

Data Mining and Machine Learning

Assignment Project Exam Help

Application of HMMs for ASR: Feature representation of speech

<https://powcoder.com>

Add WeChat powcoder

Peter Jančovič



Objectives

- Front-end analysis for ASR – feature representation of speech
 - To understand motivation and stages for ‘typical’ parameterisation of speech signals used for ASR
 - Mel Frequency Cepstral Coefficients (MFCCs)

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



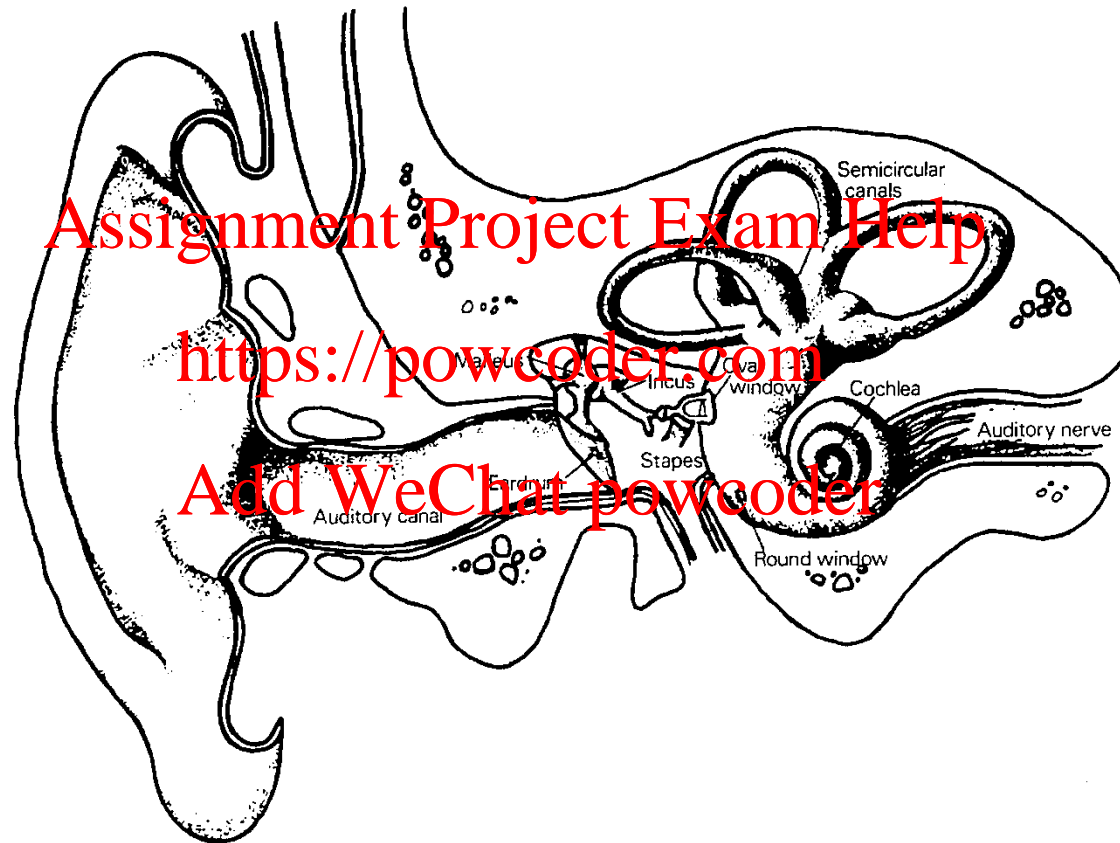
What is “Front-End Analysis”

- First stage in any speech recognition system
- Goal is to convert the raw acoustic speech waveform into a form which is suitable (or even optimal) for automatic speech recognition
- In general pattern recognition terms, front-end analysis is feature extraction
- Where do we start?

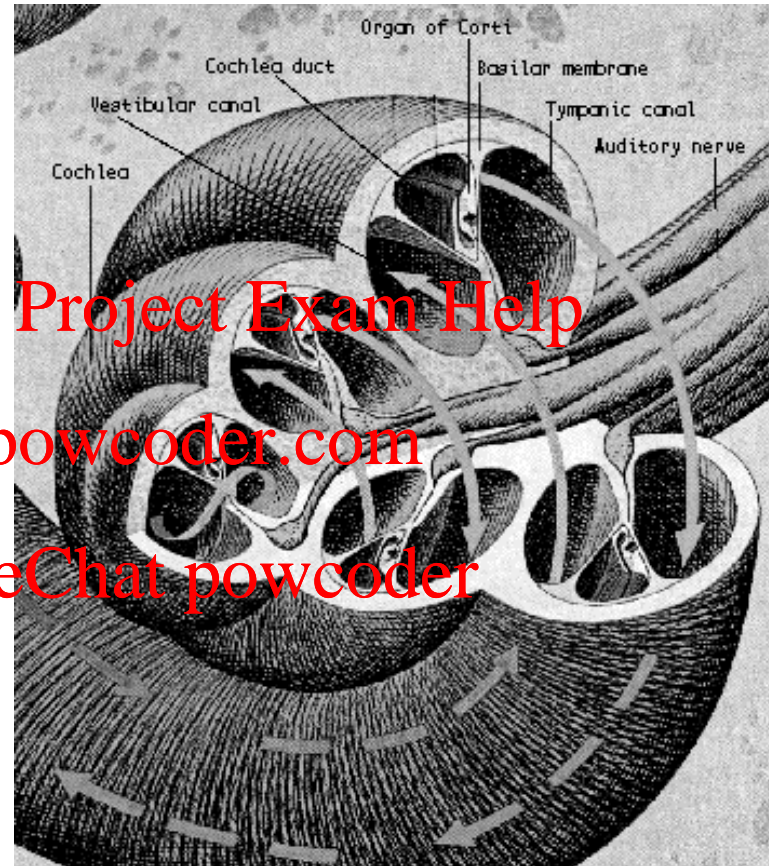


The Human Auditory System

taken from J N Holmes, "Speech Synthesis and Recognition", Van Nostrand Reinhold (1988)



The Basilar Membrane



Assignment Project Exam Help

<https://powcoder.com>

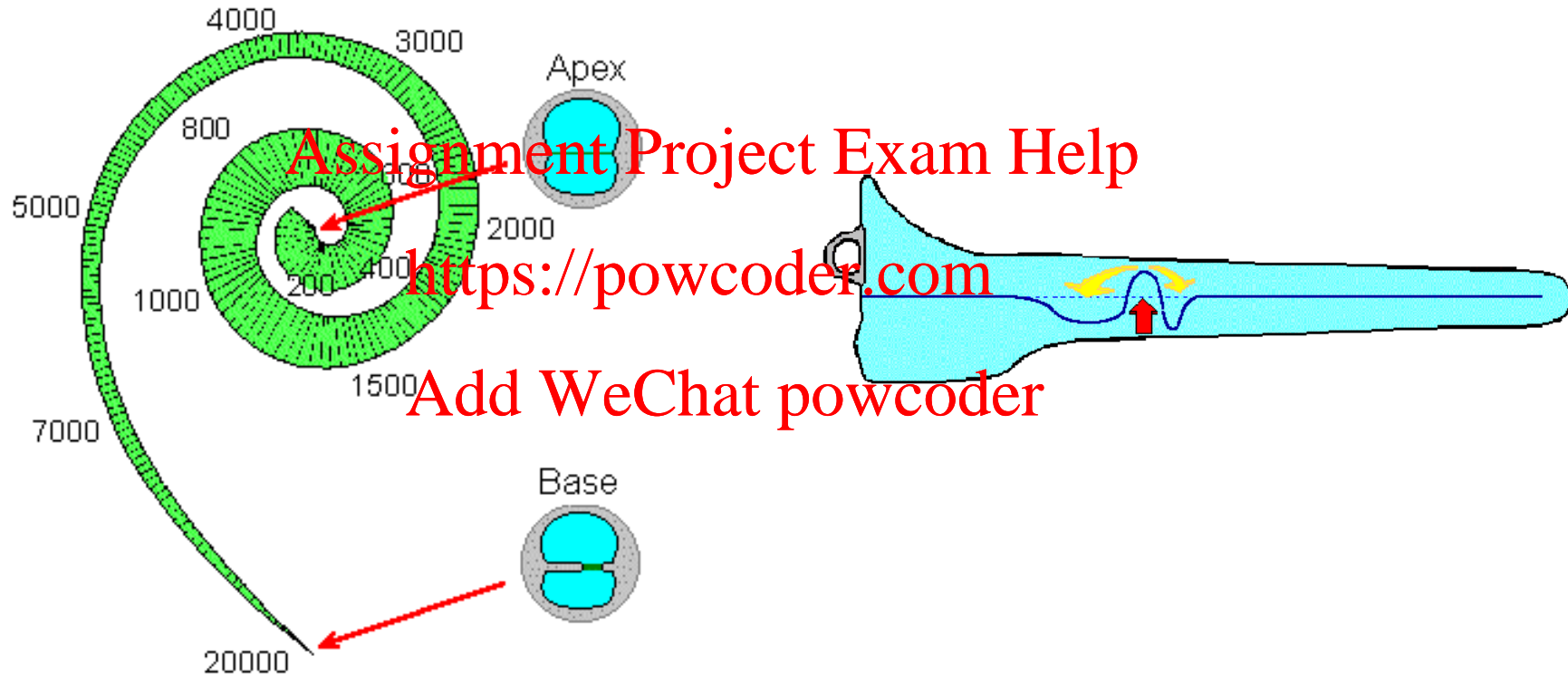
Add WeChat powcoder

Australian National
University -

<http://online.anu.edu.au/IT/A/ACAT/drw/PPofM/hearing/hearing3.html>



Frequency response of the basilar membrane



School for advanced studies, Trieste, Italy -

<http://poirot.sissa.it/multidisc/cochlea/utils/basilar.htm>

UNIVERSITY OF
BIRMINGHAM

Lessons from Psycho-Acoustics

- Human speech perception begins with frequency analysis on the basilar membrane
- Frequency is not perceived on a linear scale – hence use of non-linear perceptual frequency scales: **mel** scale, **bark** scale,...
<https://powcoder.com>
- Individual point on the basilar membrane can be modelled as band-pass filter – a **critical band** is the implicit bandwidth of such an ‘auditory filter’
Add WeChat powcoder
- Loudness perceived on logarithmic scale
- Phase of limited significance for speech recognition

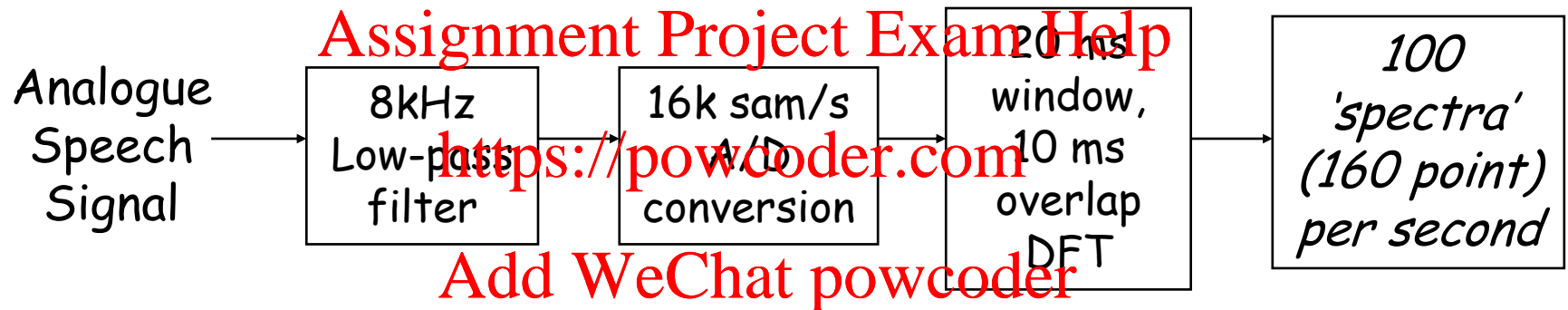


Front-end analysis for ASR

- Speech waveform typically low-pass filtered at 4kHz to 8kHz
- Sampled 8,000 to 16,000 samples per second
- Frequency analysis:
 - 20 ms analysis window
 - 10 ms overlap between windows
 - Hamming window
 - Discrete Fourier Transform



Frequency analysis for ASR

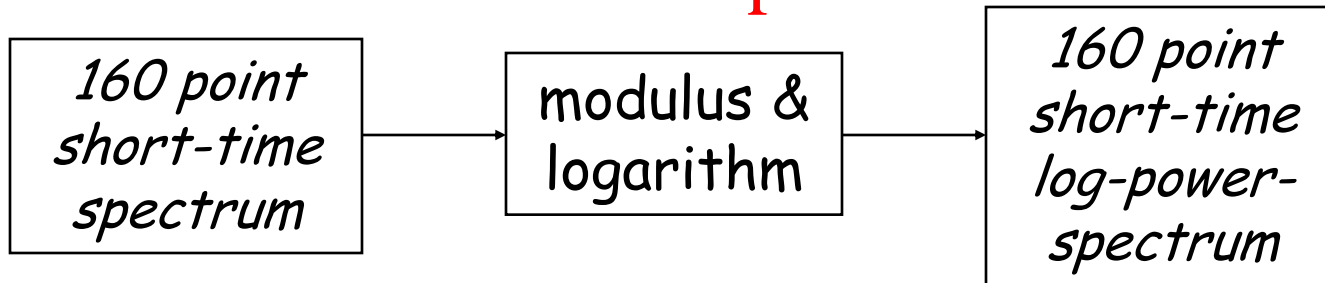


Example: 8kHz bandwidth system



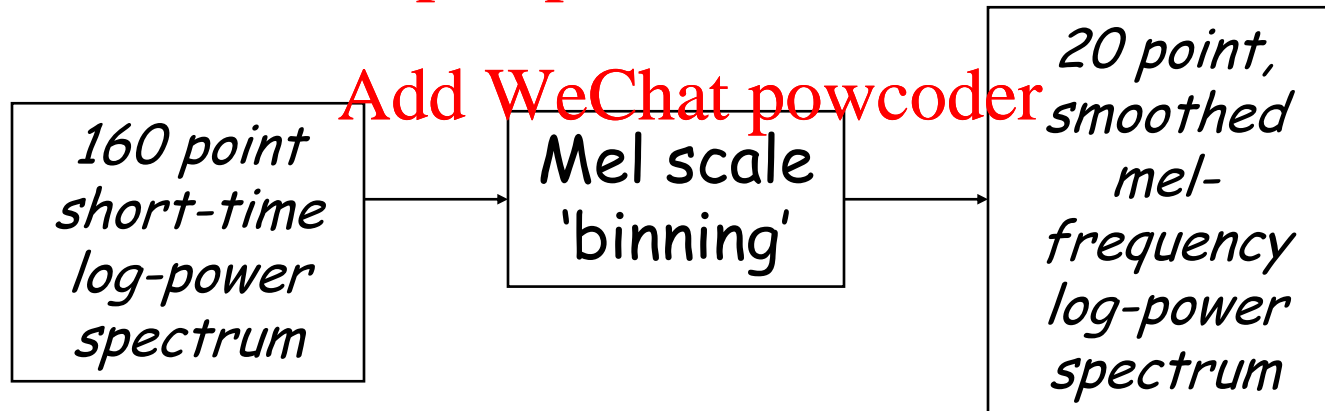
Log Power Spectrum

- Phase ignored by taking the **modulus** of the complex spectrum
- Logarithm applied
 - For consistency with psycho-acoustic results
 - To compress dynamic range



Mel-scale & smoothing

- The **mel spectrum** can be computed by **averaging** the short-time Fourier spectrum over 'bins' whose width depends on frequency...
- ...or by using band-pass filters with appropriate, frequency-dependent, band-widths



Mel Scale Filterbank

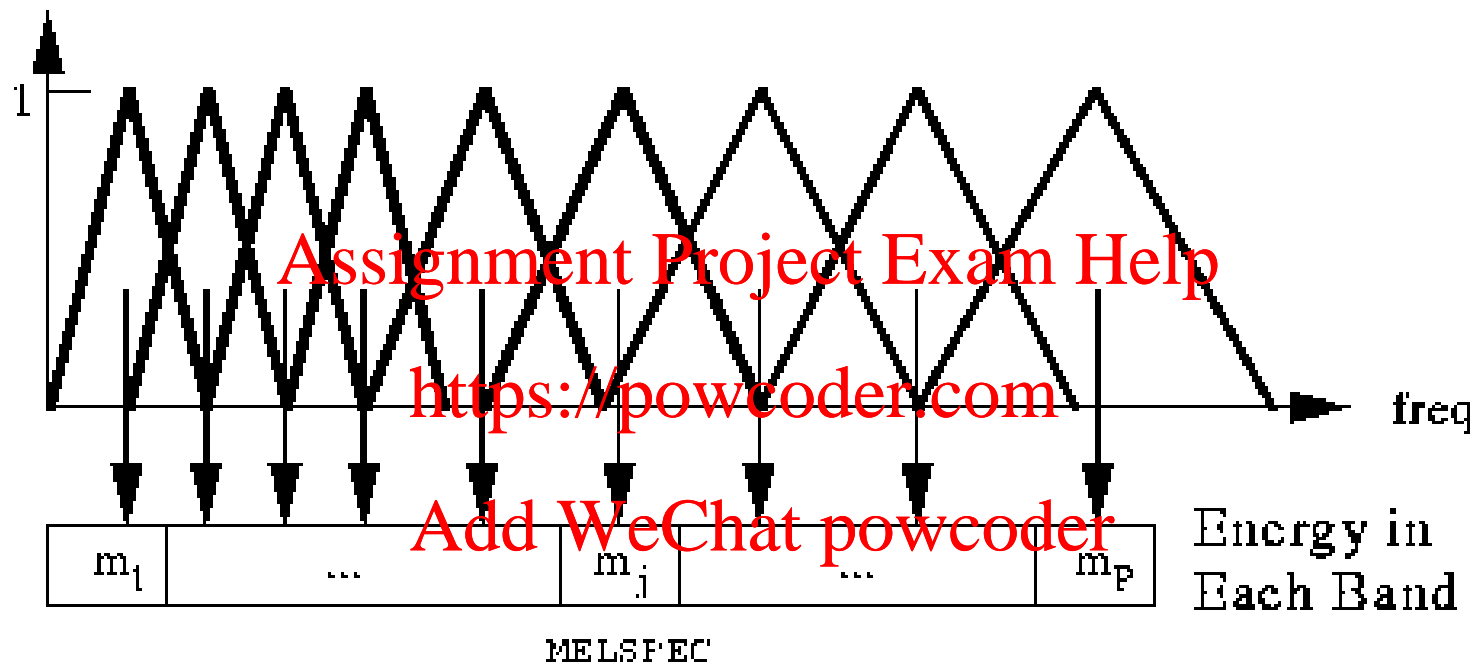


Fig. 5.3 Mel-Scale Filter Bank

From Steve Young, "The HTK Book", Cambridge University Engineering Department



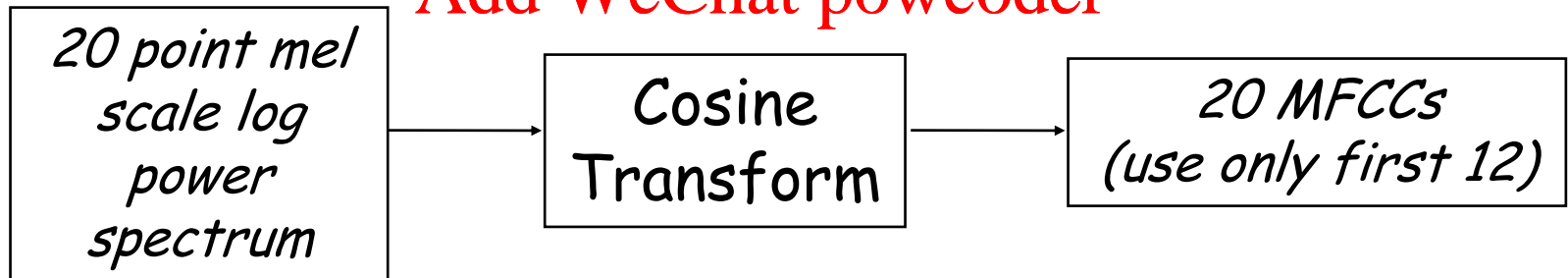
Cepstrum

- Cosine transform applied to remove correlation between components of mel-scale log power spectrum
 - Mel Cepstrum: MFCC = Mel Frequency Cepstral Coefficients
 - Mathematical expediency

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

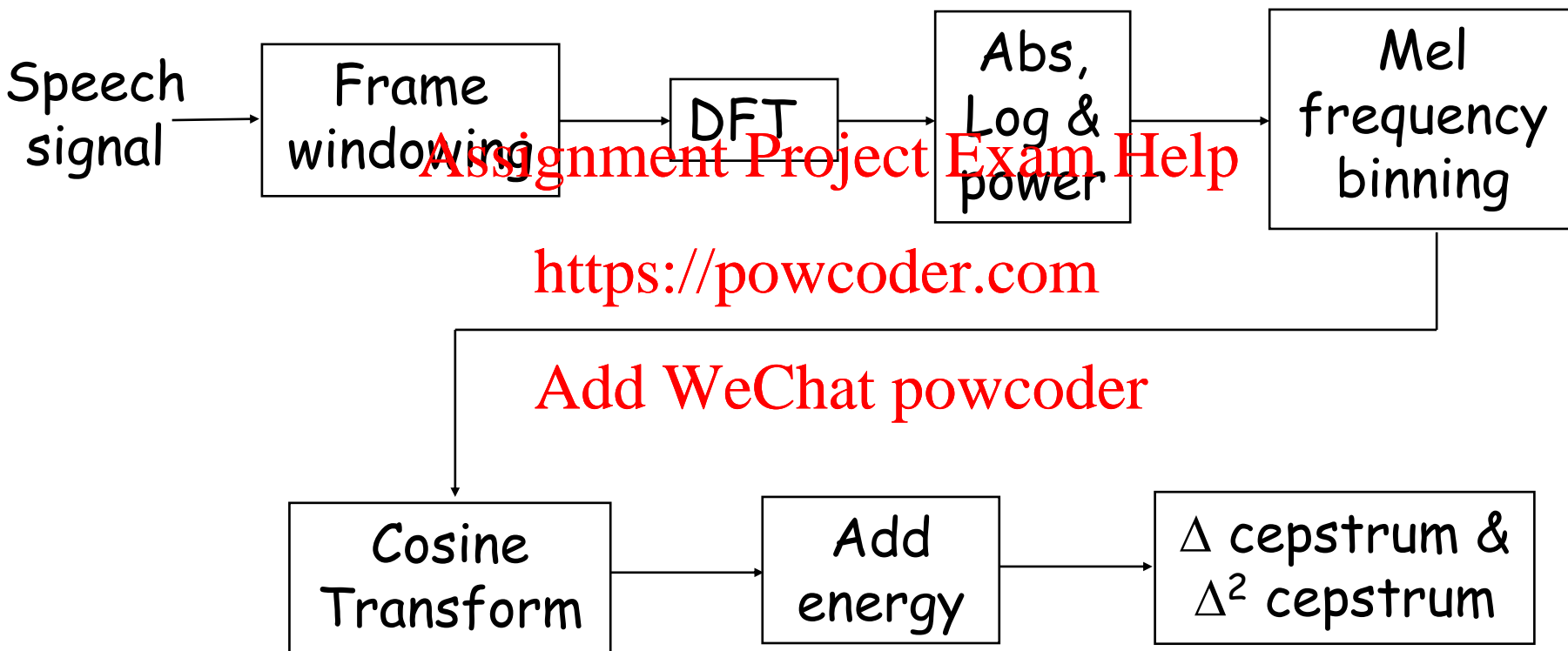


Energy & Delta Coefficients

- Add energy as 13th parameter
- Compute estimate of time-derivative of each parameter – delta cepstrum (or Δ cepstrum)
<https://powcoder.com>
- Compute estimate of time-acceleration of each parameter – delta² cepstrum (or Δ^2 cepstrum)
Add WeChat powcoder
- Cepstrum + Δ Cepstrum + Δ^2 Cepstrum = ‘standard’ 39 dimensional representation (e.g. in HTK)



Front-end analysis – summary



Summary

- Introduction to front-end speech processing for ASR
 - Motivations from human hearing
 - Description of typical front-end representation
 - Short-time log power spectrum
 - Mel scale filtering
 - Cosine transform
 - Δ and Δ^2 parameters

