HOMEWORK SOLUTIONS #6

1. (a) The two discrete-time STFTs are related by

$$\begin{aligned}
\bar{X}(e^{j\omega}, n) &= \sum_{m=-\infty}^{\infty} x[m]w[m-n+R-1]e^{-j\omega m} \\
&= \sum_{p=-\infty}^{\infty} x[n-p]w[R-1-p]e^{-j\omega(n-p)} \quad \text{via} \quad m = n-p \\
&= e^{-j\omega n} \sum_{p=-\infty}^{\infty} x[n-p]w[p]e^{j\omega p} \quad \text{via} \quad w[R-1-p] = w[p] \\
&= e^{-j\omega n} \sum_{p=0}^{R-1} x[n-p]w[p]e^{j\omega p} \\
&= e^{-j\omega n} X(e^{j\omega}, n)
\end{aligned}$$

(b) If we evaluate the first STFT definition at $\omega = \frac{2\pi}{N}k$ for $k = 0...N-1$, we see

$$X(e^{j\frac{2\pi}{N}k}, n) = \sum_{m=0}^{R-1} x[n-m] \underbrace{w[m]e^{j\frac{2\pi}{N}km}}_{h_k[m]}$$

and thus this STFT can be implemented by a bank of LTI filters $\{h_k[m]\}$ with impulse responses indicated above. If we try this for the second STFT definition, the the results of part (a) imply

$$\begin{aligned}
\bar{X}(e^{j\frac{2\pi}{N}k}, n) &= e^{-j\frac{2\pi}{N}kn} X(e^{j\frac{2\pi}{N}k}, n) \\
&= e^{-j\frac{2\pi}{N}kn} \sum_{m=0}^{R-1} x[n-m]w[m]e^{j\frac{2\pi}{N}km} \\
&= \sum_{m=0}^{R-1} \underbrace{x[n-m]e^{-j\frac{2\pi}{N}k(n-m)}}_{x_k[n-m]} w[m]
\end{aligned}$$

which means that the input is split into N different paths and modulated before the filtering. Thus, this STFT cannot be implemented as a bank of LTI filters.

(c) Taking an N-DFT of $X[k, n]$ across frequency (so that $p = 0...N-1$),

$$\begin{aligned}
\sum_{k=0}^{N-1} X[k, n]e^{-j\frac{2\pi}{N}kp} &= \sum_{k=0}^{N-1} \left(\sum_{m=0}^{R-1} x[n-m]w[m]e^{j\frac{2\pi}{N}km} \right) e^{-j\frac{2\pi}{N}kp} \\
&= \sum_{m=0}^{R-1} x[n-m]w[m] \sum_{k=0}^{N-1} e^{-j\frac{2\pi}{N}k(p-m)}
\end{aligned}$$

Note that $e^{-j\frac{2\pi}{N}k(p-m)} = N$ when $p-m = qN$ for $q \in \mathbb{Z}$, and zero otherwise. Recall that the range of p and m implies $1-R \le p-m \le N-1$, where $R \le N$. Thus, the only chance for $p-m = qN$ is when $m = p$. Since $m \in \{0...R-1\}$,

$$\sum_{k=0}^{N-1} X[k,n]e^{-j\frac{2\pi}{N}kp} \quad = \quad Nx[n-p]w[p], \quad p = 0...R-1$$

Since we know that $w[p] > 0$ when $p \in \{0...R-1\}$,

$$x[n-p] \quad = \quad \frac{1}{Nw[p]} \sum_{k=0}^{N-1} X[k,n]e^{-j\frac{2\pi}{N}kp}, \quad p = 0...R-1$$

Note that DFT of $\{X[0,n],\dots,X[N-1,n]\}$ for fixed n yields $\{x[n-R+1],\dots,x[n]\}$. This implies that it is possible to reconstruct $x[n]$ for all n given a sequence of vectors $\{X[0,mR],\dots,X[N-1,mR]\}$ for $m \in \mathbb{Z}$; it is not necessary to employ $X[k,n]$ for all k and n.

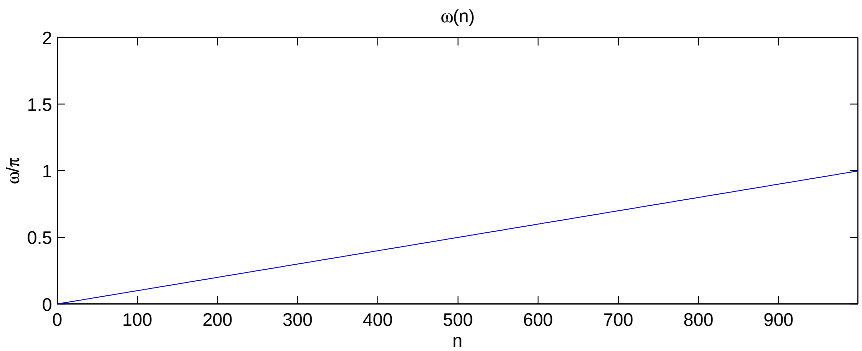
(d) Combining the results of parts (a) and (c),

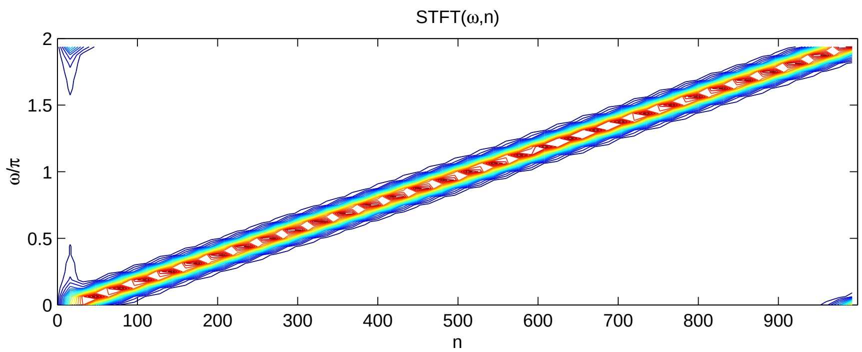
$$\begin{aligned} x[n-p] &= \frac{1}{Nw[p]} \sum_{k=0}^{N-1} X[k,n]e^{-j\frac{2\pi}{N}kp}, \quad p = 0...R-1 \\ &= \frac{1}{Nw[p]} \sum_{k=0}^{N-1} \left(\bar{X}[k,n]e^{j\frac{2\pi}{N}kn}\right) e^{-j\frac{2\pi}{N}kp}, \quad p = 0...R-1 \\ &= \frac{1}{Nw[p]} \sum_{k=0}^{N-1} \bar{X}[k,n]e^{-j\frac{2\pi}{N}k(p-n)}, \quad p = 0...R-1 \end{aligned}$$

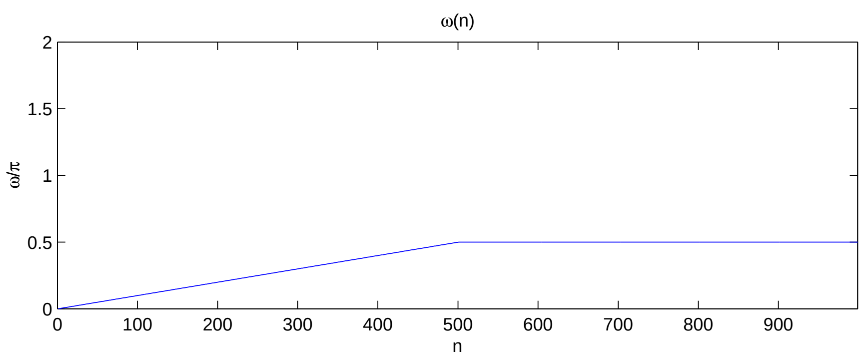
2. (a) The first figure below shows the discrete STFT magnitude as a function of time (horizontal axis) and instantaneous frequency (vertical axis). Since $x[n] = e^{-j\omega_n n}$, one might first expect that the STFT indicates an instantaneous frequency of $\omega_n = \frac{n\pi}{M}$ at time n. Instead, the STFT plot has a pronounced ridge corresponding to an instantaneous frequency of $2\frac{n\pi}{M} = 2\omega_n$. This is consistent with the phase-derivative interpretation: $\frac{\partial}{\partial n}(\omega_n n) = \frac{\partial}{\partial n}(\frac{n\pi}{M}n) = 2\frac{n\pi}{M} = 2\omega_n$.

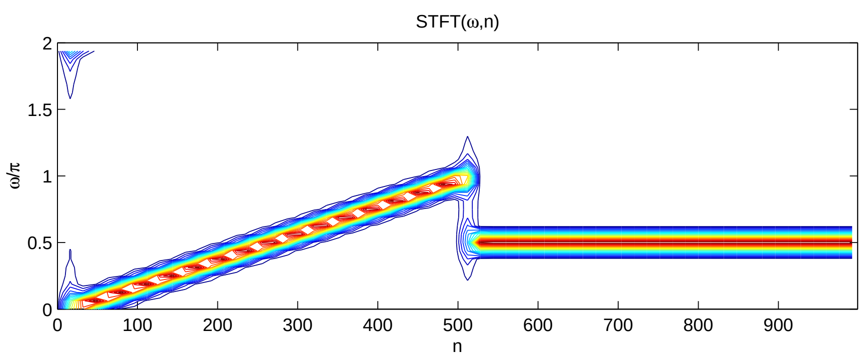
(b) The second figure below verifies the phase-derivative interpretation of instantaneous frequency. During time $n = 0..M/2-1$, the phase derivative equals $2\frac{n\pi}{M}$, while during time $n = M/2..M-1$, the phase derivative is zero. Note that ω_n is a continuous with respect to n but the derivative is not!

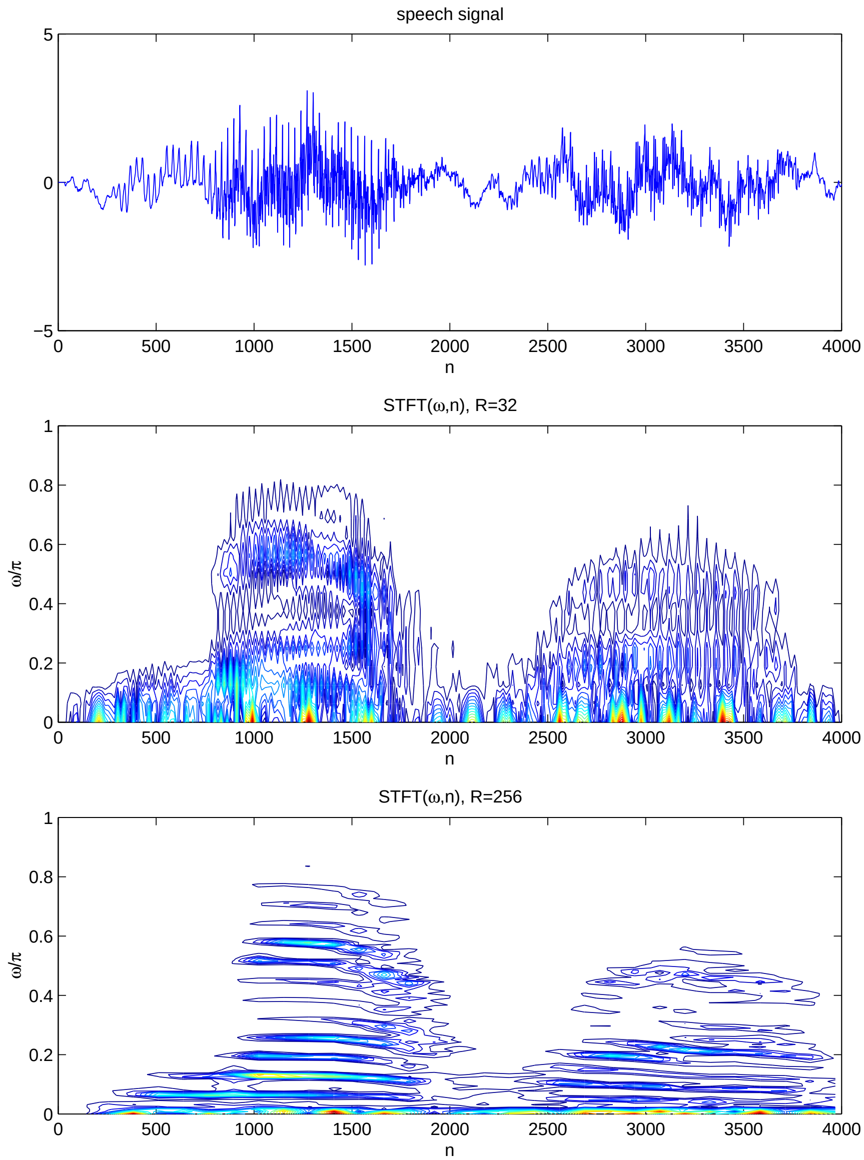
(c) The third plot below uses the STFT to analyze a speech signal. With a relatively short time duration ($R = 32$) we get good temporal resolution (as can be seen by comparing the STFT magnitudes to the time-domain signal). With a longer window ($R = 256$), we get much better spectral resolution: we can begin to see the individual resonants of the vocal tract (called "vocal formats"). Note that the resonances change frequency (slightly) over time.


ω(n)
ω/π
2
1.5
1
0.5
0
0
100
200
300
400
500
600
700
800
900
n


STFT(ω,n)
ω/π
2
1.5
1
0.5
0
0
100
200
300
400
500
600
700
800
900
n


ω(n)
ω/π
2
1.5
1
0.5
0
0
100
200
300
400
500
600
700
800
900
n


STFT(ω,n)
ω/π
2
1.5
1
0.5
0
0
100
200
300
400
500
600
700
800
900
n


speech signal
5
0
−5
0
500
1000
1500
2000
2500
3000
3500
4000
n
STFT(ω,n), R=32
1
0.8
0.6
0.4
0.2
0
ω/π
0
500
1000
1500
2000
2500
3000
3500
4000
n
STFT(ω,n), R=256
1
0.8
0.6
0.4
0.2
0
ω/π
0
500
1000
1500
2000
2500
3000
3500
4000
n