ECE 438 lecture Monday March 27, 2023

Announcements

1. No quiz this week
2. HW No. 7 due Wednesday 29 March at 11:59p EDT
3. Office Hours today:
   - @ 2:30p EDT
   - @ 4:00p EDT

Speech

Module 4.1.1 – Applications of speech processing
- Characteristics of speech

Module 4.1.2 – speech model
some overlap with Module 4.1.1

Module 4.2.0 – Short-time Fourier transform

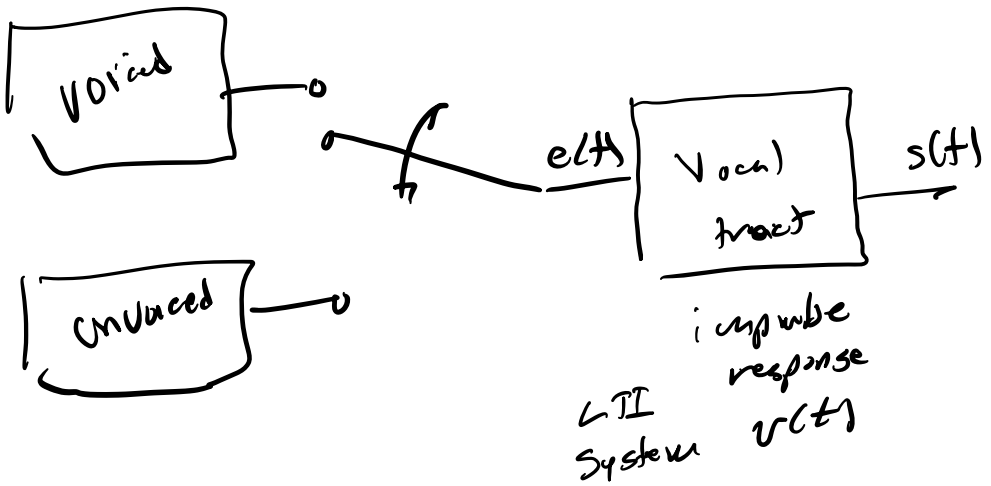
Two interpretations:

DTFT of windowed segment of speech

Module 4.2.2 filter bank

Module 4.2.1 – skip

Review model for speech generation



Voiced speech

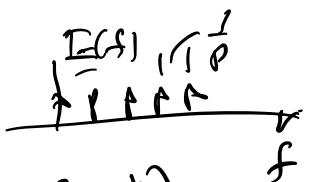
$$e(t) = rep_p[\delta(t)]$$

p: pitch period

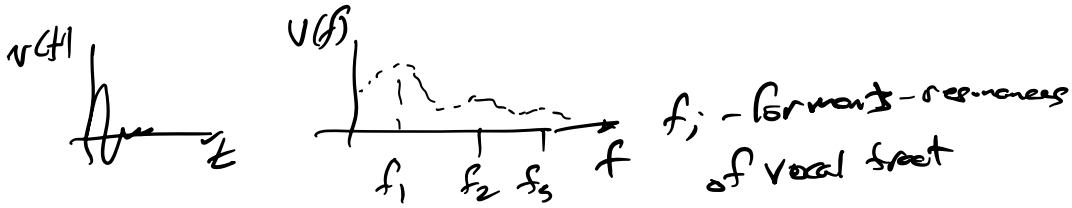
$$s(t) = v(t) * e(t) = \sum_{k=-\infty}^{\infty} v(t - kP)$$

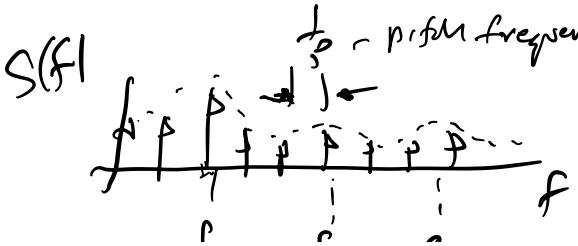
for a single phoneme

$$S(f) = V(f)E(f)$$



$$E(f) = \frac{1}{p} comb_{\frac{1}{p}}[k] = \frac{1}{p} \sum_{k=-\infty}^{\infty} \delta\left(f - \frac{k}{p}\right)$$





generally

$\frac{1}{p} \ll f_i \quad i = 1, 2, 3$

+1 +2 +3

Define STDTFT

$$\tilde{S}(\omega, n_0) = \sum_{n=-\infty}^{\infty} \tilde{s}[n, n_0] e^{-j\omega n}$$

$$\tilde{s}[n, n_0] = s[n] \cdot w[n_0 - n]$$

$$= s[n] w[-(n - n_0)]$$

different from usual w

$w[n]$ - "window" $= s[n]\, w[n - n_0]$ if $w[n]$ is symmetric

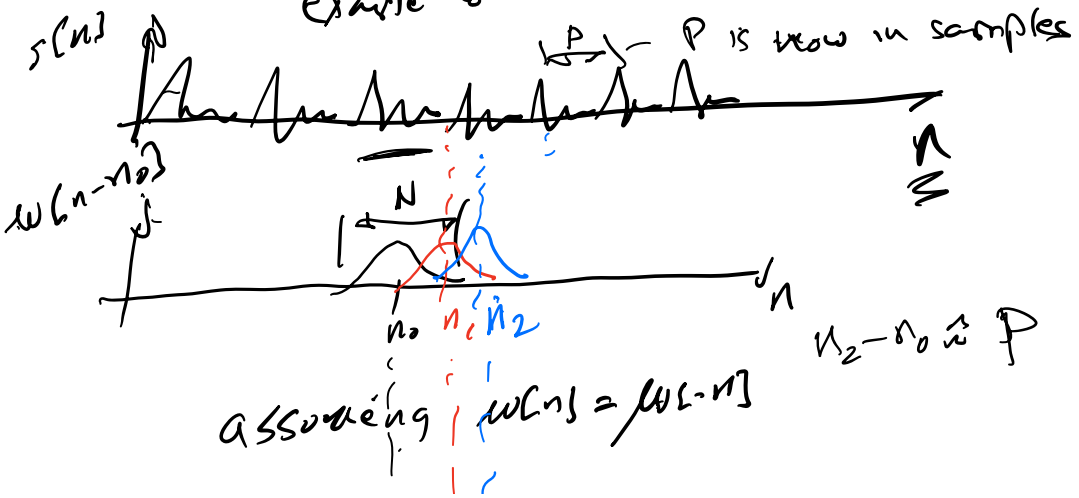
If it is symmetric $w[n] = w[-n]$

Sketches

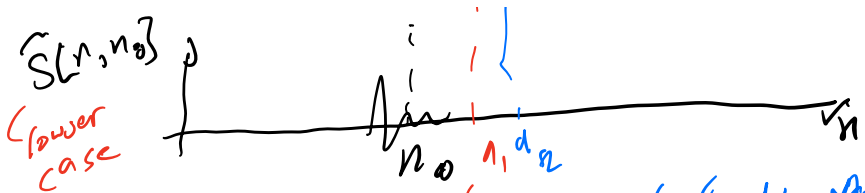
Let N be length of window

Assume $N \approx P$ or $N < P$

example for voiced phoneme



assuming $w[n] = w[-n]$

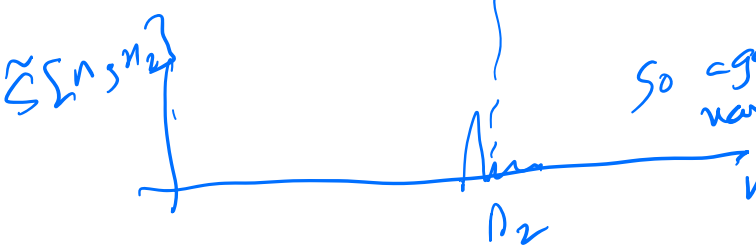


$\hat{S}(w, n_0) \approx V(w)$ — neglecting any leakage due to shift of $w[n]$ to n_0

upper case

now let's consider a different time point n_1 for the center of the window

The idea is that $\tilde{S}[n, n_2]$ has very little energy



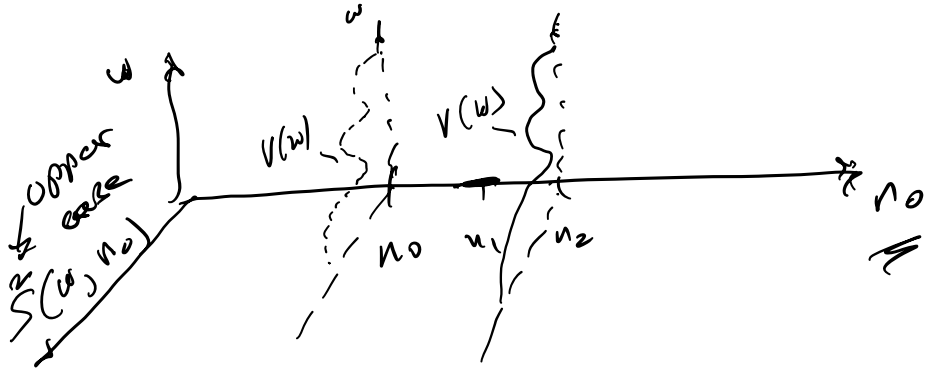
So again, we have a good replica of the vocal tract response

Now

let's consider a spectrogram

function of n_0 along horizontal axis

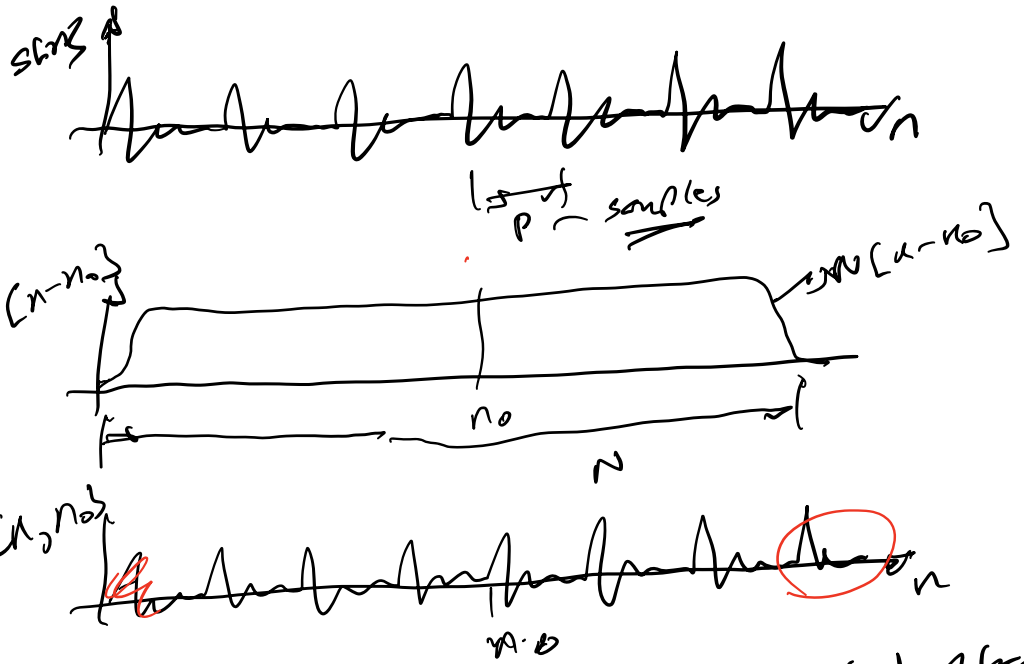
and w along vertical axis



See periodic "bands" of energy, along n_0 axis

⇒ good time resolution — can see that $s(n)$ consists of periodic repetition of vocal tract response — called a wideband spectrogram

Now consider case where $N \gg P$



$\tilde{S}(\omega, n_0) \approx$ comb$_P[V(\omega)]$ again neglected phase factor due to shift of window