MODELLING SPEECH QUALITY FOR NARROWBAND AND WIDEBAND SILK CODEC FOR VOIP APPLICATIONS

Mohammad Goudarzi, Lingfen Sun, Emmanuel Ifeachor
University of Plymouth
School of Computing and Mathematics

PIONEER WITH PLYMOUTH UNIVERSITY

Overview

- Introduction
- SILK Codec
- Overview of speech quality models
- Testbed setup
- SILK quality modelling based on E-model
- Model’s performance
- Subjective evaluation of the model
- Summary

Introduction

- VoIP networks , Skype
- Factors affect speech quality
- Proprietary software, open-source codec
- How to measure speech quality

SILK Codec

- Speech codec and compression format develo
  - Sup
  - Exp … cordless pho
  - Freely available for third party users

Mode	Fs (Hz)	BR(kbps)
Narrowband	8000	6 – 20
Mediumband	12000	7 – 25
Wideband	16000	8 – 30
Super Wideband	24000	12 – 40

- Highly scalable codec
  - Network bandwidth
  - Send bitrate
  - Complexity

Overview of measurement methods

- Subjective methods
- Objective models
  - Computational models
    - E-model
    - Regression based models
    - Machine learning
  - Signal-based models
    - PESQ (ITU-T P862) and PESQ-WB
    - POLQA (ITU-T P.863)

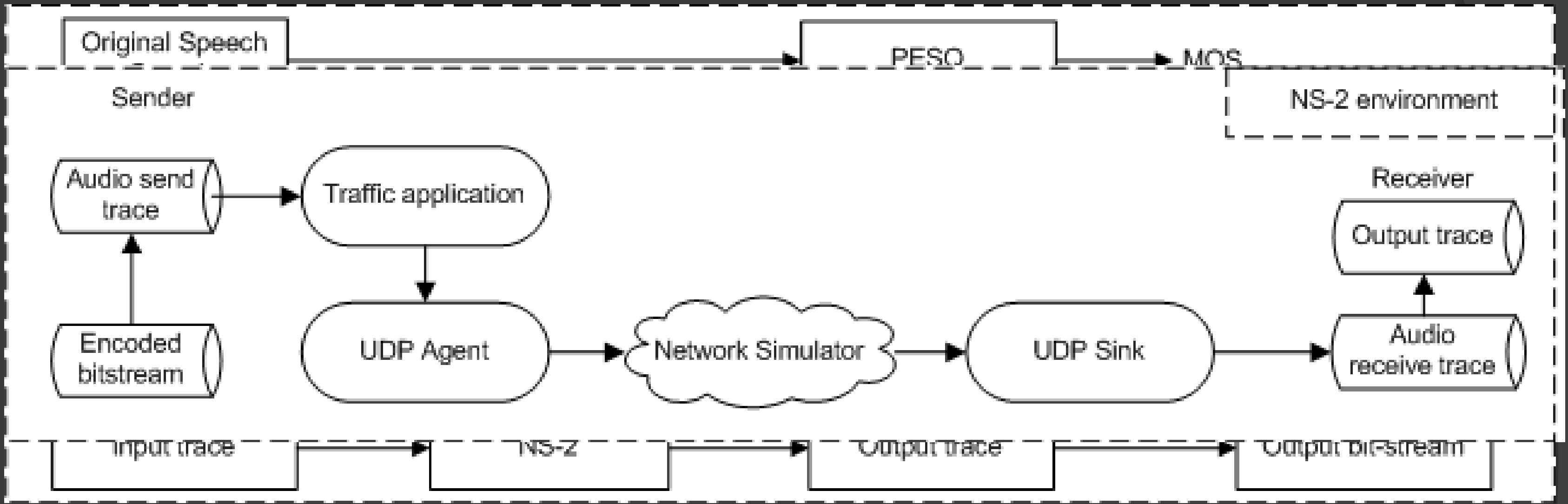
Speech quality measurement models

- Measure the end-to-end quality (MOS)
- Impact of network conditions (i.e., packet loss, delay, etc.)
- Maintain a good service quality during the call
- Application level adaptation
  - Bitrate
  - Sampling frequency
  - Packetization
  - Inband FEC
- Online monitoring and control

Aims

- Low complexity, reliable measurement model to quantify the speech quality for SILK codec
- The model can be used in adaptation, FEC tuning and user satisfaction measurement
- Use both application level and network level parameters to predict the speech quality of VoIP calls.
- Can have direct applications in monitoring and adaptation of VoIP applications

Testbed setup



- Packet loss (0% to 30% with 5% intervals)
- 16 British English speech samples (ITU-T P.50)
  - 8 male and 8 female
- 15 different initial seeds

Modelling based on E-model

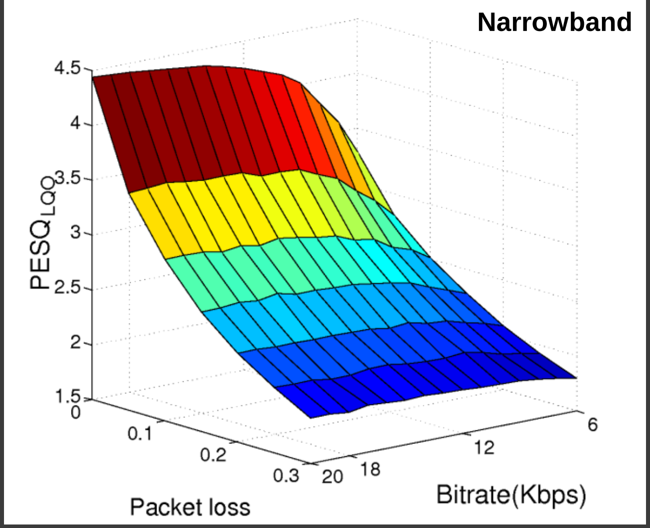
E-model (ITU-T G.107) :

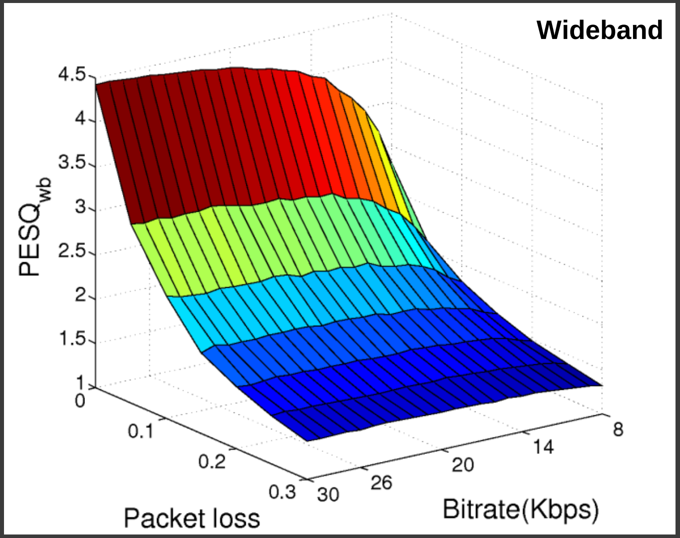
$$R = R_o - Id - Ie + A$$

- R_o : basic signal-to-noise ratio (SNR)
  - 93.2 for NB
  - 129 for WB
- Id : is the impairments caused by delay factors
- Ie : includes distortions by codec and packet loss
- A : Advantage factor.

$$R = R_o - Ie(bitrate, loss)$$

Modelling $Ie.eff$





$$R = 3.026MOS^3 - 25.314MOS^2 + 87.060MOS - 57.336$$

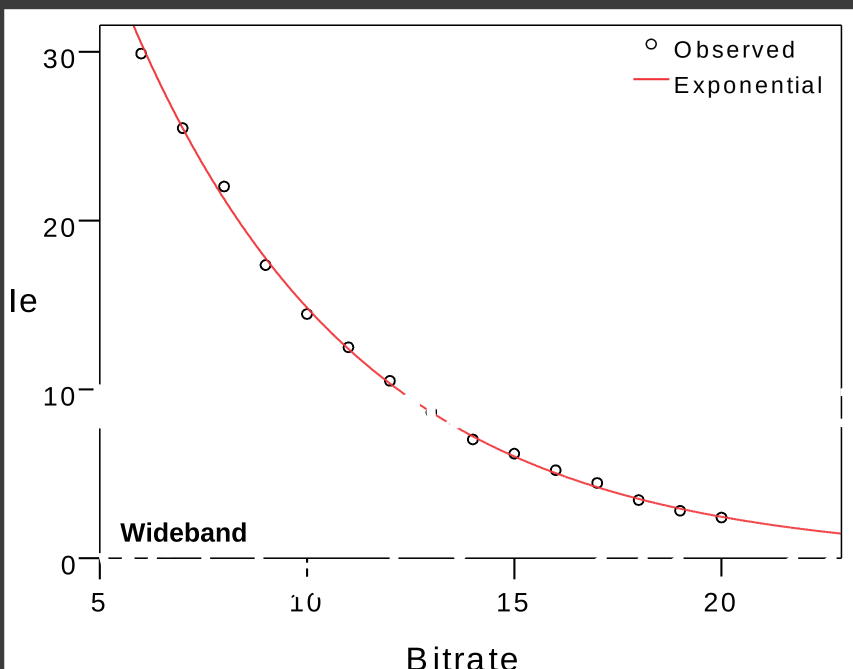
$$Ie.eff = Ro - R$$

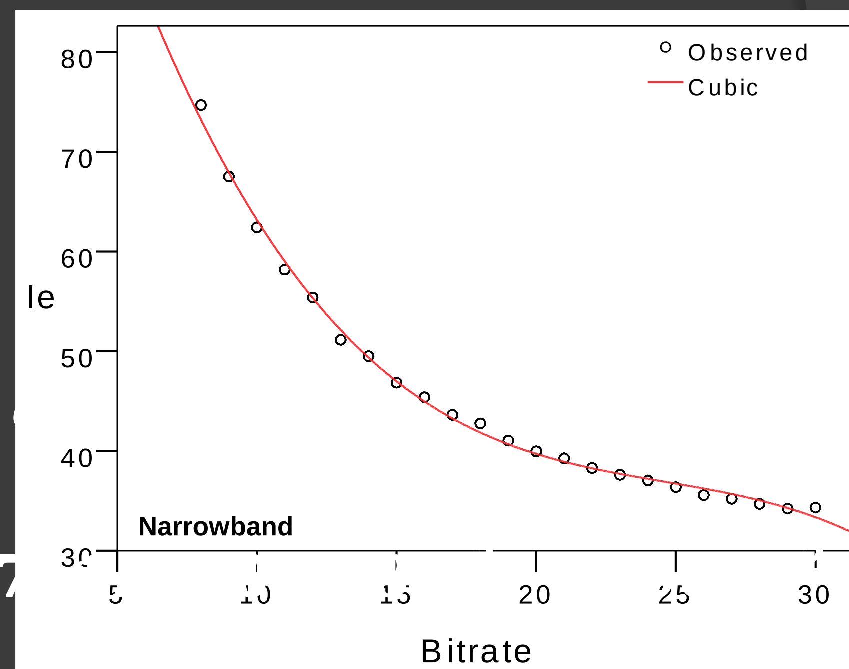
$$\mathrm{I_{e.eff}} = \boldsymbol{I_e} + (\boldsymbol{R_0} - \boldsymbol{I_e}) \frac{\boldsymbol{Ppl}}{\boldsymbol{Ppl} + \boldsymbol{Bpl}}$$

Modelling Ie

Ie values are obtained by taking the $Ie.eff$ values in 0% packet loss.

Each bitrate level, has a fixed Ie value.





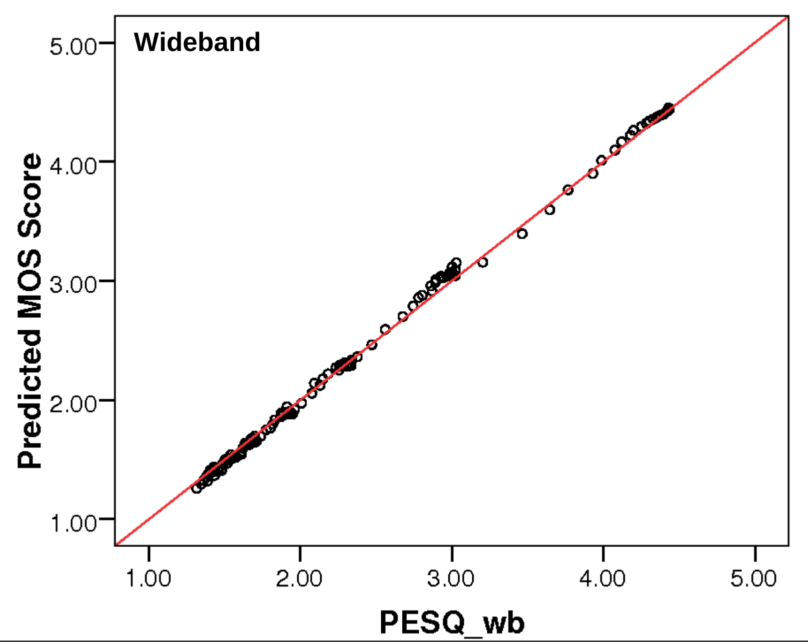
Modelling $\boldsymbol{Bpl}$

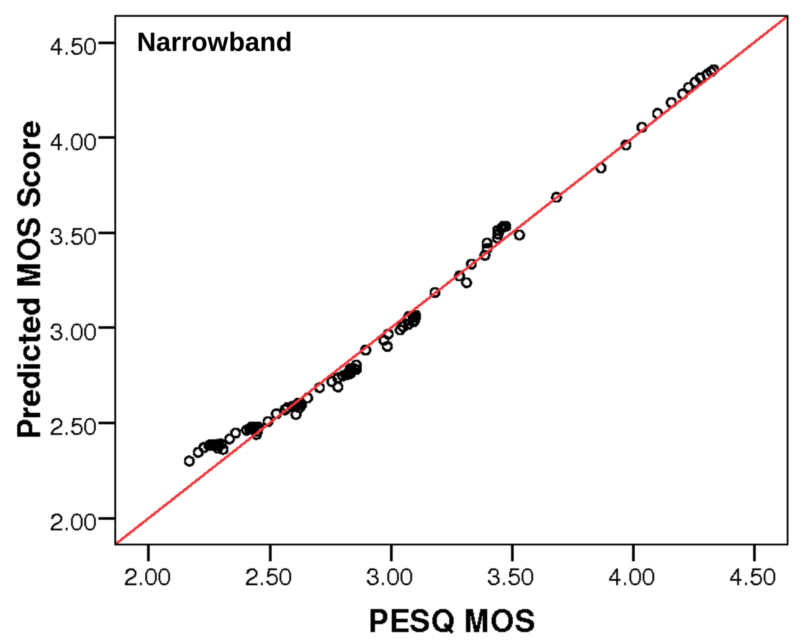
- Non-linear regression analysis of the model with our training set
- $Ie.eff$ values obtained in the previous step

$$\boldsymbol{Bpl = \propto_0 + \propto_1 B + \propto_2 B^2 + \propto_3 B^3}$$

Coeff	NB	WB
$\propto_0$	36.836	122.617
$\propto_1$	−4.893	16.391
$\propto_2$	0.287	0.011
$\propto_3$	−0.006	0
r^2	99.94%	99.8%
RMSE	0.9922	0.9871

Model vs PESQ results





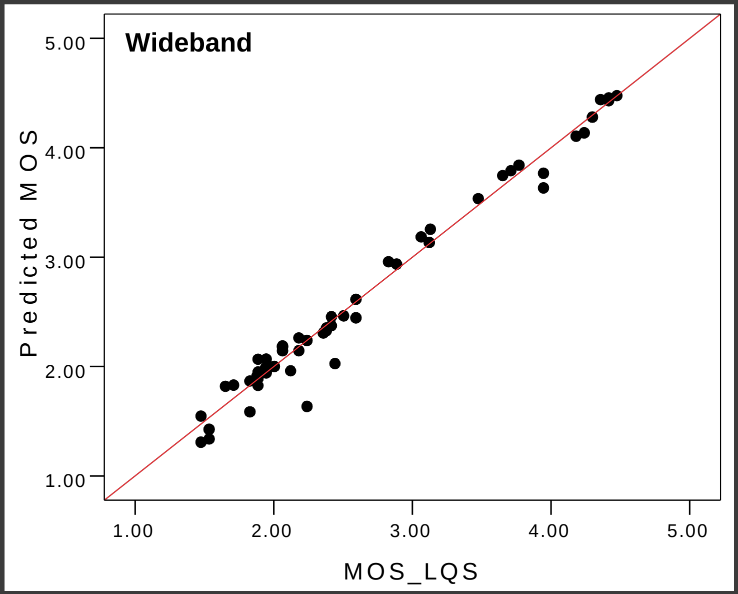
WB mode : 98%

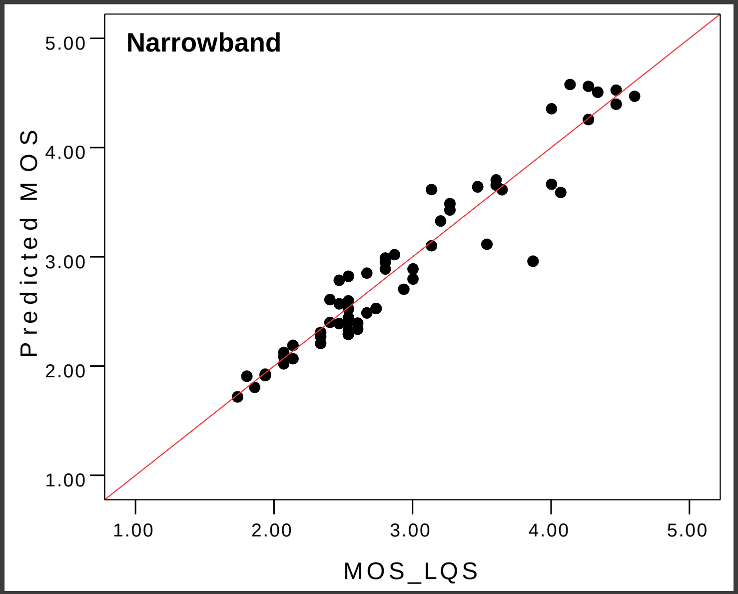
NB mode : 99%

Subjective evaluation of the model

- Two sets of subjective tests were carried out for NB and WB samples according to ITU-T P.800 recommendation.
- Sessions were conducted in 2 separate days
  - 60 Samples per session divided into 4 groups
  - Samples were randomly assigned for each user

Subjective results





Measure	WB	NB
r^2	97.83%	91.86%
Max error	0.59	0.901
Min error	−0.190	−0.486
RMSE	0.1362	0.2272

Conclusion

- Developed a regression-based model, that can measure the speech quality of VoIP calls for Skype's SILK codec for VoIP applications
- The model is a reference-free, regression-based model based on sender bitrate and packet loss parameters
- Validated our developed model with subjective test results
- Correlation of 91% for narrowband and 97% for wideband.
- Can be implemented in mobile devices or softphones
  - Monitoring, control, adaptation and tuning for optimal user satisfaction in real time
- Aspects of this work are generic and can be adopted to other multi-rate codecs

Thank you

- Q & A

- Contact information
  - WEBSITE: http://www.tech.plym.ac.uk/spmc/
  - EMAIL:
    Mohammad Goudarzi: mgoudarzi@plymouth.ac.uk
    Dr Lingfen Sun : lsun@plymouth.ac.uk
    Prof. E Ifeachor : eifeachor@plymouth.ac.uk