Lecture 3: Fourier transforms of windowed periodic signals

Mark Hasegawa-Johnson

ECE 537: Speech Processing

Outline

Fourier transform of a periodic signal

Many real-world signals (voiced speech, music) are periodic, $x[n] = x[n + T_0]$. What is the Fourier transform of a periodic signal? Can we compute it?

$$
\begin{aligned}
X(\omega) &= \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n} \\
&= \sum_{m=-\infty}^{\infty} \sum_{n=0}^{T_0-1} x[n + mT_0] e^{-j\omega(n+mT_0)} \\
&= \left(\sum_{n=0}^{T_0-1} x[n] e^{-j\omega n} \right) \left(\sum_{m=-\infty}^{\infty} e^{-j\omega m T_0} \right) \\
&= \begin{cases} \infty & \frac{\omega T_0}{2\pi} = \text{integer} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

Oops. What can we do with that?

Fourier series representation of a periodic signal

Instead of jumping in to the Fourier transform directly, let's try computing the Fourier series first.

$$x[n] = \sum_{k=0}^{T_0-1} X_k e^{j\frac{2\pi kn}{T_0}}$$

$$X_k = \frac{1}{T_0}\sum_{n=0}^{T_0-1} x[n] e^{-j\frac{2\pi kn}{T_0}}$$

Fourier transform of a complex exponential

If you apply the **inverse DTFT** formula to an impulse, you get

$$\frac{1}{2\pi}\int_0^{2\pi}\delta(\omega-\omega_0)e^{j\omega n}d\omega=\frac{1}{2\pi}e^{j\omega_0 n}$$

Scaling both sides by 2π, we get

$$e^{j\omega_0 n}\leftrightarrow 2\pi\delta(\omega-\omega_0)$$

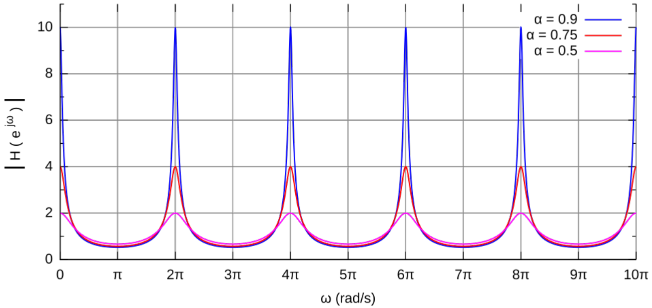
DTFT of a Periodic Signal

DTFT is a linear operator, so for any periodic signal $x[n]$ with period T_0, we can compute its DTFT as:

$$
\begin{aligned}
X(\omega) &= \text{DTFT}\left\{x[n]\right\} \\
&= \text{DTFT}\left\{\sum_{k=0}^{T_0-1} X_k e^{j\frac{2\pi kn}{T_0}}\right\} \\
&= \sum_{k=0}^{T_0-1} X_k \text{DTFT}\left\{e^{j\frac{2\pi kn}{T_0}}\right\} \\
&= \sum_{k=0}^{T_0-1} 2\pi X_k \delta\left(\omega - \frac{2\pi k}{T_0}\right)
\end{aligned}
$$

So the DTFT of a periodic signal is a sum of delta functions, each scaled by the corresponding Fourier series coefficient.

Spectrum of a periodic signal



Comb filter, showing peaks at frequencies that are integer multiples of $2\pi/T_0$. CC-BY 3.0, https://commons.wikimedia.org/wiki/File:Comb_filter_response_fb_pos.svg

Autocorrelation of a Periodic Signal

Suppose that $x[n-n_0]$ is the same signal, but starting at random offset n_0. Its autocorrelation is:

$$
\begin{aligned}
R[m] &= E\left[x[n-n_0]x^*[n-n_0-m]\right] \\
&= E\left[\left(\sum_{k=0}^{T_0-1} 2\pi X_k e^{-j\frac{2\pi k(n-n_0)}{T_0}}\right)\left(\sum_{k=0}^{T_0-1} 2\pi X_k^* e^{j\frac{2\pi k(n-n_0-m)}{T_0}}\right)\right] \\
&= \sum_{k=0}^{T_0-1} |X_k|^2 e^{-j\frac{2\pi km}{T_0}}
\end{aligned}
$$

In other words, the autocorrelation of a periodic signal is periodic.

- $x[n]$ and $R_{XX}[m]$ have the same period.
- The coefficients of $R_{XX}[m]$ are $|X_k|^2$.

Power Spectrum of a Periodic Signal

If $x[n]$ is periodic with period T_0, then

$$\begin{aligned} P_{XX}(\omega) &= \text{DTFT}\left\{R_{XX}[m]\right\} \\ &= \sum_{k=0}^{T_0-1} |X_k|^2 \text{DTFT}\left\{e^{-j\frac{2\pi km}{T_0}}\right\} \\ &= \sum_{k=0}^{T_0-1} 2\pi |X_k|^2 \delta\left(\omega - \frac{2\pi k}{T_0}\right) \end{aligned}$$

So the power spectrum of a periodic signal is:

- 0 except at integer multiples of $\frac{2\pi}{T_0}$
- Inifinite, with an area of $2\pi|X_k|^2$, at the k^{th} harmonic.

Outline

Windowing a pure tone

Multiplying in time $\equiv$ convolving in frequency:

$$x[n]w[n] \leftrightarrow \frac{1}{2\pi}X(\omega) * W(\omega)$$

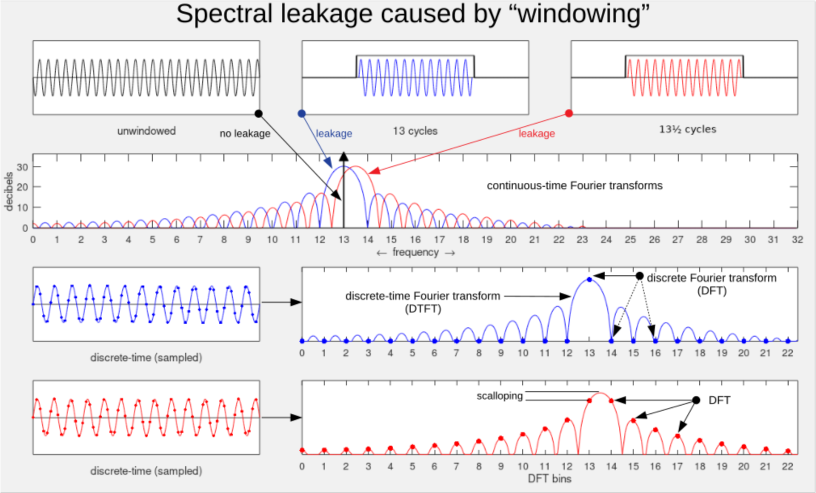
Convolution with an impulse $\equiv$ shifting:

$$\frac{1}{2\pi}\left(W(\omega) * 2\pi\delta\left(\omega - \frac{2\pi k}{N}\right)\right) = W\left(\omega - \frac{2\pi k}{N}\right)$$

Therefore, windowing a pure tone $\leftrightarrow$ shifting the window spectrum:

$$w[n]e^{-j\frac{2\pi kn}{N}} \leftrightarrow W\left(\omega - \frac{2\pi k}{N}\right)$$

Windowing a pure tone $\leftrightarrow$ shifting the window spectrum



DTFT of a Windowed Periodic Signal

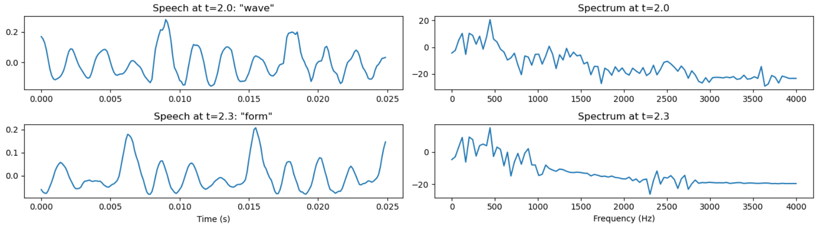
Suppose $y[n] = x[n]w[n]$, where $x[n]$ is an infinite-length periodic signal with period T_0, and $w[n]$ is a window, then

$$\begin{aligned}
Y(\omega) &= \frac{1}{2\pi}\left(W(\omega) * X(\omega)\right)\\
&= \sum_{k=0}^{T_0-1} \frac{2\pi X_k}{2\pi}\left(W(\omega) * \delta\left(\omega - \frac{2\pi k}{T_0}\right)\right)\\
&= \sum_{k=0}^{T_0-1} X_k W\left(\omega - \frac{2\pi k}{T_0}\right)
\end{aligned}$$

So the DTFT of a periodic signal is a sum of delta functions, each scaled by the corresponding Fourier series coefficient.

Examples of Vowel Spectra

On the left: roughly three pitch periods from each of two vowels ("wave" and "form"). On the right: log periodogram of segments on the left (expressed in decibels). Pitch frequency is about 115Hz (pitch period $\sim$ 8.7ms). Each harmonic is not a perfect impulse, but rather, has been blurred by windowing.



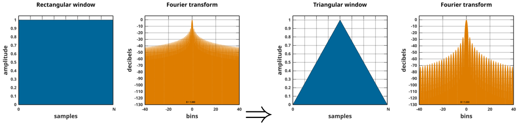
Autocorrelation of a Periodic Signal

Suppose that $w[n]x[n-n_0]$ is a windowed copy of the same signal, but shifted by some random offset n_0 prior to windowing. Its autocorrelation is:

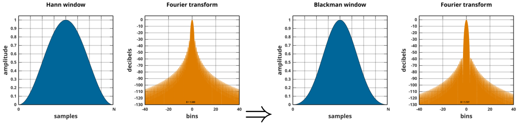
$$\begin{aligned} R[m] &= E\left[w[n]x[n-n_0]w^*[n-m]x^*[n-n_0-m]\right] \\ &= \sum_{k=0}^{T_0-1} |X_k|^2 e^{-j\frac{2\pi km}{T_0}} \sum_{n=-\infty}^{\infty} w[n]w^*[n-m] \\ &= \sum_{k=0}^{T_0-1} |X_k|^2 e^{-j\frac{2\pi km}{T_0}} \left(w[m] * w^*[-m]\right) \end{aligned}$$

- If you window the signal, then compute its autocorrelation, that's the same as. . .
- computing the autocorrelation, then windowing by $w[m] * w^*[-m]$.

Example: Autocorrelation of a rectangle is a triangle



Example: Autocorrelation of a Hann $(0.5 - 0.5\cos(2\pi n/N))$ is a type of Blackman $(0.25 - 0.5\cos(2\pi n/N) + 0.25\cos(4\pi n/N))$:



Power Spectrum of a Periodic Signal

Suppose that $w[n]x[n-n_0]$ is a windowed copy of the same signal, but shifted by some random offset n_0 prior to windowing. Its power spectrum is:

$$\begin{aligned} P(\omega) &= \text{DTFT}\{R[m]\} \\ &= \sum_{k=0}^{T_0-1} |X_k|^2 \left| W\left(\omega - \frac{2\pi k}{T_0}\right)\right|^2 \end{aligned}$$

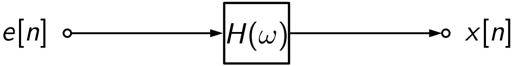
So the power spectrum of a windowed periodic signal is:

- a set of shifted squared window spectra, $|W(\omega)|^2$,
- each scaled by $|X_k|^2$.

Outline

Source-Filter Model of Speech Production

It is often useful to model speech using a source filter model:



In voiced speech:

- The excitation $e[n]$ represents periodic negative air pressure shocks, caused by sudden closure of the glottis (the hole between the vocal folds) once every T_0 samples.
- The transfer function $H(\omega)$ models filtering by the vocal tract (the resonant cavity between vocal folds and lips).

A Model of Glottal Airflow and Supraglottal Air Pressure

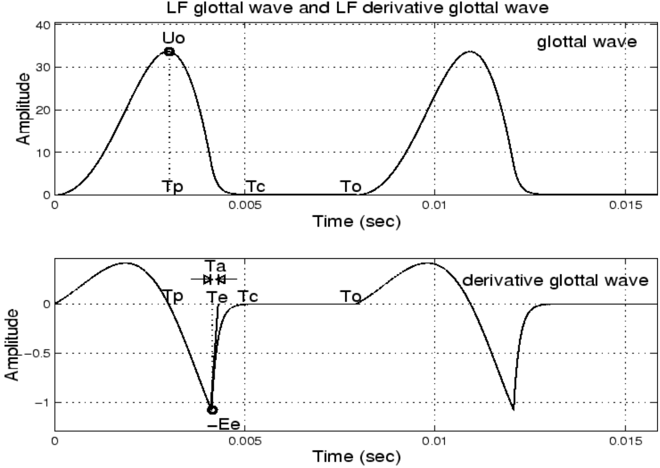


Image (c) Julius O. Smith III,
`https://ccrma.stanford.edu/~jos/SMAC03S/img120.png`

A Simplified Model of Supraglottal Air Pressure

The most important feature of supraglottal air pressure is the periodic sequence of negative air pressure spikes, occurring at glottal closure instants. We can model just this one feature using the following model:

$$e[n] = \sum_{m=-\infty}^{\infty} G\delta[n - mT_0]$$

where G is the scale of each spike (a negative number).

Outline

Fourier Series of an Impulse Train

To find the DTFT of $e[n]$, let's first find its DTFS:

$$e[n] = \sum_{k=-\infty}^{\infty} G\delta[n-kT_0]$$

$$\begin{aligned} E_k &= \frac{1}{T_0}\sum_{n=0}^{T_0-1} e[n]e^{-j\frac{2\pi kn}{T_0}} \\ &= \frac{1}{T_0}\sum_{n=0}^{T_0-1}\sum_{k=-\infty}^{\infty} G\delta[n-kT_0]e^{-j\frac{2\pi kn}{T_0}} \\ &= \frac{1}{T_0}\sum_{n=0}^{T_0-1} G\delta[n]e^{-j\frac{2\pi kn}{T_0}} \\ &= \frac{G}{T_0} \end{aligned}$$

Fourier Transform of an Impulse Train is an Impulse Train

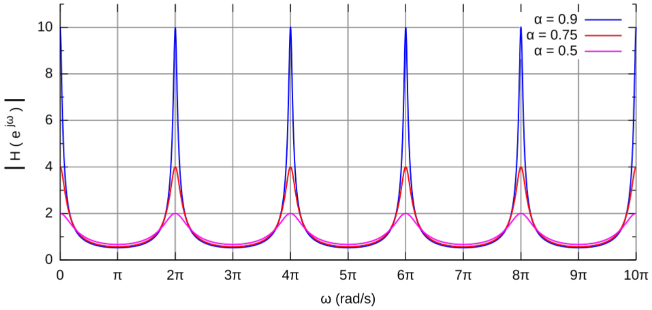
The DTFT of $e[n]$ is a sum of impulses:

$$
\begin{aligned}
E(\omega) &= \sum_{k=0}^{T_0-1} 2\pi E_k \delta\left(\omega - \frac{2\pi k}{T_0}\right) \\
&= \frac{2\pi G}{T_0} \sum_{k=0}^{T_0-1} \delta\left(\omega - \frac{2\pi k}{T_0}\right)
\end{aligned}
$$

Similarly, if the start time is random, we can still recover the power spectrum as

$$
P_{EE}(\omega) = \frac{2\pi|G|^2}{|T_0|^2} \sum_{k=0}^{T_0-1} \delta\left(\omega - \frac{2\pi k}{T_0}\right)
$$

Fourier transform of an impulse train is an impulse train



Comb filter, showing peaks at frequencies that are integer multiples of $2\pi/T_0$. `https://commons.wikimedia.org/wiki/File:Comb_filter_response_fb_pos.svg`

Outline

Source-Filter Model of Speech Production

$$e[n] \longrightarrow \boxed{H(\omega)} \longrightarrow x[n]$$

According to the model above, we should have

$$X(\omega) = E(\omega)H(\omega)$$
$$P_{XX}(\omega) = |H(\omega)|^2 P_{EE}(\omega)$$

Source-Filter Model of Speech Production

$$e[n] \longrightarrow \boxed{H(\omega)} \longrightarrow x[n]$$

Plugging in $E(\omega)$ and $P_{EE}(\omega)$, we get:

$$X(\omega) = \frac{2\pi G}{T_0} \sum_{k=0}^{T_0-1} H_k \delta\left(\omega - \frac{2\pi k}{T_0}\right)$$

$$P_{XX}(\omega) = \frac{2\pi |G|^2}{|T_0|^2} \sum_{k=0}^{T_0-1} |H_k|^2 \delta\left(\omega - \frac{2\pi k}{T_0}\right)$$

where $H_k = H\left(\frac{2\pi k}{T_0}\right)$.

Windowed Spectra

When the signal is windowed, the 2π factor is canceled, and we just have shifted windows:

$$X(\omega) = \frac{G}{T_0}\sum_{k=0}^{T_0-1} H_k W\left(\omega - \frac{2\pi k}{T_0}\right)$$

$$P_{XX}(\omega) = \frac{|G|^2}{|T_0|^2}\sum_{k=0}^{T_0-1} |H_k|^2 \left|W\left(\omega - \frac{2\pi k}{T_0}\right)\right|^2$$

where $H_k = H\left(\frac{2\pi k}{T_0}\right)$.

Outline

Examples

The vocal tract, from glottis to lips, is a resonant pipe, much like a clarinet. It has many resonances, each of which can be adjusted by adjusting the tongue, lips, and jaw. The k^{th} resonance is a second-order system with resonant frequency $\omega_k = \frac{2\pi F_k}{F_s}$ and half-bandwidth $\sigma_k = \frac{\pi B_k}{F_s}$ (expressed in radians/sample, where F_s is the sampling frequency), and its effect on speech can be modeled as:

$$x[n] = e[n] + 2e^{-\sigma_k}\cos(\omega_k)y[n-1] - e^{-2\sigma_k}y[n-2]$$

If $H(\omega)$ has only one resonance, what is $P_{XX}(\omega)$ before windowing? How about after windowing?

Outline

Summary

- The Fourier transform of an impulse train is an impulse train:

$$\sum_{m=-\infty}^{\infty} \delta[n-mT_0] \leftrightarrow \frac{2\pi}{T_0}\sum_{k=0}^{T_0-1}\delta\left(\omega-\frac{2\pi k}{T_0}\right)$$

- Voiced speech is created by sending a periodic voice source through a resonant pipe:

$$x[n] = w[n]\left(h[n] * \sum_{m=-\infty}^{\infty}\delta[n-mT_0]\right)$$

$$X(\omega) = \frac{1}{T_0}\sum_{k=0}^{T_0-1} H_k W\left(\omega-\frac{2\pi k}{T_0}\right)$$