End-to-end speech quality assessment of networks using PESQ (P.862)

Antony Rix
Psytechnics Limited
antony.rix@psytechnics.com

Overview

- Importance of speech quality
- Problems with end-to-end assessment
- Solutions and ITU competition
- Applications
- Summary
- Contact details

“The customer is always right”

» What really counts is the customer’s perception of quality

End-to-end speech quality is **the** key measure of voice QoS

Speech quality important for

- Installation
- Monitoring/fault-finding
- Service level agreements
- Optimisation of network

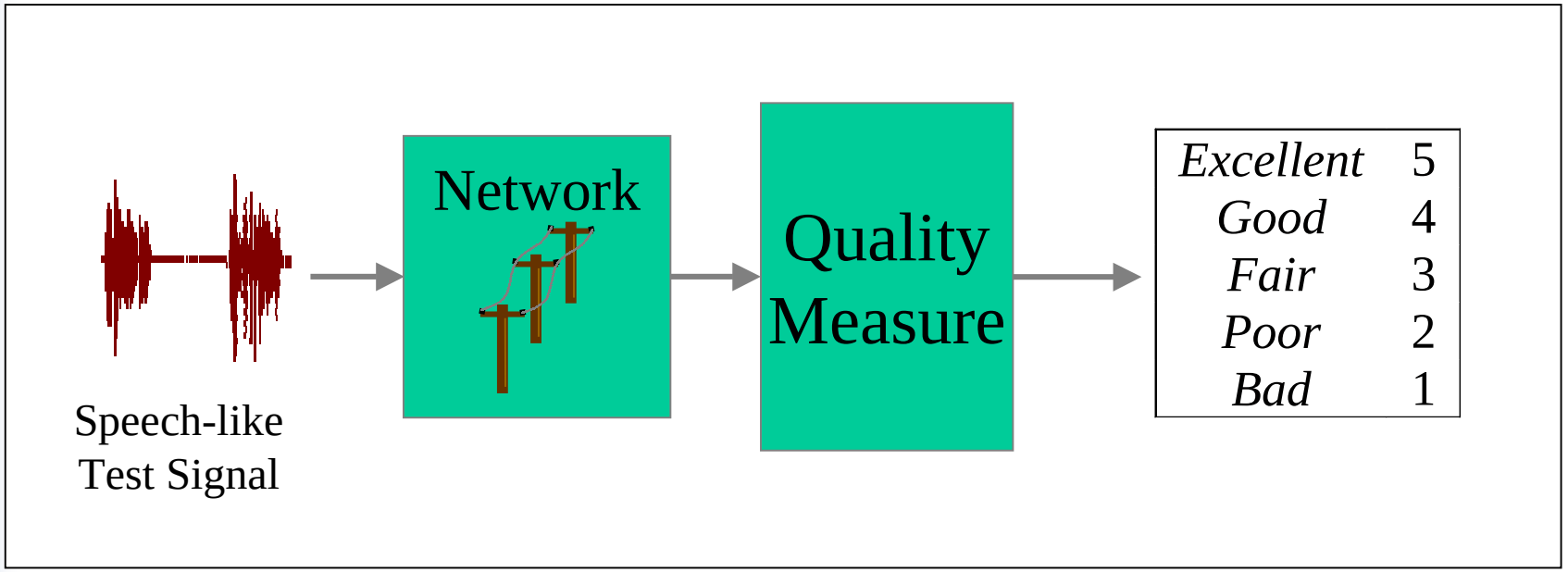
Quality will remain an issue so long as bandwidth or processing power are limited

- e.g. mobile, leased capacity

Factors that affect quality

» Background noise
» Silence suppression
» Low bit-rate coding
» Errors (mobile, packet)
» Delay
» Echo
» Handsets/access network

Objective quality model

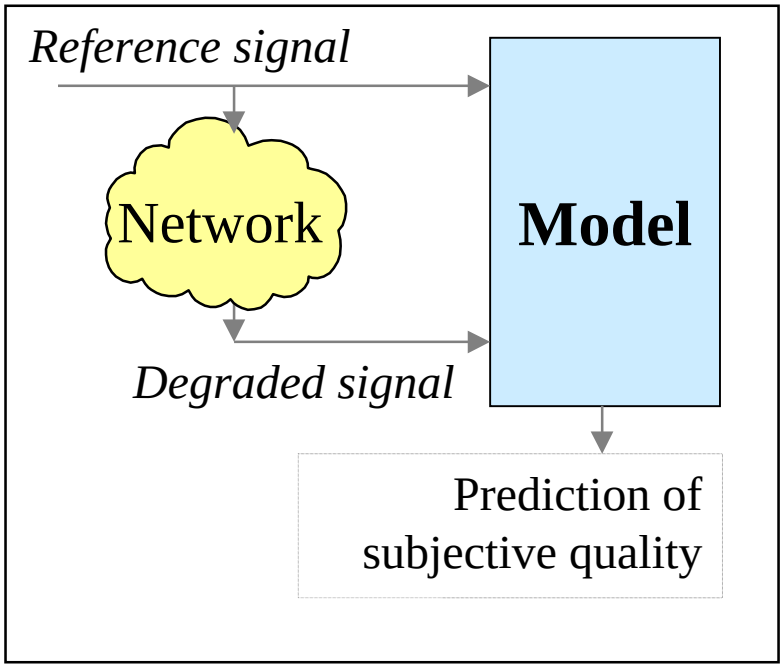


We need a repeatable, accurate measure of end-to-end speech quality, suitable for use across a wide range of codecs, network types, etc.

Objective quality model

- Measures one-way quality
- Intrusive (test signal)
- Based on speech

- Model needs to take account of much more than codecs or packet loss rates



Problems with end-to-end assessment

Previous models (e.g. BSD, P.861 PSQM, P.861 MNB, PSQM+, and others) were mainly designed for assessing speech codecs. They are unsuitable for use with today's networks, especially VoIP, because they are:

- not accurate at predicting quality with some important codecs
- unable to take proper account of noise or errors such as packet loss
- unable to deal with variable delay (common in VoIP)
- unable to account for the effect of analogue elements (e.g. 2-wire access)

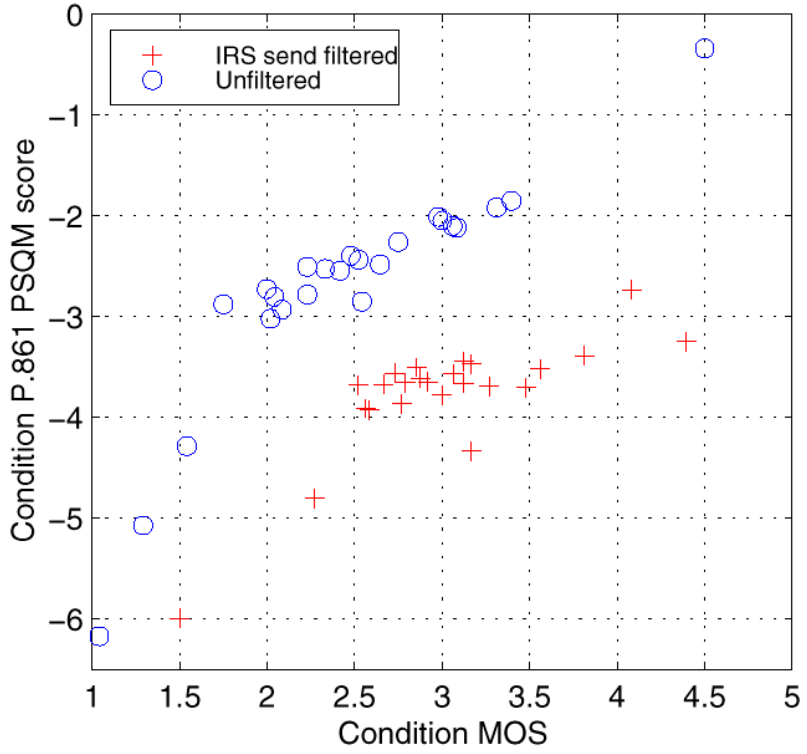
Problems with end-to-end assessment

How do we show problems with these early models?

- Compare objective quality with subjective mean opinion score (MOS) for subjective tests carried out according to ITU P.800/P.830.
- Include real-world conditions representative of those encountered in the field
  - real measurements or detailed simulation of e.g. handsets
  - wide range of codecs, transcodings, and error types
  - real systems e.g. live VoIP networks, or systems in the lab subjected to a range of representative operating conditions

Typical problem: models can't deal with filtering (analogue components)

Low bit-rate codecs, errors, filtering



Graph plots the correlation between subjective MOS and P.861 PSQM score

» P.861 gets the quality score completely wrong (by more than 1 MOS) when filtering is present

» Filter in this case is ITU reference handset model, but a typical 2-wire interface has similar effect.

» Same problem with PSQM+ and MNB

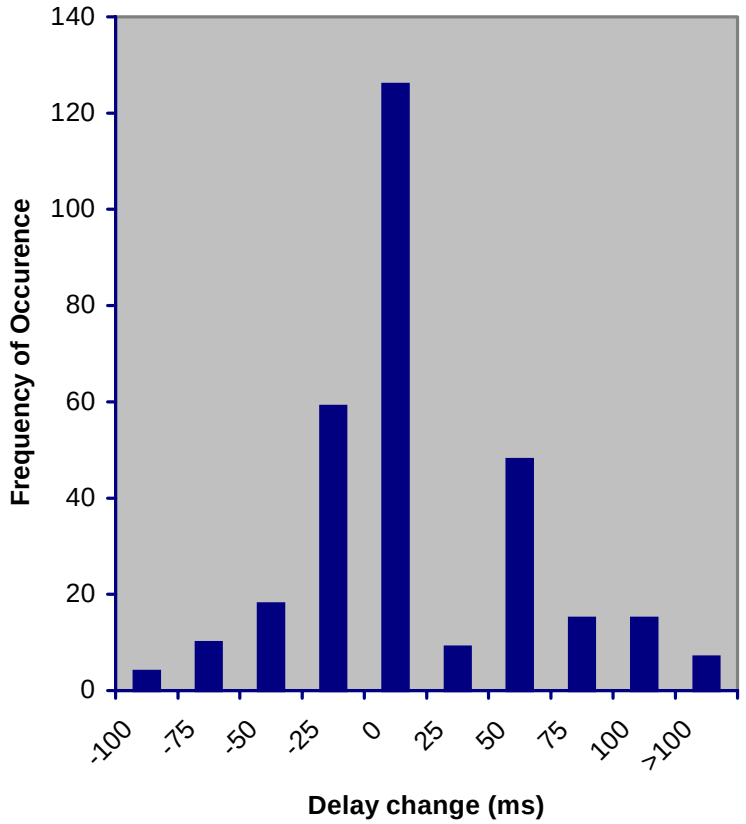
VoIP variable delay

Delay changes are common when VoIP systems are used in practical operating conditions.

Graph shows distribution of delay changes found in measurements in a VoIP field trial that we did.

The majority of delay changes occur in silent periods and are imperceptible - they have no subjective effect.

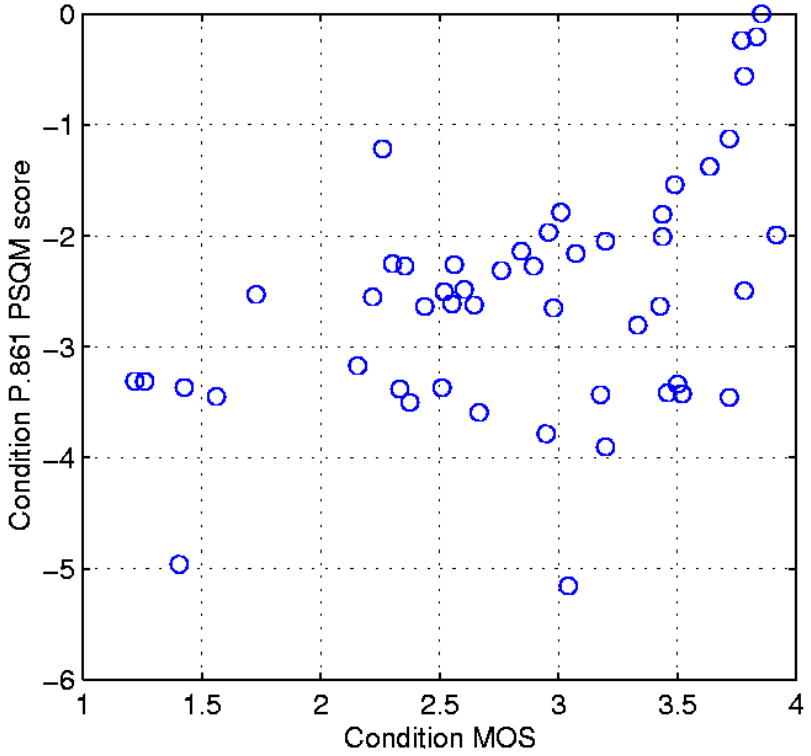
Distribution of VoIP delay change



psytechnics matching technology to people

Typical problem: models can't deal with variable delay common in VoIP

Codecs, errors and VoIP variable delay



Graph plots the correlation between subjective MOS and P.861 PSQM score

» P.861 has very low correlation (<0.5) because it gives incorrect scores to some codecs and all conditions with variable delay

» Same problem with PSQM+ and MNB

psytechnics matching technology to people

Need for a solution

These previous models are unsuitable for testing networks, especially VoIP.

Speech quality is still a key issue.

So there is a clear need for a model that can be used for end-to-end testing of a wide range of real networks: fixed, mobile and VoIP.

» PAMS and PESQ were both developed to address this need.

Replacement of P.861 (PSQM/MNB) by P.862 (PESQ)

A version of PSQM was selected as ITU-T recommendation P.861 (1996) for testing codecs **not** networks. MNB was added to P.861 in 1998, also for testing codecs only.

Since P.861 was found to be unsuitable for testing networks it was withdrawn and replaced by P.862.

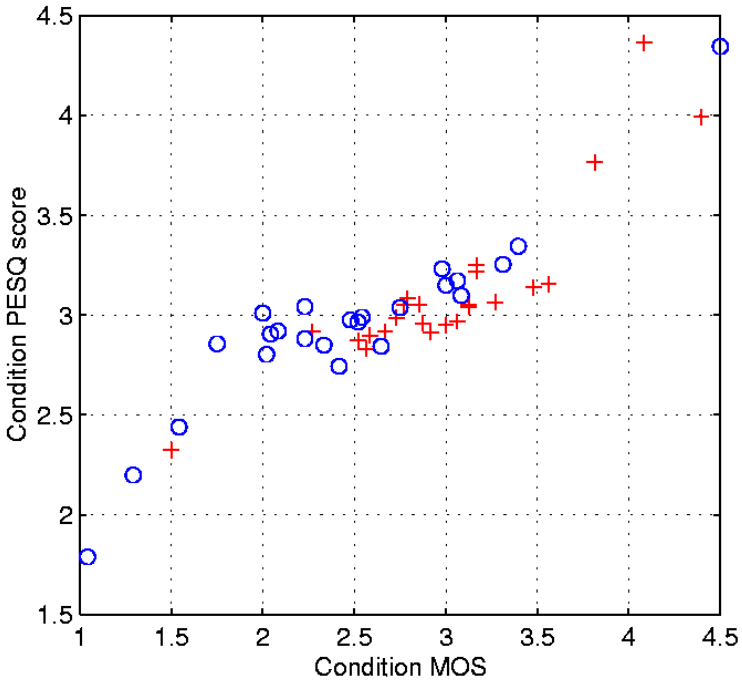
P.862 specifies PESQ (perceptual evaluation of speech quality), the result of a collaboration between BT/Psytechnics and KPN, following a 2-year competition in ITU-T study group 12 that also included Ascom, Deutsche Telekom and Ericsson.

Following extensive testing PESQ has been found to be accurate at predicting quality in a very wide range of real network conditions.

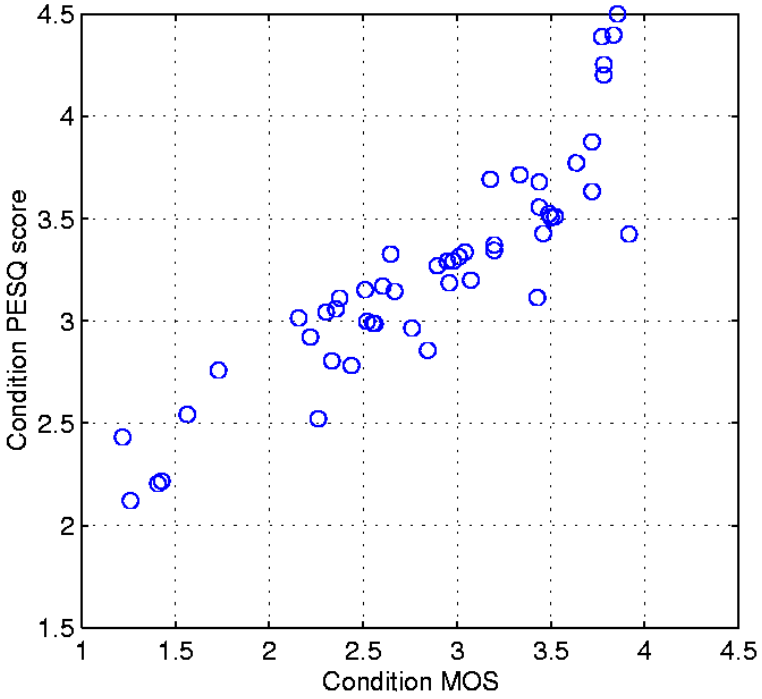
Performance of PESQ

» Same critical experiments for which results were presented before

Low bit-rate codecs, errors, filtering

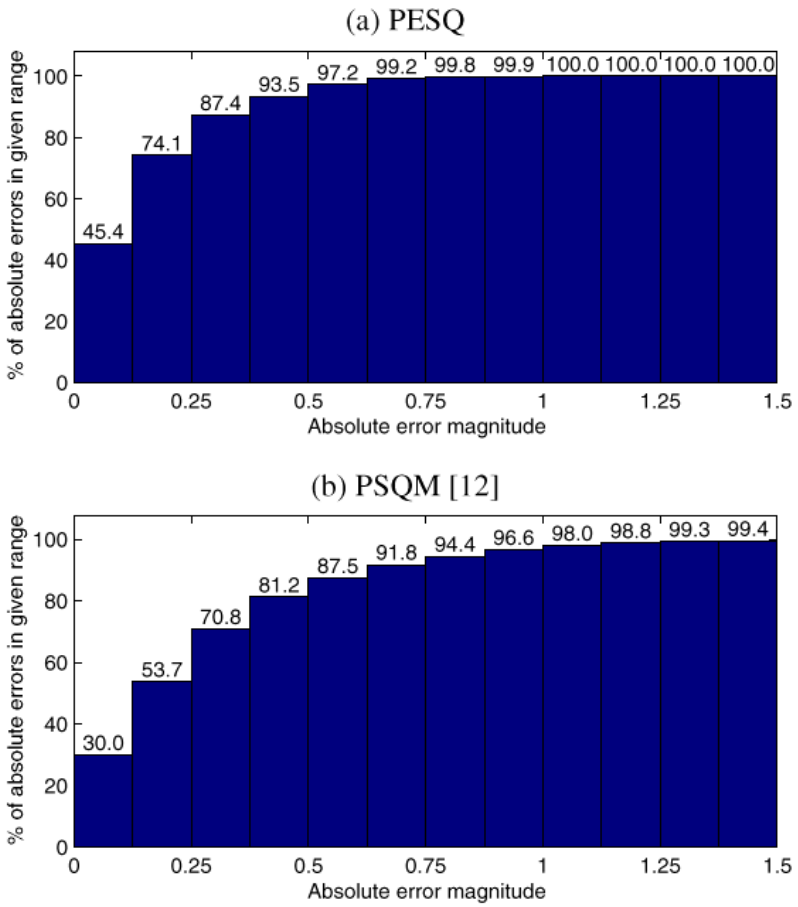


Codecs, errors and VoIP variable delay



psytechnics matching technology to people

Performance of PESQ and PSQM



Graphs show distribution of residual errors after fitting P.862 PESQ and P.861 PSQM to 40 subjective tests on the listening quality scale (P.800), with over 1900 conditions.

PESQ is much more accurate than P.861 PSQM (and other early models including PSQM+ and MNB).

For example, for 93.5% of conditions PESQ was within 0.5 MOS of subjective score.

Applications

- Routine monitoring
- Equipment selection
- Network optimisation/characterisation
- Fault-finding

- … and more

Summary

- Need measures to evaluate customers' perception of speech quality.
- The new assessment model, PESQ (P.862), solves problems that made earlier models unsuitable for use with real networks.
- Wide range of applications in designing, optimising and monitoring networks.

Further information

Antony Rix

Psytechnics Limited

Fraser House, 23 Museum Street

Ipswich IP1 1HN, UK

Tel. +44 1473 261 862

E-mail antony.rix@psytechnics.com

Web www.psytechnics.com

Psytechnics background

A spin-off from BT Labs, Psytechnics brings together a team with a 10-year track record of producing solutions such as:

Intrusive testing:

- **PAMS, PESQ (ITU-T P.862)**
- **Conversational PAMS / Conversational PESQ**
- **psyVQE**

Non-intrusive monitoring:

- **CCI (ITU-T P.562)**
- **NiQA**
- **psyVoIP**

Psytechnics is now developing network QoS solutions based on these and other new techniques.