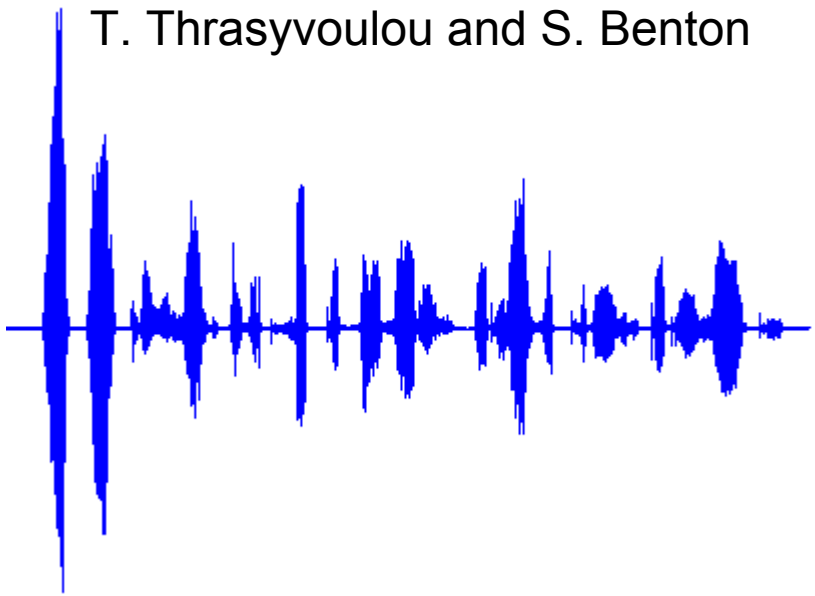
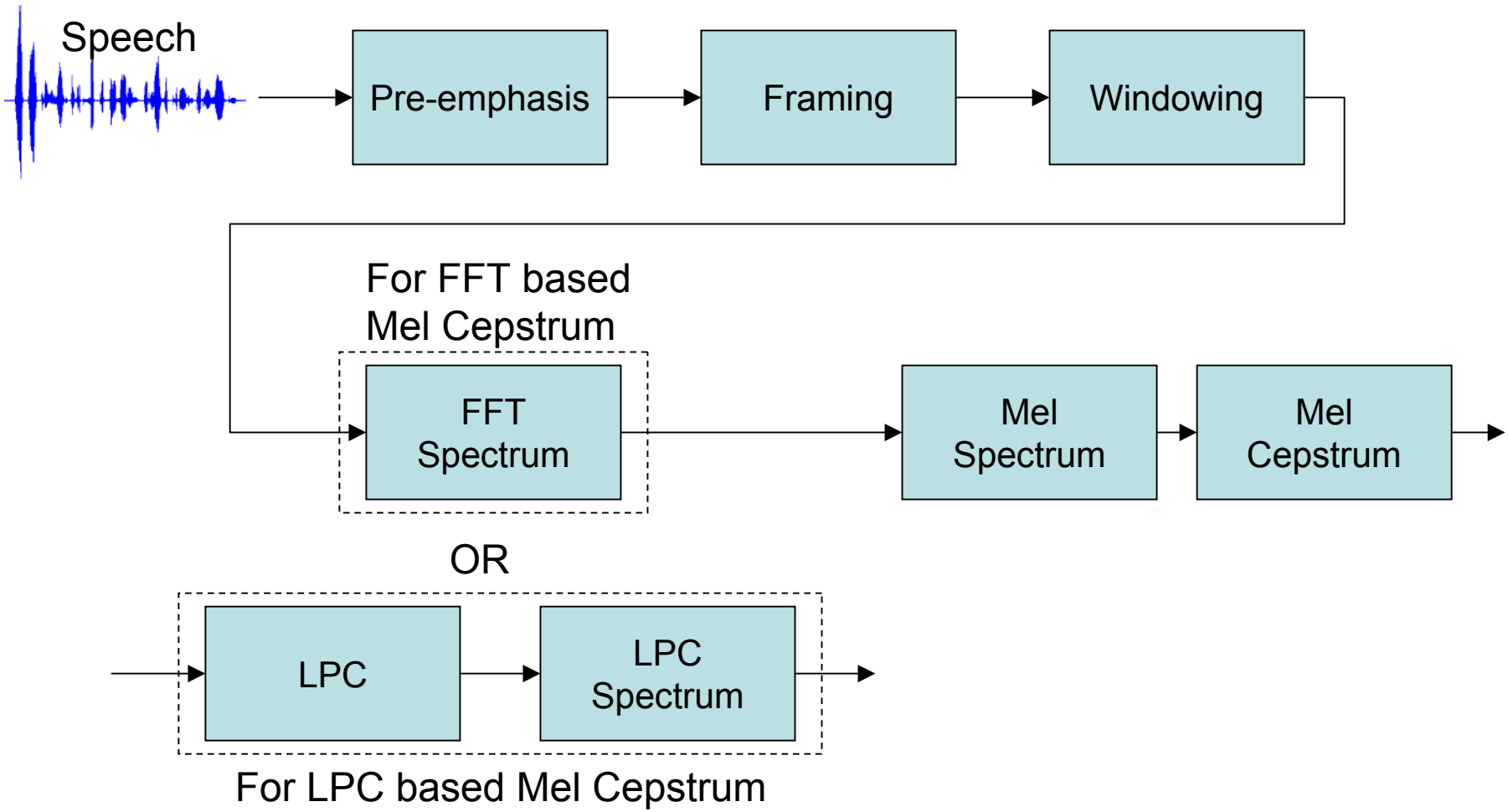
Speech parameterization using the Mel scale Part II

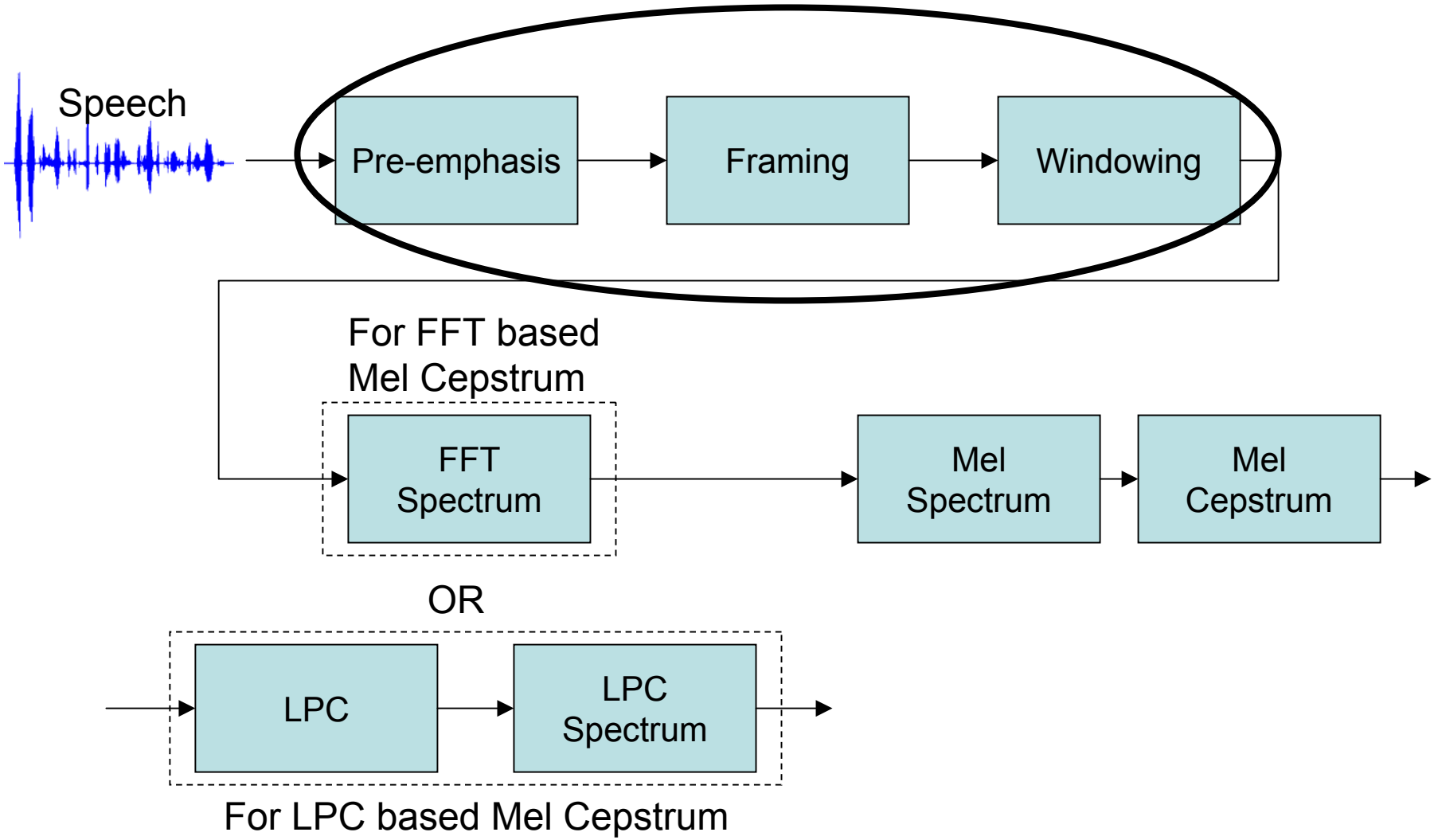
T. Thrasyvoulou and S. Benton



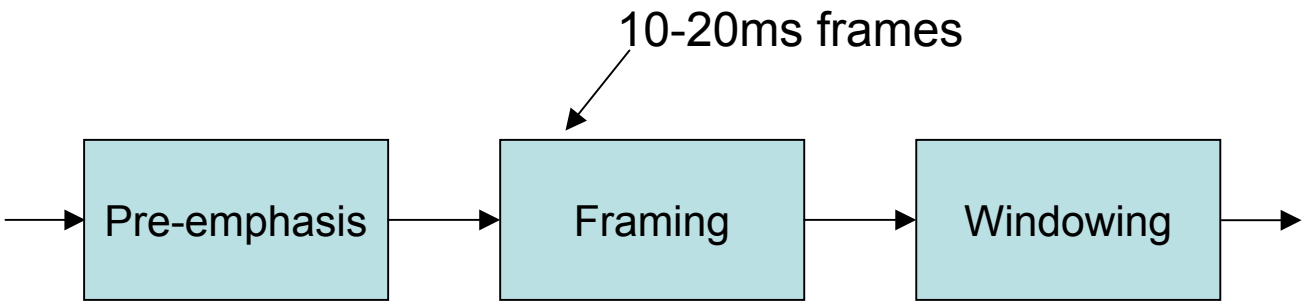
Speech Cepstrum procedure



Speech Cepstrum procedure



Pre-emphasis/Framing/Windowing



$H(z)=1+az^{-1}$
$a=-0.95$

Used to boost the signal spectrum approximately 20dB/decade

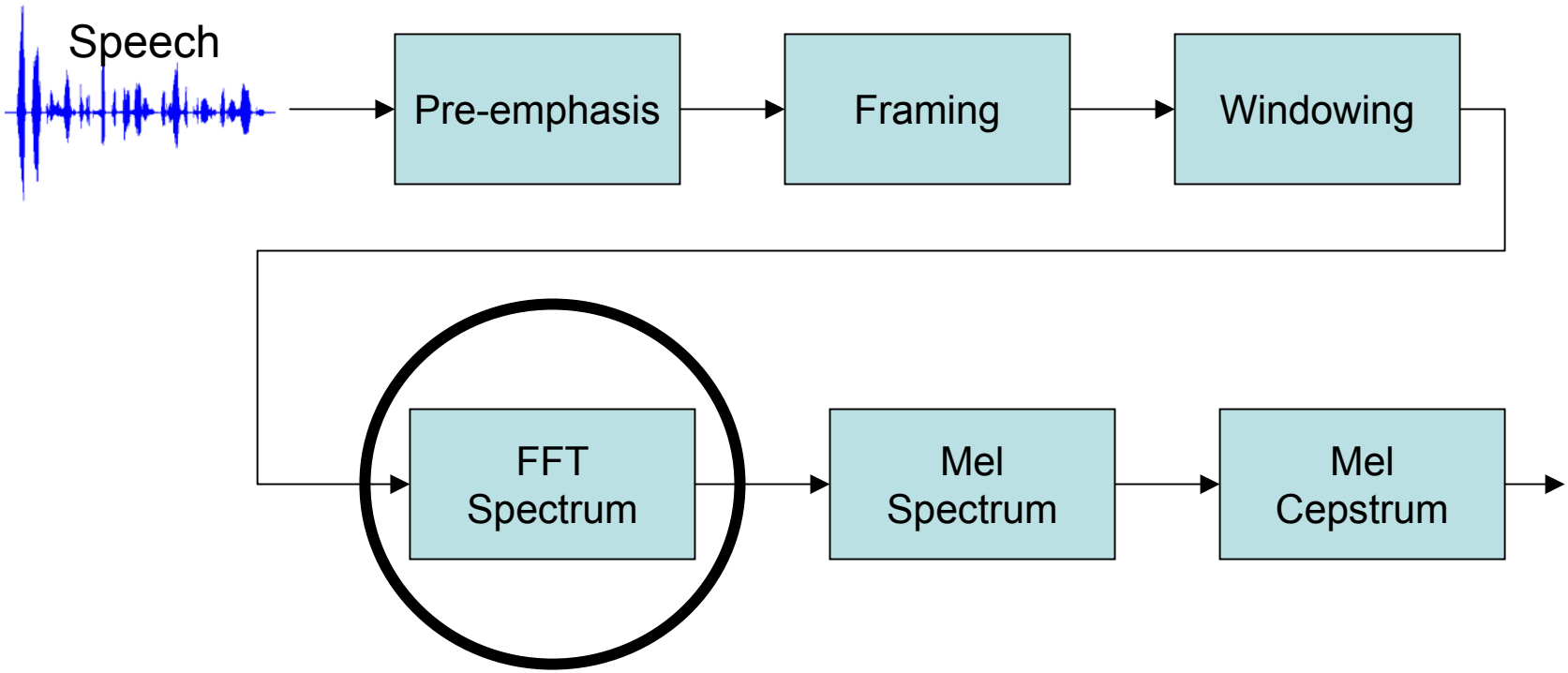
Hamming window

$$0.54-0.46\cos(\frac{2\pi n}{N-1})$$

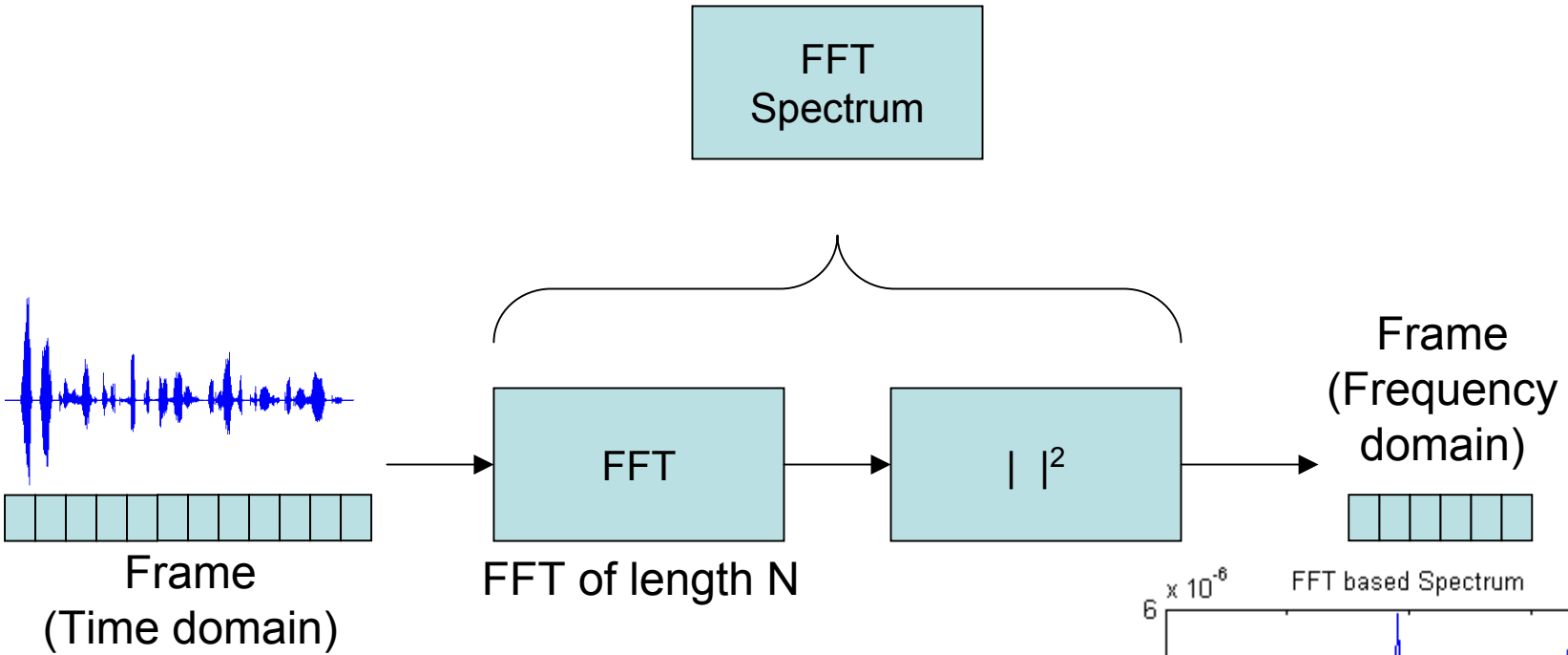
Reduces the edge effect when taking the FFT
Improves spectral estimate accuracy

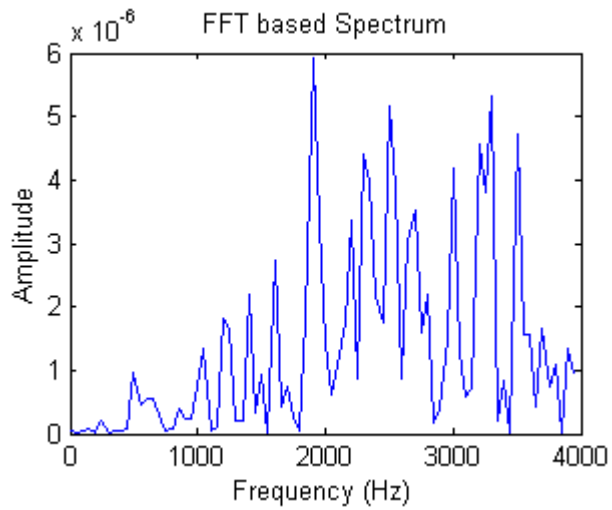
Speech Cepstrum procedure

For FFT based Mel Cepstrum



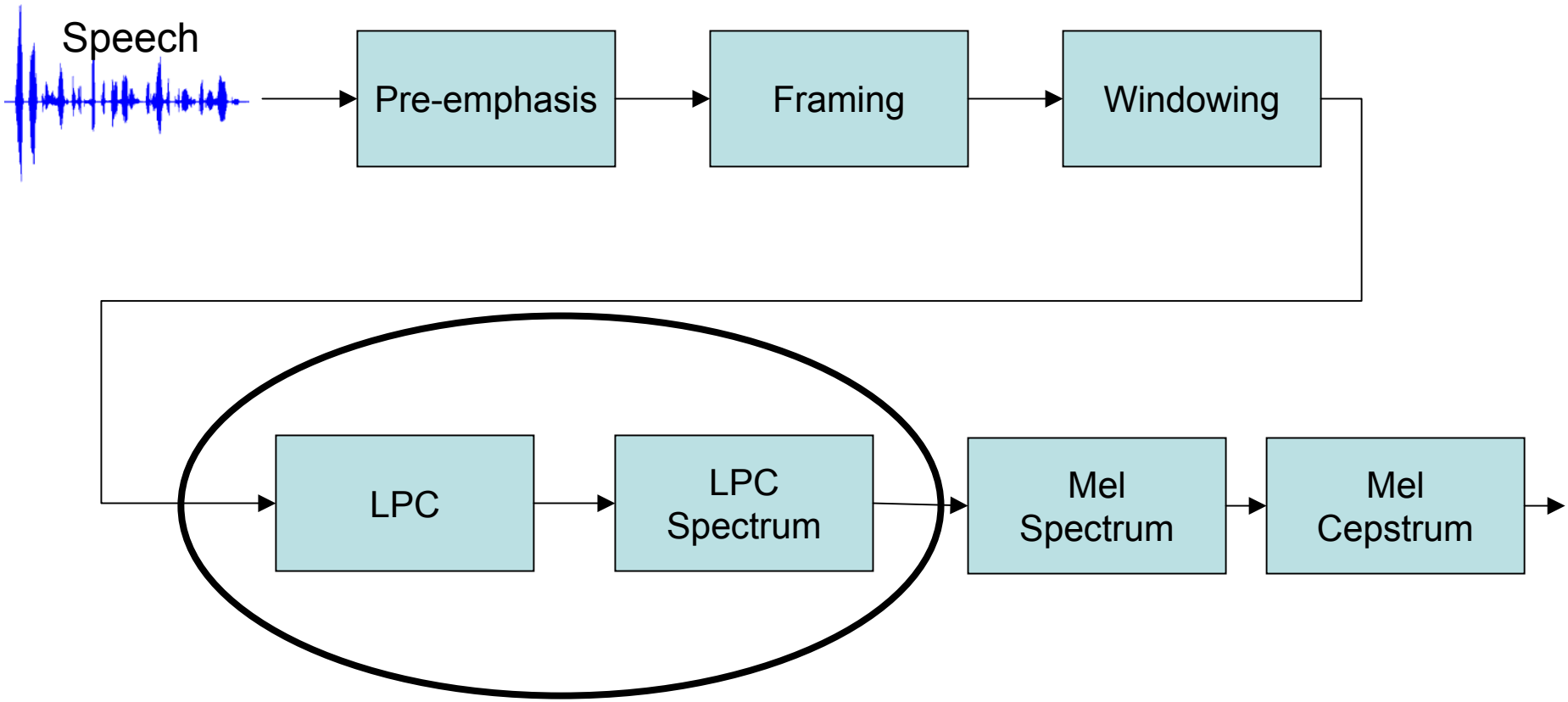
FFT Spectrum



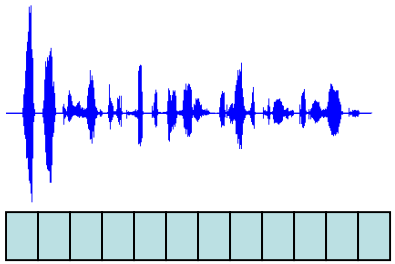


Speech Cepstrum procedure

For LPC based Mel Cepstrum

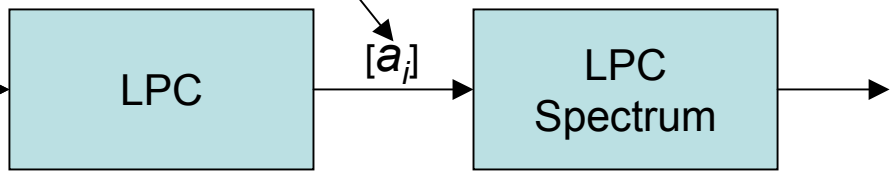


LPC Spectrum

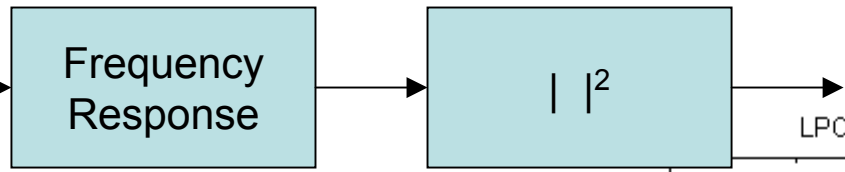


Frame
(Time domain)

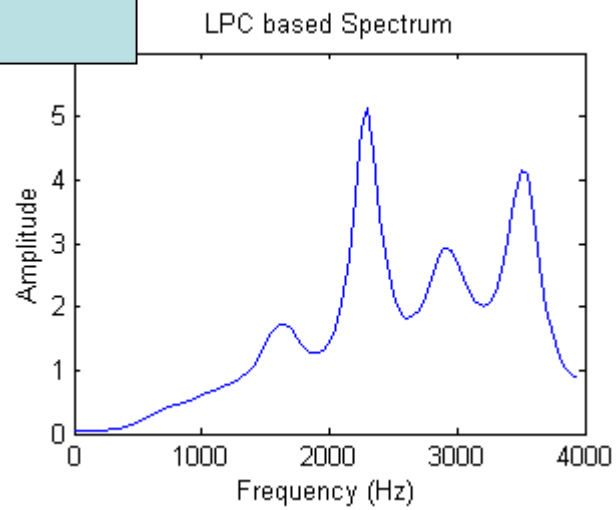
LP coefficients



LP
coefficients
$[a_i]$

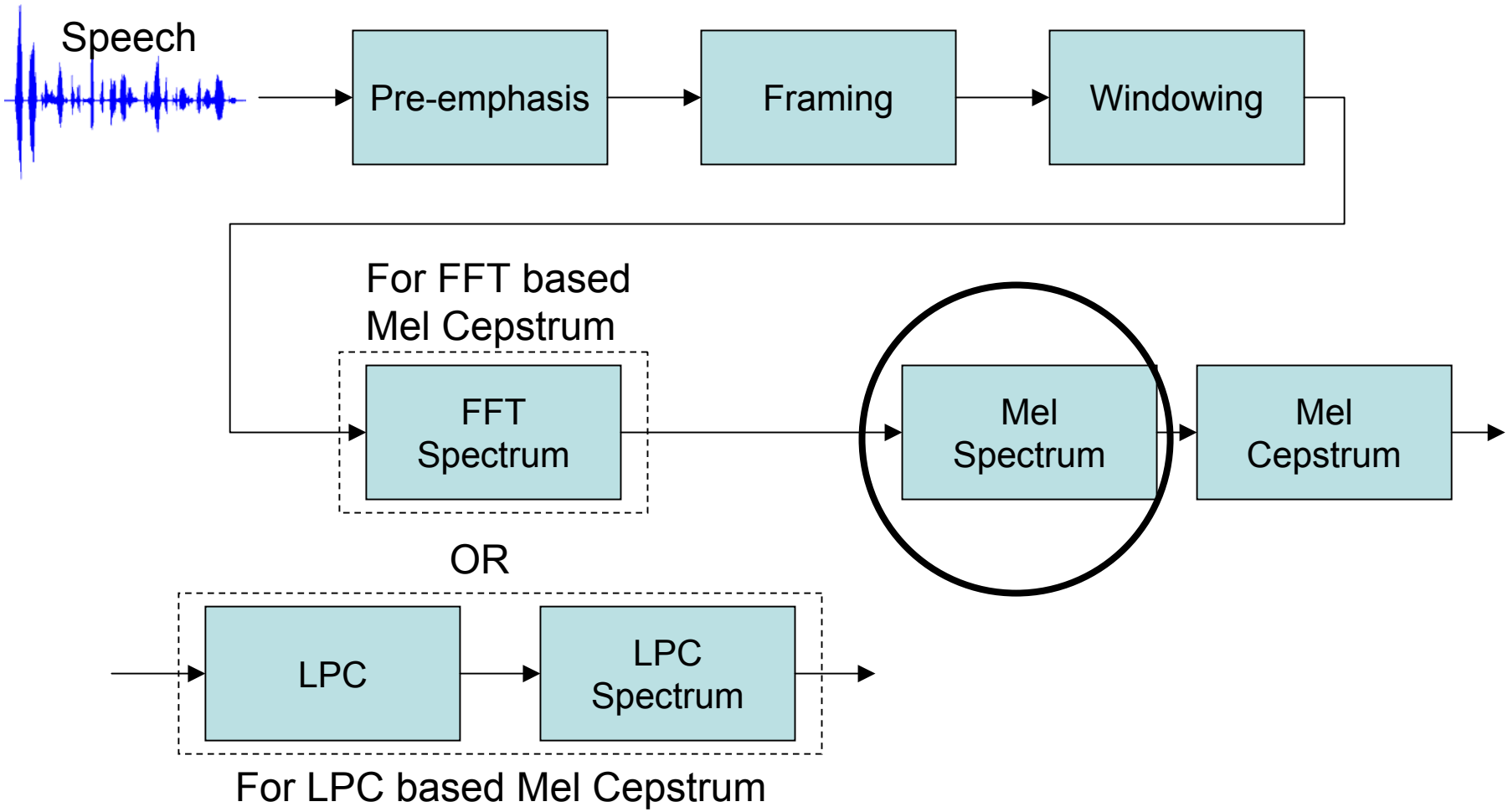


$$H(z) = \frac{G}{1 + \sum_{i=1}^{M} a_i z^{-i}}$$



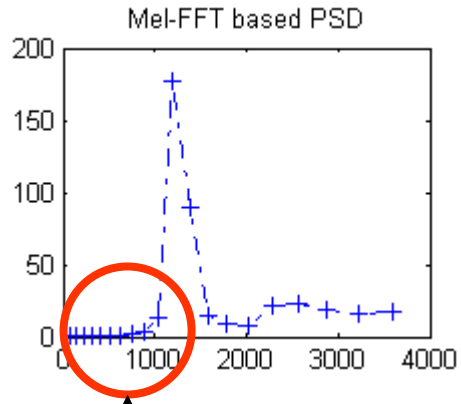
Reduces the spectrum variance
➔Eliminates various source information

Speech Cepstrum procedure

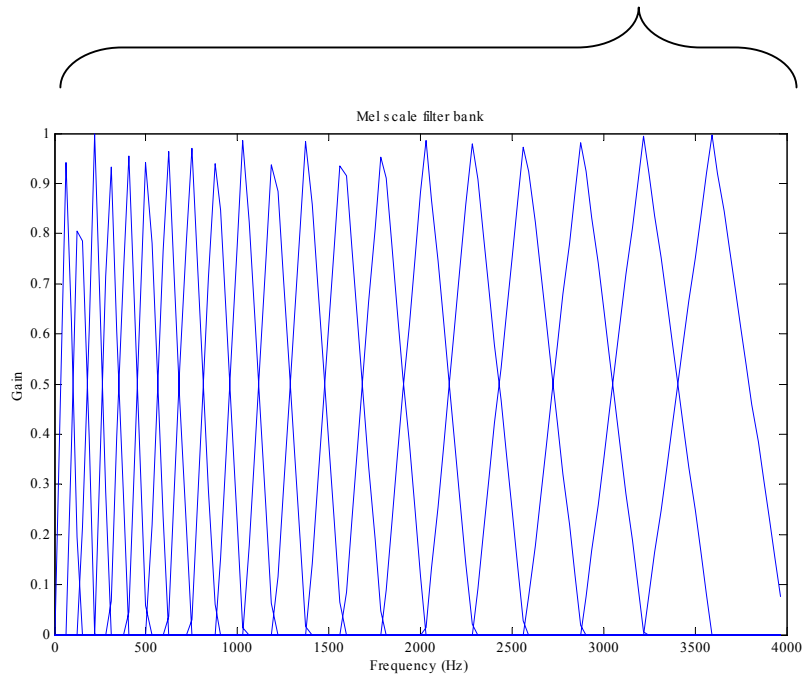


Mel Spectrum

FFT or LPC Spectrum s(k) → Mel Filter Bank →



samples =
filters in filter bank



Mel Filter bank

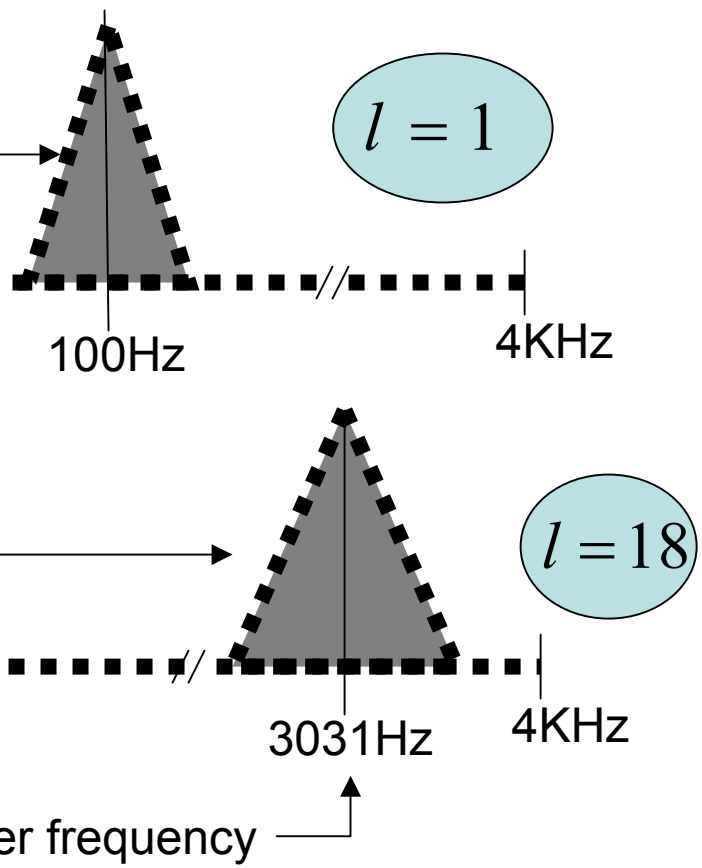
Mel Filter bank frequency table

Index	Bark Scale		Mel Scale	
	Center Freq. (Hz)	BW (Hz)	Center Freq. (Hz)	BW (Hz)
1	50	100	100	100
2	150	100	200	100
3	250	100	300	100
4	350	100	400	100
5	450	110	500	100
6	570	120	600	100
7	700	140	700	100
8	840	150	800	100
9	1000	160	900	100
10	1170	190	1000	124
11	1370	210	1149	160
12	1600	240	1320	184
13	1850	280	1516	211
14	2150	320	1741	242
15	2500	380	2000	278
16	2900	450	2297	320
17	3400	550	2639	367
18	4000	700	3031	422
19	4800	900	3482	484
20	5800	1100	4000	556
21	7000	1300	4595	639
22	8500	1800	5278	734
23	10500	2500	6063	843
24	13500	3500	6964	969

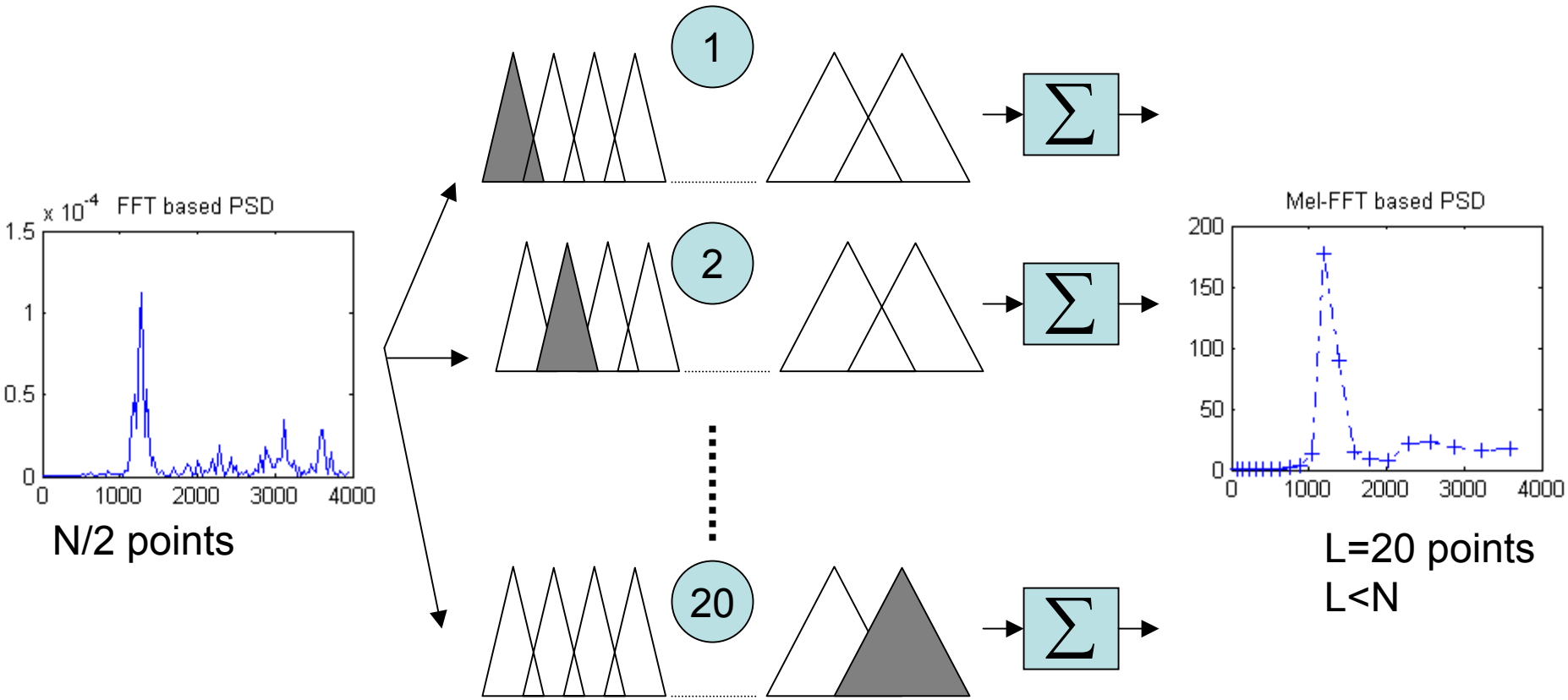
$M_l(k)$

$l = 0,1,...L-1$

$k = 0,1,...,N/2$

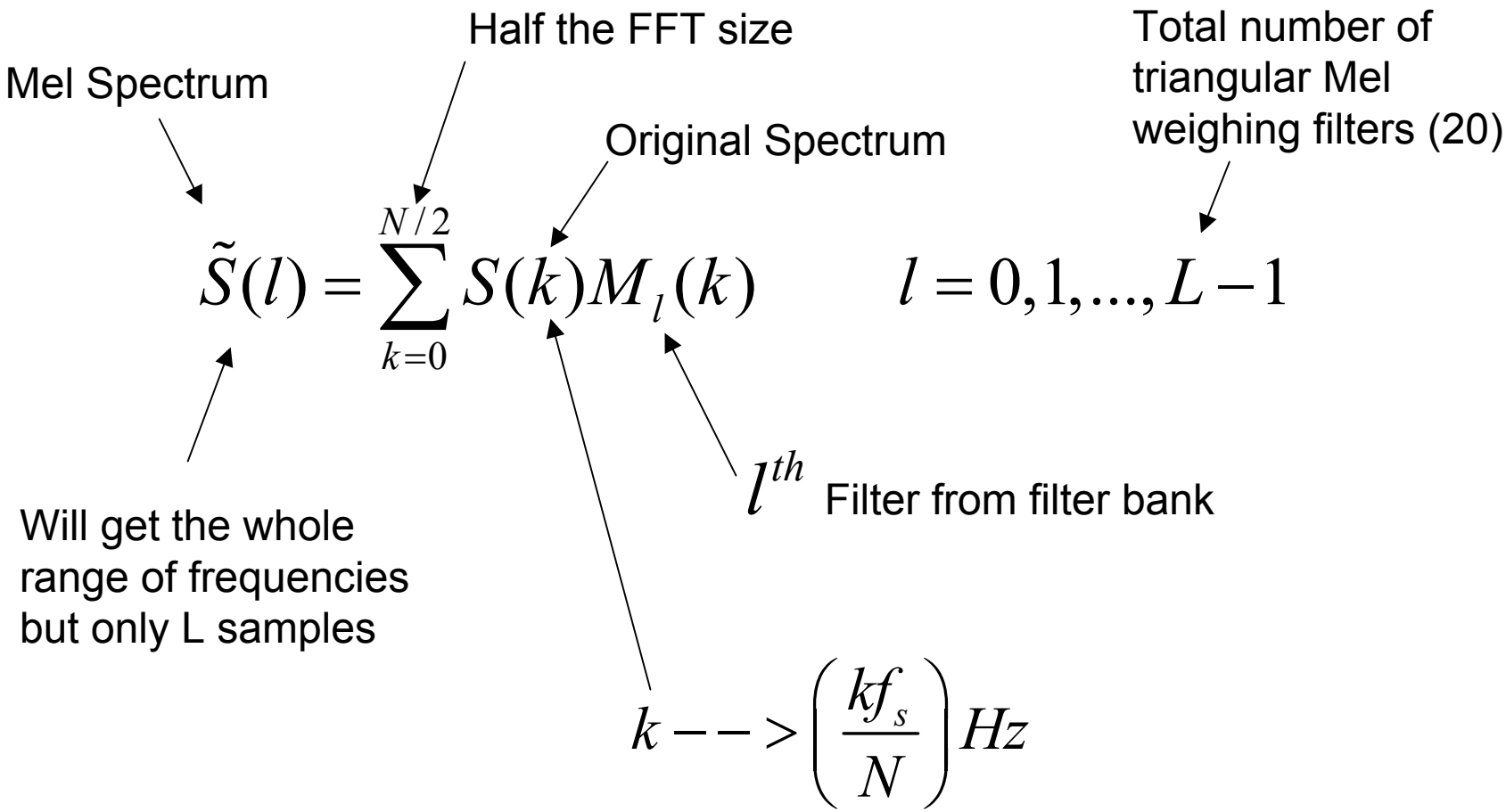


Mel Filter bank

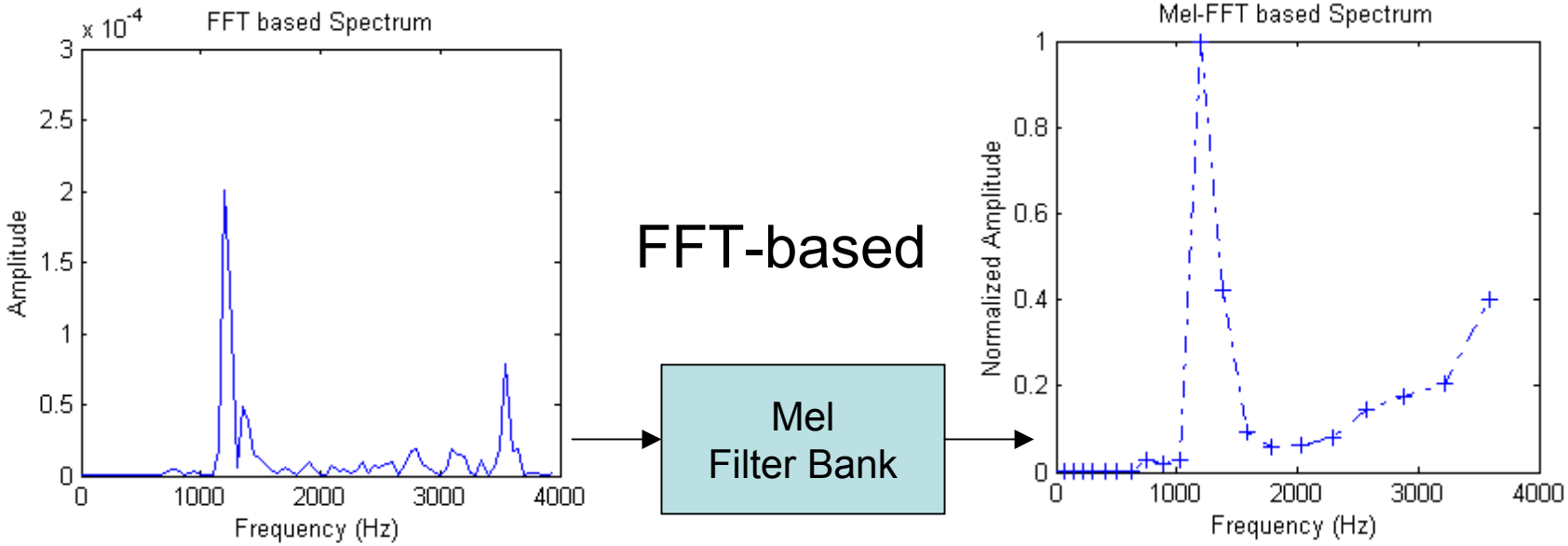


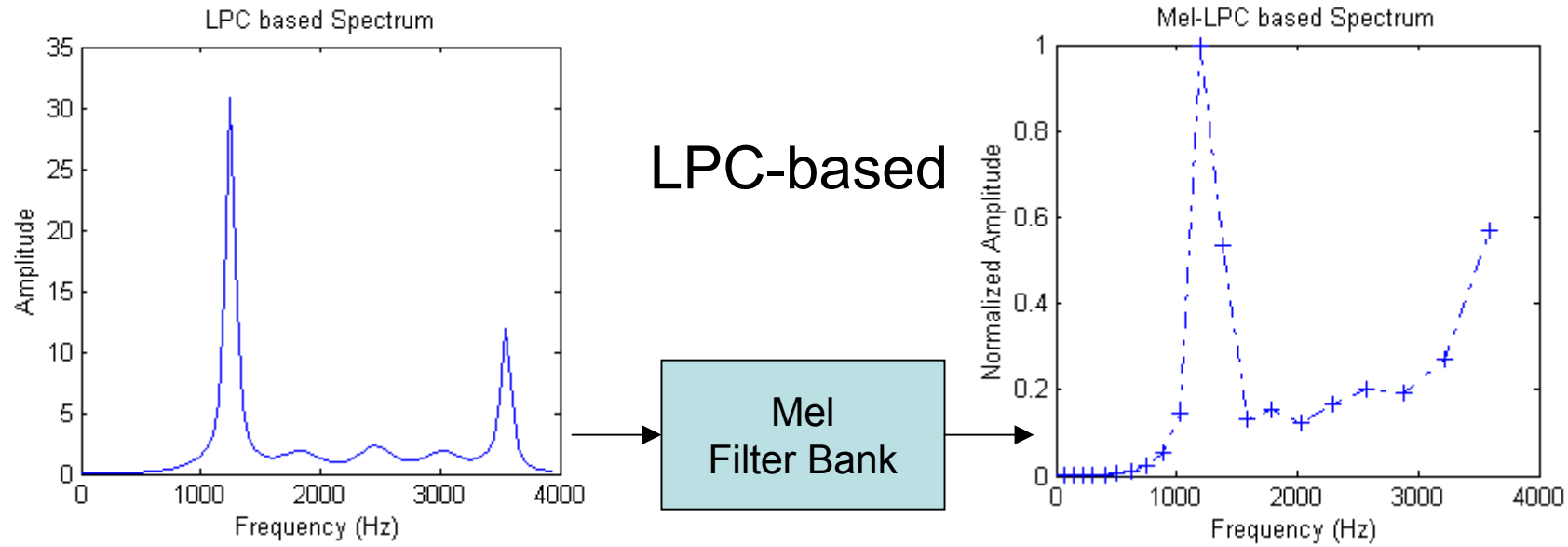
Multiply the power spectrum with each of the triangular Mel weighting filters and add the result➔ Perform a weighted averaging procedure around the Mel frequency, similar to a **weighted** Daniel periodogram

Mel Filter bank

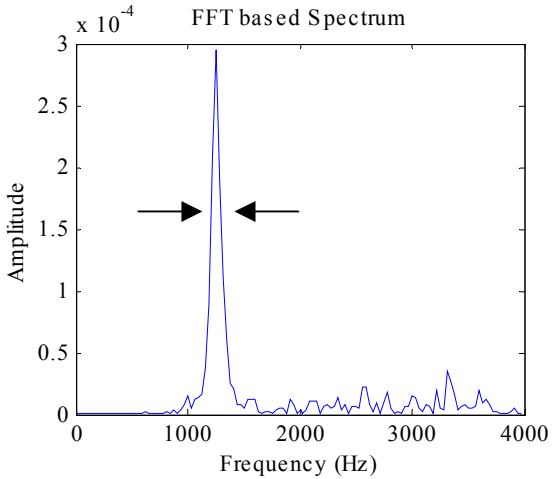


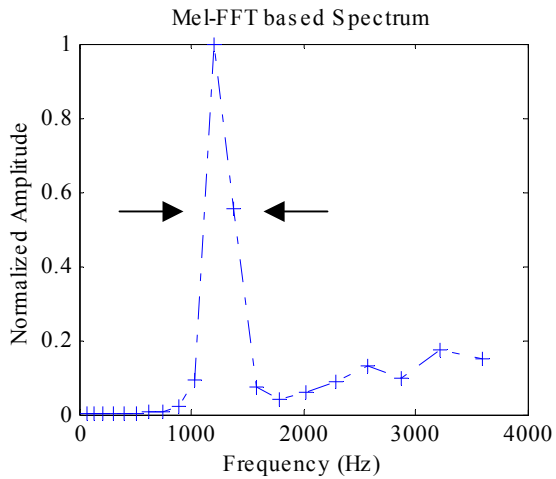
Mel spectrum plots

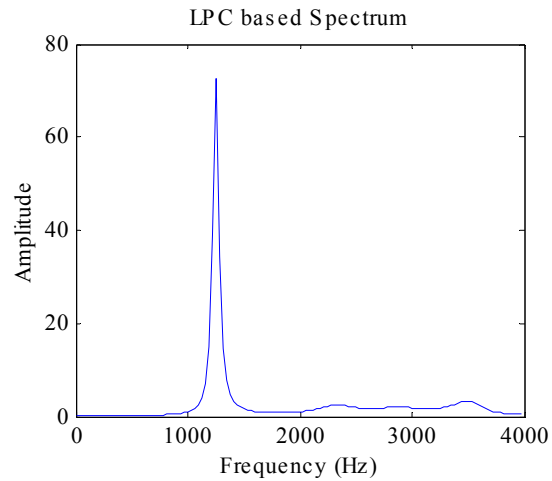


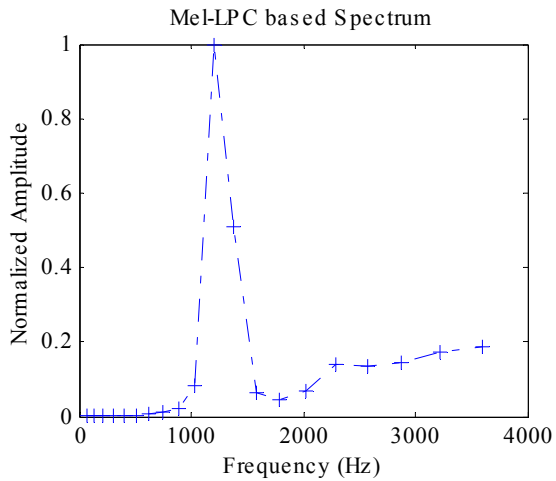


Mel spectrum plots





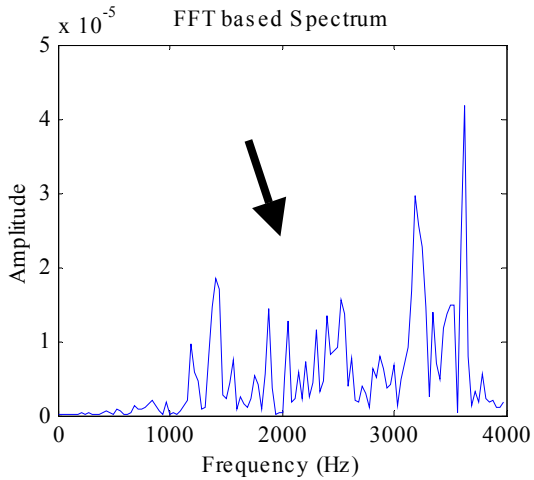


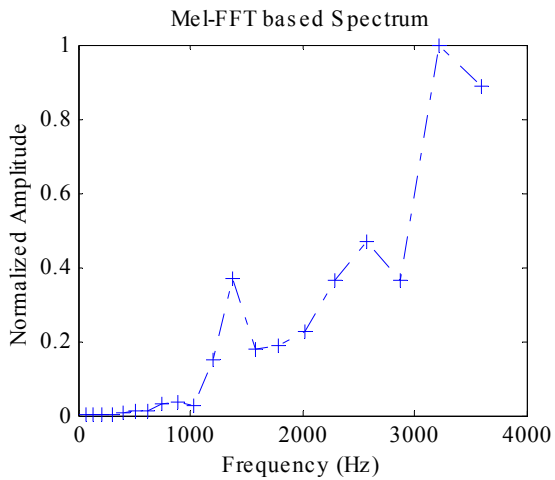


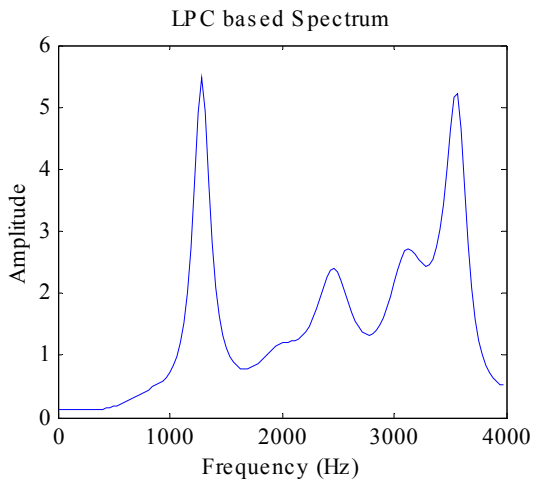
- BW of the peak expands because of the loss of resolution in Mel scale

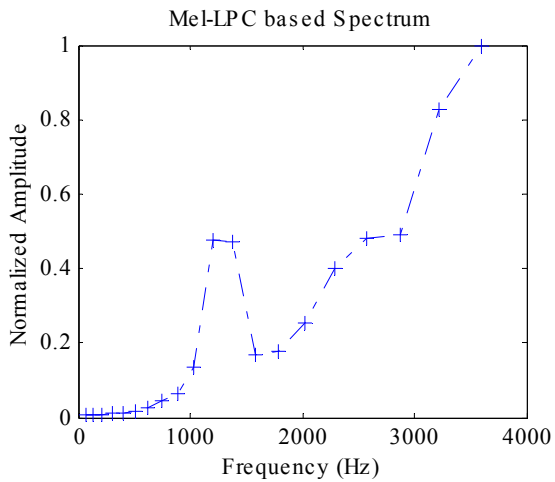
- Peaks that are not located exactly at the center Mel frequency of their corresponding filter may move to center on the Mel frequency
e.g.
110Hz peak → 100Hz

Mel spectrum plots



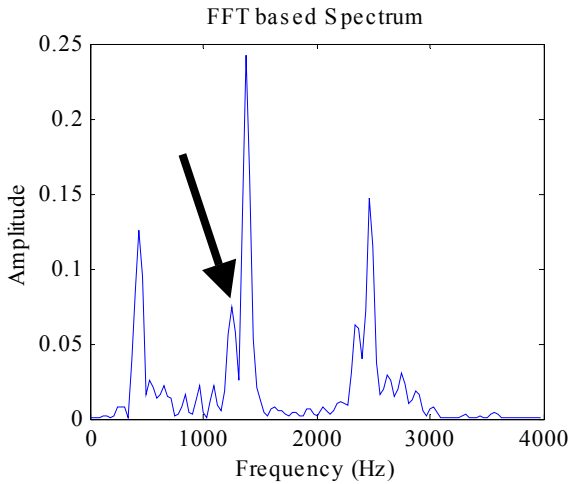


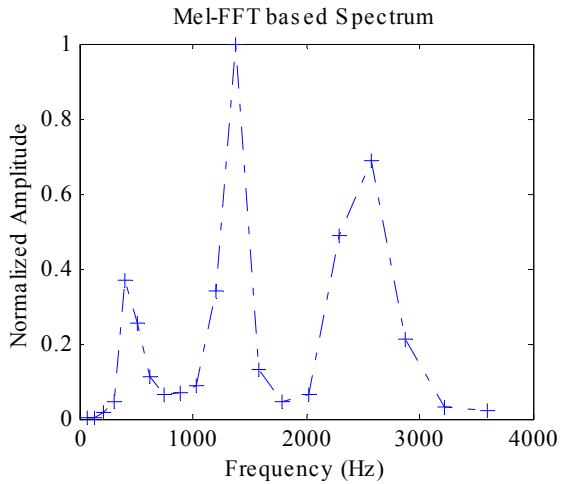


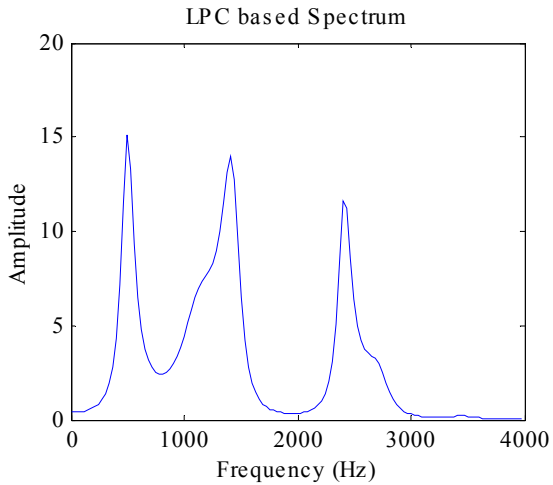


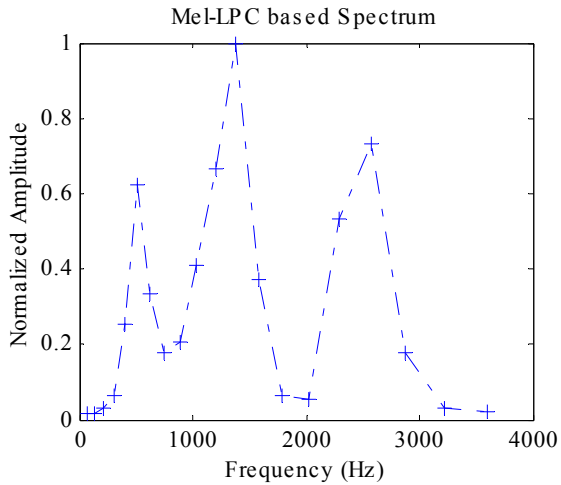
- Loss of resolution
- Maintains prominent peaks

Mel spectrum plots



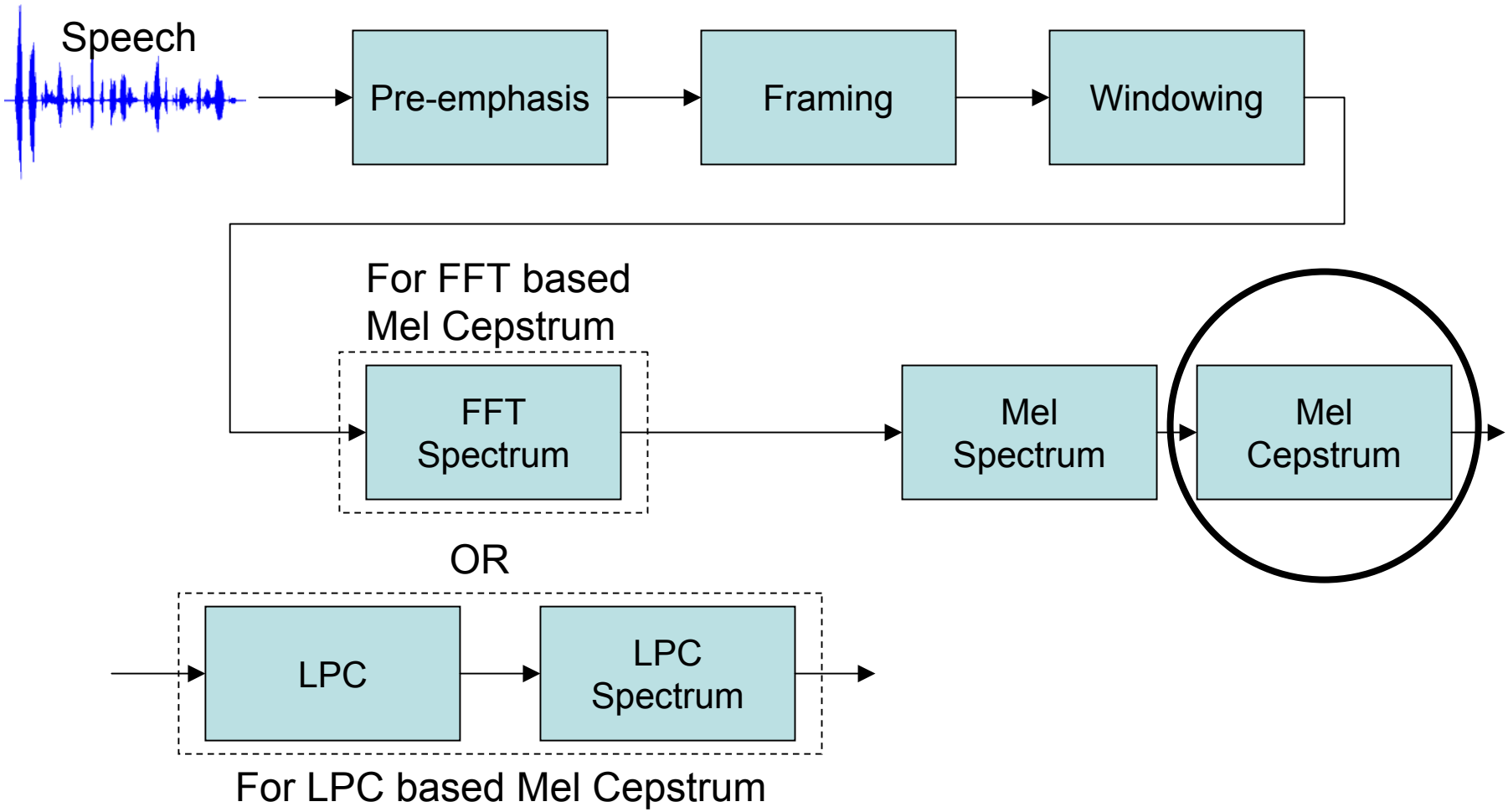




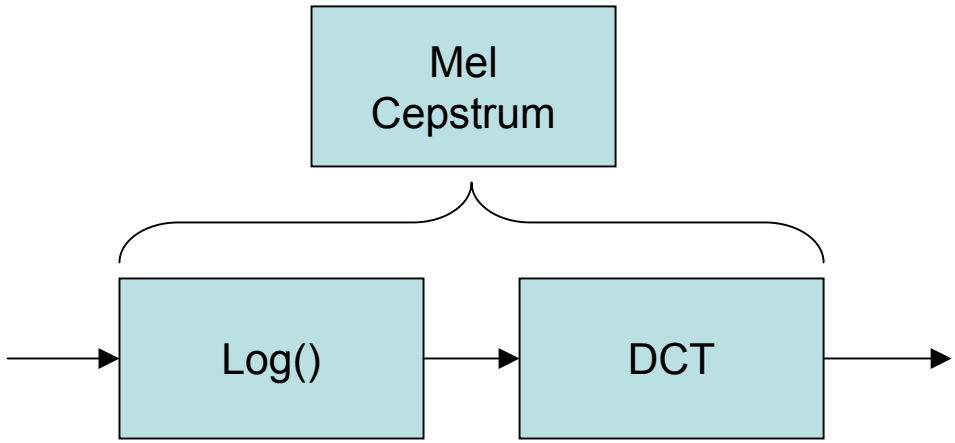


- Small peaks close to large ones are lumped into one main peak

Speech Cepstrum procedure



Speech Cepstrum parameterization



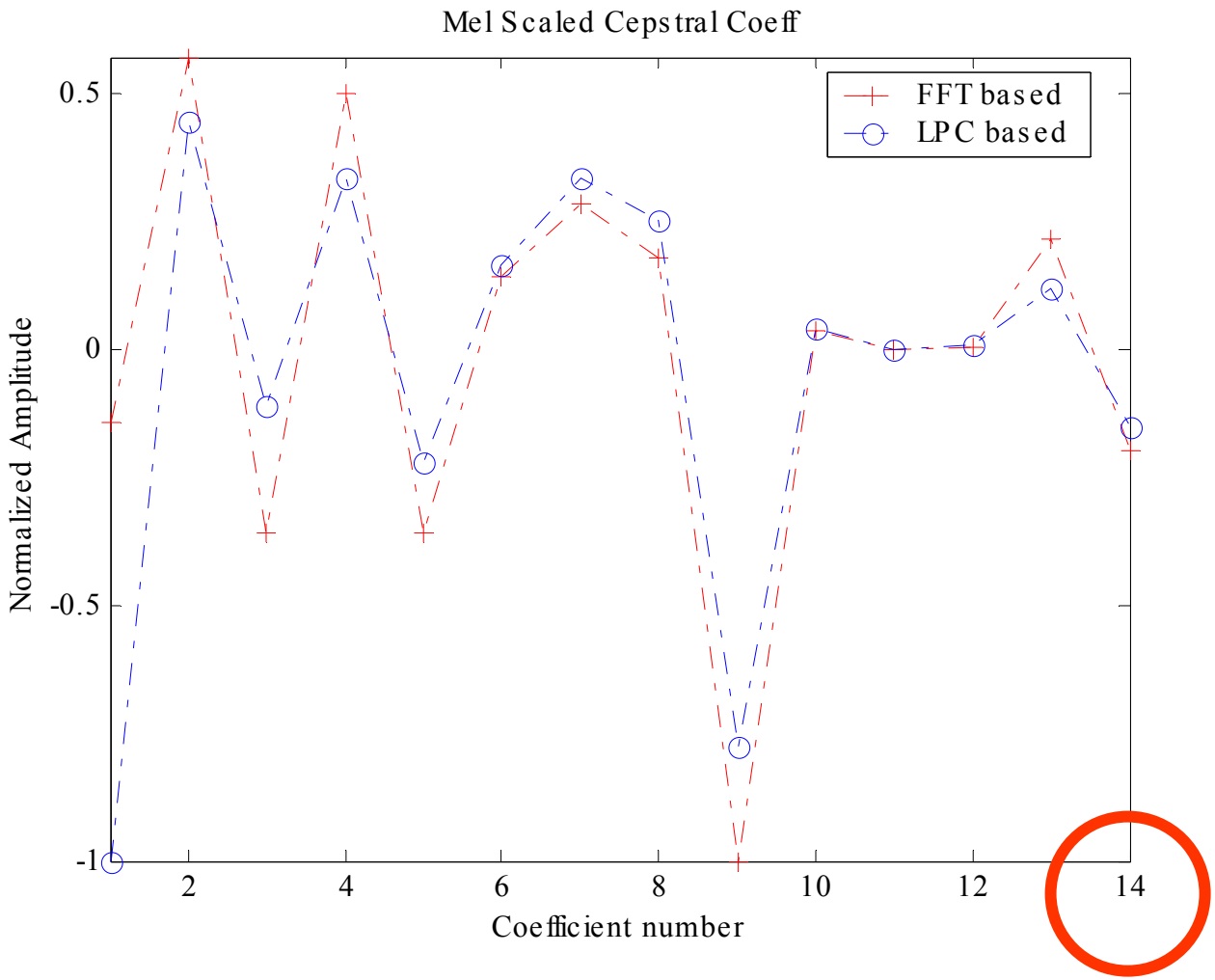
Total number of triangular Mel weighing filters (20)

DCT Equation: $$c(i) = \sqrt{\frac{2}{L}} \sum_{m=1}^{L} \log(\tilde{S}(m)) \cos\left(\frac{\pi i}{L}(m - 0.5)\right) \qquad i = 0,1,\ldots C-1$$

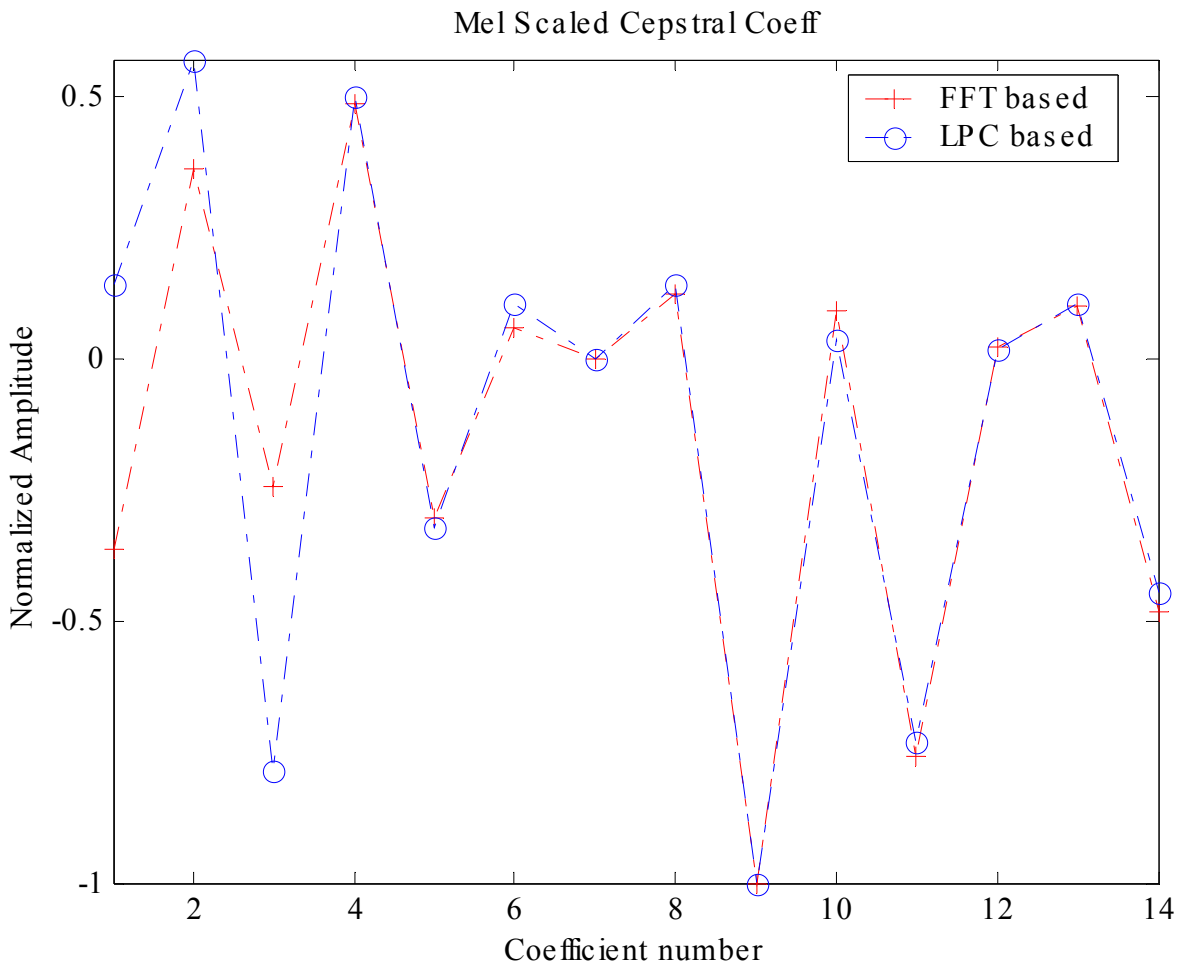
Mel Spectrum

cepstral coefficients desired

Mel Scaled Cepstral Coefficients



Mel Scaled Cepstral Coefficients



Why the DCT?

- The signal is real with mirror symmetry
- The IFFT requires complex arithmetic
- The DCT does NOT
- The DCT implements the same function as the FFT more efficiently by taking advantage of the redundancy in a real signal.
- The DCT is more efficient computationally

Conclusions

- LPC approximates speech linearly at all frequencies
- LPCC are more robust and reliable than LPC alone
- Mel-scaled LPCC and Mel-scaled FFT-CC are robust and also take into account the psychoacoustic properties of the human auditory system

References

[1] Wong, E. and Sridharan, S. “Comparison of linear prediction cepstrum coefficients and Mel-frequency cepstrum coefficients for language identification,” Intelligent Multimedia, Video and Speech Processing Proceedings, pp. 95 -98, 2001.

[2] Molau, S., Pitz, M., Schluter, R. and Ney, H., “Computing Mel-frequency cepstral coefficients on the power spectrum,” Acoustics, Speech, and Signal Processing Proceedings, Volume: 1, pp. 73 -76, 2001.

[3] De Lima Araujo, A.M. and Violaro, F., “Formant frequency estimation using a Mel-scale LPC algorithm,” ITS '98 Proceedings, Volume: 1, pp. 207 -212, 1998.

[4] Picone, J.W., “Signal modeling techniques in speech recognition,” Proceedings of the IEEE , Volume: 81, Issue: 9, pp. 1215 -1247 Sep 1993.

[5] Umesh, S., Cohen, L. and Nelson, D., “Frequency warping and the Mel scale,” IEEE Signal Processing Letters , Volume: 9, Issue: 3, pp. 104 -107, 2002.

[6] http://www-2.cs.cmu.edu/~mseltzer/sphinxman/