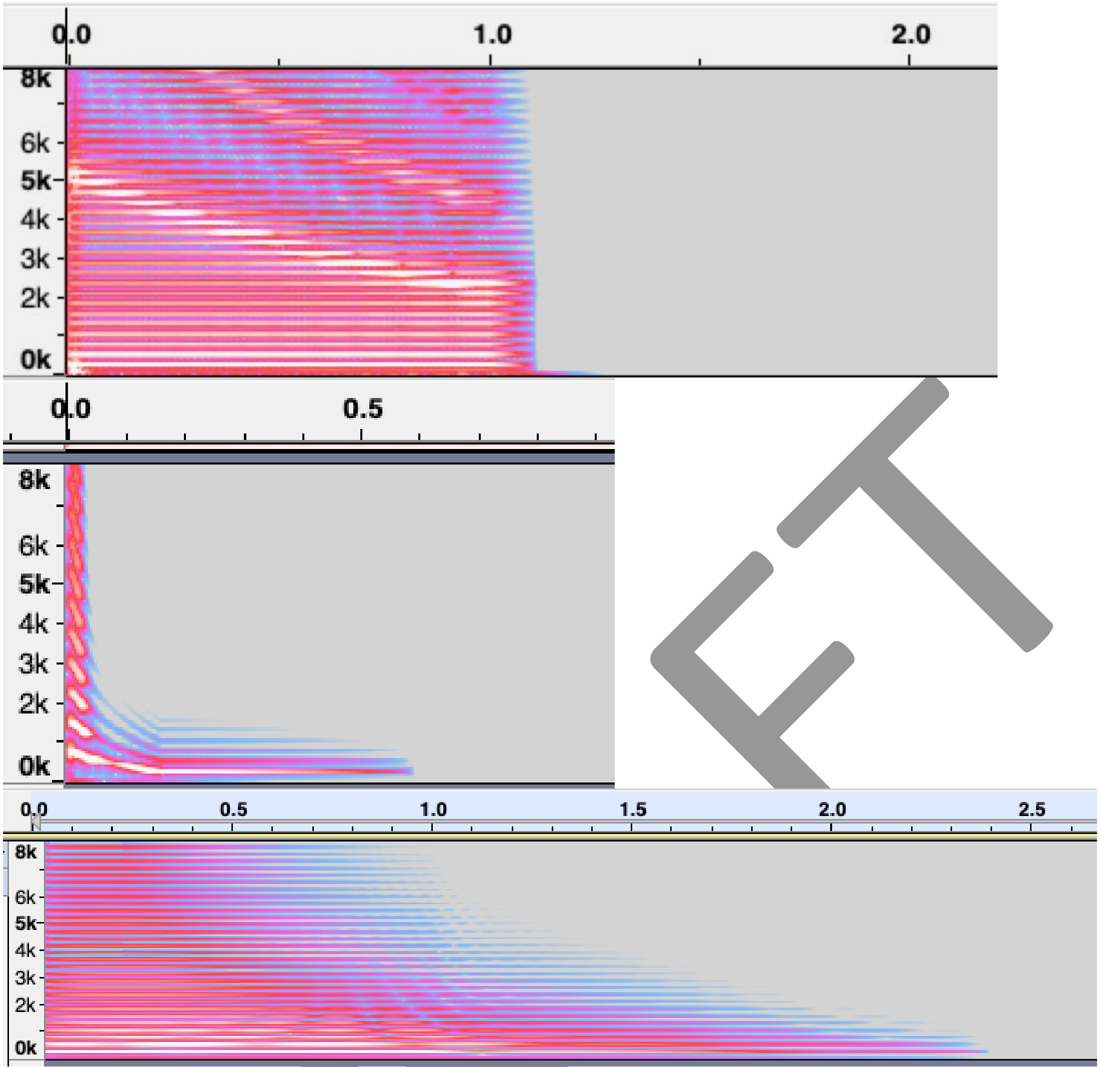



Spectrogram for Electric Double Bass:



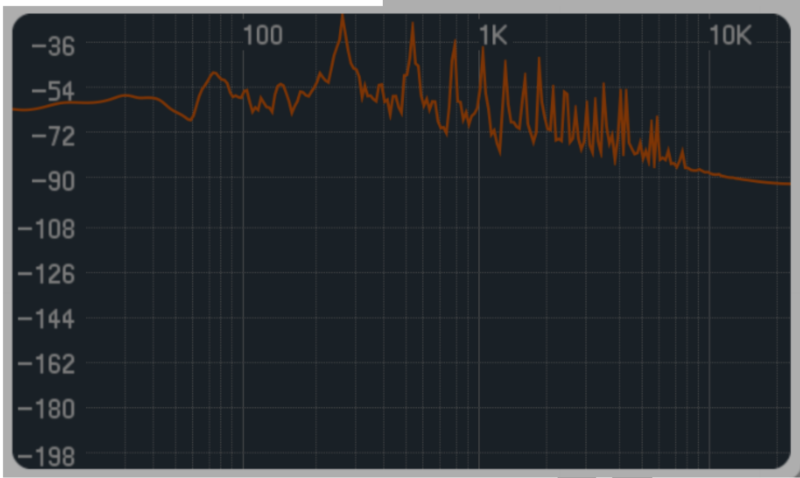

[iii] Middle C on a number of random VST and real-world instruments:



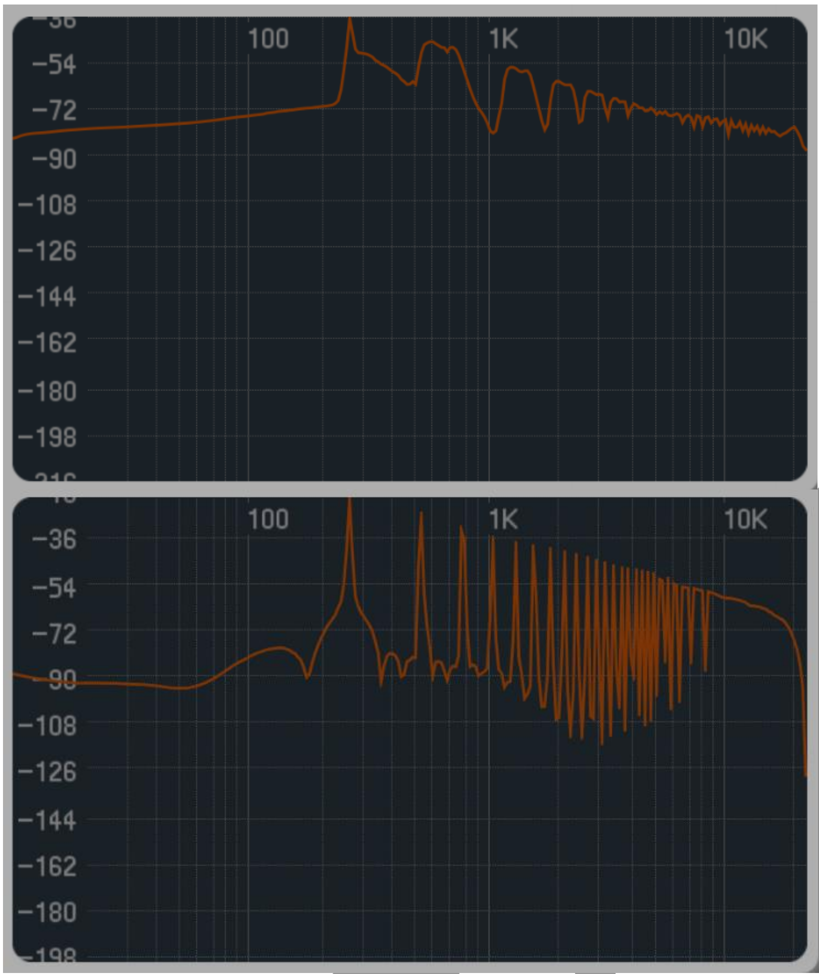



[iv]The above instruments but without time constraints (purely amplitude/frequency domains), generated in Ableton:


100
1K
10K
−36
−54
−72
−90
−108
−126
−144
−162
−180

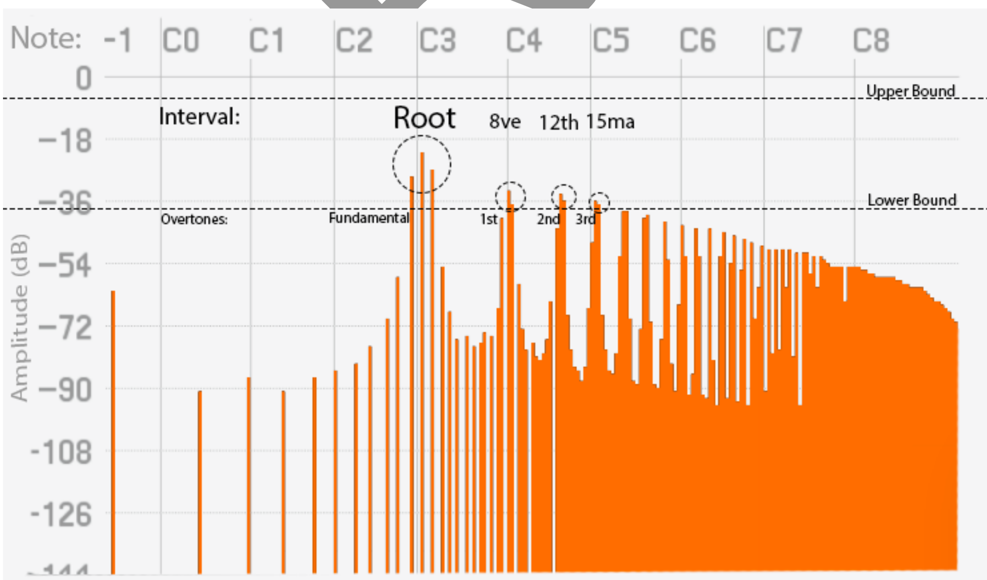

100
1K
10K
−36
−54
−72
−90
−108
−126
−144
−162
−180
−198


100
1K
10K
−36
−54
−72
−90
−108
−126
−144
−162
−180




[v] Despite using "quality" instead of "tone-colour" he made the same distinction in contrast with Johnston

[vi] Diagram to demonstrate discretisation of frequencies into pitch classes after filtering:

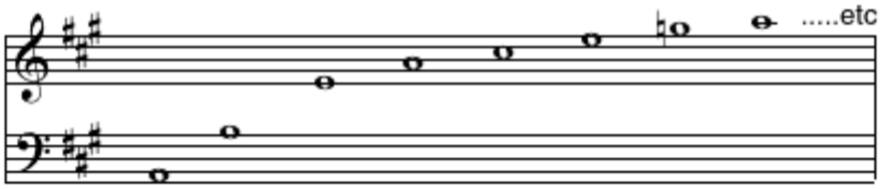


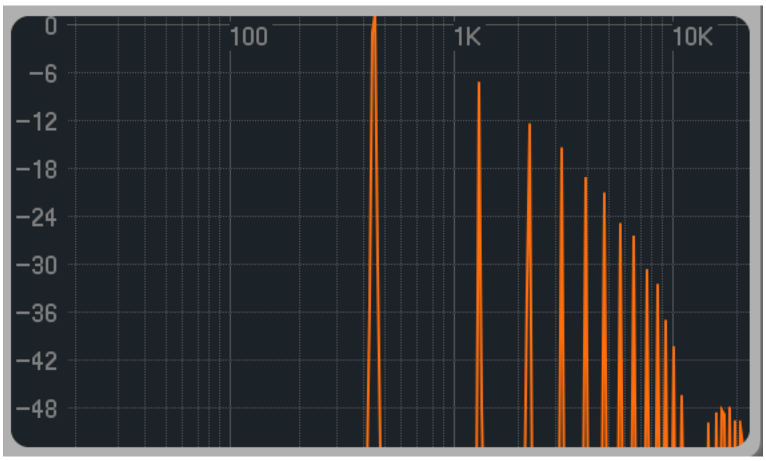

[vii] Please see *iv* for examples of tonal imprints

[viii] The harmonic series of overtones (adapted from measured tones 46 (cite))

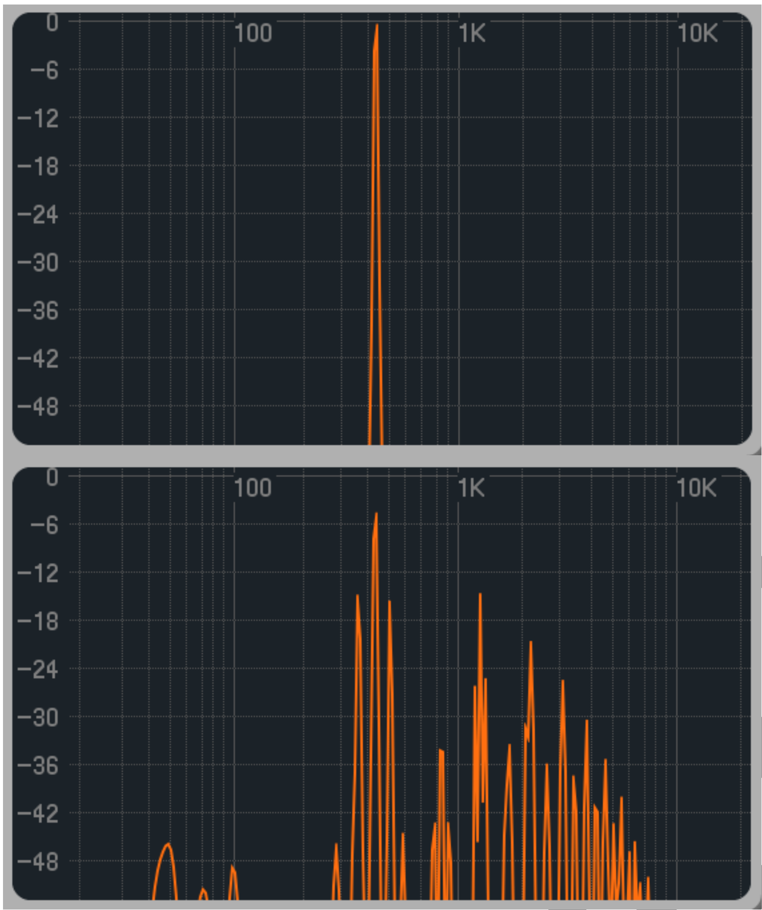
| | FREQUENCY | NOTE |
|---|---|---|
| fundamental | 110 Hz | $A_2$ |
| 1st overtone | 220 Hz | $A_3$ |
| 2nd overtone | 330 Hz | $E_4$ |
| 3rd overtone | 440 Hz | $A_4$ |
| 4th overtone | 550 Hz | $C\sharp_5$ |
| 5th overtone | 660 Hz | $E_5$ |
| ...etc. | | |

Or, on a musical stave:




[ix] Sine wave before and after tape saturation and distortion




[x] The DFT is defined by:

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{i2\pi}{N}kn}$$
$$= \sum_{n=0}^{N-1} x_n \cdot \left[\cos\left(\frac{2\pi}{N}kn\right) - i \cdot \sin\left(\frac{2\pi}{N}kn\right)\right],$$

(Rao, Kim, and Hwang 2010, 5)

Thanks to a number of optimizations (further described by Rao et al), this DFT can be performed in real time, using the Fast Fourier Transformation (FFT) and its Inverse (IFFT) (Lewis and Welch, n.d.). In order to deal with non-periodic source waves, we use a cosine wave to introduce pseudo-periodicity into the signal with a Hamming window function, the details of which are too involved for me to explain here, but are outlined wonderfully in both the max documentation (https://docs.cycling74.com/max5/tutorials/msp-tut/mspchapter25.html) and by Sanzaloni et al (2007).

After making our input signal a periodic one, then conducting our FFT, we get two sets of cartesian coordinates, a real and imaginary component, which are then converted to polar using the following transformations:

$$r = \sqrt{x^2 + y^2}$$

$$\varphi = tan^{-1}\left(\frac{y}{x}\right)$$

Where x and y are the real and imaginary outputs of the Fourier transform respectively

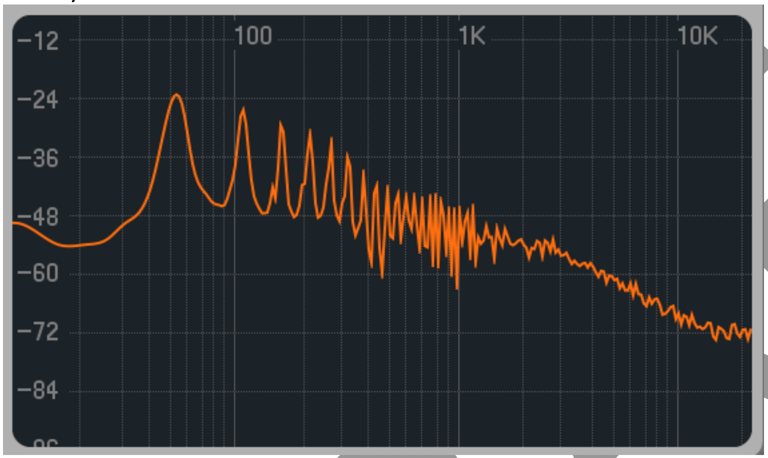
Where – for each frequency bin – r and φ correspond to the amplitude and phase (displacement of the start points of the wave in relation to a neutral 0π radians spot) of each waveform. I would recommend that one does not make this "signal" audible (unfortunately this recommendation is based on personal experience). Since the data is discrete, with a set of floating points for each frame (the number of samples in the frame is user determined and a power of 2) it makes sense to write it into an array for processing. We can do this by storing data for each frequency bin according to the following formula:

Matrix Index (MI) ⇔ "(MI +1) * Sampling rate/Frame size" Hz,

where the amplitudes for each frequency can be mapped to midi velocity logarithmically.

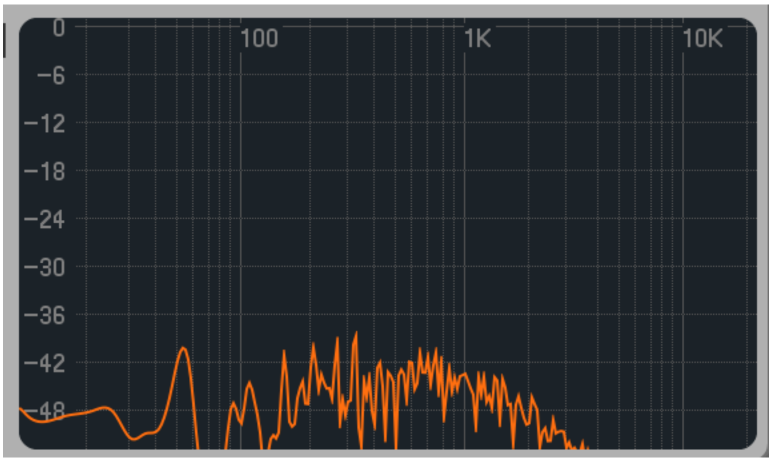
Note that, as a result of the limitations of the Nyquist frequency, the latter half of the matrix is just a reflection of the first, and so must be discarded.

[xi] Noisy sub bass before filter:




After amplitude/frequency high-pass filter:


Noise which has been filtered out:

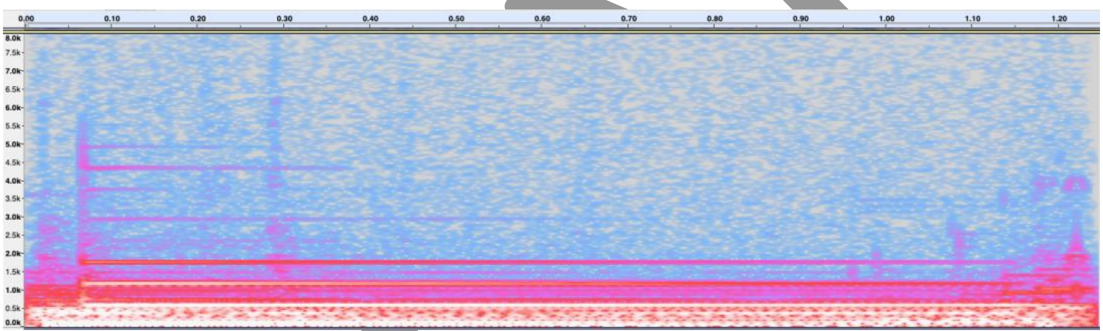



[xii] Using jit.catch. Since we are using matrix operations, our audio is technically in the video processing domain of jitter, and so, like a video, for every window of the FFT, a single matrix of data is generated.

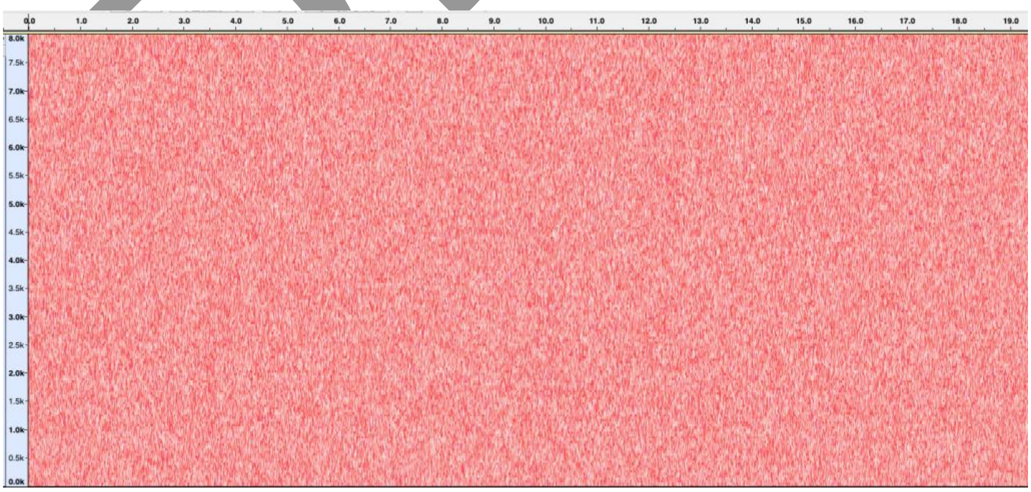
[xiii] This trick of frequency-domain matrix manipulation within the time domain is really exciting, *and definitely an opportunity for further research and experimentation! However, I did not find much on this in the reading I conducted for this essay.*

[xiv] Fig for spec accumulator: Before, white noise, after filtering

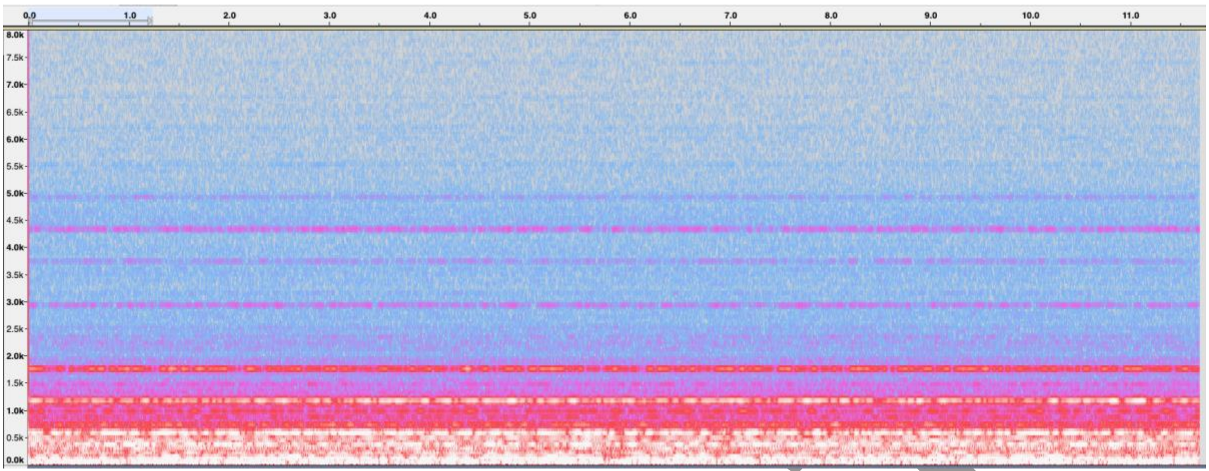
Spectrogram of source material:



White noise before subtractive filtering:



After processing – note that it has been made to last the length of the white noise, but with the tonal character of the source

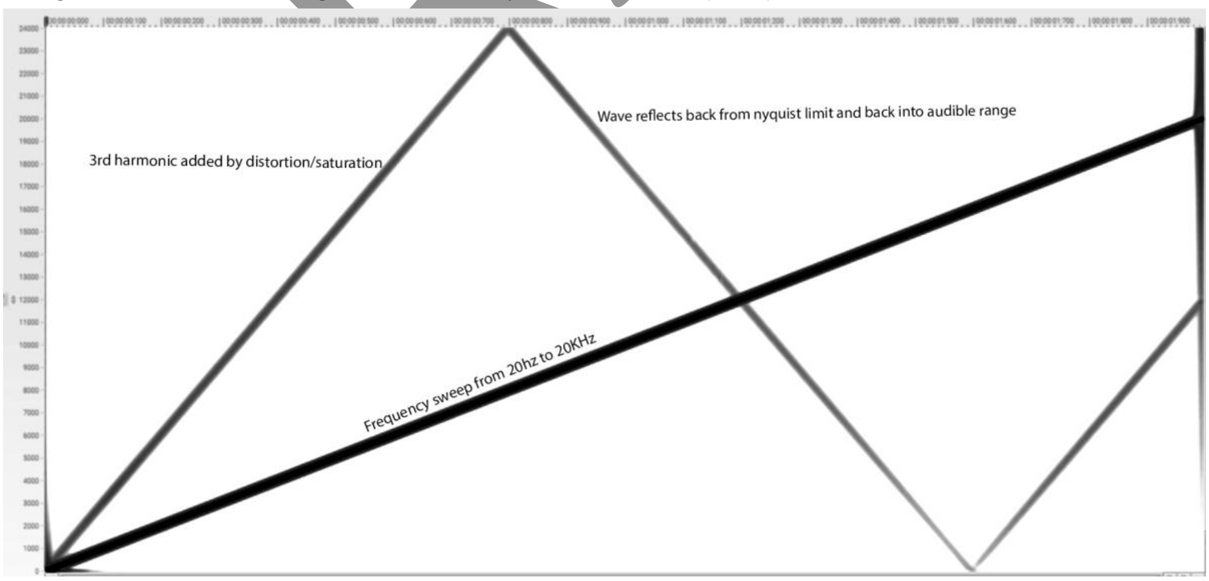


[xv] There is an issue with this approach however, when harsh eq's are used they have the potential to introduce phase inversions to the signal (Agarwal et al. 2012).

[xvi] For more discussion of phase shifts and artefacts introduced by linear and non-linear equalization, see Agarwal et al's paper (2012), as well as an excellent educational video by Dan Worral (Worrall 2013)

- [xvii] The average matrix of size M was to be mapped to a midi matrix of size 128 (one channel of midi notes from 0 to 127), with the amplitudes being mapped from the logarithmic decibel to the linear/exponential (depending on the synthesiser) midi velocity scale (from 0 to 127). This was achieved using dynamic floor/ceiling functions for our audio range, and then applying an inverse log and appropriate scaling to the midi domain

[xviii] Fig to demonstrate aliasing/reflections, adapted from Worral (2020)




[xix] If we sum a number of sine waves (which are all below our Nyquist frequency), and their frequencies and phase are such that in a non-linear (or somewhat distorted, as most systems are) system, they sum to produce an additional wave – which has a frequency above the Nyquist frequency, then this wave would be reflected down, eventually back into the audible/processable domain (as per the diagram above) – AND survive the low-pass filtering at all stages of processing. Aliasing distortion and inter-modulation are resultantly extremely difficult to avoid with complex analogue signals, regardless of the sampling rate.