

# Subjective wideband speech quality and modelling issues

Dr Antony Rix  
Psytechnics Limited

# Agenda

## Background

### Test method

- What is “wideband”?
- Subjective methods for wideband
  - P.800 (ACR etc)
  - BS.1534 (MUSHRA)
  - BS.1116 (Broadcast)
- Example test design and results

## Subjective test issues

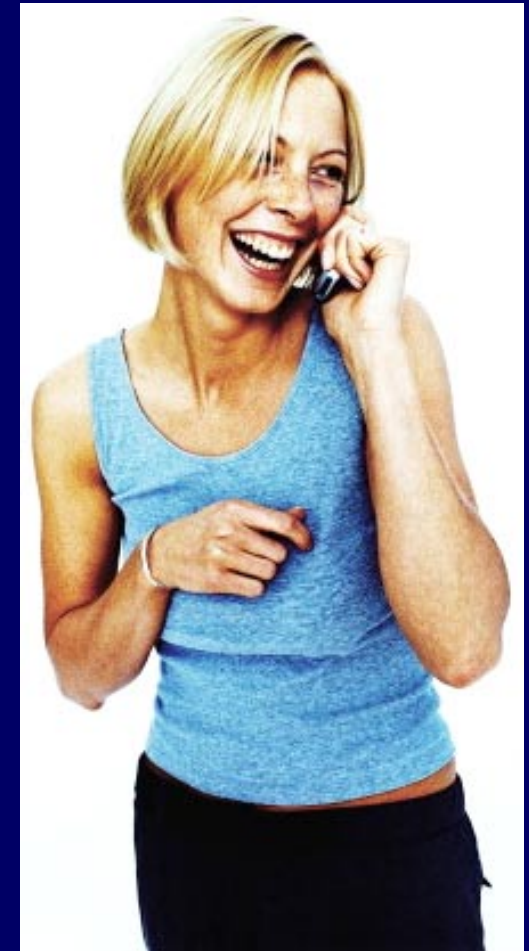
- Agree method (P.800?)
- Listening set-up
- Distortions/interaction
- Narrow/wideband

## Modelling challenges

- Bandwidth
- Interactions
- Content dependency

# Psytechnics background

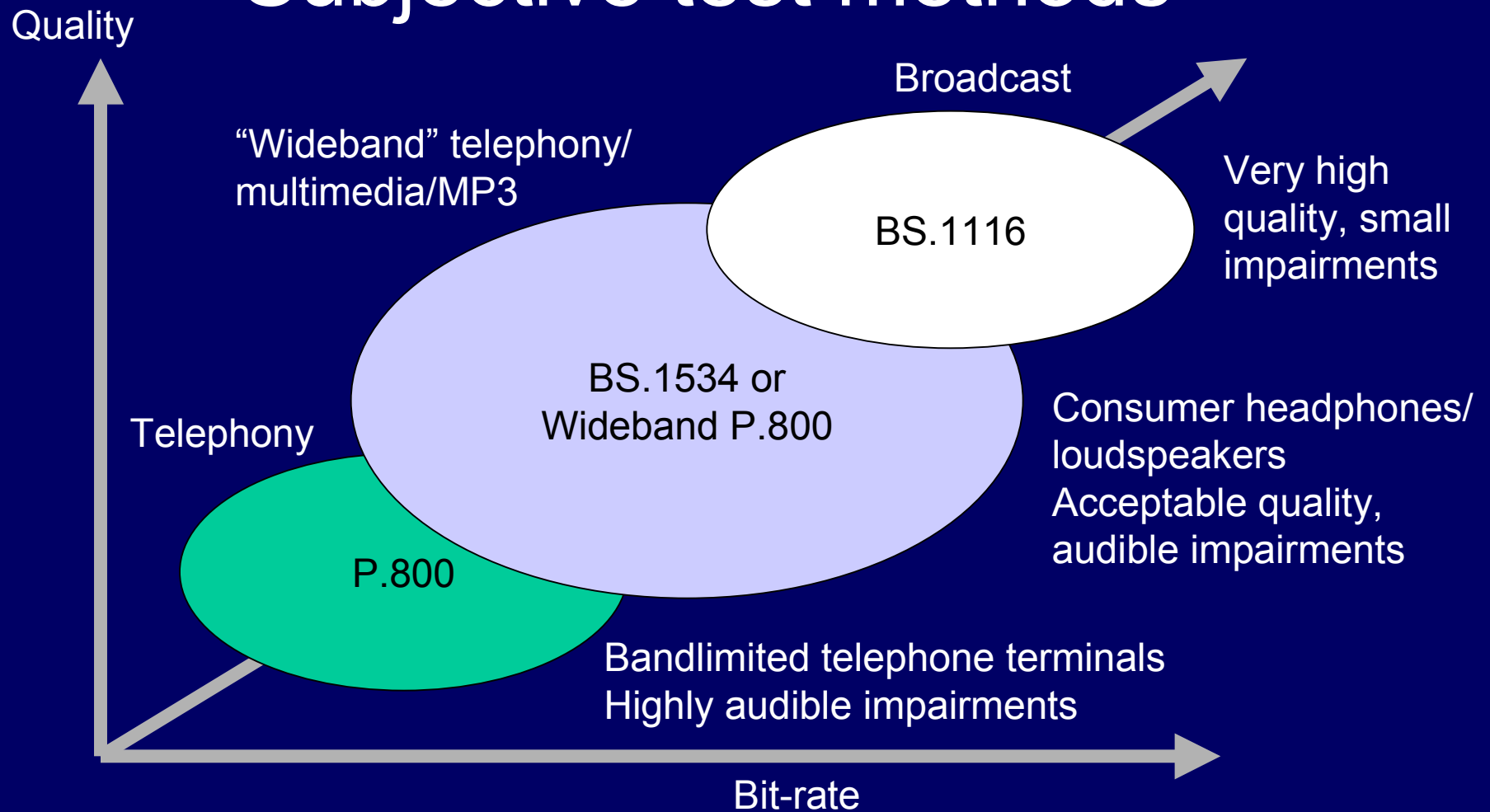
- Solutions for measuring/monitoring speech, audio, video quality
- Extensive subjective testing background
- Main products are objective quality models (software)
  - Intrusive (P.862 PESQ, ...) – for testing
  - Non-intrusive (P.VTQ/psyVoIP, P.563 SEAM/NiQA, P.562 CCI) – for monitoring
- Experience in wideband in both subjective testing and objective models (PAMS, PESQ).



# What is “wideband”?

- Anything better than narrowband 3.1kHz telephony?
  - Up to 20kHz, 16bit, stereo (or higher...)
- In practice:
  - 7kHz (16kHz sampling) often enough for speech/multimedia
  - Mono is fine for speech and many audio signals
- Applications:
  - ISDN telephony/conferencing
  - “Wideband” telephony (16kHz sampling, mono) – VoIP, 3G
  - PC/Internet/mobile audio and multimedia

# Subjective test methods



# P.800 telephony tests

- Listen using narrowband handset
- ACR quality scale (excellent...bad) most common, but many others available
- No or hidden reference
- 16-32 naïve subjects
- Random presentation
- No repeat listening

## Advantages

- Fast and cheap
- Many conditions

## Disadvantages

- Needs 16+ subjects
- Sensitivity issues
- No anchor so MOS varies – can be confusing

# BS.1116 small impairment tests

- Wideband, equalised headphones
- Impairment scale (Imperceptible ... Very annoying), hidden reference
- 6-8 screened & trained expert subjects
- Random presentation
- Repeat listening

## Advantages

- Very sensitive
- Anchors zero distortion

## Disadvantages

- Slower & more costly
- Experts can be 20dB “over-sensitive”
- Not approved for large impairments

# BS.1534 (MUSHRA) tests

- Wideband, equalised headphones
- Continuous quality scale (excellent...bad)
- ~16 naïve subjects
- Initially random presentation
- Re-order & repeat listening
- Bandlimited anchors

## Advantages

- More focused on quality
- Repeat listening improves sensitivity

## Disadvantages

- Still slow
- Repeat listening – subjects over-train
- Anchors can skew results

# P.800 wideband tests

- Wideband, equalised headphones
- Like telephony, uses naïve subjects, no repeats, discrete scale
- Successfully used for wideband speech & audio quality

## Advantages

- Fast and cheap
- Many conditions
- Can model acoustics

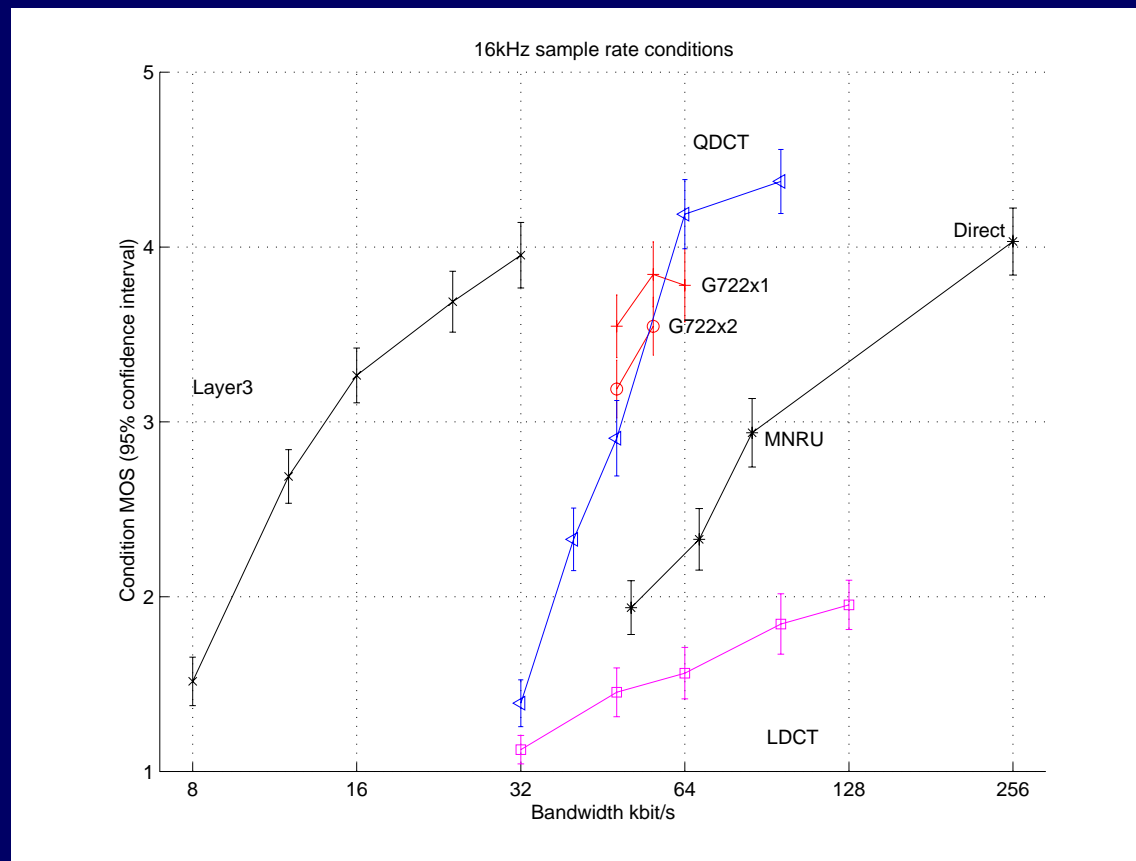
## Disadvantages

- Needs 16+ subjects
- No anchor so MOS varies – can be confusing

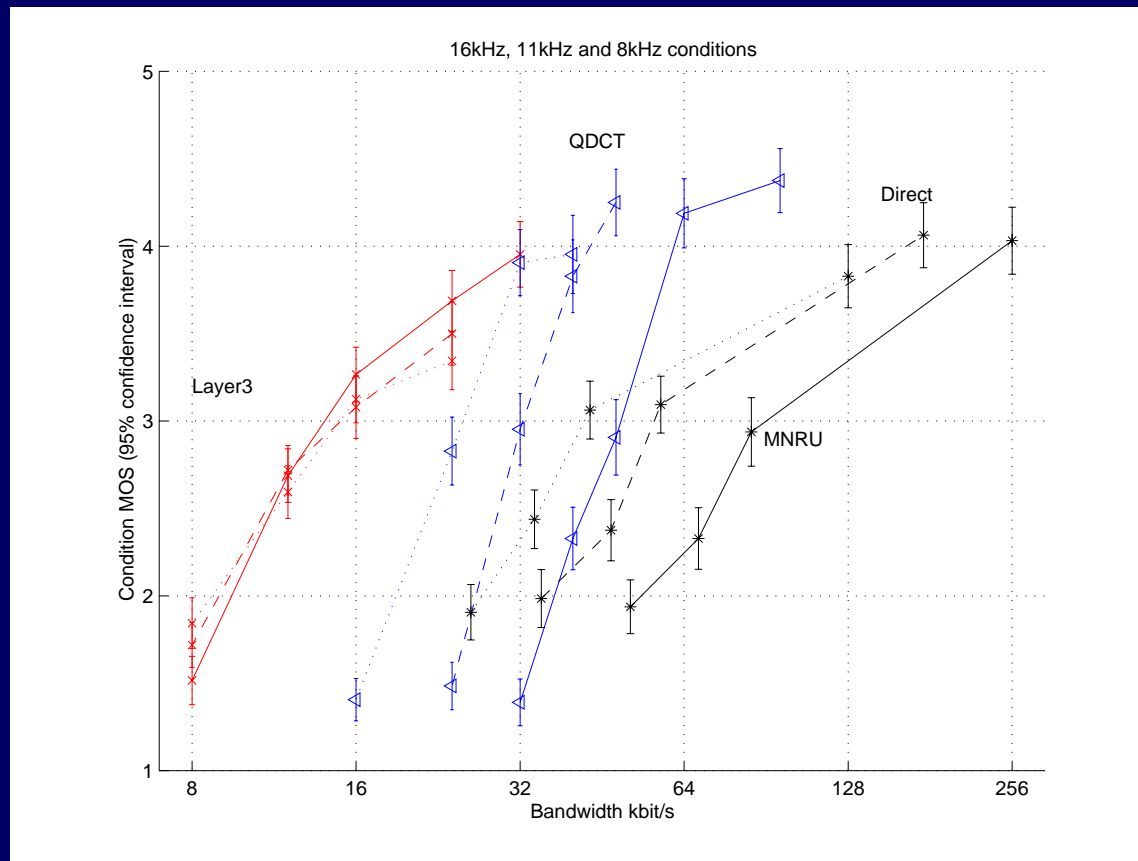
# Example P.800 wideband test

- Wideband circumaural headphones (Sennheiser HD265)
- Equalised to flat ERP response, 73dB SPL, binaural presentation
- 16kHz unfiltered mono speech (2 male, 2 female talkers)
- Conditions:
  - Codecs: G.722, MP3, QDCT, LDCT, MNRU
  - Bit-rates: 8-256kbit/s
  - Coded sampling rates: 8, 11.025, 16kHz
- 8-way partially factorial design to control material dependency
- More details: Rix 1999 paper

# Example P.800 wideband test



# Example P.800 wideband test



Key:  
— 16kHz  
- - - 11.025kHz  
· · · 8kHz sample rate.

# Issues & challenges

Subjective methods  
Modelling issues

# Subjective test issues (1)

## Opinion scale for wideband

- MUSHRA vs. P.800 – need to decide
  - Repeat listening and comparison options in MUSHRA make for an artificial test context
  - We favour the P.800 ACR method
  - P.800 method is most commonly used in ETSI/3GPP and ITU-T for clean speech

# Subjective test issues (2)

## Listening equipment & process

- Type of headphones – supra-aural/ circumaural, closed/open
- Listening level (73dB SPL is slightly quiet)
- Monaural/binaural
- Equalisation – ERP is unusual; free field or diffuse field more typical

# Subjective test issues (3)

## Distortions and interaction

- What do we use as a reference (like MNRU for narrowband) – it affects results?
- Need to set minimum quality and include hidden reference
- Noise reduction problem (e.g. QDCT codec in example results above)
- Bandwidth extension codecs – intelligibility vs. quality
- Interaction between bandwidth & coding distortion – often better to reduce bandwidth at low bit-rates
- Content dependency is significant (talker gender, non-speech signals)

# Subjective test issues (4)

## Narrowband vs. wideband

- Some evidence of effect of subjective context on perception of bandlimiting – e.g. Eurescom P905
- Particular issue with hearing-impaired subjects – some adults lose 4-7kHz range
- Our results show no problems including multiple audio bandwidths in the test
- **Bandwidth is not a major factor!**
- Effect of linear pre-filtering (like IRS send) is yet to be explored
- Terminal dependency – acoustics, filtering

# Modelling issues

- Strong interactions between factors – can't separate these out for a model:
  - Bandwidth and coding distortion
  - Pre-filtering and distortion perception
  - Effect of the content – some signals are much more critical than others for certain codecs
  - Listening level and distortion audibility
- How much of an advantage is wideband?
  - Is 100-3800Hz good enough in many applications?
  - IRS-type telephone handset can perform better in noise

# Conclusions

- P.800 wideband method is workable and has advantages over BS.1534 MUSHRA
- Need further standardisation on key issues of subjective testing method
- Wideband shows strong interaction effects that may be difficult to model parametrically

# References

- ITU-T P.800. *Methods for subjective determination of transmission quality.* August 1996.
- ITU-R BS.1534. *Method for the subjective assessment of intermediate quality levels of coding systems.* Jan 2003.
- ITU-R BS.1116. *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems.* Oct 1997.
- ITU-T P.58. *Head and torso simulator for telephony.* August 1996.
- Rix, A. W. and Hollier, M. P. *Perceptual speech quality assessment from narrowband telephony to wideband audio.* 107th AES Convention, New York, preprint 5018, September 1999.

Dr Antony Rix  
Psytechnics Limited  
[antony.rix@psytechnics.com](mailto:antony.rix@psytechnics.com)  
[www.psytechnics.com](http://www.psytechnics.com)