Lecture 7: Fourier Transform;
Speech Sources and Filters
Today

- Fourier Series (Discrete Fourier Transform)
  - Sawtooth wave example
  - Finding the coefficients
- Spectrogram
- Frequency Response
- Speech Source-Filter Theory
- Speech Sources
  - Transient
  - Frication
  - Aspiration
  - Voicing
  - Formant Transitions
Topic #1: Discrete Fourier Transform
Sawtooth Wave

- Sawtooth first harmonic:
  - Amplitude: 0.64
  - Phase: 0.52\pi\text{ radians}
Sawtooth Second Harmonic

- Sawtooth second harmonic:
  - Amplitude: 0.32
  - Phase: $0.53\pi$ radians
Sawtooth Third Harmonic

- Sawtooth third harmonic:
  - Amplitude: 0.21
  - Phase: $0.55\pi$ radians
Sawtooth: Calculating the Fourier Transform

Average over one full period of \((125\text{Hz sawtooth} \times 125\text{Hz cosine})\) = -0.03
Sawtooth: Calculating the Fourier Transform

- Average over one full period of \((125\text{Hz sawtooth} \times 125\text{Hz sine}) = 0.636\)
Sawtooth: Amplitude and Phase of the First Harmonic

First Fourier Coefficient of the Sawtooth Wave:
- \(-0.03 + 0.636j\)
- \(|-0.03 + 0.636j| = 0.64\)
- \(\text{phase}(-0.03 + 0.636j) = 0.52\pi\)
- \(X_1 = -0.03+0.636j = 0.64 \, e^{j0.52\pi}\)

First Harmonic:
- \(h_1(t) = 0.64 \, e^{j(250\pi t+0.52\pi)}\)
- \(\text{Re}\{h1(t)\} = 0.64 \cos(250\pi t+0.52\pi)\)
Sawtooth: Calculating the Fourier Transform

Average over one full period of \((125\text{Hz sawtooth} \ast 250\text{Hz cosine})\) = -0.03
Sawtooth: Calculating the Fourier Transform

- Average over one full period of \((125\text{Hz sawtooth} \times 250\text{Hz sine}) = -0.3173\)
Sawtooth: Amplitude and Phase of the Second Harmonic

- Second Fourier Coefficient of the Sawtooth Wave:
  - \(-0.03 + 0.3173j\)
  - \(|-0.03 + 0.3173j| = 0.32\)
  - \(\text{phase}(-0.03 + 0.3173j) = 0.53\pi\)
  - \(X_1 = -0.03+0.3173j = 0.32 \: e^{j0.53\pi}\)

- First Harmonic:
  - \(h_2(t) = 0.32 \: e^{j(500\pi t+0.53\pi)}\)
  - \(\text{Re}\{h_2(t)\} = 0.64 \cos(250\pi t+0.53\pi)\)
How to Reconstruct a Sawtooth from its Harmonics

- At sampling frequency of $FS=8000\,Hz$ ($8000$ samples/second), the period of a $125\,Hz$ sawtooth is
  - $\frac{(8000\,\text{samples/second})}{(125\,\text{cycles/second})} = 64\,\text{samples}$
- The “0”th harmonic is DC ($0\times125\,Hz = 0\,Hz$)
- So we need to add up 63 harmonics:
  - $x(t) = h_1(t) + h_2(t) + \ldots + h_{63}(t)$
Why Does This Work???

- It’s a property of sines and cosines
- They are a *BASIS SET*, which means that...
- Let <> mean “average over one full period:”

\[
<\cos^2(m\omega_0 t)> = 0.5 \text{ for any integer } m
\]
\[
<\cos(m\omega_0 t)\cos(n\omega_0 t)> = 0 \text{ if } m \neq n
\]

above two properties also hold for sine

\[
<\cos(m\omega_0 t) \sin(n\omega_0 t)> = 0
\]

- \(\omega_0 = 2\pi F_0 = 2\pi/T_0\) is the “fundamental frequency” in radians/second
Topic #2: Spectrogram
How to Create a Spectrogram

- Divide the waveform into overlapping window
  - Example: window length = 6ms
  - Example: window spacing = one per 2ms
  - Result: neighboring windows overlap by 4ms

- Fourier transform each frame to create the “short-time Fourier transform” $X_k(t)$
  - Read: kth frequency bin, taken from the frame at time t seconds
  - Could also write $X(t,f) = \text{Fourier coefficient from the frame at time=}t \text{ seconds, from the frequency bin centered at } f \text{ Hz.}$

- In pixel $(k,t)$, plot $20\times\log_{10}(\text{abs}(X_k(t)))$
  - The Fourier transform expressed in decibels!
Reading a Spectrogram

Each vertical stripe is a glottal closure instant (BANG – high energy at all frequencies)
Inter-bang period = pitch period ≈ 10 ms

Stop
Turbulence (Frication or Aspiration)
Vowel
Nasals
Glide (Extreme Formant Frequencies)
Topic #3: Frequency Response and Speech Source-Filter Theory
Frequency Response

- Any sound can be windowed
  - Call the window length “N”
- Its windows can be Fourier transformed:
  \[ x(t) = h_1(t) + h_2(t) + \ldots \]
  \[ x(t) = X_1 e^{j\omega_0 t} + X_2 e^{2j\omega_0 t} + \ldots \]
- What happens when you filter a cosine?
  - Input = \( e^{j\omega_0 t} \)  ---  Output = \( H(\omega_0) e^{j\omega_0 t} \)
  - Input = \( e^{2j\omega_0 t} \)  ---  Output = \( H(2\omega_0) e^{2j\omega_0 t} \)
    ...
If \( x(t) \) goes through any filter:

\[
x(t) \rightarrow \text{ANY FILTER} \rightarrow x(t)
\]

Then we can find \( y(t) \) using freq response!

\[
x(t) = X_1 e^{j\omega_0 t} + X_2 e^{2j\omega_0 t} + ... \\
y(t) = H(\omega_0)X_1 e^{j\omega_0 t} + H(2\omega_0)X_2 e^{2j\omega_0 t} + ...
\]
In general, we can write

\[ Y(\omega) = H(\omega)X(\omega) \]
Examples of Filters

- **Low-Pass Filter:**
  - Eliminates all frequency components above some cutoff frequency
  - Result sounds muffled

- **High-Pass Filter:**
  - Eliminates all frequency components below some cutoff frequency
  - Result sounds fizzy

- **Band-Pass Filter:**
  - Eliminates all components except those between $f_1$ and $f_2$
  - Result: sounds like an old radio broadcast

- **Band-Stop Filter:**
  - Eliminates only the frequency components between $f_1$ and $f_2$
  - Result: usually hard to hear that anything’s happened!
Examples of Filters

- **A Room:**
  - Adds echoes to the signal

- **A Quarter-Wave Resonator:**
  - Emphasizes frequency components at 
    \[ f = \left( \frac{c}{4L} \right) + \left( \frac{nc}{2L} \right), \text{ for all integer } n \]
  - Components at other frequencies are passed without any emphasis

- **A Vocal Tract:**
  - Emphasizes any part of the input that is near a formant frequency
  - Any energy at other frequencies is passed without emphasis
The Source-Filter Model of Speech Production

- **Input: Weak Sources**
  - Glottal closure
  - Turbulence (frication, aspiration)

- **Filter: The Vocal Tract**
  - Energy near formant frequency is emphasized by a factor of 10 or a factor of 100
  - Energy at other frequencies is passed without emphasis
  - Speech = (Vocal Tract Transfer Function) X (Excitation)
  - $S(\omega) = T(\omega) \cdot E(\omega)$
Topic #4: Speech Sources.
Presented as:
Events in the Release of a Stop Consonant
Events in the Release of a Stop

"Burst" = transient + frication (the part of the spectrogram whose transfer function has poles only at the front cavity resonance frequencies, not at the back cavity resonances).
Pre-voicing during Closure

To make a voiced stop in most European languages:

- Tongue root is relaxed, allowing it to expand so that vocal folds can continue to vibrating for a little while after oral closure.

- Result is a low-frequency “voice bar” that may continue well into closure.

In English, closure voicing is typical of read speech, but not casual speech.
Burst = Transient and Frication
Frication = Turbulence at a Constriction

Turbulence striking an obstacle makes noise

Front cavity resonance frequency:
\[ F_R = \frac{c}{4L_f} \]

The fricative “sh” (this ship)
Aspiration = Turbulence at the Glottis (like /h/). Transfer Function During Aspiration...
Formant Transitions: Labial Consonants

First Sensitivity Function: $-dF_1/d \log A(x)$

Second Sensitivity Function: $-dF_2/d \log A(x)$

Third Sensitivity Function: $-dF_3/d \log A(x)$
Formant Transitions: Alveolar Consonants

First Sensitivity Function: $-dF1/d \log A(x)$

Second Sensitivity Function: $-dF2/d \log A(x)$

Third Sensitivity Function: $-dF3/d \log A(x)$

FREQ (kHz)

TIME (ms)
Formant Transitions: Post-alveolar Consonants

First Sensitivity Function: $-dF_1/d \log A(x)$

Second Sensitivity Function: $-dF_2/d \log A(x)$

Third Sensitivity Function: $-dF_3/d \log A(x)$
Formant Transitions: Velar Consonants

First Sensitivity Function: $-\partial F_1 / \partial \log A(x)$

Second Sensitivity Function: $-\partial F_2 / \partial \log A(x)$

Third Sensitivity Function: $-\partial F_3 / \partial \log A(x)$
Formant Transitions: A Perceptual Study

Fig. 1. Synthetic spectrograms showing second-formant transitions that produce the voiced stops before various vowels.