

# JACO - HIGH QUALITY SINUSOID TRACKS AND ENVELOPES

Christoph Reuter, Herbert Griebel

Musicological Institute  
 University of Vienna, Austria  
[christoph.reuter@univie.ac.at](mailto:christoph.reuter@univie.ac.at)  
[griebel@gmx.at](mailto:griebel@gmx.at)

## ABSTRACT

This paper describes ways how to accurately identify sinusoids and how to calculate and transfer spectral envelopes from one sound to another, both with high quality.

## 1. INTRODUCTION

There are many tools that can analyze sound signals using algorithms. However, algorithms are sometimes not perfect, not because they are implemented badly, but because the algorithm is inferior when compared with a human doing the analysis by ear and visual inspection. Real world sinusoids of a trombone for example are not steady, are intermittent, but the overall pattern still makes a perceptual sinusoidal tone. Because we are lacking an algorithm that is capable of tracking such sinusoids reliably, the only resort is sometimes to identify such sinusoids with human interaction. Tools that close the gap between human interaction and working on the signal manually do almost not exist. Jaco Visual Signal is a tool that tries to fill this gap [1]. Two functionalities that are based on human interaction are described in the following.

## 2. TRACKING SINUSOIDS

Identifying sinusoids and tracking them is a difficult task in dense environments or when the sinusoid is far from the ideal mathematical formula. We show how this is done manually using a tracking algorithm.

### 2.1 An example

The figure below shows sinusoids of a trombone at the beginning of the sound using a Fourier spectrum in decibel and a Kaiser window with a length of 100 milliseconds (see Figure 1).

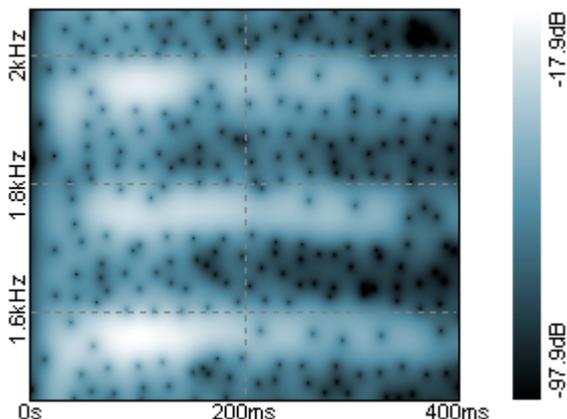


Figure 1: Sinusoids of a trombone, black colors correspond to small amplitudes.

The sinusoids near 1.5 kHz and near 2 kHz are both starting at time  $t=0$  seconds and have no steady amplitude. Both of them even have a gap in the amplitude and continue somehow steady for only a short time before they exhibit gaps again. If sinusoids are somehow steady in their amplitudes, identifying them is simply solved by interactively clicking on the sinusoid in a spectrogram view and a simple tracking algorithm with constraints for amplitude and frequency changes.

### 2.2 Analysis window

The analysis window for the Fourier transform is crucial to get a good time-frequency representation of the sound signal. Interference terms should be a minimum. The length of the window and the window function are equally important. A long window can actually be short if the window function has very low values at the tapering ends. It is important to understand that a long window will smooth out in time, and a broad window in the frequency domain will smooth out in frequency. A window can be designed to smooth out in time and frequency, e.g., by multiplying the window with a *si* function (or *sinc* function) to make it broader:

$$\tilde{w}(n) = w(n) \frac{\sin(an)}{an} \quad (1)$$

The figure below shows the same sound as in Figure 1, but using a longer window with 300 milliseconds.

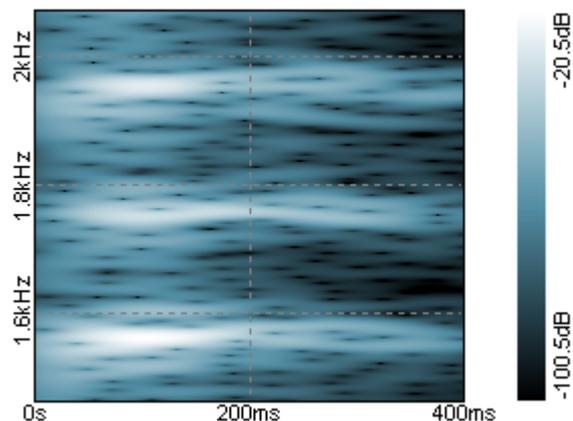


Figure 2: Sinusoids with a longer window, black colors correspond to small amplitudes.

Now with the longer window the beginning of the sinusoid looks much better and is track-able, i.e., a simple tracking algorithm with constraints will be able to follow the frequency path. However the sinusoid exhibits now intertwined multiple

frequency tracks after 200 milliseconds. A tracking algorithm will fail here again.

### 2.3 Wigner distribution details

If we take a closer look at the sinusoid which is a little below 2 kHz, we see an almost steady sinusoid which splits up into 2 sinusoids a little before  $t = 200$  milliseconds (see figure 3).

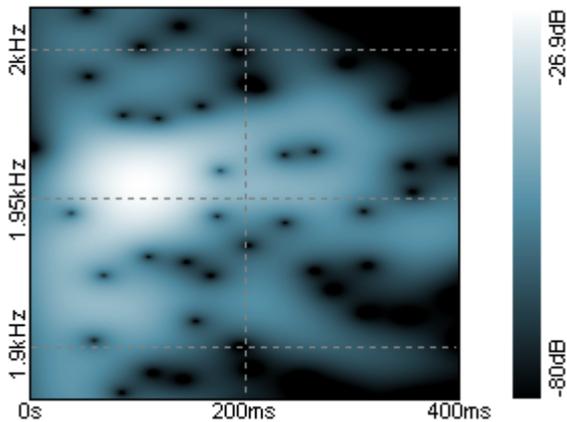


Figure 3: Splitting of one sinusoid into two.

To analyze this part we are using the Wigner distribution. The Wigner distribution has much higher resolution and accuracy but also much more cross terms (interference terms). To avoid cross terms we can isolate the sinusoid by copying it with a simple time-frequency region around the sinusoid track. The figure below shows the Wigner distribution with a very long rectangular window using 2 seconds, which is the mathematically perfect Wigner distribution (see figure 4). We can now see the splitting very accurately.

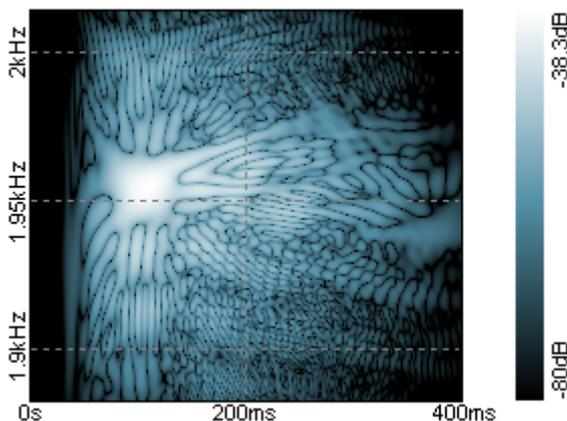


Figure 4: Wigner distribution of the splitting into two sinusoids.

### 2.4 Selecting the sinusoid

After we have identified the sinusoid we can select it by clicking on the two sinusoids and also on the common part before (see figure 5). Now we can remove the sinusoid to separate harmonics from noise. If all harmonics have been removed from the signal by deleting them, all is left is noise, which we can subtract from the original signal to get harmonics only.

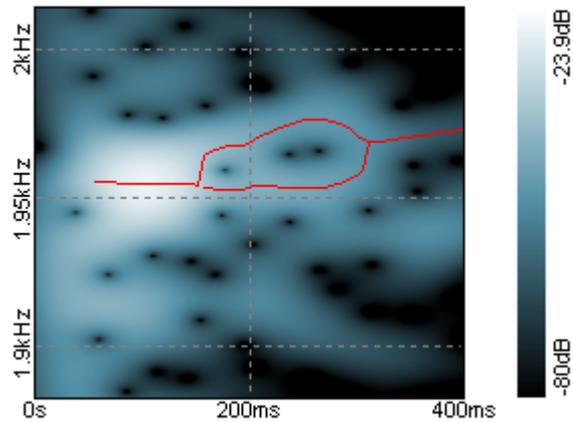


Figure 5: Sinusoid track in red.

## 3. COPYING ENVELOPES

Identifying, extracting and transferring envelopes from one signal to another can be solved using manually identified sinusoids in source and destination signal.

### 3.1 Identifying envelopes

There are several methods on how to calculate envelopes. LPC envelopes will tend to have sharp peaks pointing up at the poles of the signal model but otherwise follow the envelope quite good. Cepstrum envelopes also follow the envelope quite good, but fail in giving the envelope at the right amplitude level: they are often far below the envelopes at the peaks. Another method using a window function for the spectral analysis that has a very broad frequency characteristic works well, but also fails because of the interference terms of the now overlapping frequency components.

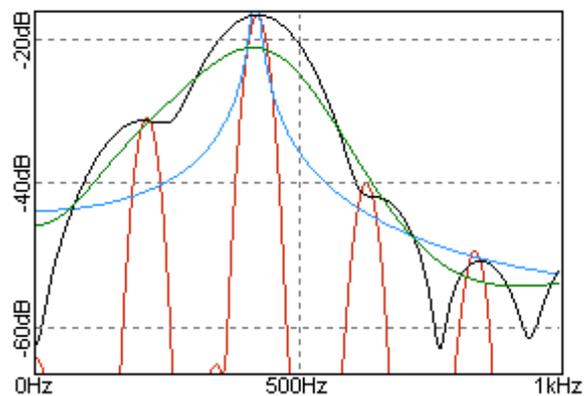


Figure 6: Different envelopes: LPC (sharp peaks pointing up), cepstrum, broadband window (sharp peaks pointing down).

To improve these methods additional information about the signal would be necessary, e.g., the cepstrum envelope could be improved if the fundamental frequency is known and the cut-off frequency of the low-pass filter in the cepstrum domain is adapted (lifter).

### 3.2 Defining the envelope

If a signal consists of two basic components, noise and harmonics (sinusoids), they are shaped both by the frequency characteristic of the given system (the envelope), when assuming a source/filter model. Looking at the spectrum of some sounds show that the noise and harmonics sometimes

follow different envelopes. All of the methods above will follow the energy distribution of noise and harmonics together, giving an overall value for the envelope.

### 3.3 Different envelopes for harmonics and noise

Identifying the harmonics in a signal will allow to handle noise and harmonics in a signal separately. To get harmonics only we first select all sinusoids and remove them to get noise only, and then subtract the noise from the original signal. Adding both parts again will exactly give the original signal, modifying the envelope of one or both of them and adding them up will give a new signal which is perfectly aligned with the original. The cepstrum envelope works very well for noise signals, for harmonics however, as explained above, it does not.

### 3.4 Envelopes for tracks

To avoid the shortcomings of the methods above, identified sinusoid tracks can be used to calculate the envelope using an appropriate interpolation algorithm (see figure 7).

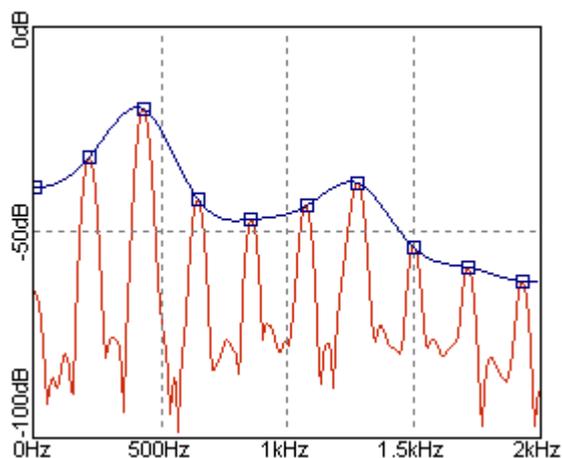


Figure 7: Interpolated envelope using sinusoid tracks.

A quick solution with good results that only needs sinusoid tracks in the target sound is to calculate the overall envelope of the source sound, e.g., with a cepstrum using no harmonic tracks, and applying the envelope using a linear time-variant filter only to identified harmonics in the target sound. This will not change the noise parts, especially in the upper frequencies where no harmonics may exist.

### 3.5 Transferring the envelope

To be able to compare or relate envelopes of two different sounds, the way the envelope is calculated in both sounds must be exactly the same. Therefore, to transfer the envelope you apply the difference between two envelopes using a linear time-variant filter to the target sound signal.

## 4. MUSICOLOGICAL APPLICATIONS

### 4.1 Recognizing and editing harmonics

One of the main and powerful advantages of this program is the recognition of harmonics by selecting just one single partial. This function enables one to edit only the harmonics of a sound independently of all other present noises. It is possible to isolate or amplify or attenuate a single instrument in an already recorded ensemble or to separate the harmonic parts from the noisy parts of one single instrument. For

example, with the help of this function it is possible to listen to a bowing noise of a violin only or to the belting noise of a trombone playing a crescendo.

### 4.2 Spectral analysis and resynthesis

Another very valuable function of this tool is the possibility to use spectral information as a complex filter function using single envelopes or the complete series of slices of a waterfall spectrum. So spectral editing is possible in various ways like flattening, morphing or transferring spectral information from one sound to another. For stimuli generation it is easily possible to modify a bassoon sound with an oboe spectrum in a way that it sounds like an oboe and vice versa. It is also possible to clear the formant structure of a bassoon sound, so that it loses its characteristic sound features (see figure 8a and b).

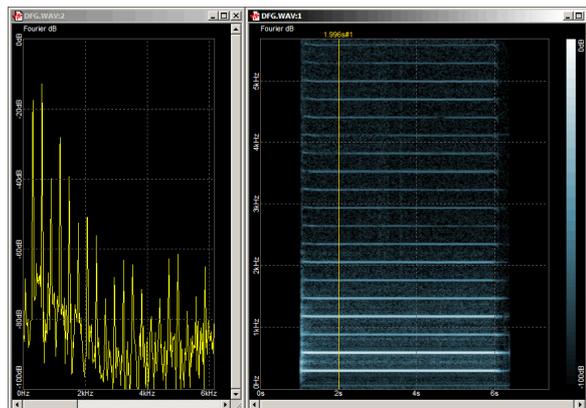


Figure 8a: Typical bassoon spectrum.

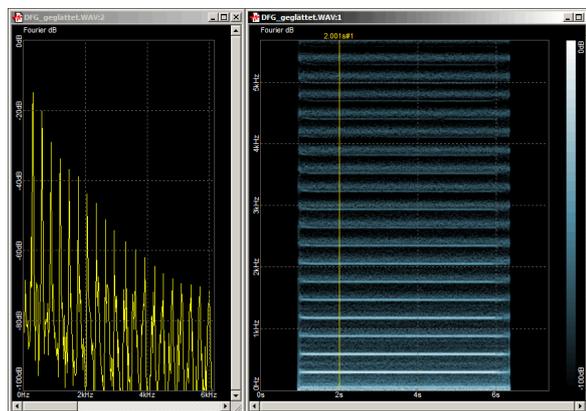


Figure 8b: Bassoon spectrum with flattened structure.

### 4.3 Moving the spectral threshold

With the threshold function it is possible to look and listen to free selectable spectral regions (e.g. only partials with amplitudes between -20 and -35 dB). This audiovisual spectral magnifying glass is very helpful on different occasions, like in an actual sports science project [2]: Very good bowmen can predict the quality of a shoot by just listening to the noise of the arrow when it leaves the bow. Comparing shooting sounds of Austrian and German cadre athletes together with their quality judgements and hitting successes it turned out that they listen especially to a frequency region between 3000 and 3500 Hz: If there is an energy maximum in this area, the shoot got mostly a bad rating and the hitting score were pure. This holds also for the opposite: shoots get a good quality rating (and a high score) if there is no spectral maximum in this area (see figure 9a and b).

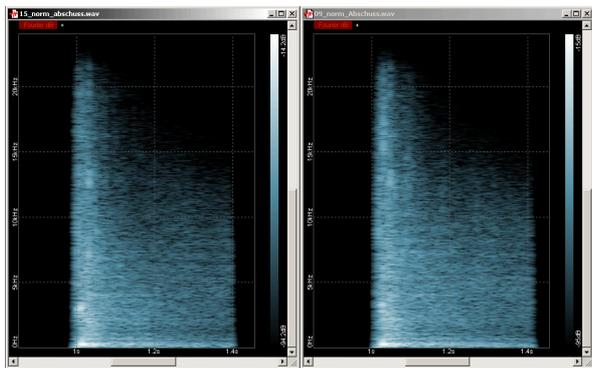


Figure 9a: Typical spectrum of a bad rated shoot (left) and typical spectrum of a good rated shoot (right), both original recordings).

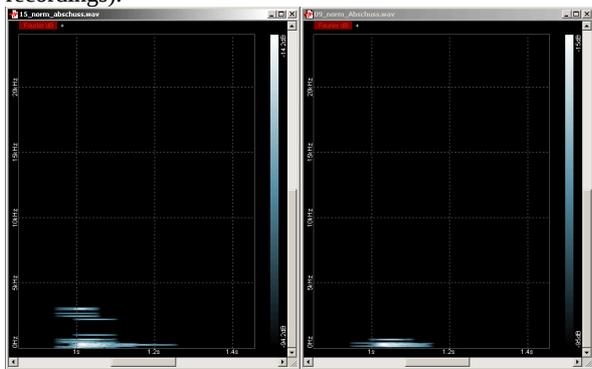


Figure 9a: Typical spectrum of a bad rated shoot (left) and typical spectrum of a good rated shoot (right), both recordings with partials above -40dB only (shifted spectral threshold).

## 5. CONCLUSIONS

Manually defining components of a signal by visual inspection alone yields very good results and is sometimes the only option to get results at all. Having a tool that allows manual interaction for the analysis and resynthesis of sound is very helpful and superior to algorithm based analysis and gives new perspectives to acoustical research and systematic musicology.

## 6. REFERENCES

- [1] H. Griebel. (2010, Sep. 21). Jaco Visual Signal Sound Samples. *Soundboot Homepage* [Online]. Available: <http://www.soundboot.com>
- [2] Horsak, Brian; Heller, Mario; Reuter, Christoph (2010). Ein akustischer Ansatz zur Bestimmung der Durchführungsqualität von Bogenschüssen. *Fortschritte der Akustik, DAGA 2010*, Berlin 2010, p. 487-488.