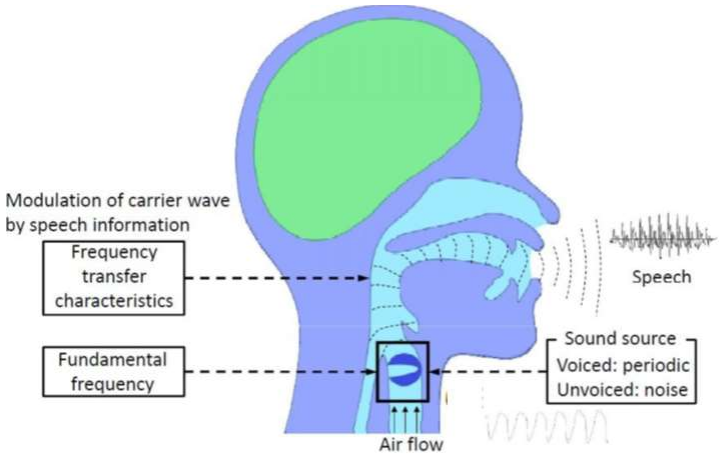
# Introduction To Mel Frequency Cepstral coefficients (MFCC)

Dr. Aparna P.
Dept of EC Engineering
NITK, Surathkal

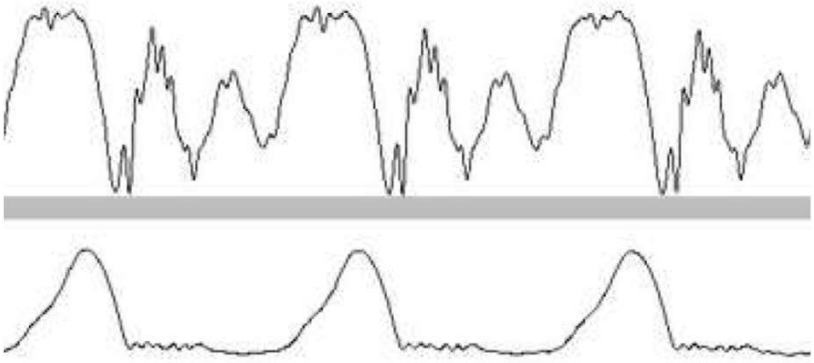


## Speech Production Model




Source: Tomoki Toda; WiSSAP 2013

# Voiced Sound Production



vowel अ



# Glottal Impulses and Resonances of VT

Uniform tube model


glottal vibration vocal tract speech wave

# Source-Filter model of speech production


glottal vibration vocal tract speech wave

$$s(n) = e(n) * h(n)$$

$$S(k) = E(k)H(k)$$

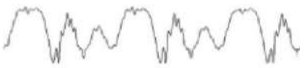
$$log(|S(k)| **2) = log(|E(k)| **2) + log(|H(k)| **2)$$



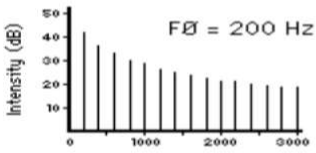
# Source-Filter model of speech production

Glottal air flow

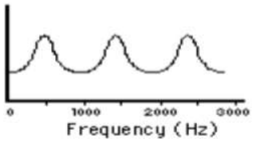
vocal tract: vowel /a/



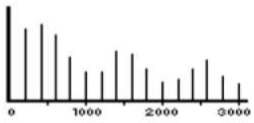
time waveform




SOURCE SPECTRUM



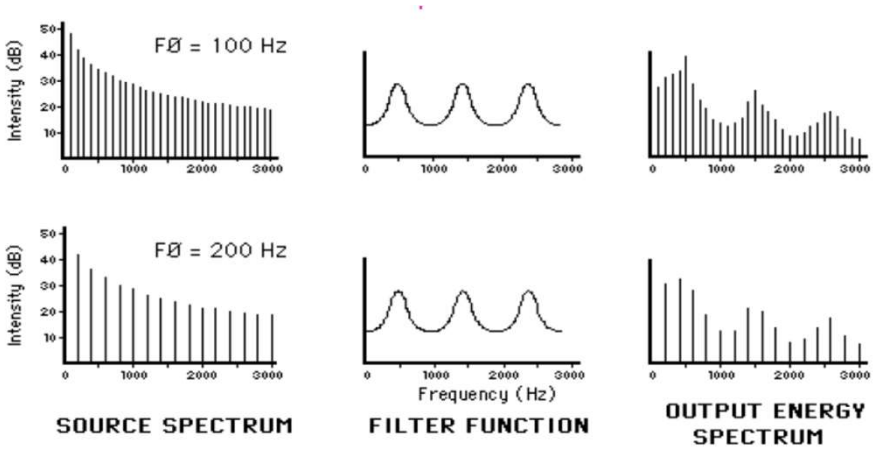

FILTER FUNCTION




OUTPUT ENERGY SPECTRUM

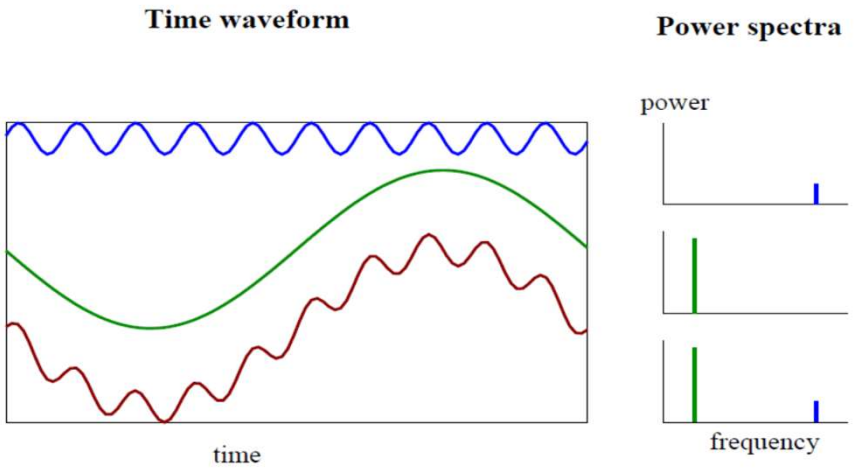
$$log(|S(k)| **2) = log(|E(k)| **2) + log(|H(k)| **2)$$

# Illustration in Spectral Domain





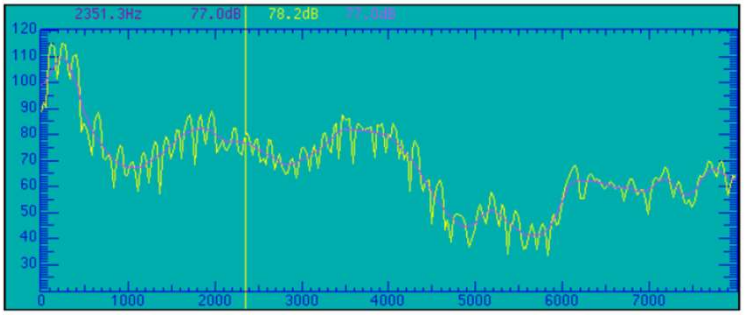

# Cepstral Analysis



# Cepstral Analysis


Waveform  Power spectrum  Log spectrum  Cepstrum

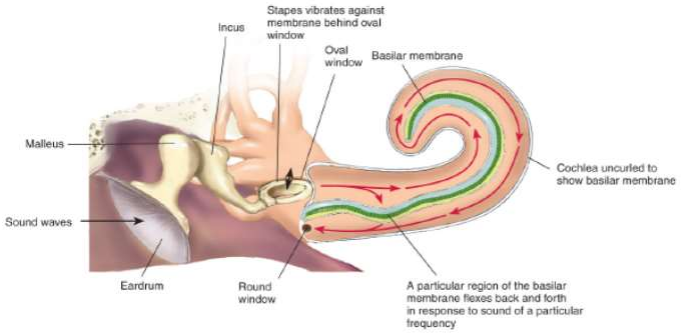
$$cep(q) = IFFT\{log(|S(k)| * *2)\} \qquad q = 0, 1, \ldots N-1$$




Captures not only resonances but also anti-resonances.



# Cochlea & Hearing



# Basilar membrane: Bark/mel scale

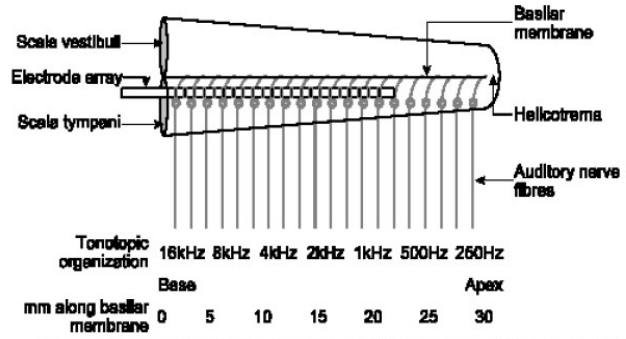
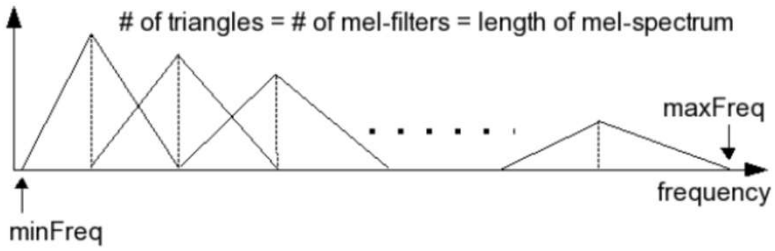



Figure A simplified unrolled representation of the cochlea showing the auditory nerve fibres, the tonotopic organization of these nerve fibres and an intracochlear electrode array in the scala tympani.



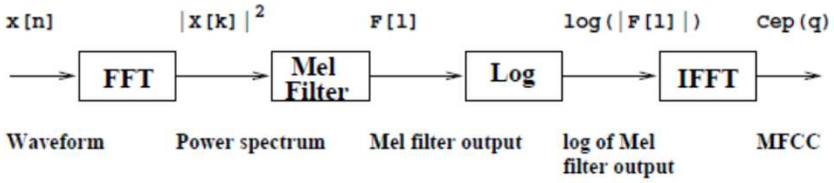
# MFCC




$$B(m) = \sum_{k=lo(m)}^{hi(m)} |X(k)|^2$$

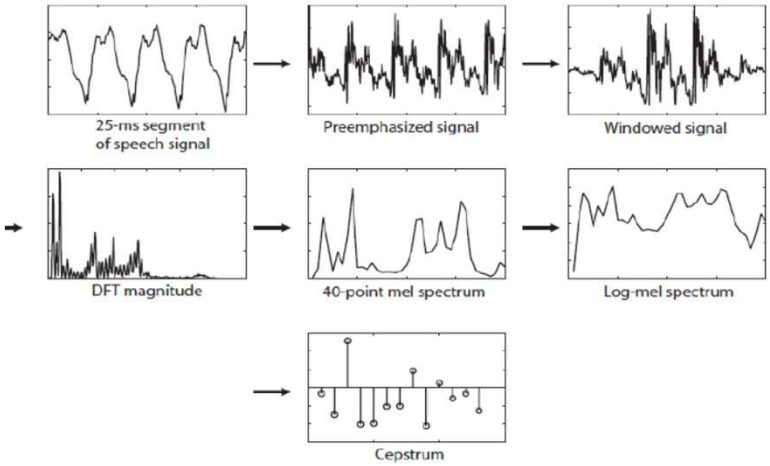
$$cep(q) = IFFT\{log(|B(m)|^2)\} \qquad q = 0, 1, ...N$$

# MFCC-Contd






# Wave to MFCC (Illustration)



• Acknowledgement: Prof. Samudra Vijaya K., IIT Guwahati.