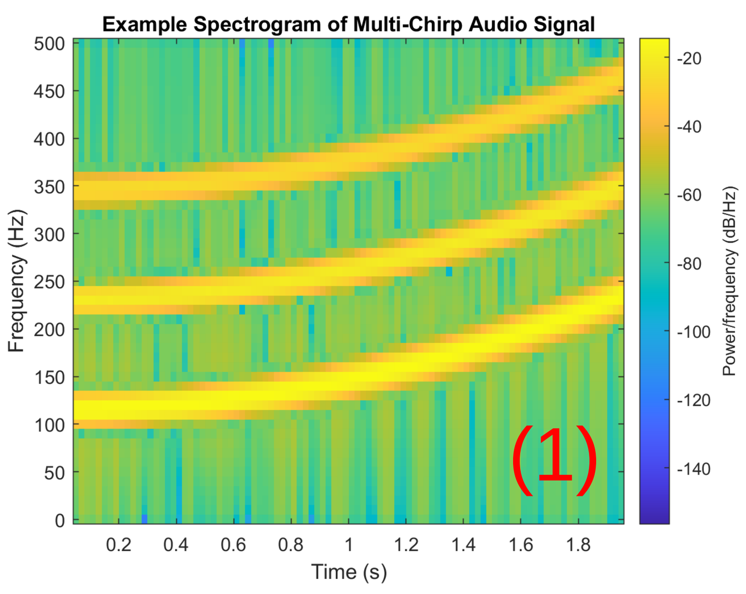
Exploring Spectrogram-based Audio Signal Representation and Synthesis

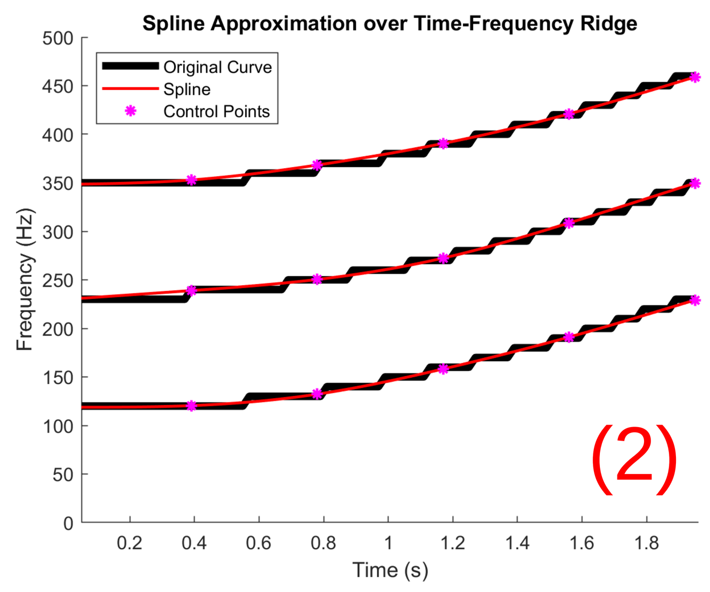
Josh Falejczyk
Research Mentor: Dr. Paul Cuffe
School of Electrical and Electronic Engineering
University College Dublin, Dublin, Ireland

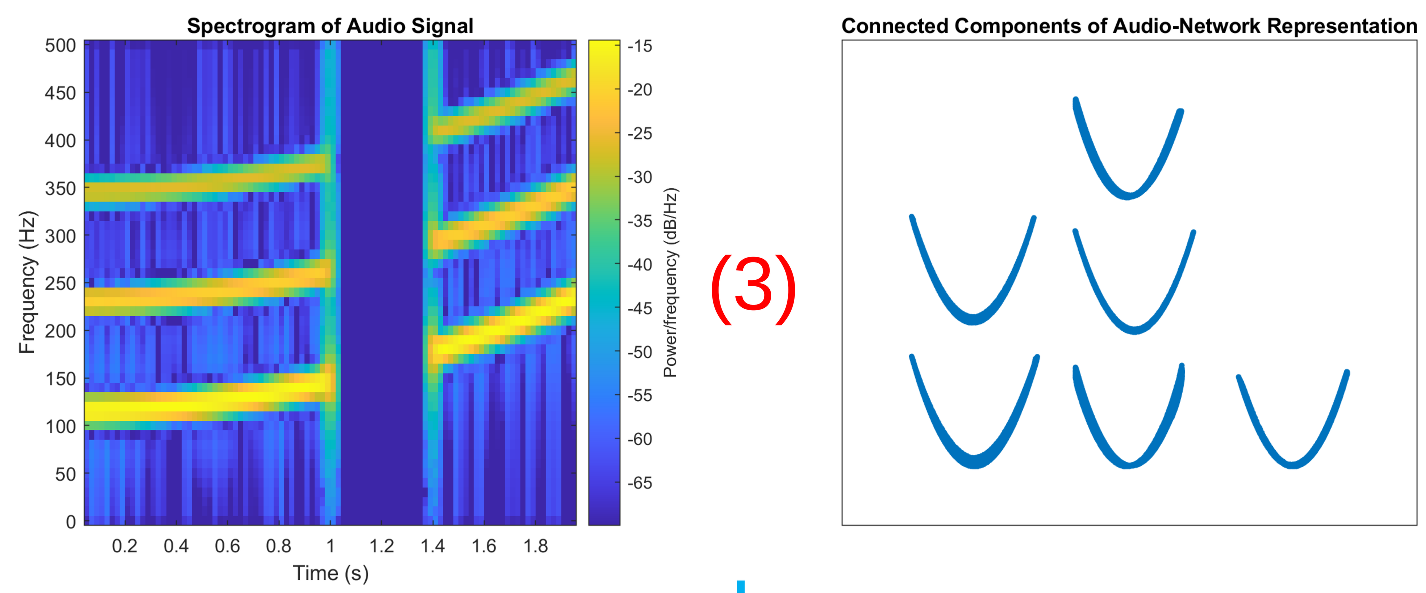
Introduction

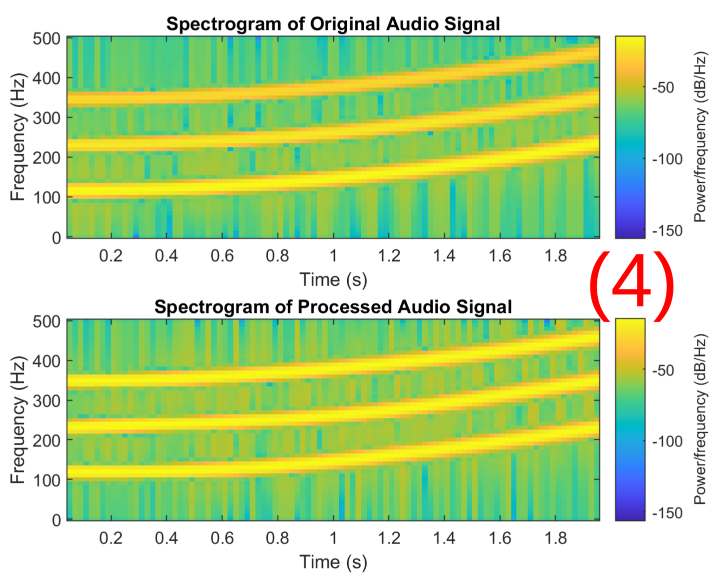
Audio signals can be represented and analyzed in a variety of ways, one of which being a spectrogram. Typical audio representations show amplitude varying on a time domain, whereas spectrograms visualize frequency varying on a time domain (**Figure 1**). This research explored what occurs to an audio signal if its spectrogram data was extracted, converted from a raster representation to a vectorized one, and synthesized back into audio. The implications of this research may not only offer unique audio signal augmentations, but also alternative methods of audio signal storage.

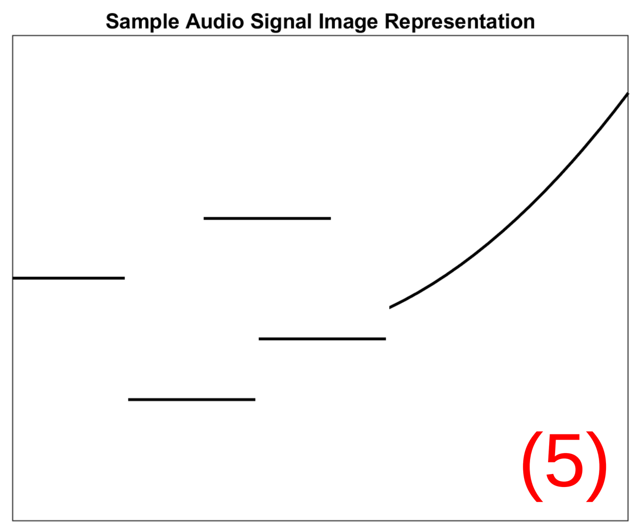
Figures











Conclusion

Overall, the basic questions of this research are answered by stage one, and expanded upon by stages two and three. Looking towards the future for image-audio representation, stage three's image input support opens the doors for image-based storage of audio signals that emulate spectrograms (**Figure 5**). While the storage space efficiency of these images vs regular audio signals was not within the research's scope, **Figure 5** also serves as an example for alternative musical notation. As for stage three's future, our working goal is precise reconstruction from complex realistic audio signals.

Methodology

Development Stage One:

1. Generate spectrogram from basic artificial audio signal.
2. Receive curve data by extracting time-frequency ridge of each harmonic.
3. Vectorize curve data by generating splines from input points (**Figure 2**).
4. Reconstruct audio from spline approximations and compare to original.

-Good start, but what about audio signals whose spectrograms have gaps of silence within each harmonic?

Development Stage Two:

1. Similar approach to stage one, but abandon time-frequency ridge extraction in favor of representing power density data as nodes of a graph/network **(Figure 3)**.
2. Vectorize curve data by generating splines from network connected components input points.
3. Continue from stage one step four.

-Another good improvement, but what about complex, realistic audio signals with hundreds of connected components?

Development Stage Three (plus):

1. Similar approach to stage two, but first use image processing segmentation to break up spectrogram into connected components.
2. Create networks for each connected component and filter out unwanted irregularities within these networks.
3. Continue from stage two step two.

Results

After ten weeks of development, stage one has been fully completed, stage two has been halted in favor of stage three, and stage three will continue to be worked on as its features are subject to change. One initial goal of this research was to see what would occur to an audio signal if ran through the specified process, and with stage one's completion, processed sample input has been achieved (**Figure 4**). As seen in the comparative spectrograms, very little has changed with the audio signals, likely due to the precision of the splines. Lowering precision does bring drastic signal augmentations to the degree of qualifying as a musical effect, though this was a byproduct not within the scope of the research. As for alternative methods of audio signal storage, stage three supports image input, leaving that goal partially achieved too.

Acknowledgements

I would like to truly thank Dr. Paul Cuffe for his willingness in taking on a project that pushes the boundaries of both of our areas of expertise, openness towards unorthodox ideas, and steadfast patience and guidance throughout this semester. I would additionally like to thank UCD and Dr. Kevin Nolan for this wonderful research opportunity in my time abroad.