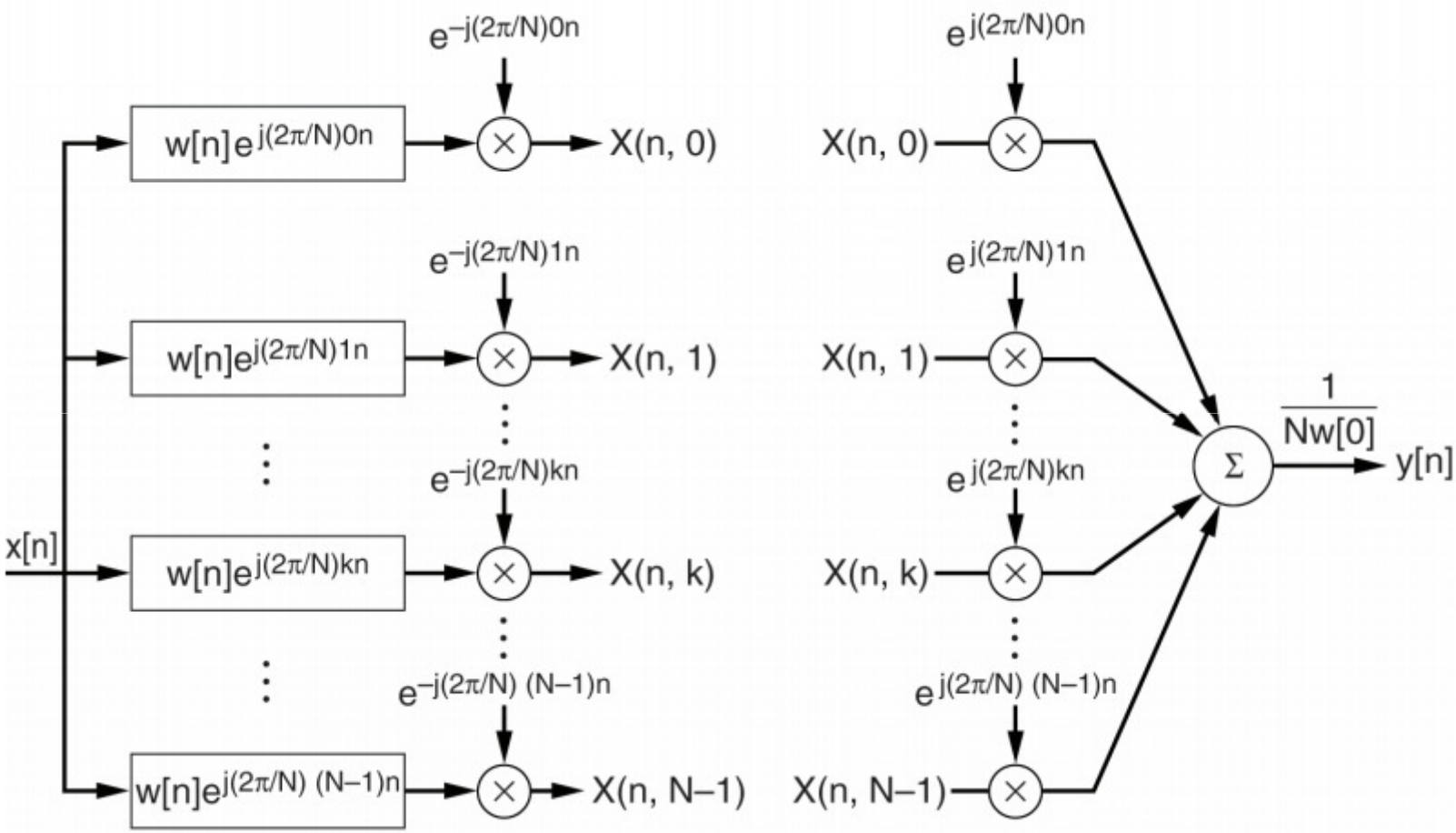

x[n]
w[n]e^j(2π/N)0n
w[n]e^j(2π/N)1n
w[n]e^j(2π/N)kn
w[n]e^j(2π/N) (N–1)n
e^–j(2π/N)0n
e^–j(2π/N)1n
e^–j(2π/N)kn
e^–j(2π/N) (N–1)n
X(n, 0)
X(n, 1)
X(n, k)
X(n, N–1)
e^j(2π/N)0n
e^j(2π/N)1n
e^j(2π/N)kn
e^j(2π/N) (N–1)n
X(n, 0)
X(n, 1)
X(n, k)
X(n, N–1)
Σ
1/Nw[0]
y[n]

Time varying Signal $g(t)$

$\Downarrow$

Joint Time-Freq (TF) representation

$\equiv$ "information diagram"

$\equiv \underbrace{G(t,f)}$

will have spread of time-events as well as frequency-events.

$$\int_{-\infty}^{\infty} G(t,f)\,df \longrightarrow \frac{|\Psi(t)|^2}{\int |\Psi(t)|^2 dt}$$

$|\Psi(t)|^2$ ← energy distribution along time

$\int |\Psi(t)|^2 dt$ ← time marginal.

$$\int_{-\infty}^{\infty} G(t,f)\,df \longrightarrow \frac{|\Phi(f)|^2}{\int |\Phi(f)|^2 df}$$

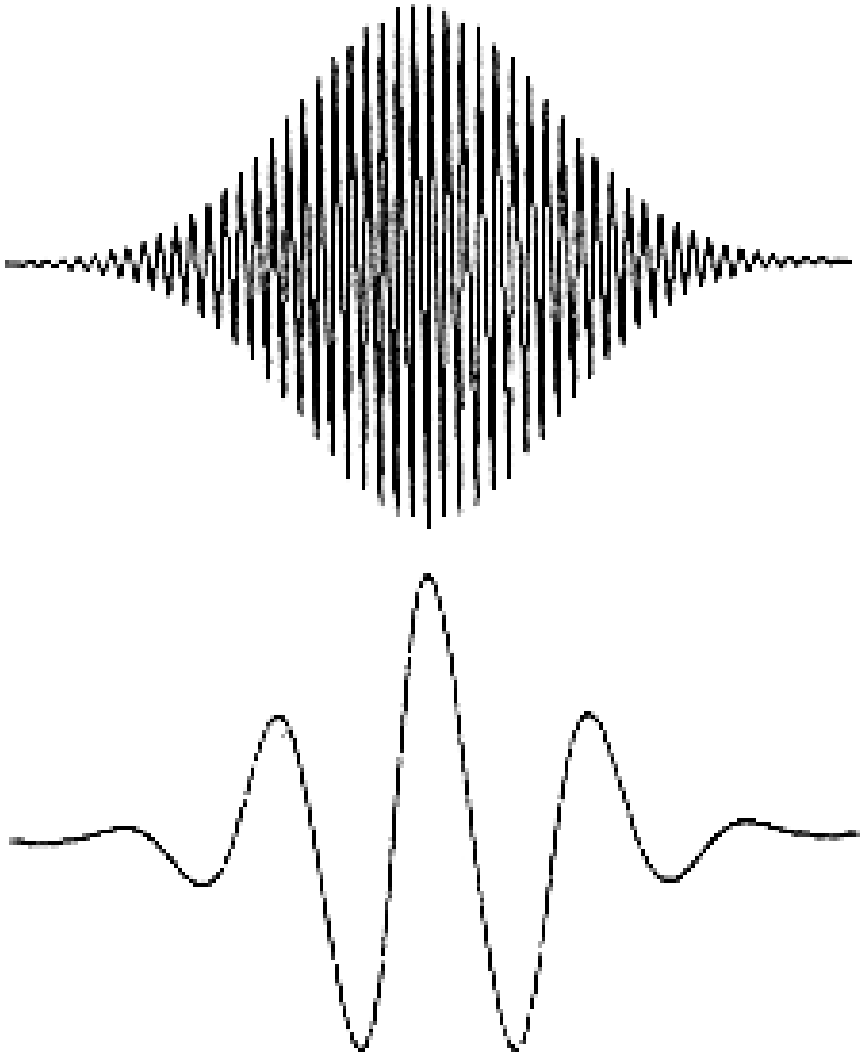
$|\Phi(f)|^2$ ← energy distribution along freq.

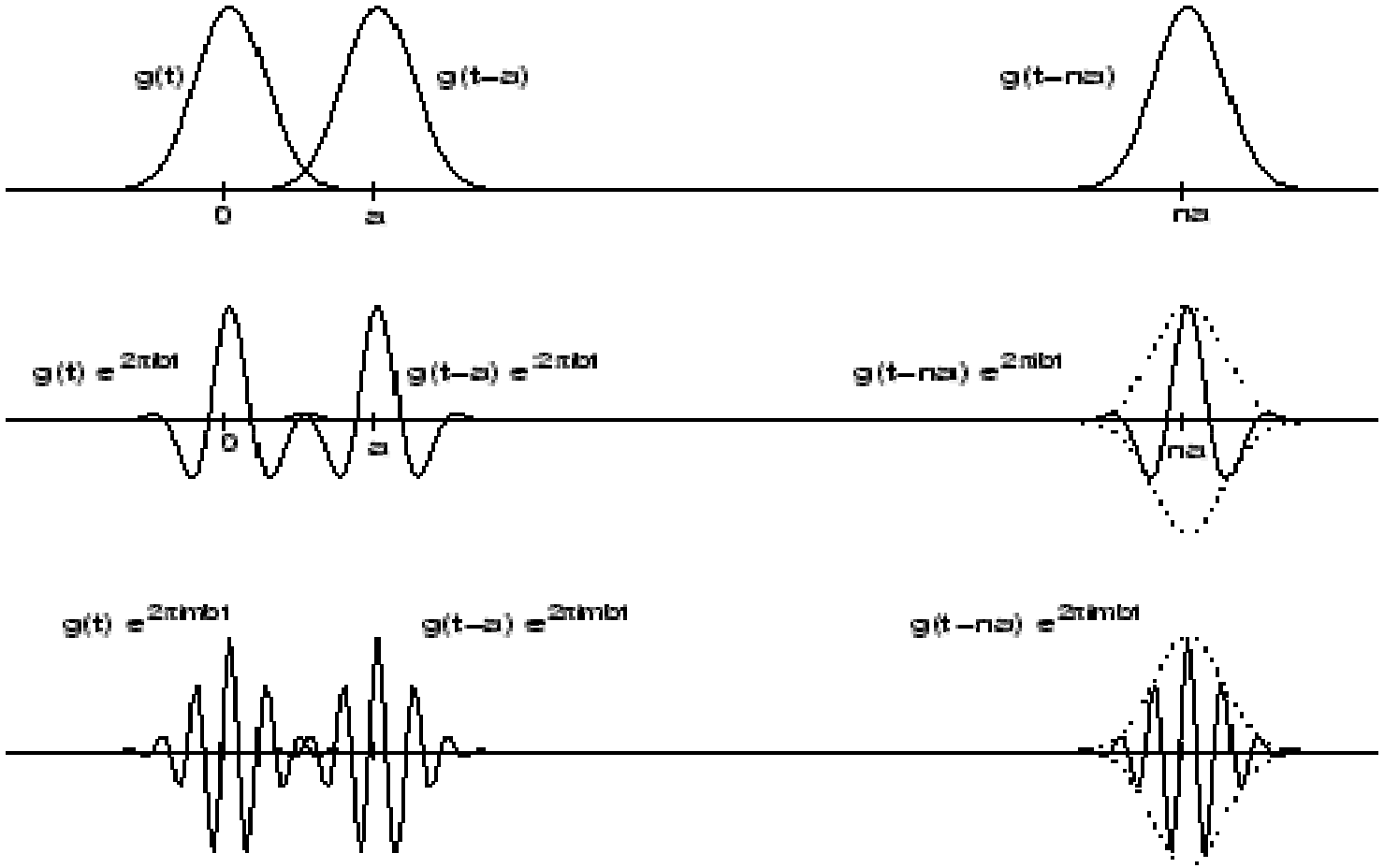
$\int |\Phi(f)|^2 df$ ← freq marginal.

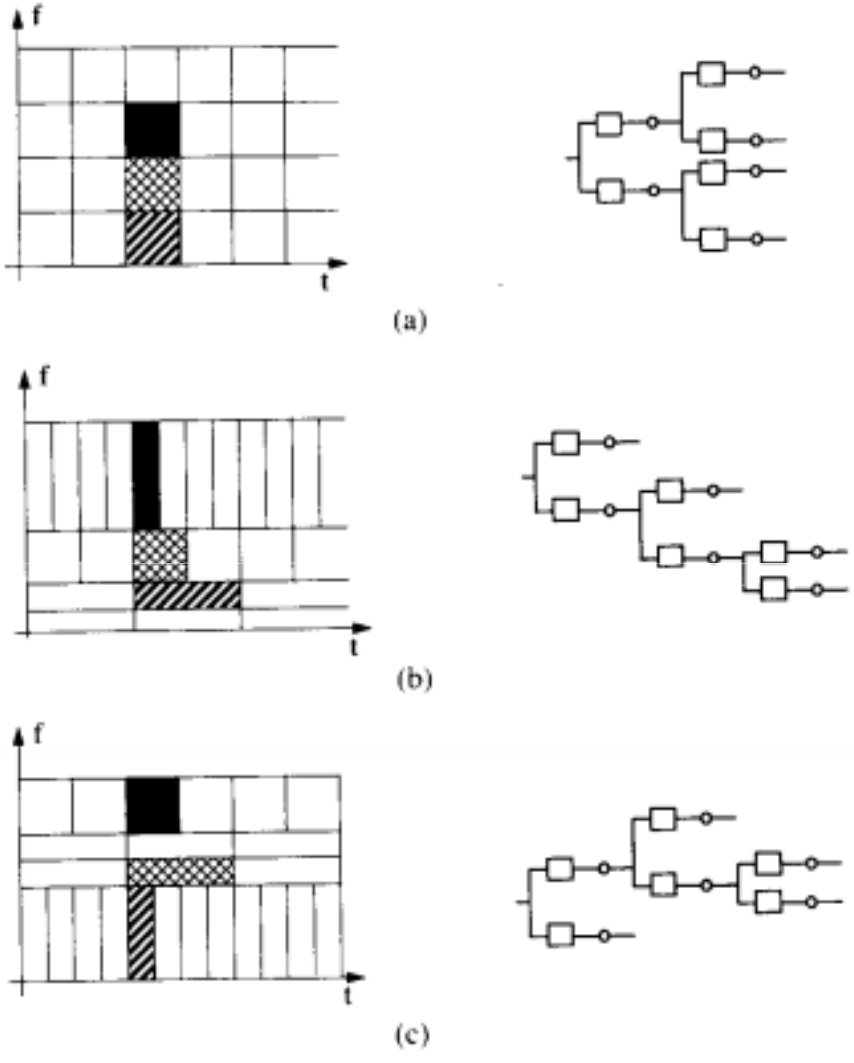
Time spread (T), Freq spread (B)

$$T^2 = \int\int (t-\bar{t})^2 G(t,f)\,dt\,df, \quad \bar{t} = \int\int t\, G(t,f)\,dt\,df$$

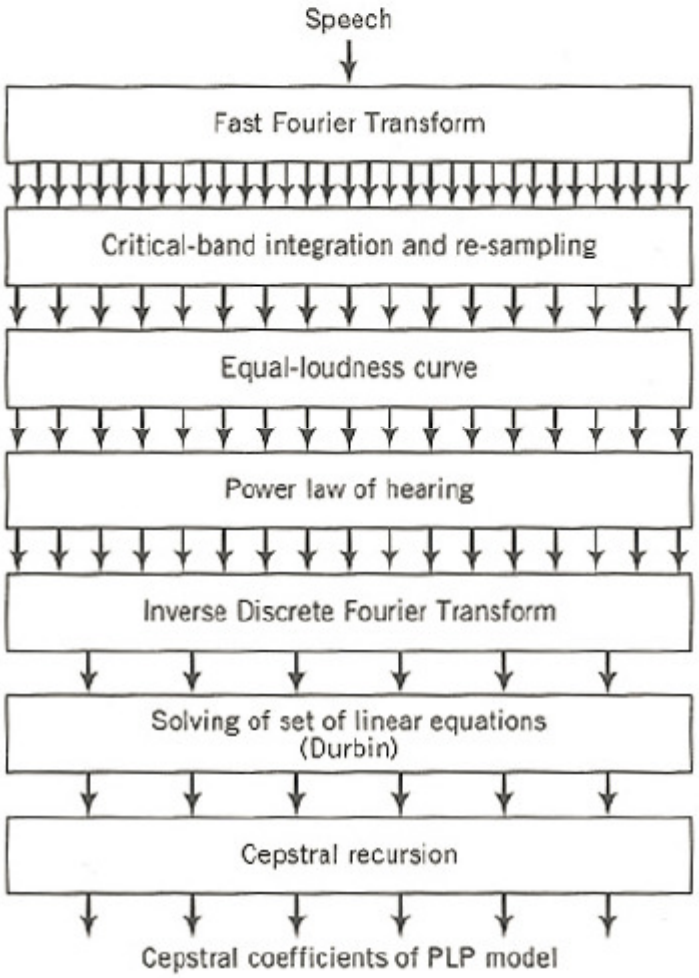
$$B^2 = \int\int (f-\bar{f})^2 G(t,f)\,dt\,df, \quad \bar{f} = \int\int f\, G(t,f)\,dt\,df.$$




g(t)
g(t−a)
g(t−na)
0
a
na
g(t) e^2πibt
g(t−a) e^2πibt
g(t−na) e^2πibt
0
a
na
g(t) e^2πimbt
g(t−a) e^2πimbt
g(t−na) e^2πimbt

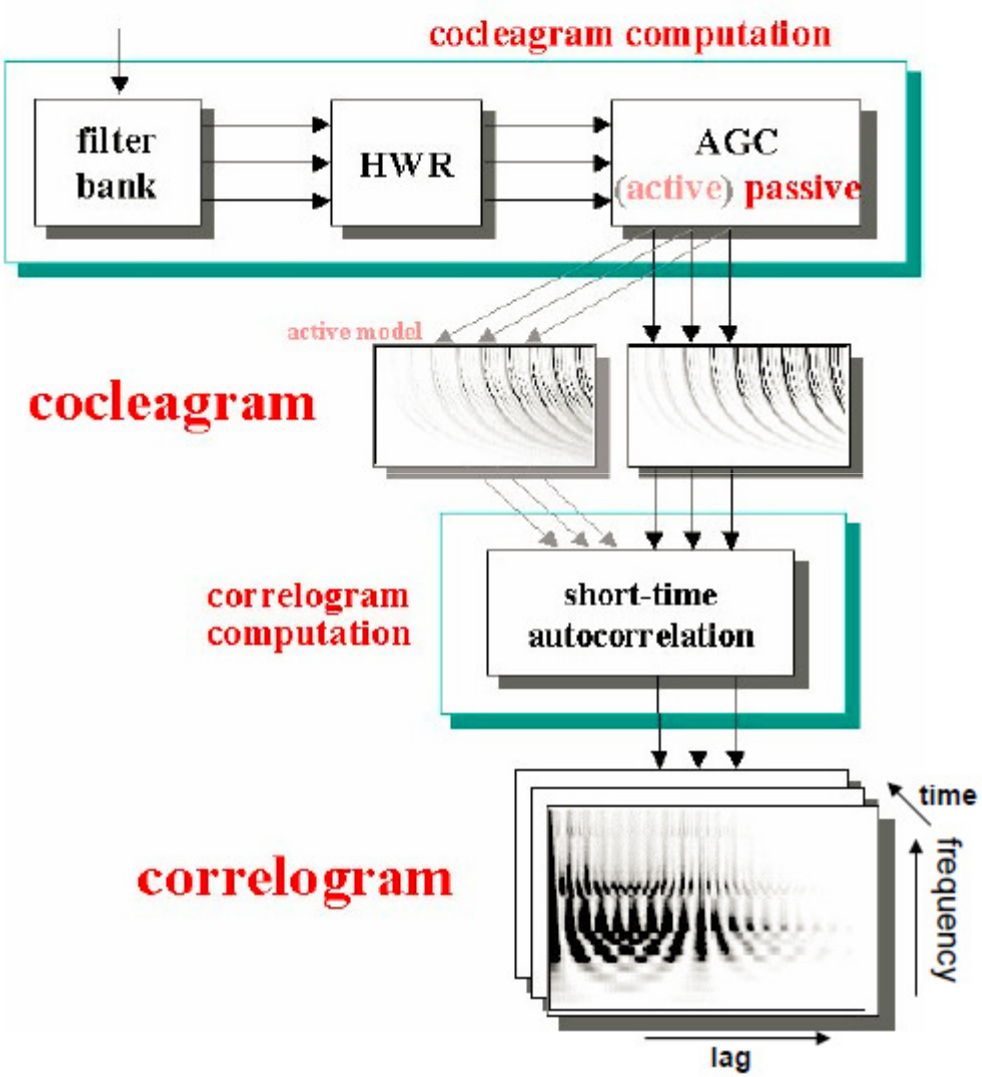

f
t
(a)
f
t
(b)
f
t
(c)

Perceptual Linear Prediction



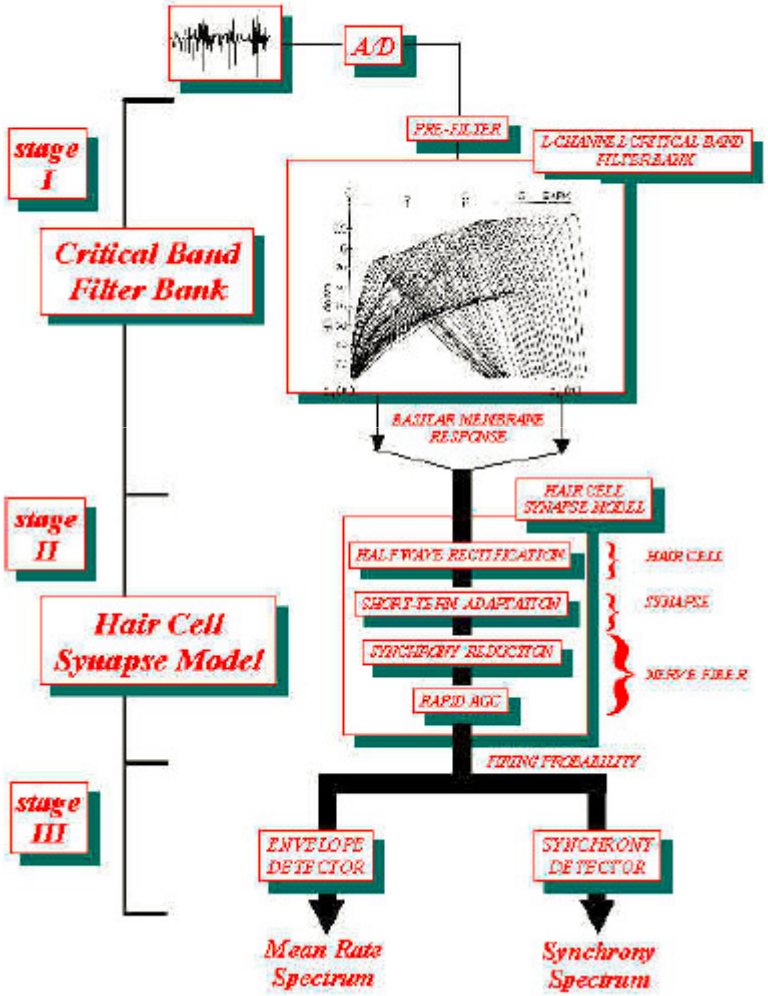
1. Compute power spectrum estimate using FFT
2. Use either a triangular window (mel-based) or a trapezoidal window (Bark-based) to integrate the power spectrum within overlapping critical bands
3. Pre-emphasize the spectrum to approximate the unequal sensitivity of human hearing versus frequency (approximates equal loudness curves)
4. Compress the spectral amplitudes with a logarithmic compressor (approximates power-law relation between intensity and loudness)
5. Perform an inverse DFT to give cepstral coefficients
6. Perform spectral smoothing (e.g., using cepstral lifter for mel-cepstral processing, or using an autoregressive model for the compressed critical band spectrum)
7. Use an orthogonal representation (e.g., convert autoregressive components back to cepstral components
8. Perform liftering using an n^{α} lifter

Lyon's Auditory Model



Models behavior of the cochlea as a non-linear, compressive, filter bank

Seneff Auditory Model



Mean Rate Spectrum: firing rates of auditory nerve fibers; measure of spectral energy

Synchrony Spectrum: synchrony of fine temporal structure in each channel

Lateral Synchrony (Ali): repetition cues from adjacent channels

Stages I and II: peripheral transformations occurring in the early stages of the hearing process

Stage III: extract information relevant to perception, such as formants and enhanced sharpness of onset and offset of speech segments