Estimating the parameters of an envelope that decays all exponential like

Dr. Bob L. Sturm, Universitetlektor
School of Electronic Engineering and Computer Science
KTH, Sweden

January 22, 2019

We are given a sound recording $x(t)$ of a tone produced by a musical instrument, and are told to approximate it by a superposition of N sinusoids (*overtones*), also known as "additive synthesis". This means we are building the model:

$$x(t) \approx \sum_{n=0}^{N-1} \mathrm{env}_n(t) \sin(2\pi f_n t) \tag{1}$$

where $\mathrm{env}_n(t)$ is the amplitude envelope of the nth overtone with a frequency f_n. When tones are produced by striking, such as a mallet on a xylophone bar, a good envelope to choose is exponential decay:

$$\mathrm{env}_n(t) = A_n e^{-\gamma_n t}, t \geq 0 \tag{2}$$

where A_n is the initial amplitude of the overtone struct at time $t = 0$, and γ_n is the rate at which it decays. Now we just need to estimate for each overtone its frequency, amplitude and decay rate.

We can easily estimate each overtone frequency by looking at the location of peaks in a magnitude spectrum of $x(t)$; but how do we estimate A_n and γ_n? First, we know from our model that the power of the overtone in any portion is only going to depend on the envelope of the sinusoid. The sinusoid has a constant power no matter when we look. Hence, we can relate the dB amplitude of the nth overtone to its envelope by the following:

$$\begin{aligned}\mathrm{dB}_n(t) := 20\log_{10}\mathrm{env}_n(t) &= 20\log_{10} A_n e^{-\gamma_n t} = 20\log_{10} A_n + 20\log_{10}(e^{-\gamma_n t}) \\ &= 20\log_{10} A_n + 20\frac{\log_e e^{-\gamma_n t}}{\log_e 10} = 20\log_{10} A_n - 20\gamma_n t/\log_e 10\end{aligned} \tag{3}$$

where we have used the fact that

$$\log_b x = \frac{\log_a x}{\log_a b}. \tag{4}$$

Considering the envelope at two different times t_1 and t_2 produces two equations:

$$\mathrm{dB}_n(t_1) = 20\log_{10} A_n - 20\gamma_n t_1/\log_e 10 \tag{5}$$

$$\mathrm{dB}_n(t_2) = 20\log_{10} A_n - 20\gamma_n t_2/\log_e 10. \tag{6}$$

Subtracting the two equations gets rid of the unknown A_n:

$$\mathrm{dB}_n(t_1) - \mathrm{dB}_n(t_2) = (20/\log_e 10)\gamma_n(t_2 - t_1). \tag{7}$$

Solving for the exponential decay parameter gives:

$$\gamma_n = \frac{\log_e 10}{20}\frac{\mathrm{dB}_n(t_1) - \mathrm{dB}_n(t_2)}{t_2 - t_1}. \tag{8}$$

Having computed γ_n, we can then plug it into (5) or (6) and solve for the amplitude:

$$A_n = 10^{\mathrm{dB}_n(t_1)/20} 10^{\gamma_n t_1/\log_e 10}. \tag{9}$$

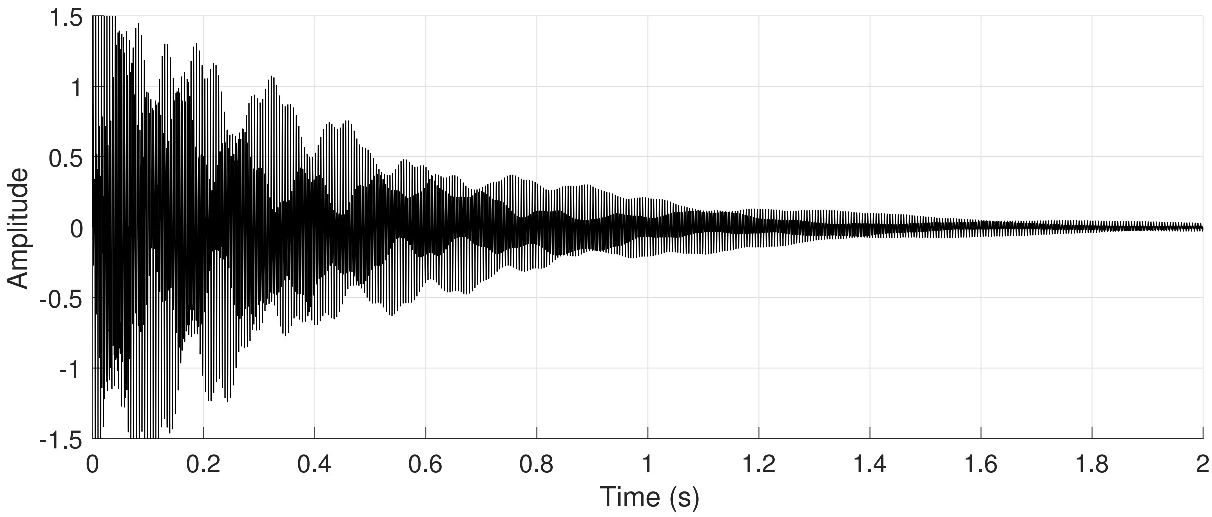
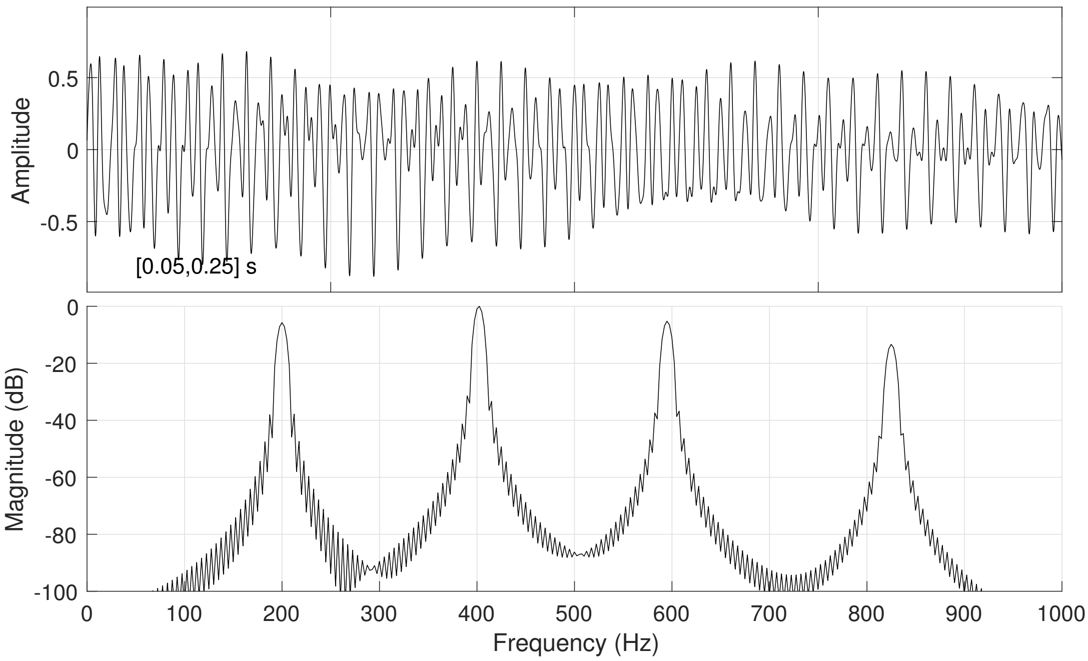


Figure 1: Our synthetic test signal in time.

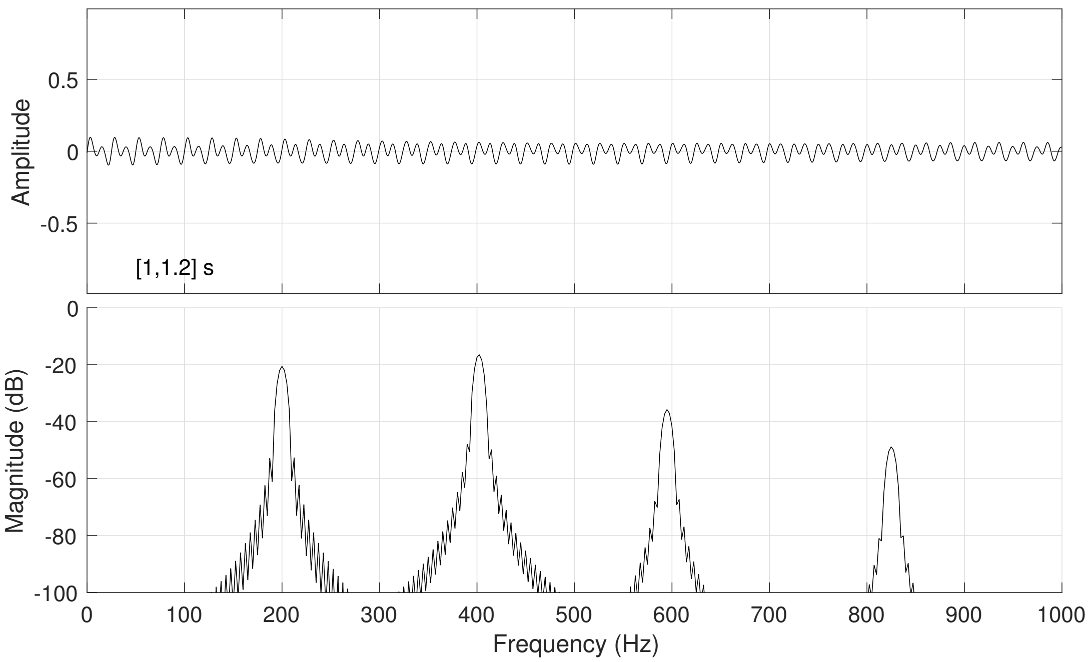
Let's apply this using a signal that we synthesize with known parameters. This will allow us to compare our estimates with the truth. Figure 1 shows our signal, synthesized using the following code:

```
f = [200 402 595 825];
a = [0.5 1 0.7 0.3];
gamma = [1.8 2.0 3.7 4.3];
fs = 22050; dur = 2;
t = [0:dur*fs-1]'/fs;d
x = zeros(length(t),1);
for ii=1:length(f)
  x = x + a(ii)*exp(-gamma(ii)*t).*sin(2*pi*t*f(ii));
end
x = x./max(abs(x)+0.01);
```

Now we look at the dB magnitude spectrum of this signal at two different points in time. This is computed with the discrete Fourier transform. Figure 2 shows two 200 ms portions of the signal, one taken at $t_1 = 0.05$ s and the other at $t_2 = 1.0$ s. We see four significant overtones. Using the peak picking tool on the MATLAB plot, we find the frequency of these peaks to be: 200, 402, and 595 and 825 Hz. These agree perfectly with the frequencies we used to generate this signal.



(a) Top: Waveform; Bottom: dB magnitude spectrum



(b) Top: Waveform; Bottom: dB magnitude spectrum

Figure 2: 200 ms segments of our synthetic test signal at two times and the dB magnitude spectra.

We also find that each overtone decays in dB magnitude the following amounts from t_1 to t_2. The zeroth overtone starts at -5.7 dB and decays to -20.6 dB. The first overtone starts at 0 dB and decays to -16.5 dB. The second starts at -5.2 dB and decays to -35.7 dB. The third starts at -13.3 dB and decays to -48.8 dB. This gives us all we need to know to estimate the

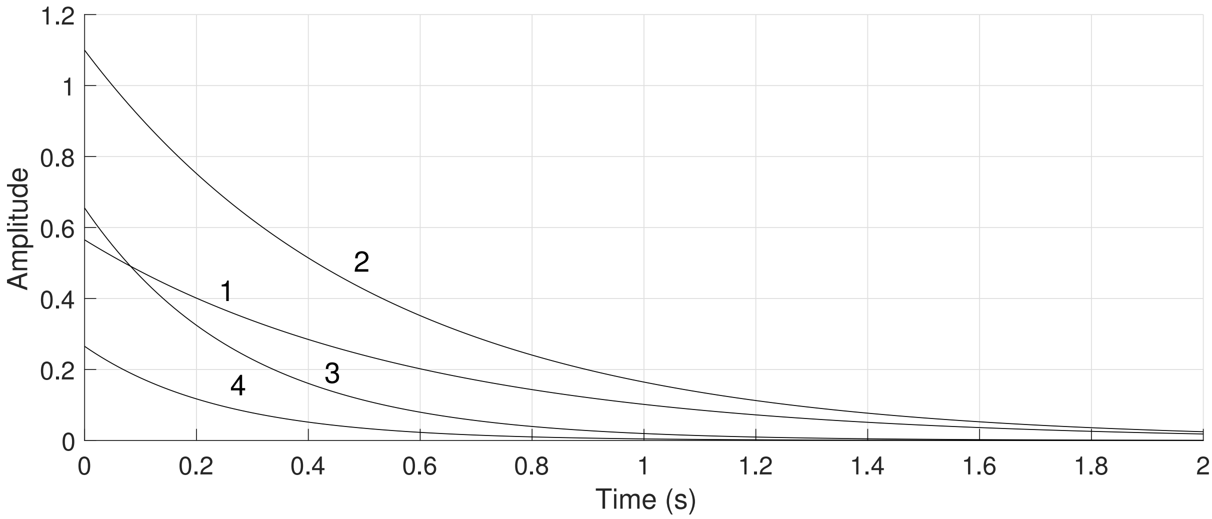


Figure 3: The estimated envelopes of the overtones (labeled) of our synthetic test signal.

decay parameters for each overtone: $\hat{\gamma}_1 = 1.8057$, $\hat{\gamma}_2 = 1.9996$, $\hat{\gamma}_3 = 3.6963$, and $\hat{\gamma}_4 = 4.3022$. These all agree very closely with the truth. Our estimates of the amplitudes are: $\hat{A}_0 = 0.5678$; $\hat{A}_1 = 1.1051$; $\hat{A}_2 = 0.6611$; $\hat{A}_3 = 0.2682$. These are close to the ground truth, but have some deviation.[1] Figure 3 shows the estimated envelopes of each overtone in this test signal.

[1] Can you suggest some reasons for this discrepancy?