Frequency Analysis over Time

1. Introduction

In these notes, we introduce the concept of short-time frequency analysis for signals whose frequency content changes over time, and the *spectrogram* plot for visualizing such analysis.

2. Short-time spectral analysis

A. Introduction

Often when analyzing signals, we are interested in the local frequency content of a signal, rather than the global frequency properties of that signal, since the frequency content of real signals, such as speech or music, typically is not static, but rather changes over time. Consider, for example the following two signals:

$$x_1(t) = \cos(2\pi(10)t) + \cos(2\pi(25)t) + \cos(2\pi(50)t) + \cos(2\pi(100)t),\ 0 \le t < 1 \tag{1}$$

$$\begin{aligned} x_2(t) = {} & \cos(2\pi(10)t)[u(t) - u(t-0.3)] + \cos(2\pi(25)t)[u(t-0.3) - u(t-0.6)] + \\ & \cos(2\pi(50)t)[u(t-0.6) - u(t-0.8)] + \cos(2\pi(100)t)[u(t-0.8) - u(t-1)] \end{aligned},\ 0 \le t < 1\,. \tag{2}$$

Note that the first signal's frequency content does not change over the interval $t \in [0, 1]$, while the second signal's frequency content changes at times $t = 0.3$, $t = 0.6$ and $t = 0.8$. We now sample both signals at a sampling frequency $f_s = 1000$ Hz over a one-second interval and compute the magnitude FFT for the resulting discrete time sequences,

$$x_1[n] = x_1(n/f_s) \text{ and } x_2[n] = x_2(n/f_s),\ n \in \{0, 1, \ldots, 998, 999\}\,. \tag{3}$$

In Figure 1 below, we plot $x_1[n]$ and $x_2[n]$, as well as the magnitude FFTs for $x_1[n]$ and $x_2[n]$. From Figure 1, we make the following observations. Even though the two discrete-time sequences $x_1[n]$ and $x_2[n]$ are entirely different, their FFTs show similar properties, with peaks at the expected frequencies (±10 Hz, ±25 Hz, ±50 Hz and ±100 Hz). Also note, that the FFT for $x_2[n]$ does not show us where (in time) each of the dominant frequency components occurs; it simply tells us that somewhere in the signal, each of the four dominant frequency components is present. Therefore, if we are interested in how the frequency content of a signal changes over time, we cannot simply compute the FFT of the entire signal; rather, we must compute the FFT over short time periods. Below, we illustrate this type of analysis for two examples — a synthetic signal (similar to equation (2) above) and a speech signal.

B. Synthetic example

Here we consider the changing frequency content of the following signal,

$$\begin{aligned} x_3(t) = {} & \cos(2\pi(100)t)[u(t) - u(t-0.3)] + \cos(2\pi(300)t)[u(t-0.3) - u(t-0.6)] + \\ & \cos(2\pi(500)t)[u(t-0.6) - u(t-0.8)] + \cos(2\pi(1000)t)[u(t-0.8) - u(t-1)] \end{aligned} \tag{4}$$

sampled at $f_s = 10,000$ Hz for one second to yield the discrete-time signal $x_3[n]$:

$$x_3[n] = x_3(n/f_s),\ n \in \{0, 1, \ldots, 9998, 9999\}\,. \tag{5}$$

We now compute short-time FFTs for $x_3[n]$ in segments of length 500 (50msec), with an overlap of 250 samples. For examples, the first FFT is computed for the first 500 samples ($n \in \{0, 1, \ldots, 489, 499\}$), the second FFT is computed for samples corresponding to $n \in \{250, 251, \ldots, 748, 749\}$, and so on. In all, for a discrete-time sequence of length 10,000, this generates a total of 39 FFTs, each of length 500. Note that each FFT will tell us something about the localized frequency content for its corresponding segment of $x_3[n]$.

We can visualize the above computations in a *spectrogram*, with time t on the horizontal axis, and frequency f on the vertical axis, as shown in Figure 2 below. The columns in this plot correspond to magnitude FFTs of 50msec segments, with darker colors indicating larger values and lighter colors indicating smaller values; that is, black indicates the maximum value in the spectrogram, while white indicates zero. Note, that in the spec-

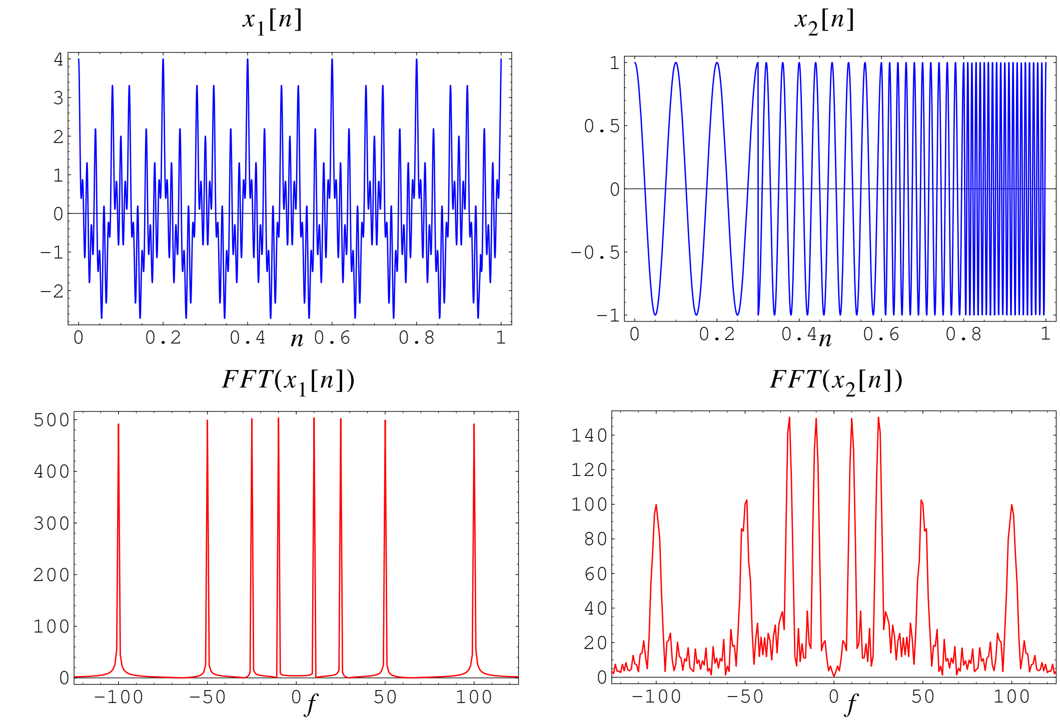


Figure 1

trogram plot, it is very easy to visualize the changing frequency content of $x_3[n]$ over time; also note that when the signal changes from one frequency to another (e.g. $t = 0.3$ sec), the FFT corresponding to that transition indicates both frequencies.

For comparison, in Figure 3 we plot the spectrogram for the following constant-frequency signal:

$$x_4(t) = \cos(2\pi(100)t) + \cos(2\pi(300)t) + \cos(2\pi(500)t) + \cos(2\pi(1000)t),\ 0 \le t < 1 \quad (6)$$

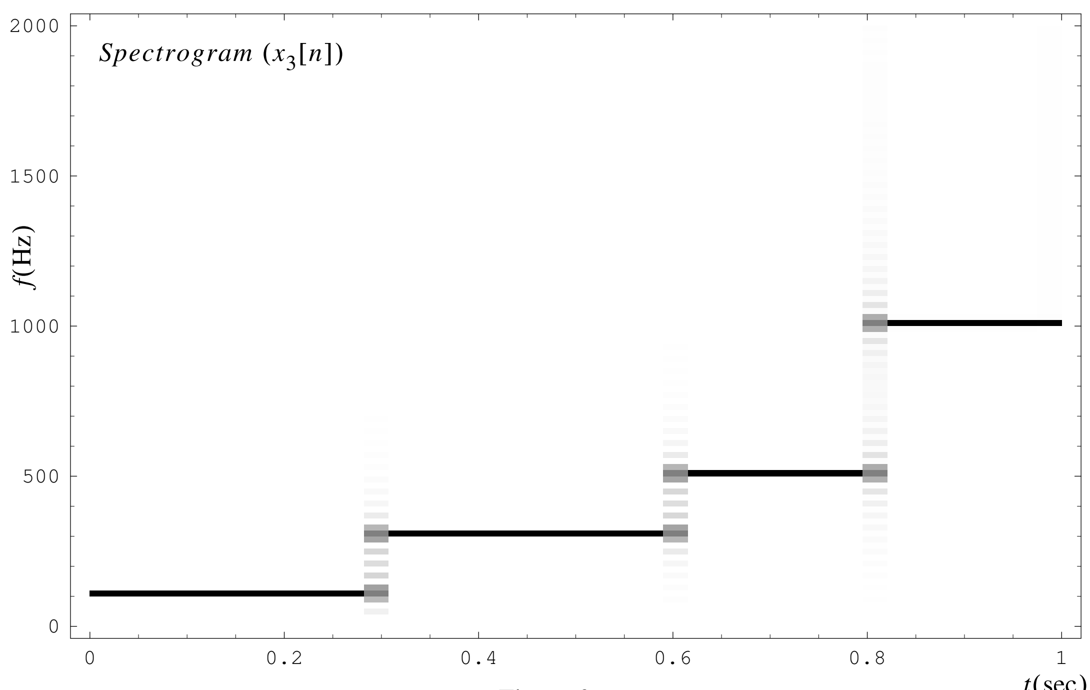


Figure 2

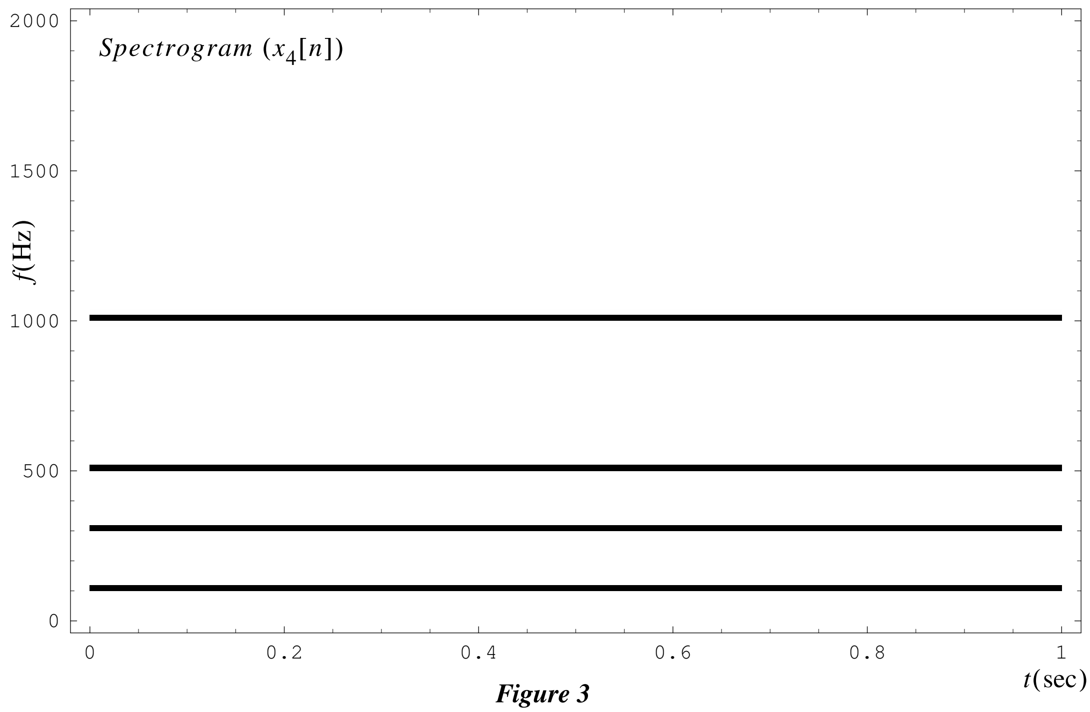


Figure 3

sampled at $f_s = 10,000$ Hz for one second to yield the discrete-time signal $x_4[n]$:

$$x_4[n] = x_4(n/f_s), \; n \in \{0, 1, \ldots, 9998, 9999\}. \tag{7}$$

Note that the spectrogram in Figure 3 clearly shows that the frequency content of $x_4[n]$ does not vary over time. That is, the spectrograms for signals $x_3[n]$ and $x_4[n]$ illustrate the differences between the two signals much more clearly than the full-length FFTs for signals $x_1[n]$ and $x_2[n]$ in Figure 1.

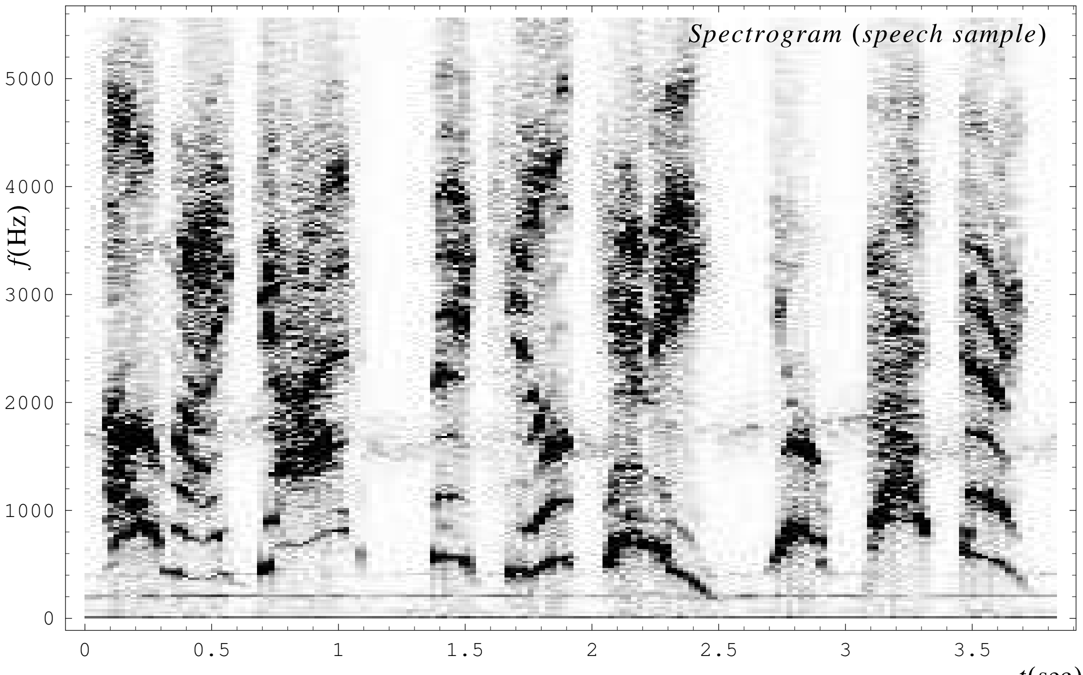
C. Speech example

In Figure 4, we show what a spectrogram looks like for a segment of digitized speech. The speech sample is from the television show "South Park,"[1] is 3.83 seconds long, and was digitized at a sampling rate of $f_s = 11,127$ Hz. Note that each syllable of the speech sample corresponds to one of the bands in the spectrogram with a lot of frequency content, while pauses in the speech signal (i.e. silence) correspond to relatively little spectral power. If you listen to the speech sample, you can also correlate higher-pitched syllables (e.g. "Oh") with bands in the spectrogram that contain large higher frequency components.

3. Conclusion

The *Mathematica* notebook "time_varying.nb" was used to generate the examples in this set of notes.

1. See the web site or corresponding Mathematica notebook to listen to the speech example — "Oh my God, they killed Kenny — you bastards!"



Spectrogram (*speech sample*)
f(Hz)
5000
4000
3000
2000
1000
0
0
0.5
1
1.5
2
2.5
3
3.5
t(sec)

Figure 4