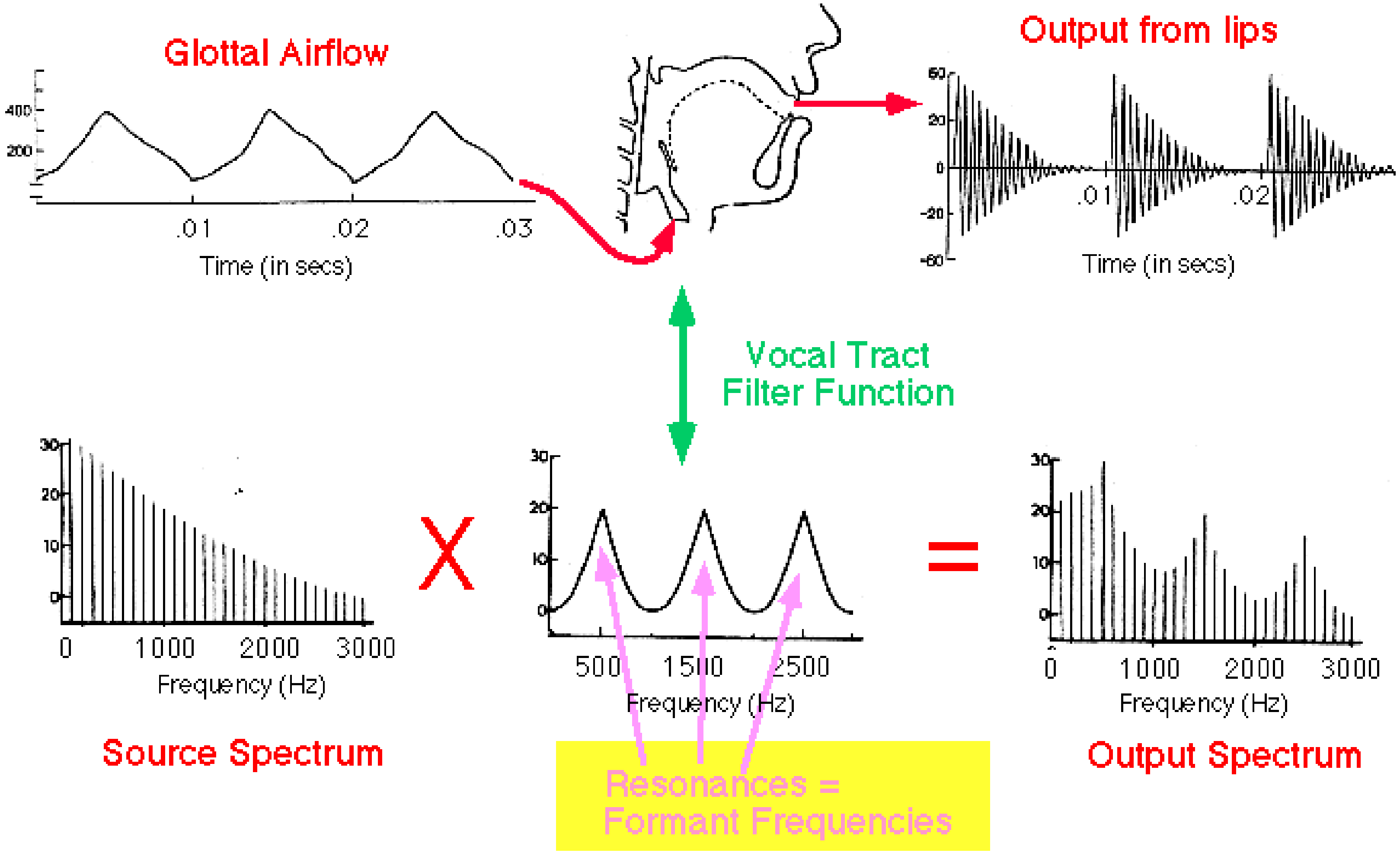
# E9 205 Machine Learning for Signal Processing
21-8-17

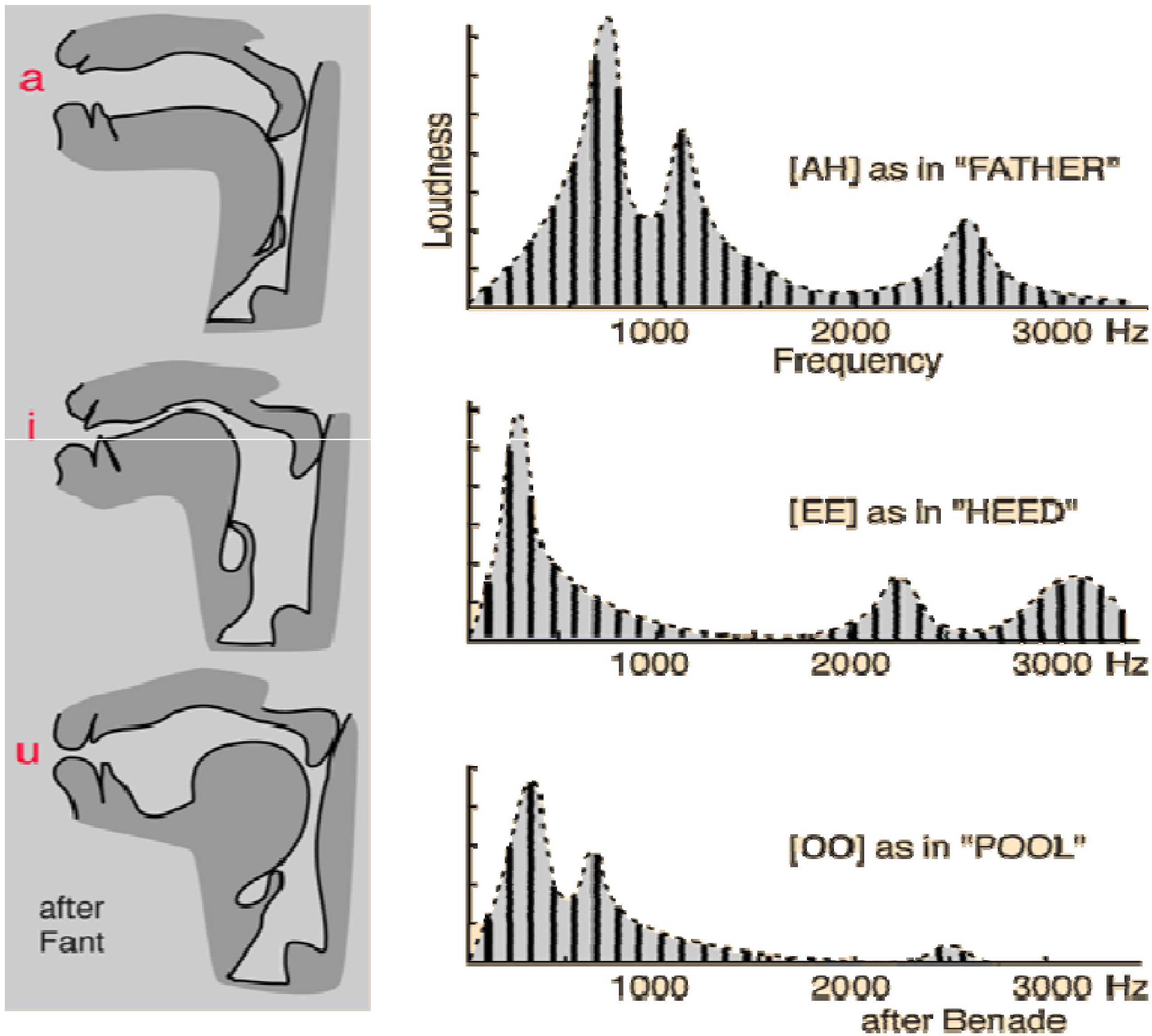
# Outline

- Melfrequency cepstral Coefficients (MFCC)

- Linear Prediction

# Source filter model of speech

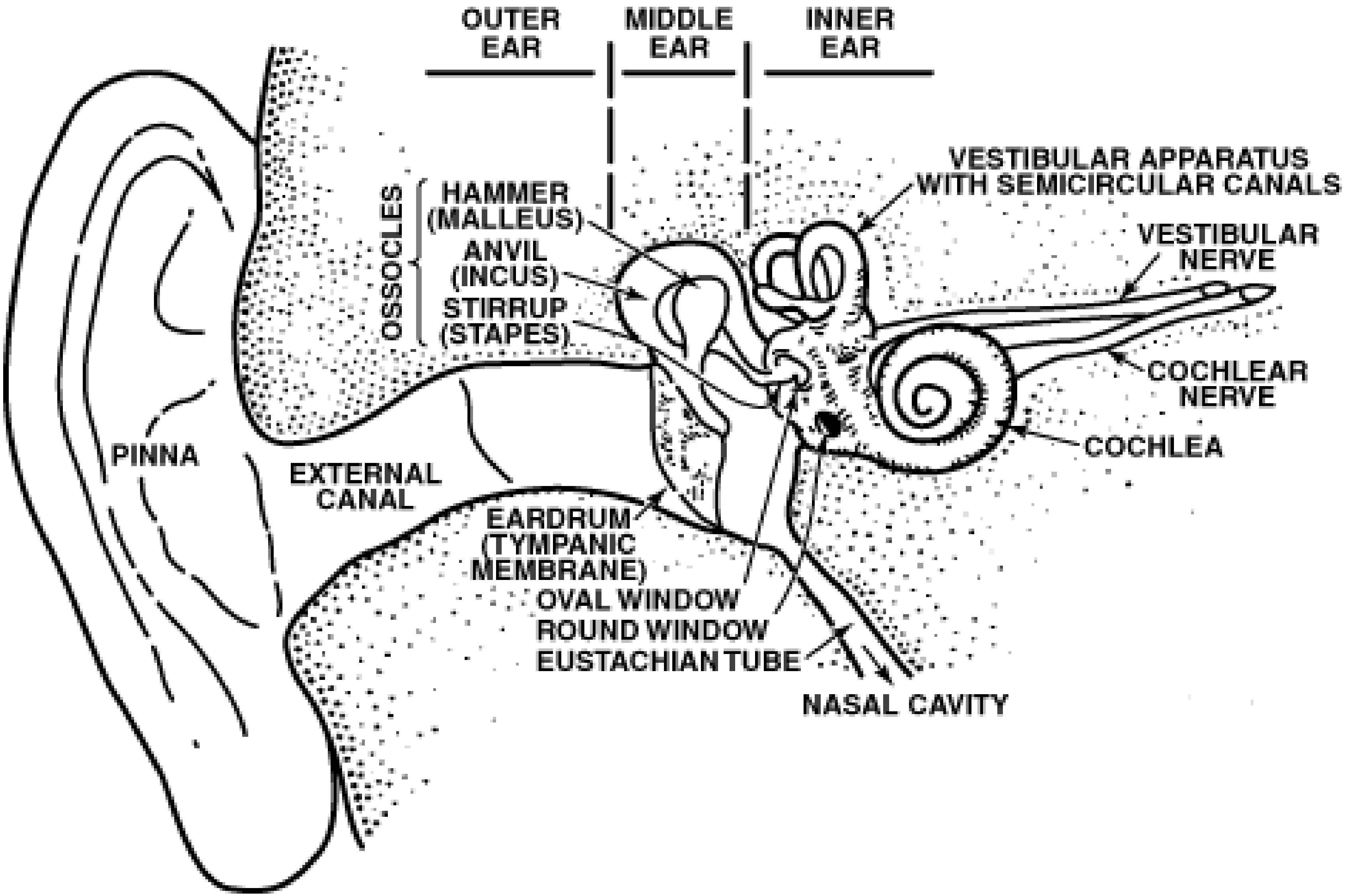



a
i
u
after
Fant
Loudness
[AH] as in "FATHER"
1000
2000
3000 Hz
Frequency
[EE] as in "HEED"
1000
2000
3000 Hz
[OO] as in "POOL"
1000
2000
3000 Hz
after Benade

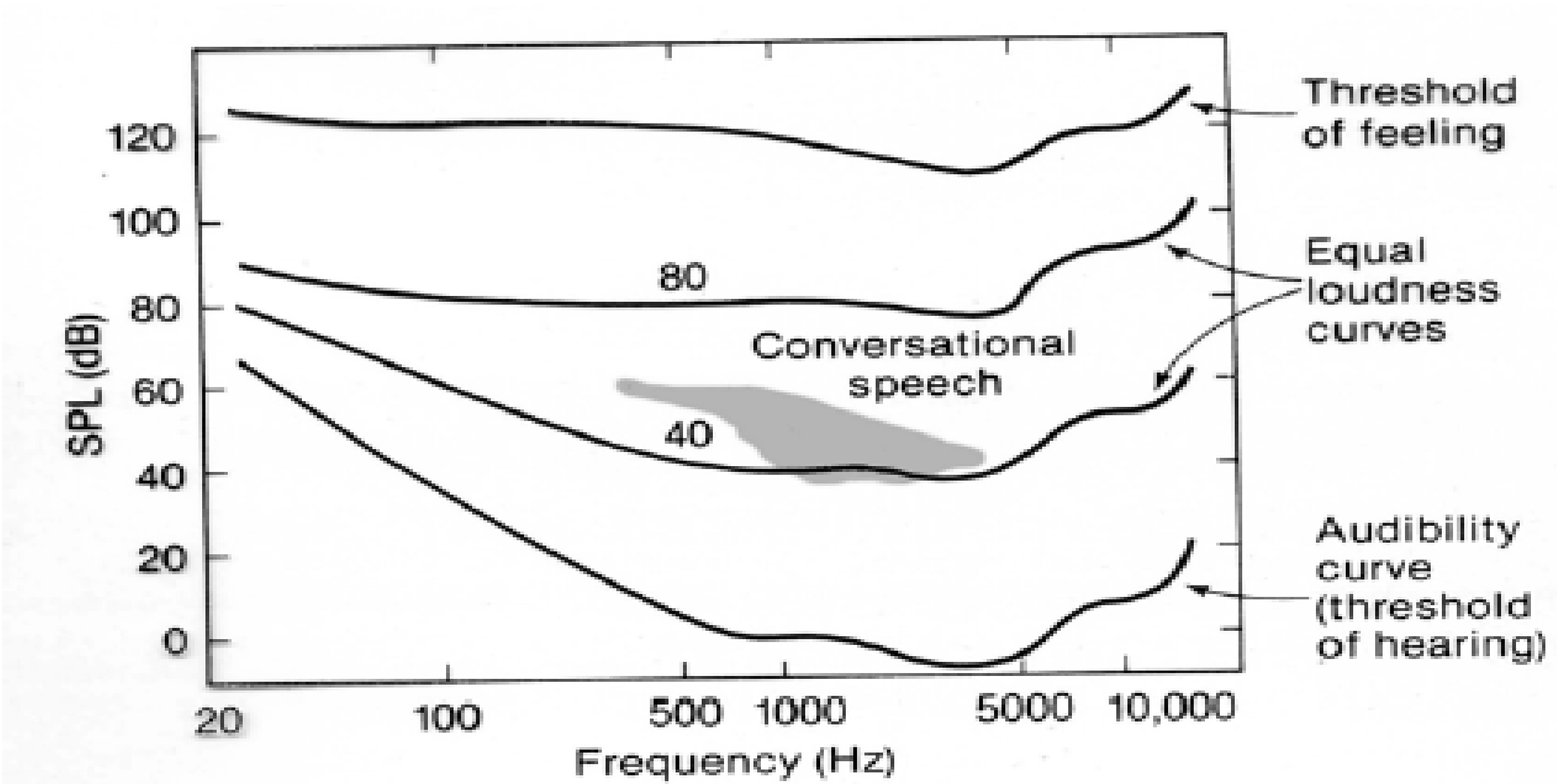
# MFCC

- MFCC coefficients model the spectral energy Distribution in a perceptually meaningful way
- Why do we need?
  - Automatic speech recogonition
  - Speaker Identification
  - Audio classification

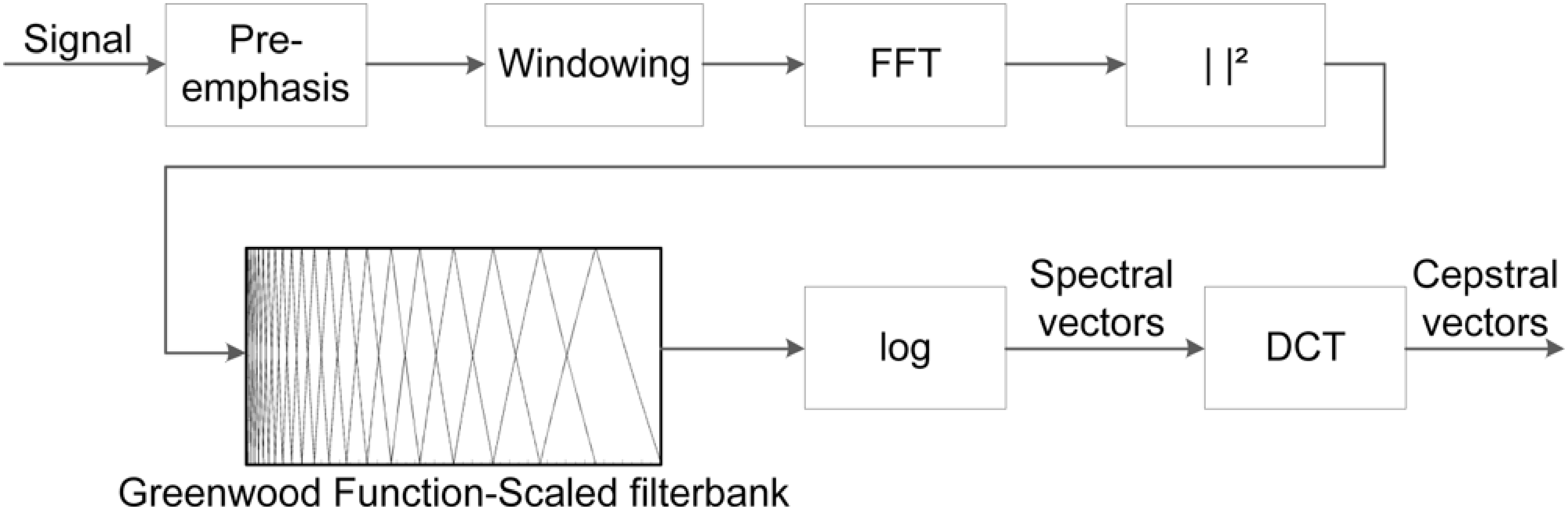
# How do we hear?



- Equal loudness contours (loudness ≠ Intensity)

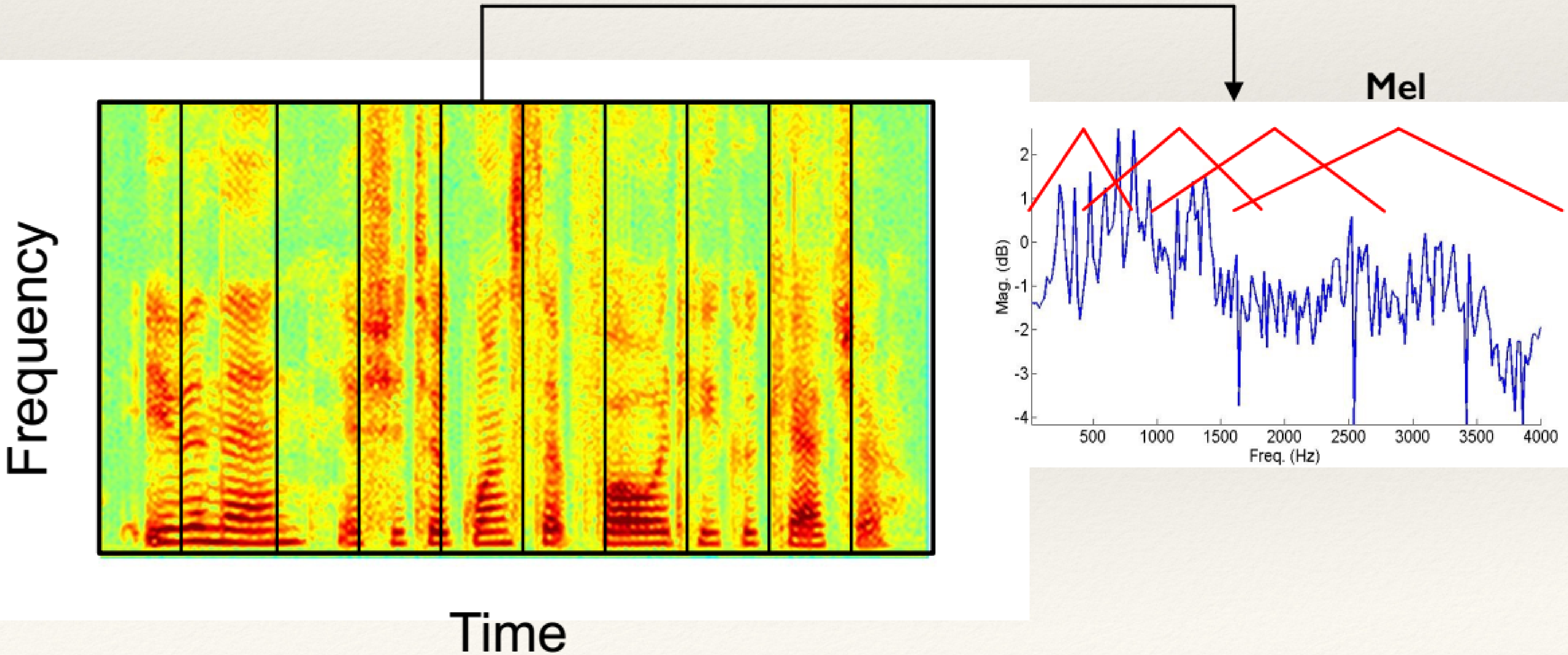


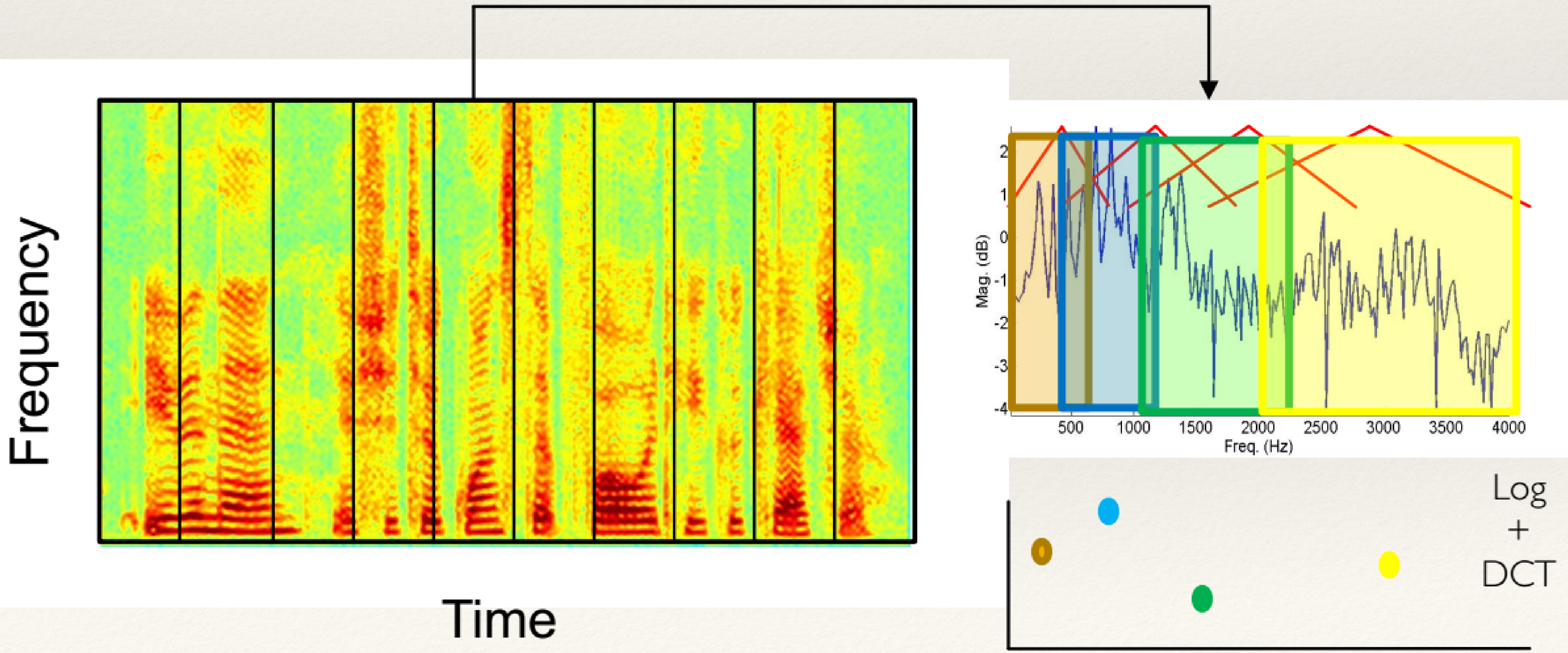
- Implementation steps





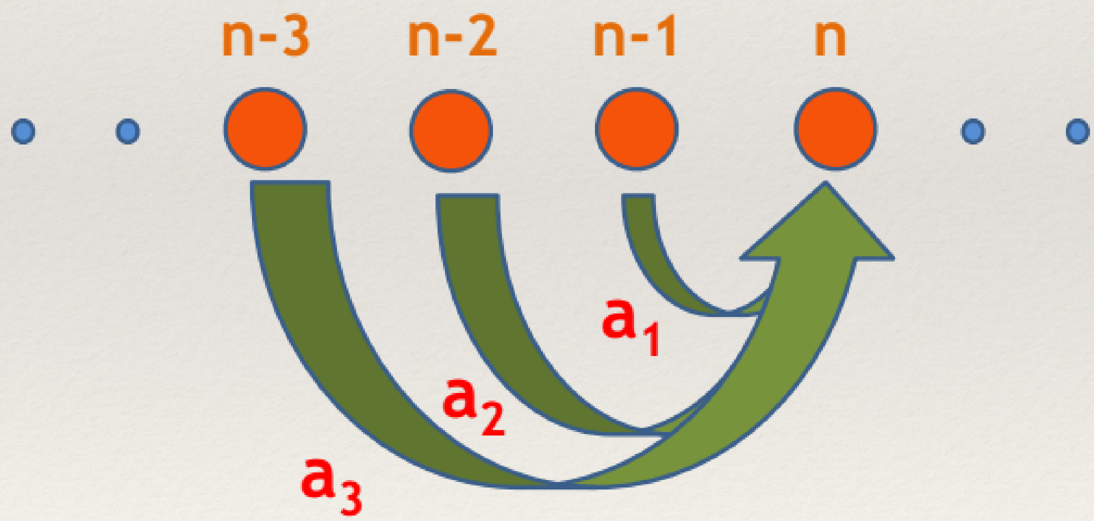
References:
[1]Chapter 4: Mohamed, Abdel-rahman. "Deep neural network acoustic models for asr." PhD diss., 2014.
[2] http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/


Mel
Frequency
Time
Mag. (dB)
2
1
0
-1
-2
-3
-4
500
1000
1500
2000
2500
3000
3500
4000
Freq. (Hz)


Frequency
Time
Mag. (dB)
500
1000
1500
2000
2500
3000
3500
4000
Freq. (Hz)
Log
+
DCT

# Linear Prediction

- Current sample expressed as a linear combination of past samples



# Properties of LP

Error signal (for the optimal predictor) is orthogonal to the samples used in the predictor.

$$e[n] \perp \{x[n-1], .., x[n-N]\}$$

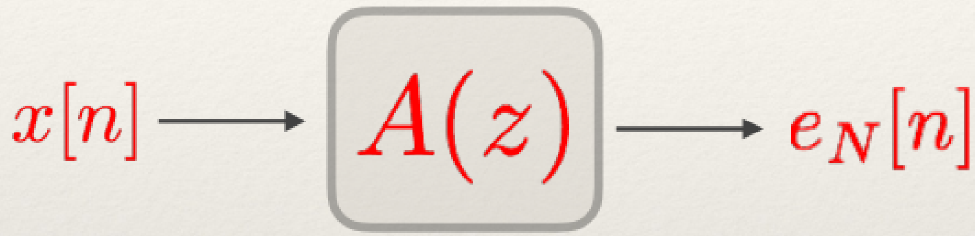
Using the orthogonality property -> normal equations

$$\mathbf{Ra} = -\mathbf{r}$$

Autocorrelation matrix is Hermitian symmetric.

# Properties of LP

Forward linear prediction filter




Properties of $A(z)$ - stability (all roots $q$ )

$$|q| < 1$$

except for line spectral process $|R(k)| = R(0) \quad for\ some\ k$

# Properties of LP

AR(N) process - Any WSS process which satisfies

$$y[n] \longrightarrow \boxed{A(z)} \longrightarrow e[n]$$

Filter is stable - error signal is white

$$e[n] \longrightarrow \boxed{\frac{1}{A_N(z)}} \longrightarrow y[n]$$

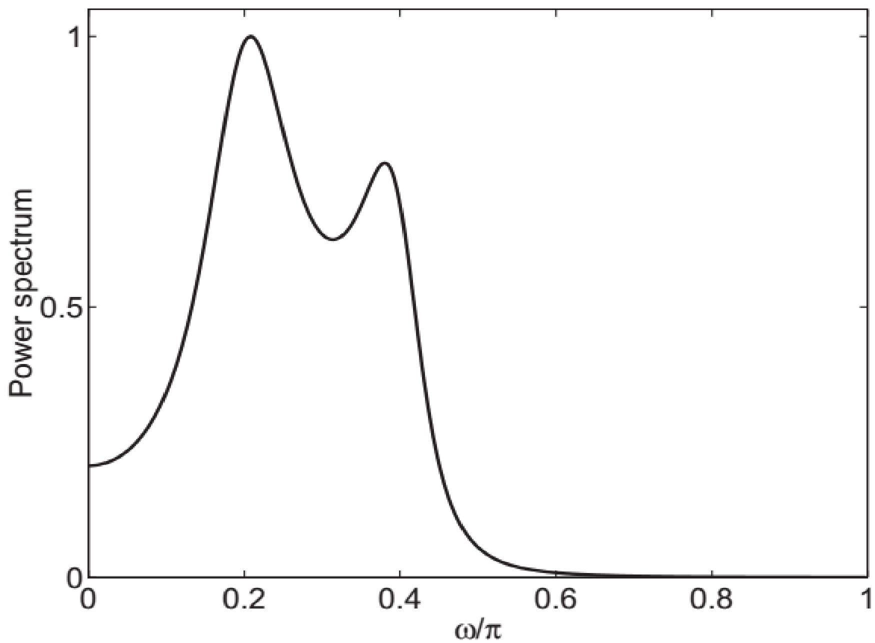
$$S_{yy}(f) = \frac{\epsilon_N}{|1 + \sum_{n=1}^{N} a_{N,n}^{*} e^{-j2\pi fn}|^2}$$

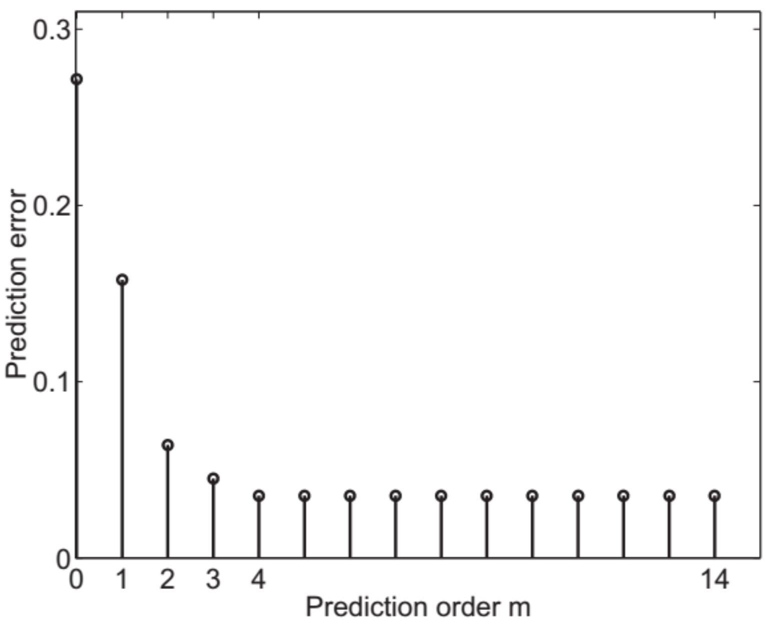
Approximating $x[n]$ by $y[n]$ i.e. $S_{xx}(f)$ with $S_{yy}(f)$

Autoregressive modeling

- AR(4) process:
  - For AR process, error decreases monotonically and then becomes a constant for m ≥ 4.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $A_0(z)$ | 1.0 | | | | | | |
| $A_1(z)$ | 1.0 | −0.6473 | 0 | 0 | 0 | 0 | 0 |
| $A_2(z)$ | 1.0 | −1.1457 | 0.7701 | 0 | 0 | 0 | 0 |
| $A_3(z)$ | 1.0 | −1.5669 | 1.3967 | −0.5469 | 0 | 0 | 0 |
| $A_4(z)$ | 1.0 | −1.8198 | 2.0425 | −1.2714 | 0.4624 | 0 | 0 |
| $A_5(z)$ | 1.0 | −1.8198 | 2.0425 | −1.2714 | 0.4624 | 0 | 0 |
| $A_6(z)$ | 1.0 | −1.8198 | 2.0425 | −1.2714 | 0.4624 | 0 | 0 |



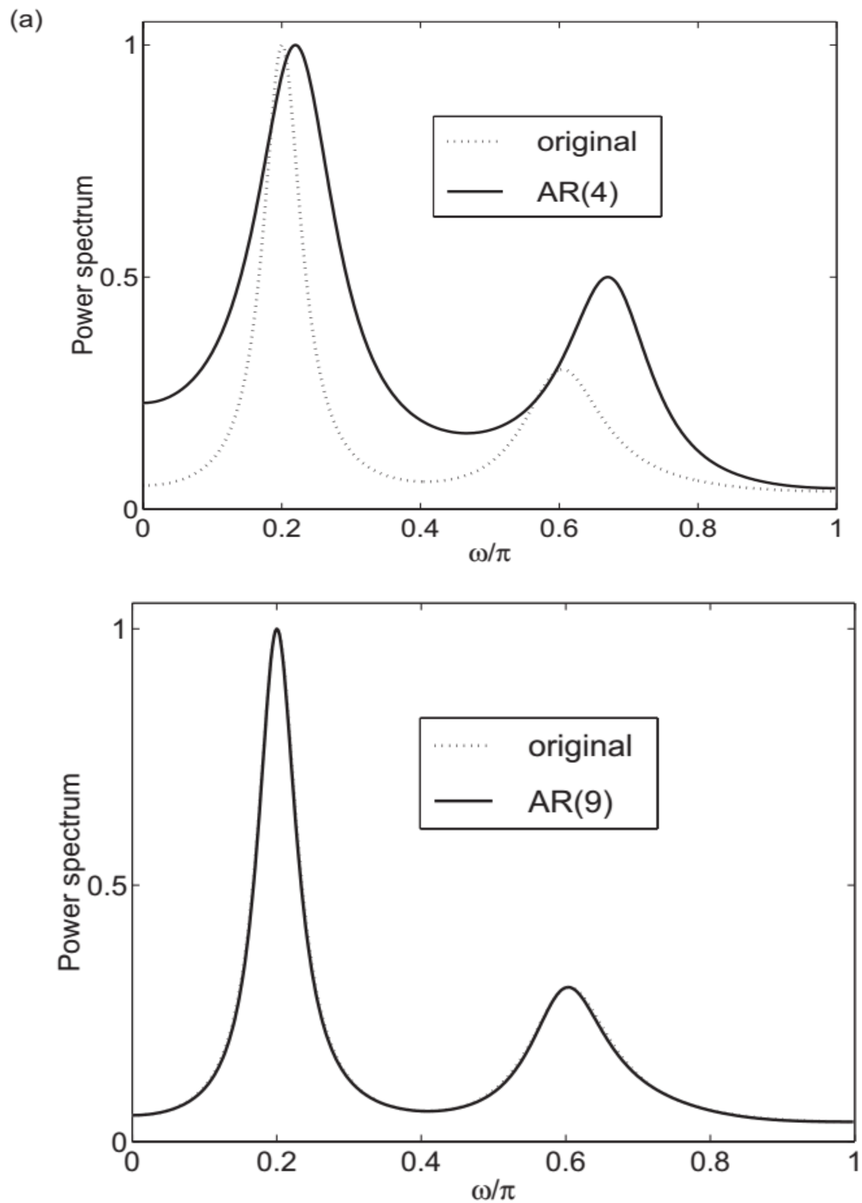
# AUTOCORRELATION MATCHING PROPERTY

- Let R(k) and r(k) be the autocorrelations of x(n) and the AR(N) approximation y(n), respectively. Then,

$$R(k) = r(k), |k| \leq N$$

Thus, the AR(N) model y(n) is an approximation of x(n) in the sense that the first N + 1 autocorrelation coefficients for the two processes are equal to each other.

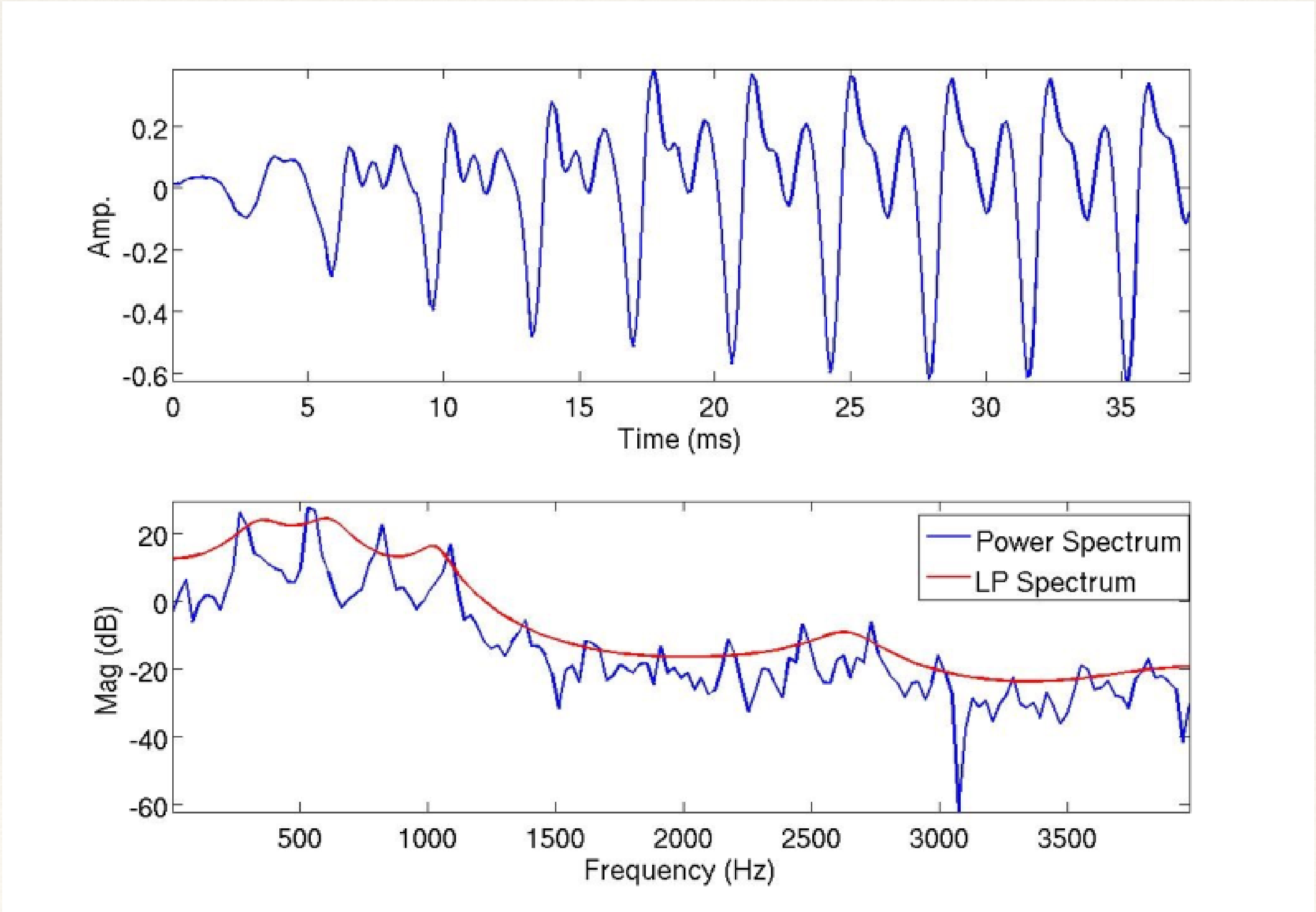
$$G_4(z) = \frac{3.9 - 2.7645z^{-1} + 1.4150z^{-2} - 0.5515z^{-3}}{1 - 0.9618z^{-1} + 0.7300z^{-2} - 0.5315z^{-3} + 0.5184z^{-4}}$$




| $m$ | $R(m)$ | $r(m)$ |
|---|---|---|
| 0 | 0.1667 | 0.1667 |
| 1 | 0.0518 | 0.0518 |
| 2 | −0.0054 | −0.0054 |
| 3 | 0.0031 | 0.0031 |
| 4 | −0.0519 | −0.0519 |
| 5 | −0.0819 | −0.0819 |
| 6 | −0.0364 | −0.0364 |
| 7 | −0.0045 | −0.0045 |
| 8 | 0.0057 | 0.0057 |
| 9 | 0.0318 | 0.0318 |
| 10 | 0.0430 | 0.0441 |
| 11 | 0.0234 | 0.0241 |

# Linear Prediction

## AR Model of the Power Spectrum of the Signal



# References

[1] Vaidyanathan, P. P. "The theory of linear prediction." Synthesis lectures on signal processing 2.1 (2007): 1-184.

- Chapter 2, 5
- Appendix A, B, C

[2] Makhoul, John. "Linear prediction: A tutorial review." Proceedings of the IEEE 63.4 (1975): 561-580.

Thank you