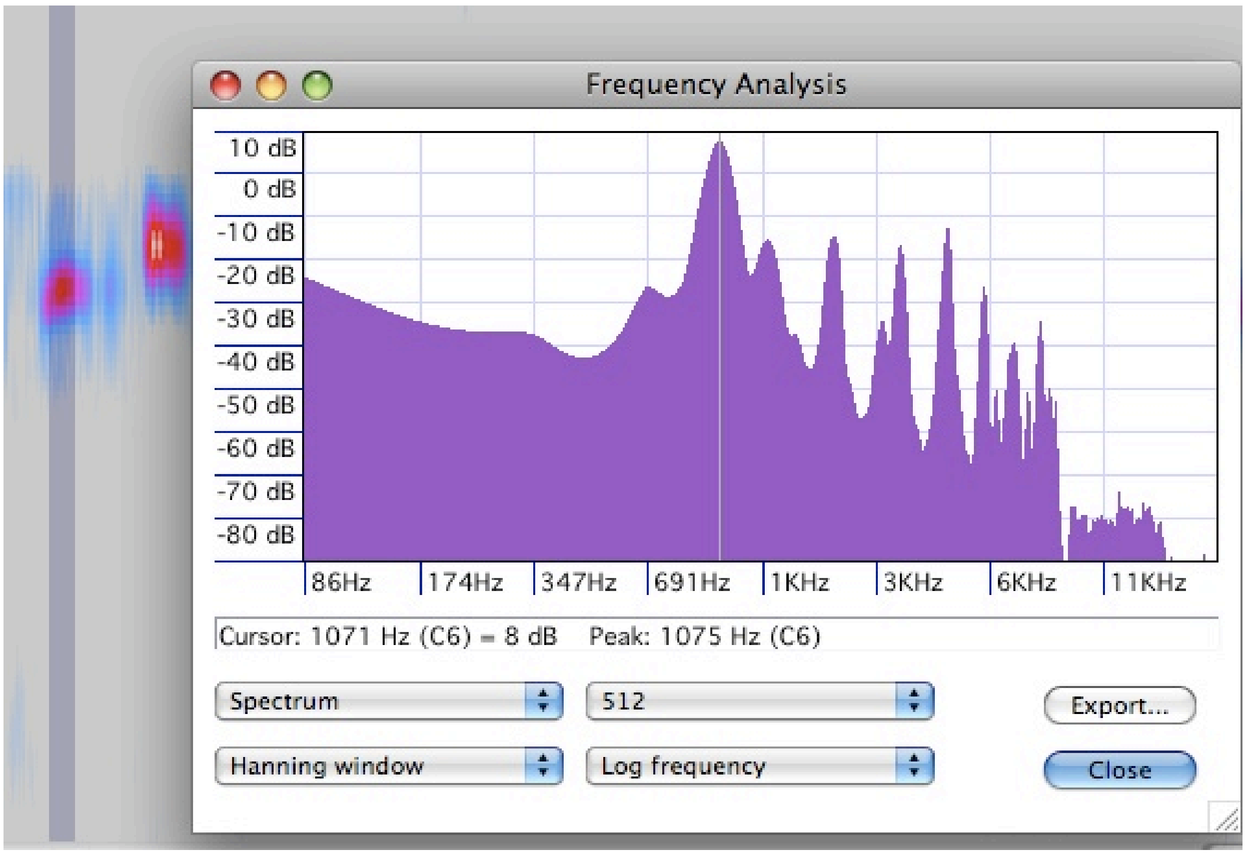
Mini Project 2: Singing Synthesis

Jonathan Ward

The objective of this mini project was to improve the quality of opera singing synthesized by a Pure Data patch. The patch uses sinusoidal modeling and synthesis to create an artificial version of Mozart's "Magic Flute" sung by Diana Damrau. A short phrase, "ah" sung as an A5 note was analyzed with the SPEAR sinusoidal analysis tool. SPEAR takes in sound files and uses Fourier analysis to separate them into sinusoidal tracks with frequency and amplitude information to obtain a good approximation of the source. To synthesize "Magic Flute," the Pure Data patch took the sinusoidal tracks from the A5 note and scaled their frequencies to create different pitched versions of the same phrase.

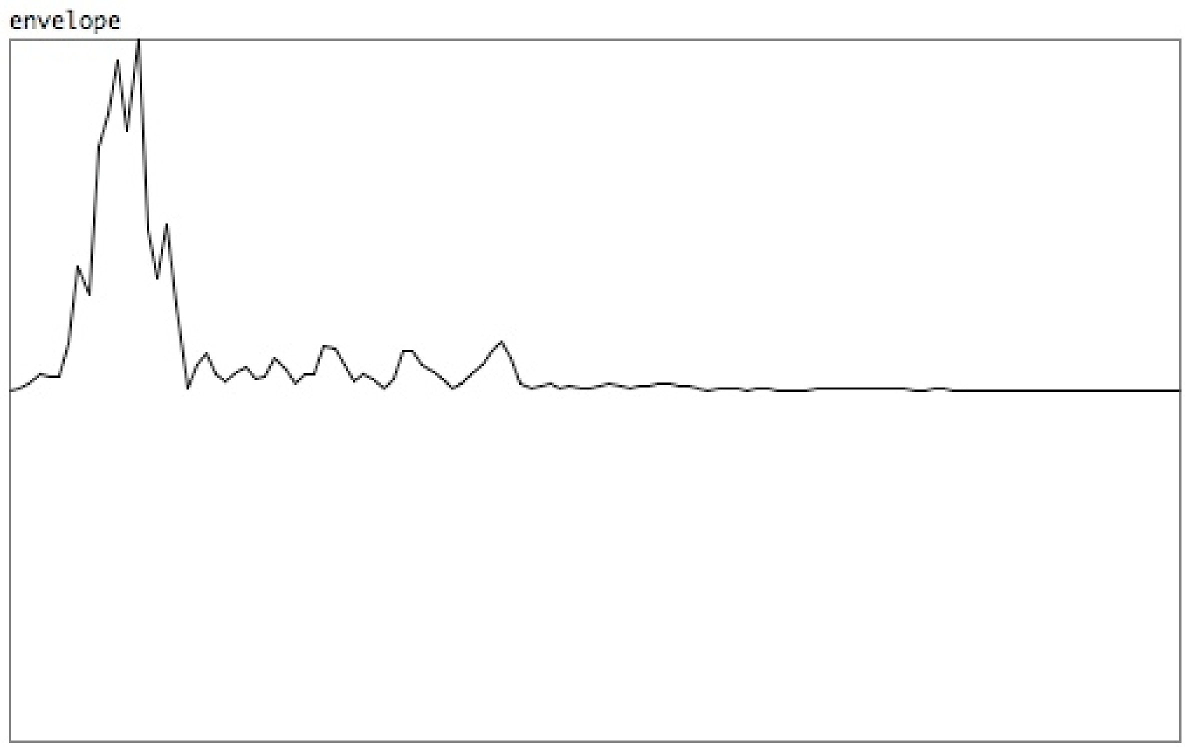
This sounds like a reasonable approach, but the resulting synthesis doesn't sound very lifelike. The problem is that scaling the sinusoidal frequencies while keeping the amplitudes constant doesn't do a good job of modeling how sound is produced by humans. To the first order, the vocal chords produce a comb of equal amplitude, equally spaced tones that are harmonics of some fundamental frequency. These tones are then shaped by the throat and mouth to create different sounds. Each short phrase or phoneme can be associated with a different filter shape. One can't simply scale the frequency of a phrase to generate a new pitch because the filter formed by the throat and mouth would change the amplitudes of the harmonics. This is why the original singing synthesis patch doesn't sound real. A different pitch sung by a real singer would expose different parts of the throat and mouth filter.

To create a better synthesis, one has to analyze more than just one phrase sung at a single pitch. To synthesize a phrase at all pitches, the filter shape of the throat and mouth for that phrase should be extracted somehow. This is essentially what I did to improve the original patch. To extract this filter shape, I used the audio program Audacity to record the spectrum of different pitches sung by Diana Damrau. To form the final filter envelope, I used the maximum amplitude contained in the frequency bins of several different pitches.




While crude, this seemed to be the best method given the available data. The absolute best way to extract the filter envelope would be to analyze the singer's voice as she sings a phrase over her whole vocal range. The method I used has the drawback that it also includes some of the accompanying symphony.

Once I had this filter envelope, I still had to find some way to apply it to the signal. Originally, I discarded the amplitude information of the SPEAR analysis file altogether and just used the amplitude of my filter envelope. This didn't sound any better or worse, just different and equally artificial. I think this is because SPEAR extracted subtle changes in amplitude with A5 phrase that my envelope wasn't be able to recreate.




To keep these subtle changes while still adding filtering for different pitches, I made the compromise of multiplying the SPEAR analysis amplitude with the amplitude of my extracted filter envelope. This method produced decent results, but still isn't optimal. In a sense, the filter is getting applied twice. For the A5 pitch, it gets applied first by the amplitude table from the SPEAR analysis file and then by the extracted envelope. For other pitches, an erroneous filter is applied by scaling the frequency data from the analysis file without changing the amplitude and then the extracted throat and mouth filter gets added after.. The extracted filter doesn't undo the damage done by original bad filter, though. What I really need to do is to apply a filtering function that takes the amplitude information from the SPEAR file and scales it so that the pitch harmonics have the correct amplitude in relation to the A5 pitch.

I actually tried this on my first pass at improving the synthesis, but ran into trouble. I tried to create a harmonic map of different pitches with amplitudes relative to the A5 pitch harmonics. This was an arduous process that required a lot of manual extraction of harmonic frequencies and amplitudes. I soon realized that besides being very tedious work, this technique wouldn't be able to help with pitches that I might have overlooked.

I tried looking for a better to do this, but stopped when I stumbled upon the technique that forms the basis of my submission.