Demystifying QoS - Measurement:
Monitoring, MOS scores and R factors

Alan Clark, CEO, Telchemy

Key Points

- VoIP Performance Monitoring Architecture
- MOS Scores and R Factors
- E Model and VQmon
- Some MOS misconceptions

telchemy
Actively managing multimedia

MOS Scores and R Factors

- MOS = Mean Opinion Score
  - 1-5 range
  - Actually a subjective score but we use it for objectively measured quality
  - MOS may be listening (MOS-LQ) or conversational (MOS-CQ)

- R Factors - from G.107 (E Model)
  - 0-95 range for narrowband codecs
  - 0-120 range for wideband codecs
  - R-LQ and R-CQ are often used for listening and conversational quality

telchemy
Actively managing multimedia

Ways of measuring MOS

- Subjective - use a listening panel

- Full reference approach - compare speech output with speech input
  - P.862 - PESQ

- No reference approach - use measurements from the receiving end to estimate MOS
  - VQmon, E Model, P.563, P.564
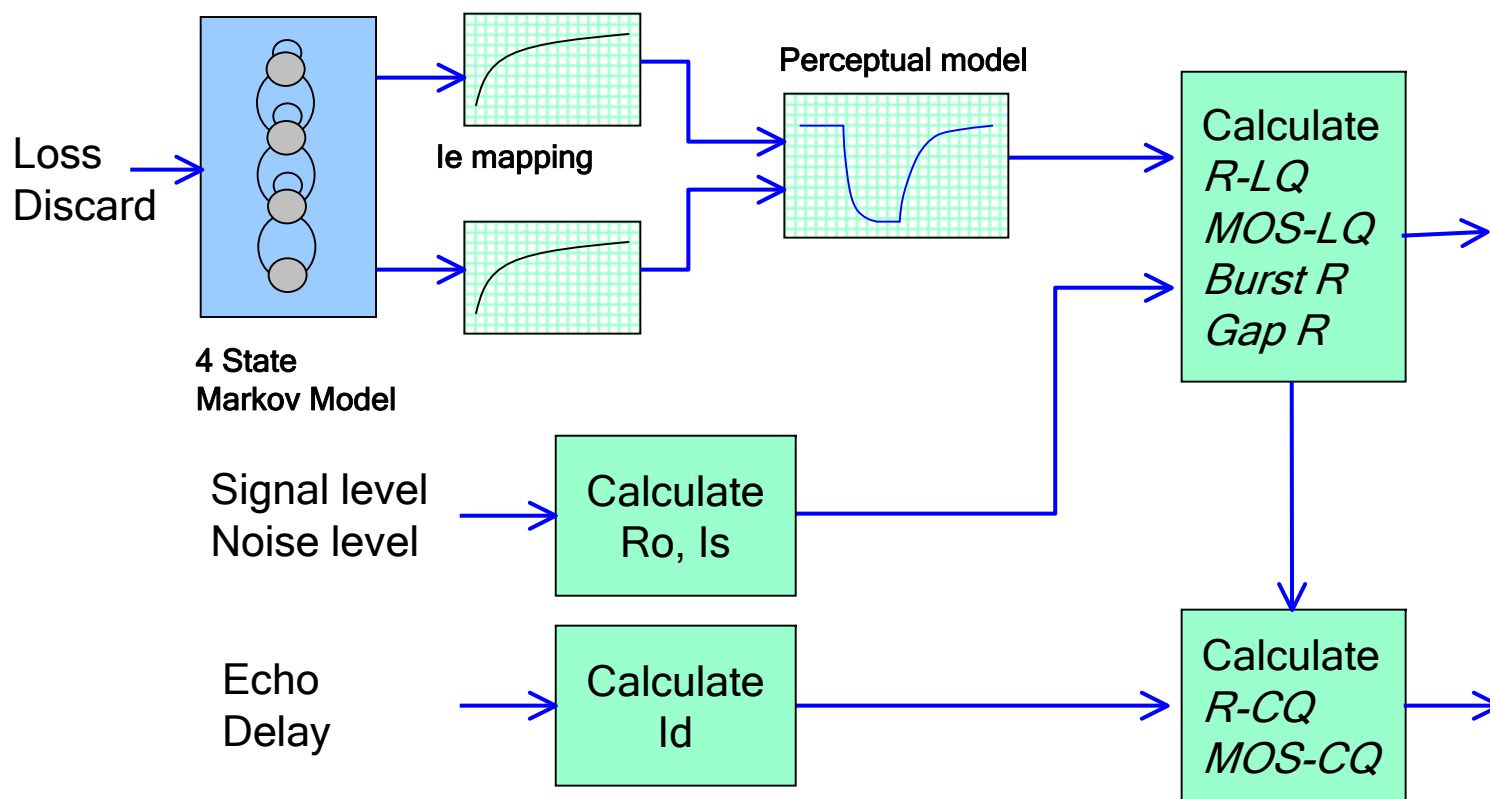
VoIP Performance Monitoring
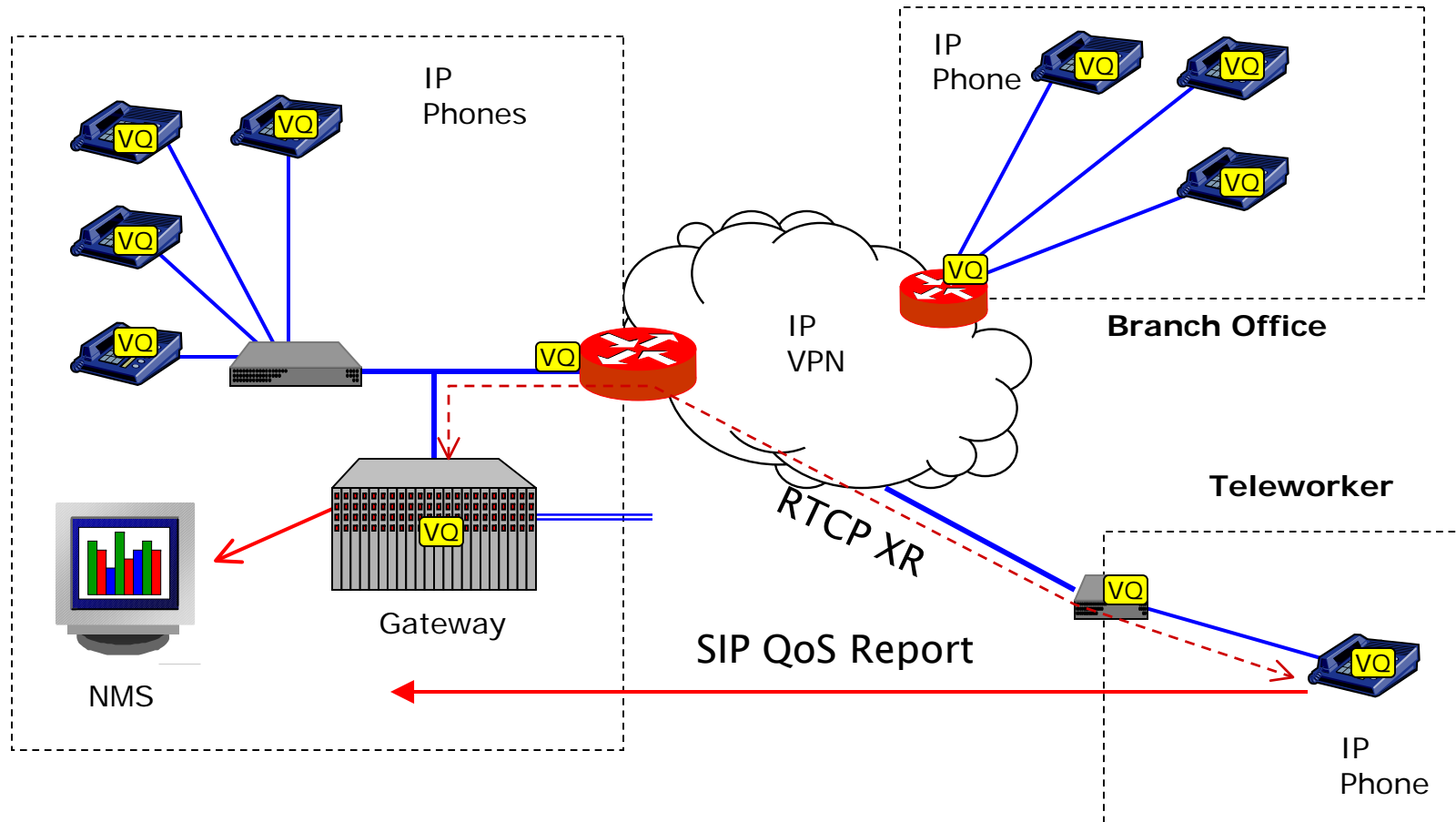
E Model - Simple but Inaccurate

- ITU Recommendation G.107
- Additive model .. R = Ro - Is - Id - Ie

telchemy
Actively managing multimedia

E Model - Simple but Inaccurate

- ITU Recommendation G.107
- Additive model .. $R = Ro - Is - Id - Ie$
- But...
  - Additive assumption is known to be invalid
  - Some implementations only calculate Ie but still claim "E Model"
  - Relies on pre-defined parameters for codecs, only a few published by ITU and some of these are wrong
  - Does not consider time varying impairments (typically due to congestion)
  - Does not consider effects of extended consecutive loss periods

telchemy
Actively managing multimedia

VQmon

Most widely used VoIP performance monitoring algorithm
Only algorithm to properly model time varying impairments

MOS Scores compared
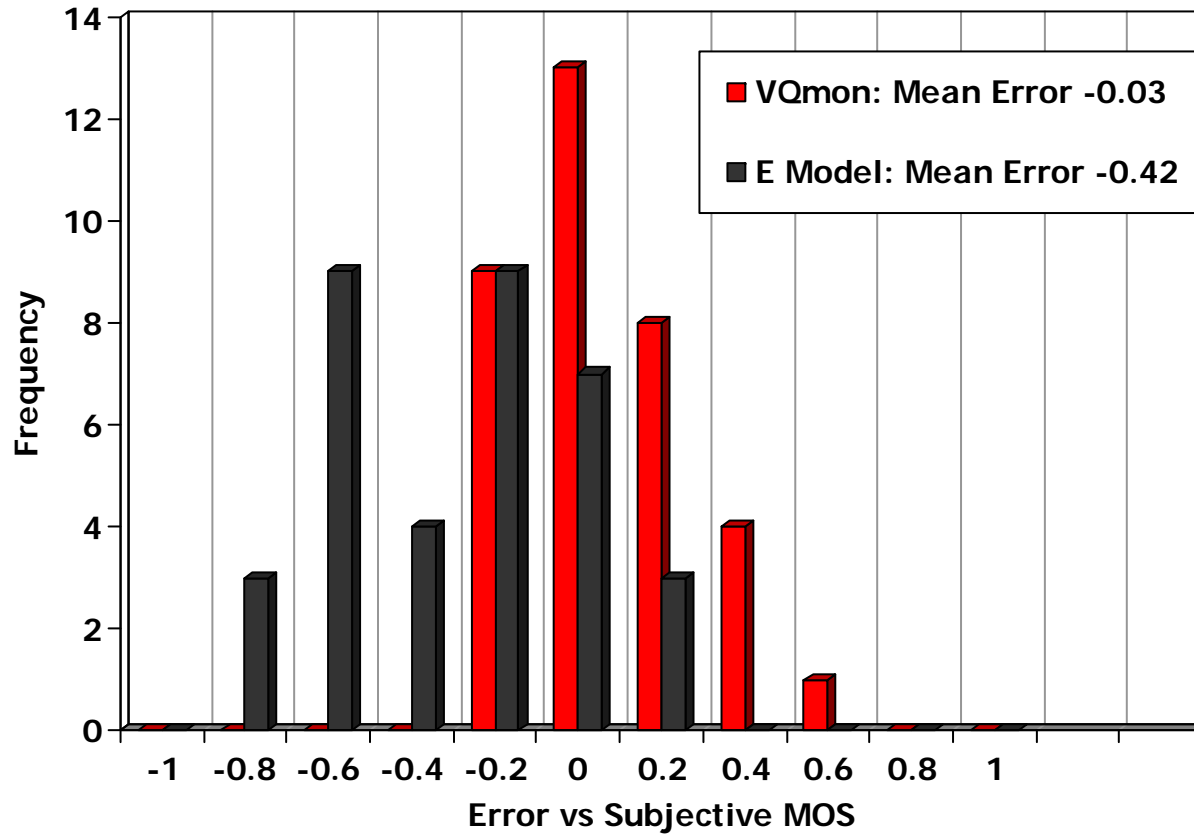
Source ITU, data from France Telecom and University of Bochum

New!! ITU P.564

- New ITU Recommendation (June 2006) which describes how voice quality monitoring algorithms are tested and sets performance criteria

- Very narrow scope
  - Narrowband only
  - Codec specific test - I.e. conforms for G.xxx
  - Only permited inputs - loss & jitter
  - Listening quality only - no conversational quality
  - Defines testing against PESQ, not against subjective test results

- Results?
  - VQmon achieves Class 1 compliance for G.711
  - E Model does not meet requirements of P.564

telchemy
Actively managing multimedia

Some common MOSunderstandings

- MOS scores are actually relative scores (even "Absolute Category Rating")
  - There is no "official MOS" for G.711 - we tend assume a value (e.g. 4.2)

- Narrowband MOS and Wideband MOS use the same 1-5 range
  - E.g. a wideband codec with a MOS of 3.9 may sound much better than a narrowband codec with a MOS of 4.2

- R Factors don't have a 0-100 scale
  - R is generally 0-93 for narrowband codecs but can go up to 120 or more for wideband codecs

- Average MOS is of limited use - Burst/ Gap metrics are more informative
  - Average per-call MOS scores are useful but users are aware of transient problems - typically 1-2 seconds in length

telchemy
Actively managing multimedia

Summary

- Avoid over-simplistic implementations of the E Model, be sure that MOS scores are dependable and accurate

- Be aware of the differences between MOS-LQ and MOS-CQ, and Narrowband vs Wideband MOS

- Implement the VoIP Performance Management Framework - RTCP XR and SIP QoS reporting