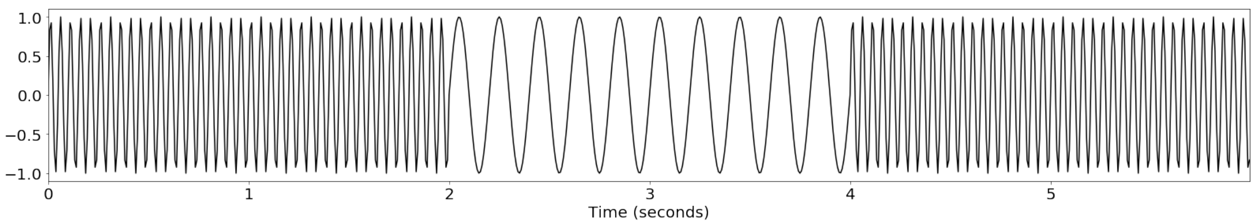
ECS7006 Music Informatics

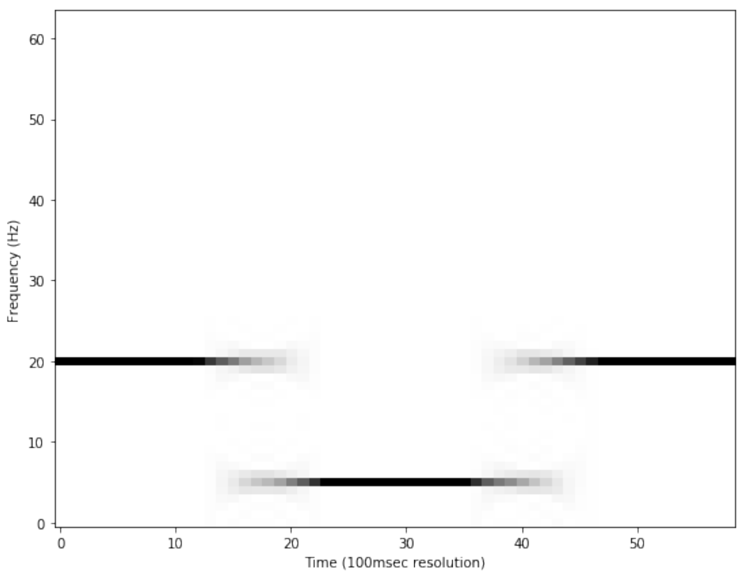
Tutorial - Time–Frequency Representations

1. Consider the below signal:



Sketch the STFT spectrogram for the above signal.

Solution:



2. Compute the time resolution in msec, frequency resolution in Hz, and the Nyquist frequency of a discrete STFT based on the following parameter settings: $F_s = 44100$ Hz, $N = 2048$, and $H = 512$.

Solution:
Time resolution in msec: $T_{coef} = \frac{m \cdot H}{F_s} = \frac{512000}{44100} \approx 11.6$ msec.
Frequency resolution: $F_{coef} = \frac{k \cdot F_s}{N} = \frac{1 \cdot 44100}{2048} = 21.53$ Hz
Nyquist frequency: $F_s/2 = 22050$ Hz

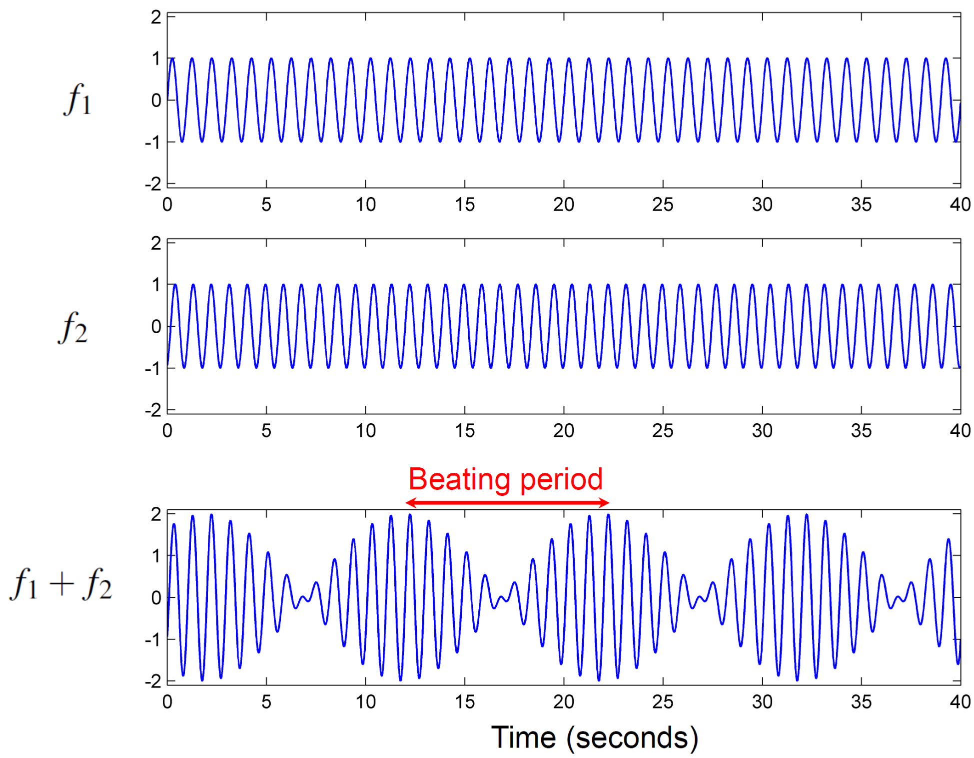
3. Let $F_s = 48000$ Hz, $N = 2048$ and $H = 512$ for a discrete STFT. What is the physical meaning of STFT coefficients $\mathcal{X}(1000, 500)$ and $\mathcal{X}(16, 1024)$? Why is the coefficient $\mathcal{X}(16, 1024)$ problematic?

   Solution:
   The coefficient $\mathcal{X}(1000, 500)$ corresponds to time $T_{coef} = \frac{1000 \cdot 512}{48000} = 10.66$ sec and physical frequency $F_{coef} = \frac{500 \cdot 48000}{2048} = 11.71$ kHz. For $\mathcal{X}(16, 1024)$ it corresponds to $T_{coef} = \frac{16 \cdot 512}{48000} = 0.17$ sec and $F_{coef} = \frac{1024 \cdot 48000}{2048} = 24$ kHz. 24 kHz corresponds to the Nyquist frequency for this signal, and this coefficient yields a poor approximation of the actual frequency of the underlying analogue signal.

4. Assume that the STFT of exercise 3 is used as input to a mel filterbank with 128 coefficients using the mel scale proposed by Fant: $m = 1000 \cdot \log_2\left(1 + \frac{f}{1000}\right)$. What is the physical meaning of the mel spectrogram bin $\mathcal{X}_{mel}(30, 64)$ and to which unit in the mel scale does the frequency correspond to?

   Solution:
   Time coefficient: $T_{coef} = \frac{30 \cdot 512}{48000} = 0.32$ sec.
   The 128th bin in the Mel spectrogram corresponds to the Nyquist frequency of 24 kHz, thus the mel unit corresponding to that frequency is equal to: $m = 1000 \cdot \log_2\left(1 + \frac{24000}{1000}\right) = 4644$ mel.
   Thus, the 64th coefficient in the mel spectrogram could correspond to $4644/2 = 2322$ mel.
   Thus: $2322 = 1000 \cdot \log_2\left(1 + \frac{f}{1000}\right) \Rightarrow 1 + \frac{f}{1000} = 2^{2.322} \Rightarrow \frac{f}{1000} = 4.002 \Rightarrow f = 4002$ Hz.

5. **Beating** - Two sinusoids of similar frequency may add up (constructive interference) or cancel out (destructive interference). Let $f_1(t) = \sin(2\pi\omega_1 t)$ and $f_2(t) = \sin(2\pi\omega_2 t)$ be two such sinusoids with distinct but nearby frequencies $\omega_1 \approx \omega_2$. In the following figure, for example, $\omega_1 = 1$ and $\omega_2 = 1.1$ is used.



The figure also shows that the superposition f_1+f_2 of these two sinusoids results in a function that looks like a single sine wave with a slowly varying amplitude, a phenomenon also known as *beating*. Determine the rate (reciprocal of the period) of the beating in dependency on ω_1 and ω_2. Compare this result with the plot of $f_1 + f_2$ in the figure.
Hint: Use the trigonometric identity $\sin(a) + \sin(b) = 2\cos\left(\frac{a-b}{2}\right)\sin\left(\frac{a+b}{2}\right)$ for $a, b \in \mathbb{R}$.

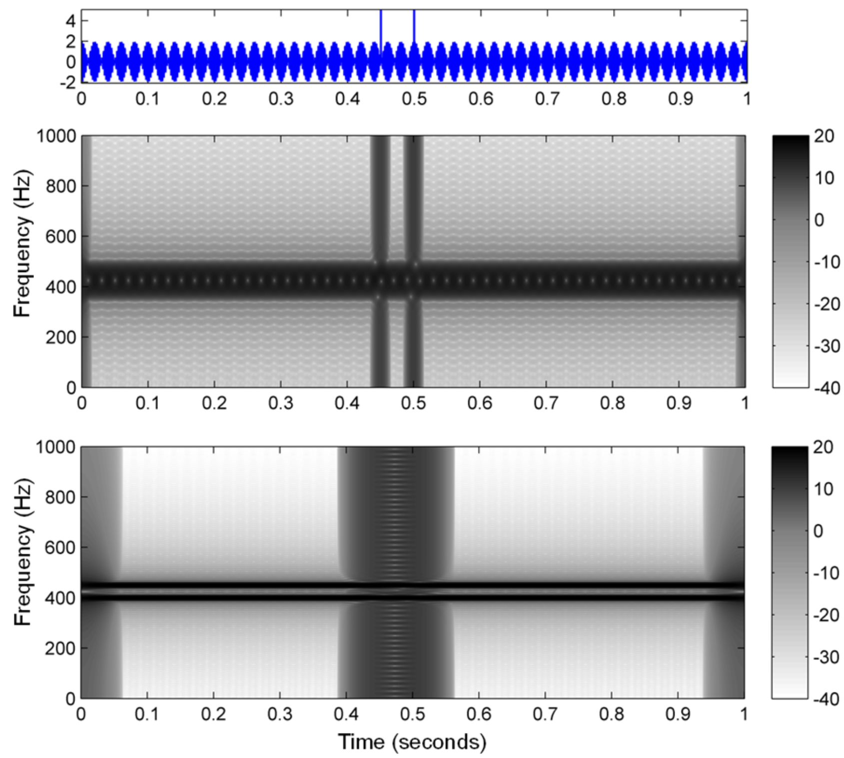
Solution:
Setting $a = 2\pi\omega_1 t$ and $b = 2\pi\omega_2 t$, one obtains:

$$\sin(2\pi\omega_1 t) + \sin(2\pi\omega_2 t) = 2\cos\left(2\pi\frac{\omega_1 - \omega_2}{2}t\right)\sin\left(2\pi\frac{\omega_1 + \omega_2}{2}t\right)$$

This shows that if the difference $\omega_1 - \omega_2$ is small, the cosine term has a low frequency compared with the sine term. As a result the signal $f_1 + f_2$ can be seen as a sine wave of frequency $(\omega_1 + \omega_2)/2$. While the frequency of the cosine term is $(\omega_1 - \omega_2)/2$, every second burst in the modulation pattern is inverted. Each peak is replaced by a trough and vice versa. Therefore, subjectively, the frequency of the envelope seems to have twice the frequency of the modulating cosine, which means the audible beat frequency is: $|\omega_1 - \omega_2|$.

In the example with $\omega_1 = 1$ and $\omega_2 = 1.1$ shown in the figure, the beating rate is 0.1 Hz and the beating period is 10 sec.

6. In the below spectrogram one can notice vertical stripes at $t = 0$ and $t = 1$. Why?



Solution:

The signal is defined in the time interval $[0, 1]$. Furthermore, it is assumed to be zero outside this interval. Now, in a neighborhood of $t = 0$, the signal is zero for $t < 0$ and it is a superposition of two sinusoids for $t > 0$. In the Fourier representation, two exponential functions are needed to represent the signal for $t > 0$. However, for $t < 0$ these oscillations need to compensated to generate the zero function. To this end, based on the principles of destructive interference, many different frequency components spread over the entire spectrum are needed, which explains the vertical stripe in the spectrogram at $t = 0$. The same explanation applies for $t = 1$.