

# Real-Time Formant Tracker

CLAYTON KARDAS

University of Rochester  
clayton.kardas@gmail.com

## Abstract

*Formant filters can be used to help emulate vowel sounds on broadband signals such as impulses, noise, and multi-track mixes. This project was developed using MATLAB's Audio Test Bench feature for the real-time implementation and pre-existing information about the spectral envelopes of formants to shape the spectral output signal.*

## I. INTRODUCTION

Formants have many different meanings in the audio industry [Wolfe], but the definition closest to this project would be the four or five frequencies that serve as the resonant frequencies of the human vocal tract when pronouncing vowel sounds [Sundberg,1977]. Formant frequencies very much differ from the pitch of the voice, which is partly caused by the vocal folds. So, one could use the source-filter model to describe the production of vowel sounds through the human voice with the vocal folds as what causes the excitation signal and the resonant frequencies as the peaks of the spectral envelope. Tracking these frequencies and determining them in real-time has been a great challenge for many researchers that have been trying to improve devices such as hearing aids that boost the input signal at the presumed formant frequencies [Mustafa and Bruce, 2006]. Fortunately, there has been enough research to determine a protocol for formant frequencies that correspond to different vowel sounds [Piché and Peter, 1994]. It is this already solid information that is used to shape the spectral envelope of the formant filter.

## II. METHODS

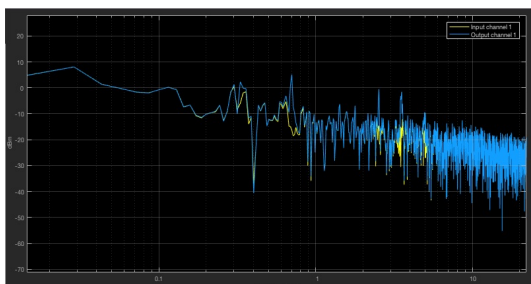
This formant filter is based around MATLAB and the Audio Test Bench for developing audio plug-in MATLAB classes. The plug-in also uses a look-up table from the Csound Manual [Piché and Peter, 1994]. The formants are cate-

gorized by the vocal range and the vowel sound they are associated with. The five vocal ranges are Soprano, Alto, Countertenor, Tenor, and Bass, whereas the vowels are A, E, I, O, and U. This gives 25 different combinations of vowel sounds. The Csound Manual also provides peak amplitudes for the five frequencies of each setting. In reference to Table 1, which has been copied from the Csound Manual, it is clear that the provided amplitudes begin with 0 for the lowest frequency and a negative number for the highest. A MultibandParametricEQ system object with 5 bands was used to filter the sound, along with sliders to vary each bands quality factor (Q), but when using the peak amplitudes provided from the table the results were not as expected. To solve this problem, I added the absolute of the highest frequencies amplitude suggestion and added the scalar to each of the amplitudes of their respective formant frequencies. A single slider for gain was also added to this absolute to control the filter's overall gain variability.

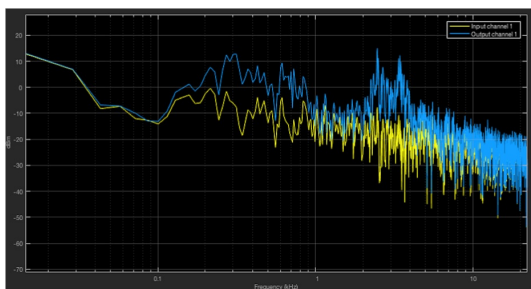
## III. RESULTS

Results that properly emulate vowel sounds are best seen with broadband signals such as noise, impulses, and multi-track audio mixes. Pure and harmonics tones do not react well with this filter because those frequencies do not always fall under the filter band's center frequencies. However, when they do, a vowel sound is produced. Compared to an impulsive signal such as a drum sample, vowel sounds are much more

noticeable in the difference of the input and output frequency spectrum and while listening to the output because the impulse frequencies are affected by each of the filter bands, thus creating a vowel sound. The formant effects are most noticeable with pink noise, which provides a fairly uniform spectral envelope. When filtered with the formant bands, the frequency spectrum and output signal clearly boost the formant frequencies that are sought after. Figure 1 shows an example of pink noise with an Alto U with high valued quality factors that represent minimal bandwidth. Out of the several different combinations of vocal ranges and vowels, Alto U showed the clearest difference for each band that was boosted by the formant filter. Figure 2 shows the same settings but with a lower quality factor to show broader bandwidth for each band.



**Figure 1:** *Alto U: [325, 700, 2530, 3500, 4950] Hz, High Quality Factor*



**Figure 2:** *Alto U: Low Quality Factor*

#### IV. FUTURE WORK

Only having 25 different combinations of formants is limiting. To further innovate on the amount of formant sounds that can be produced

it would be well to implement dynamic interpolation between the various sounds when moving the slider bars. Currently, the top two sliders that represent the vocal range and the vowel sounds are represented by the integer interval (1, 5). If the integers were turned to doubles and there was interpolation between the frequencies based on the sliding doubles, then this would enable to user to explore more combinations of possible formant sounds.

Another goal is to track the formants from a user's own vocal sample. The formants from the user sample could then be placed as the center frequencies for the bands and allow the user to alter the formants to achieve the ideal sound design of their output.

The best way of doing this would be to take the sample, isolate the section of the sample that represents the formants, track the formant frequencies, and then use those frequencies as the center frequencies of the parametric bands. These steps have already been done, but there needs to be a bridging of the gap between the methods of collecting frequencies and implementing them in an audio plug-in. Because most audio plug-ins only have one input and output, whereas this proposed plug-in would need two inputs, one from the vocal microphone, and the other from the source that is meant to be filtered.

#### REFERENCES

- [Mustafa and Bruce, 2006] Mustafa, K. and Bruce, I. C. (2006). Robust Formant Tracking for Continuous Speech With Speaker Variability *IEEE Transactions on Audio, Speech and Language Processing* *IEEE Trans. Audio Speech Lang. Process.*, 14.2:435–444.
- [Piché and Peter, 1994] Piché, Jean, and Peter J. Nix (1994) FORMANT TABLE *HTML Csound Manual*, Web.
- [Sundberg,1977] Sundberg, Johan (1977). The Acoustics of the Singing Voice *Sci Am Scientific American*, 236.3:82–91

[Wolfe] Wolfe, Joe. Formant: What Is a Formant? *The University New South Wales*, Web.