Speech Quality in Noise

"Method for the evaluation of a speech enhancement system in terms of the perceived improvement in the quality of a noisy speech signal"

Dushyant Sharma
Imperial College

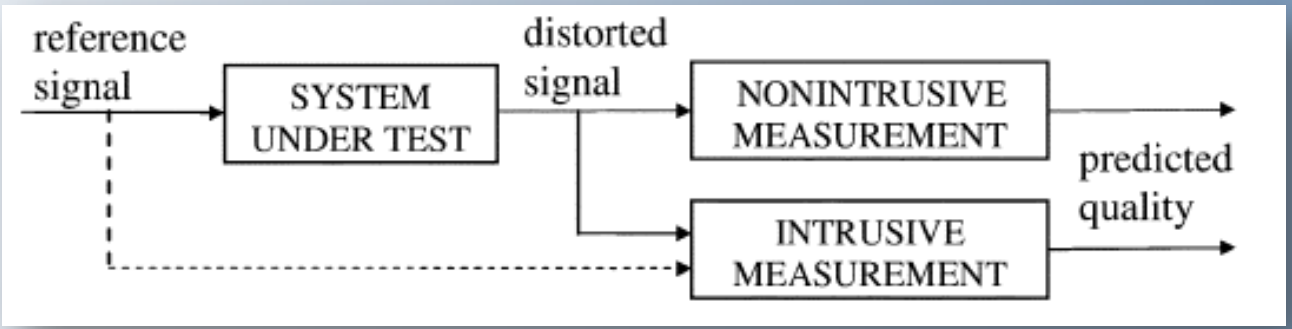
Speech Quality

- **Speech Quality Measurement**
  - **Subjective**
  - **Objective**
    - **Intrusive**
    - **Non- Intrusive**
- **Listening Tests : Baseline Vs Commercial**
  - **Methodology**
  - **Results & Conclusions**
- **Low Complexity Quality Assessment Algorithm**
  - **Overview**
  - **Performance**

Speech Quality :: **Subjective Measurement**

- Assessment of speech quality harder than intelligibility, as quality is a less formal measure.
- Broadly two types of subjective speech quality tests:
  - Quality Rating Tests (ITUT, 1998a)
    - Listeners assign absolute ratings to individual speech stimuli
    - Mean Opinion Score – scale from 1 (bad) to 5 (excellent)
  - Preference Tests
    - Listeners exhibit preference for one speech stimulus over one or more others

Speech Quality :: **Objective Measurement**



- Intrusive :
  - PESQ (ITU-T P.862) - complex sequence of processing steps to generate distortion scores as a function of frequency and time
- Non-Intrusive :
  - ITU-T P.563 – standard for non-intrusive quality assessment
  - LCQA - low complexity algorithm for non-intrusive quality assessment

Pair-wise Listening Tests

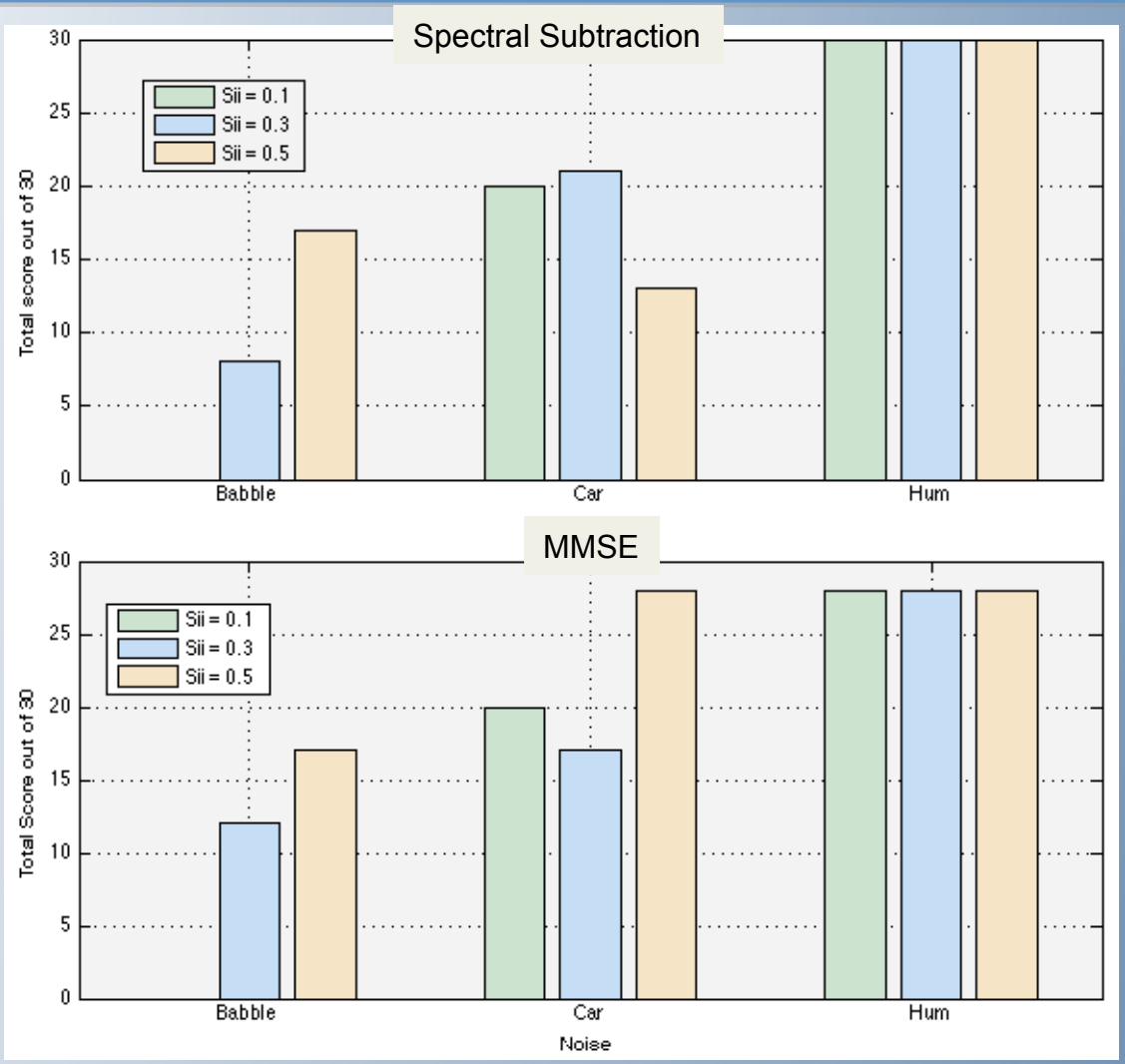
- Pair-wise tests performed to get a subjective view for the effect of different algorithms
- Experiment Design
  - 30 Listeners
  - Two categories:
    - Preference of enhanced over noisy signal
      - Conducted for baseline algorithms : SS, MMSE
      - 3 Noise Types: Car, Hum, Babble
      - 3 SII levels: 0.1, 0.3 and 0.5
    - Preference of enhancement algorithm
      - Conducted for Com1, Com2, SS and MMSE
      - 3 Noise Types : Car, Hum, Babble
      - 1 SII : 0.3

Soundjudge Interface

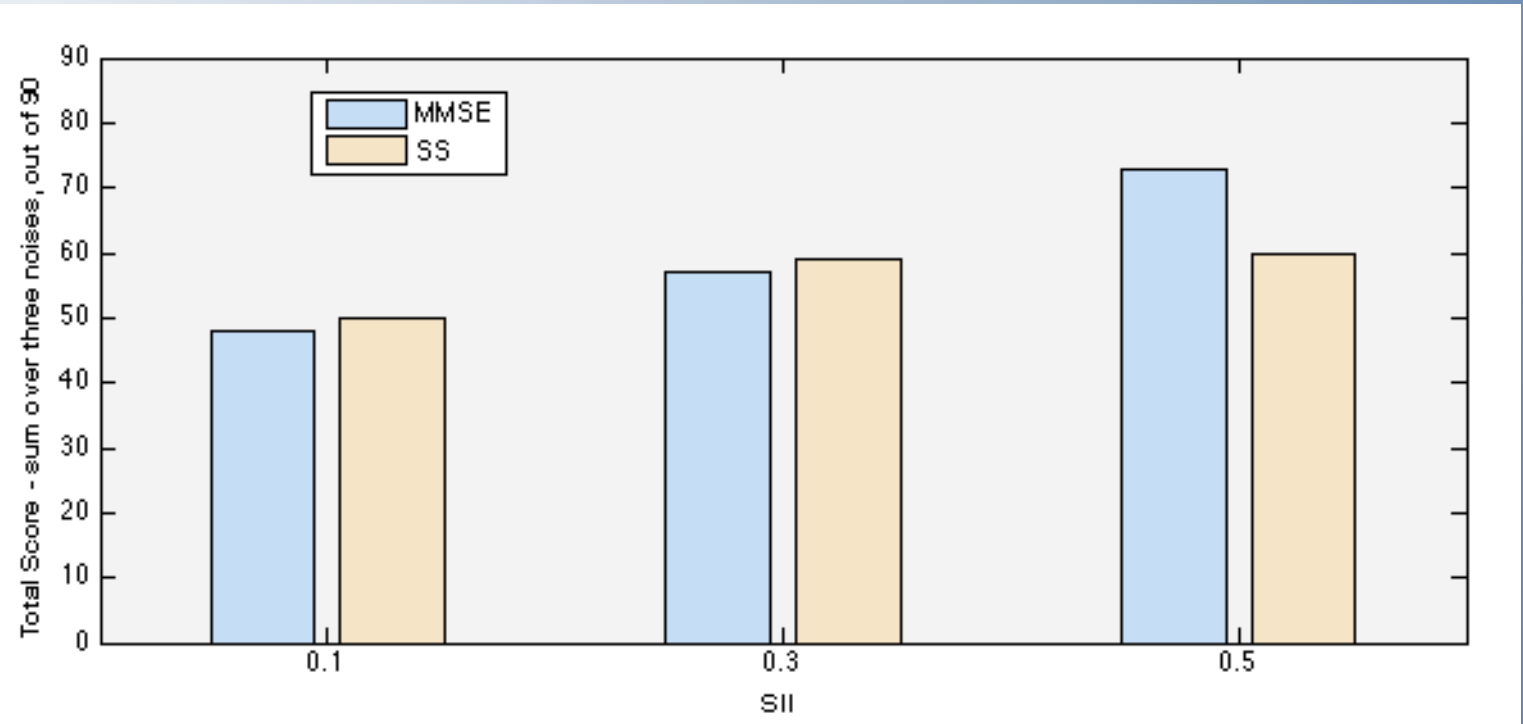
Results

- Comparing Enhancement over Original Noisy:
  - Results yielded a total score, representing the number of times the enhancement was chosen over the noisy original
    - reference metric used for comparison
  - Results for Spectral Subtraction (SS) and Minimum Mean Square Estimator Algorithms (MMSE)

Preference of Enhanced Signal



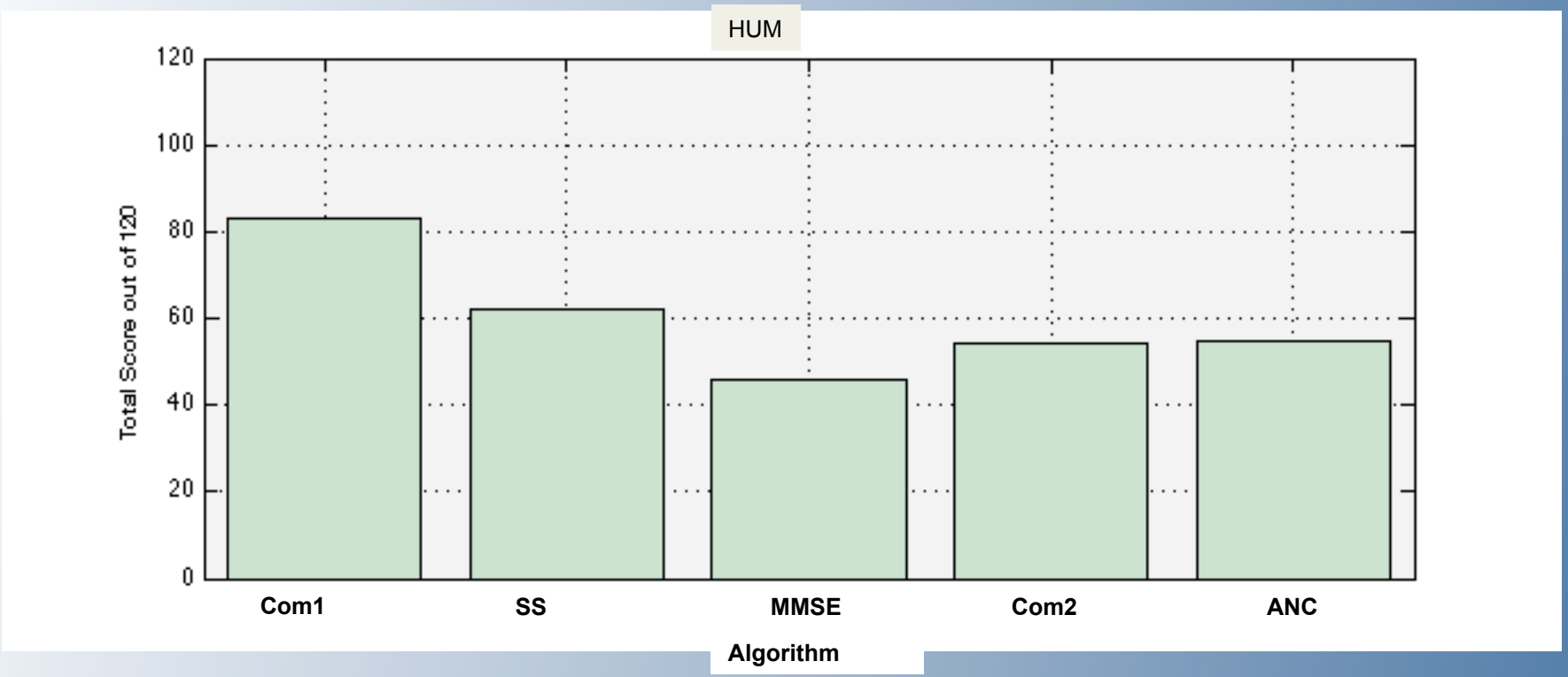
SII Comparison : MMSE/SS



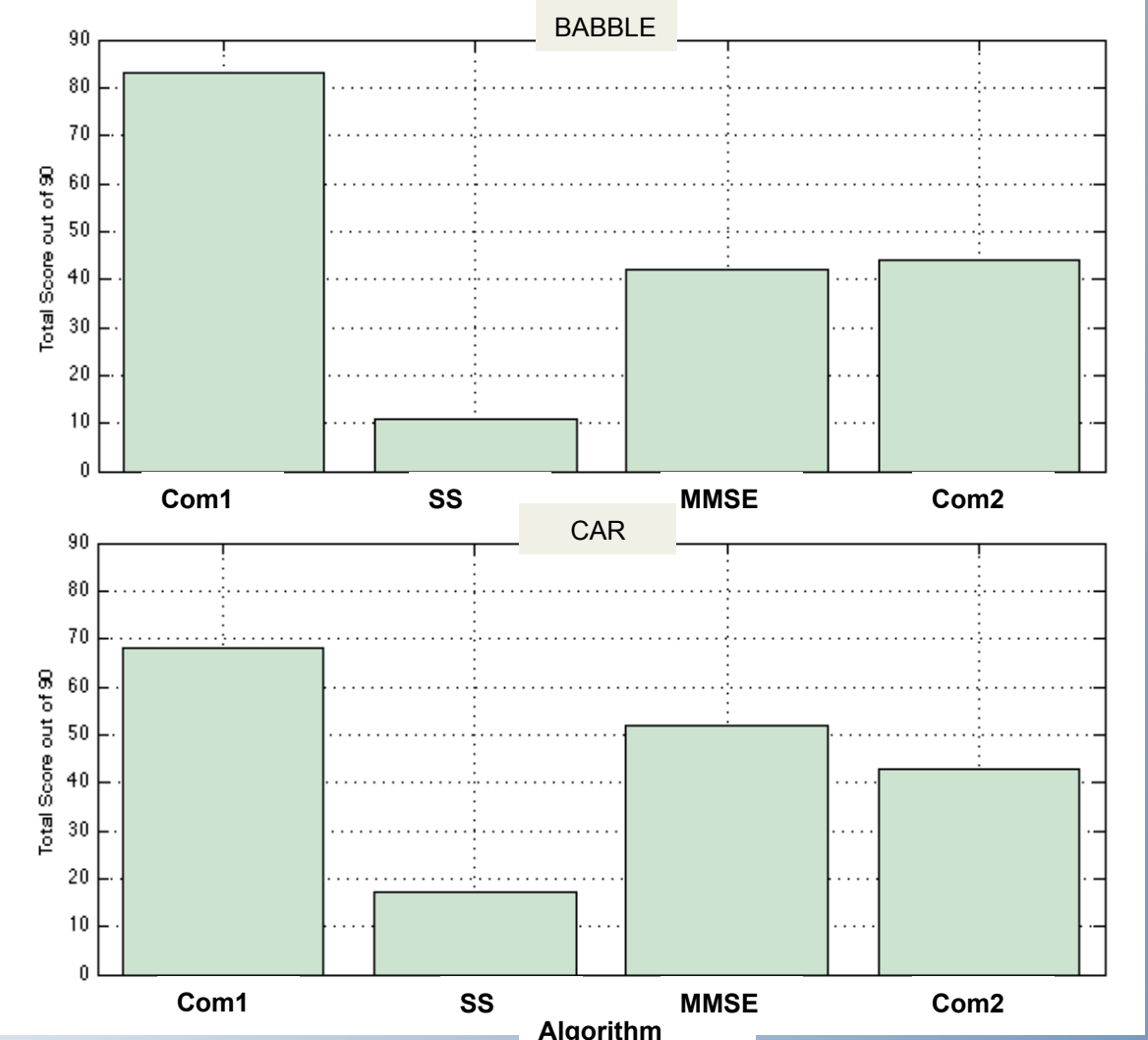
Results

- Comparing Different Enhancement Algorithms:
  - The results from the tests yielded a total score, representing the number of times one enhancement technique was chosen over another
  - This is the reference metric used for comparison
  - Results for SS, MMSE, Com1, Com2
  - ANC tested for Hum noise only

Comparing Algorithms :*Hum Noise*



Comparing Algorithms : *Car and Babble*



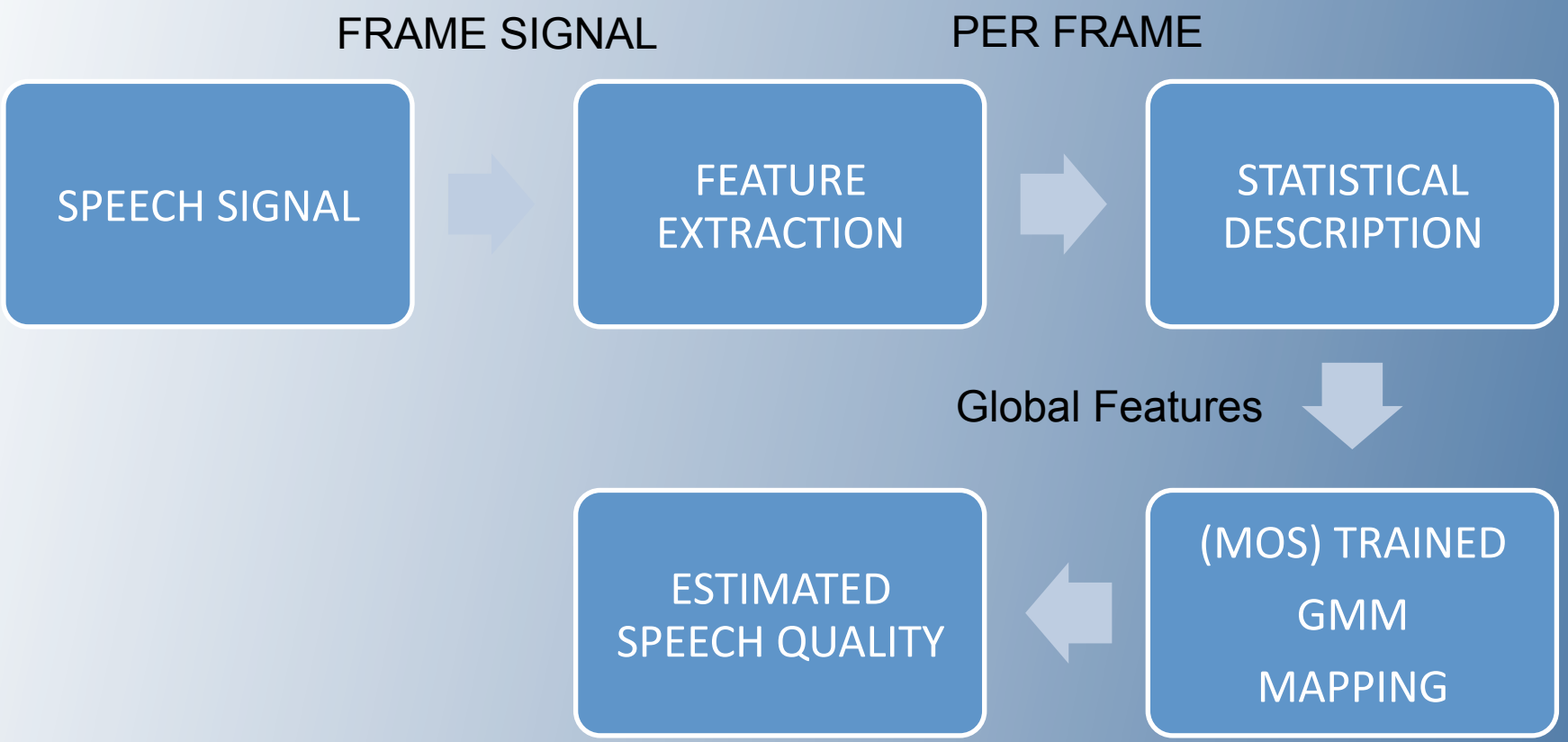
Conclusions

- Com1 has best overall performance for all SII and Noise Types
  - Uses a 'clever' combination of enhancement algorithms
  - Applies Post-Processing to sound more natural
- SS/MMSE work best on Hum
  - SS, MMSE, Com2, ANC : Similar Performance
- Car and Babble
  - MMSE and Com2: Similar Performance
  - SS : Poor Performance due to corruption by musical noise
- Need for a nonintrusive objective quality classifier for fine tuning of baseline algorithms

Low Complexity, Nonintrusive Speech Quality Assessment *[Grancharov et al]*

- Designed as a tool for QoS monitoring of Communication Networks
- Extracts a set of features and GMM mapping to obtain a measure of quality
- Lower Complexity than the ITUT P.563 standard algorithm
  - Generates quality assessment ratings without explicit distortion modeling
  - No perceptual transformation of the signal
  - Features can be computed from commonly used speech-coding parameters

LCQA : Algorithm Overview



LCQA : Per Frame Features

- Feature Set aims to capture the structural information from the speech signal
  - **Spectral Flatness** : related to the strength of the resonant structure in the power spectrum. High spectral flatness indicates that the spectrum has a similar amount of power in all frequency bands (white noise).
  - **Spectral Dynamics** : rate of change of the power spectrum
  - **Spectral Centroid** : frequency area around which most of the signal energy is concentrated
  - **Var. of excitation of a 10^{th} order AR model**
  - **Speech Signal Var.**
  - **Pitch Period** : Speaker Dependant feature (quality is speaker dependant)

LCQA : Global Features

ELEMENTS OF PER-FRAME FEATURE VECTOR

Description	Feature	Time derivative of feature
Spectral flatness	Φ_1	Φ_7
Spectral dynamics	Φ_2	-
Spectral centroid	Φ_3	Φ_8
Excitation variance	Φ_4	Φ_9
Speech variance	Φ_5	Φ_{10}
Pitch period	Φ_6	Φ_{11}

Mean, Variance and Skewness of the 11 per frame features gives Global Feature Vector of size 14

$$\tilde{\Psi} = \{s_{\Phi_1}, \sigma_{\Phi_2}, \mu_{\Phi_4}, \mu_{\Phi_5}, \sigma_{\Phi_5}, s_{\Phi_5}, = \mu_{\Phi_6}, s_{\Phi_7}, \mu_{\Phi_8}, \mu_{\Phi_9}, \sigma_{\Phi_9}, s_{\Phi_9}, \mu_{\Phi_{10}}, \mu_{\Phi_{11}}\}.$$

LCQA : GMM Mapping

- A Gaussian Mixture Model of 12 mixtures is trained on MOS score labeled data
  - Gives a trained GMM

- Uses the stored GMM to predict subjective MOS score from global features of the test data

- LCQA needs lots of training data to model the system effectively

- Claims correlation up to 0.95 with subjective score