Quality aspects of real-time voice communication

Ian Marsh, SICS/KTH

Summary of research
February 2009

Vision

The Internet will carry a significant proportion of the world's voice traffic *and* the quality should be *no worse* than that offered by the traditional telephony system.

What does quality mean for a network researcher?

- Delay
  - Strictly it is not quality, rather interactivity
  - Under 180 ms rated as "good", between 180 ms and 400 ms "acceptable" and over 400 ms "unacceptable"
  - Note: once delay is in the system it cannot be removed
- Loss
  - For G.711 (POTS quality) losses can be 1-10% (if random)
  - Losses can be concealed, unlike delay
- Jitter
  - Jitter either results in delay, or loss, but not both

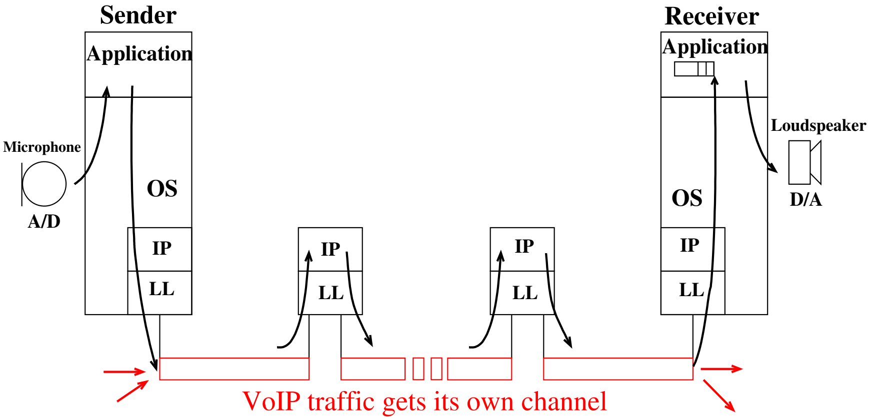
If you ask a subjective quality assessment person they will tell you something very different!

Problems facing VoIP quality

First a *non-problem*. There is nothing inherently problematic with sending real-time voice across a packet switched network, but:

- Disruption caused by competing TCP traffic
  - Probes available capacity by inducing loss
  - Traffic varies on different time scales
  - An increase in P2P media transfers
- Wireless communication
  - Interference, the environment & users (movement) can induce problems
- Other obstacles to the VoIP vision
  - Poor quality infrastructure (in developing countries)
  - The end-systems (can add) considerable delay
  - Human tolerances, the *real* end-systems

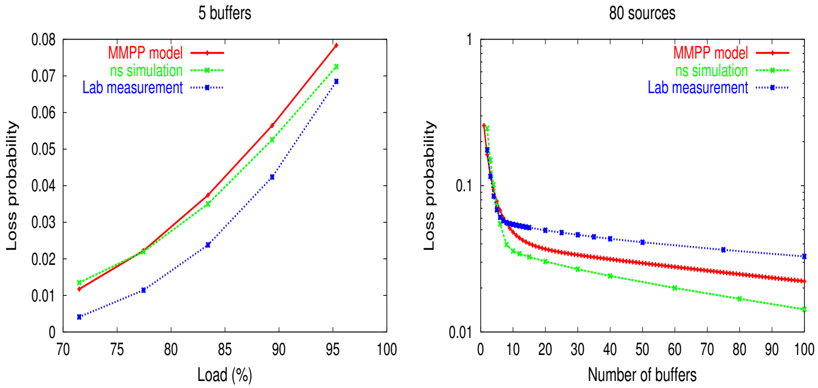
Protecting VoIP traffic I



Protecting VoIP traffic II

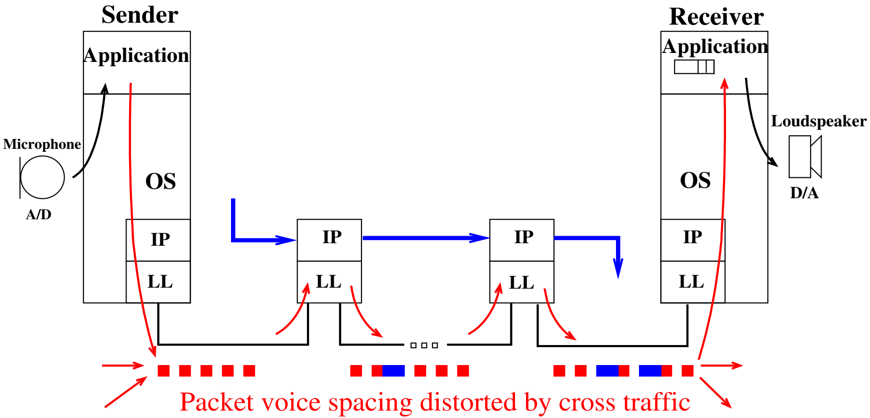
- Problem becomes capacity allocation for a required quality
  - Telephony and ATM research fields have suggested solutions
  - Largely ignored by the IP community
- We investigated an existing proposal and applied it to IP networks
- Implemented a computationally efficient model
- We modelled the superposition of independent sources as a Markov Modulated Poisson Process (MMPP)
- Measured the loss probability through a finite buffer (for different loads and buffer sizes)
- Compared the model, simulation *and a* laboratory setup
- Quite unusual to try all three approaches

Loss probabilities of different loads & buffer sizes

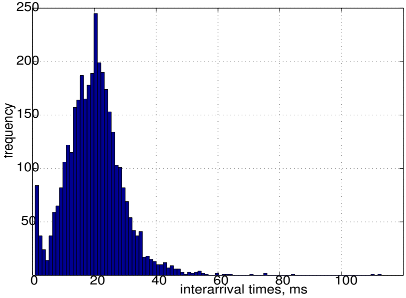


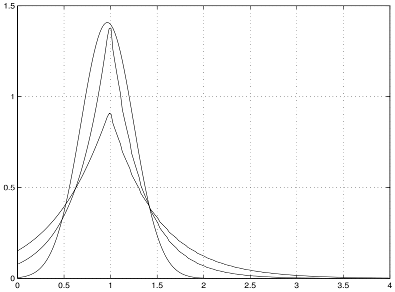
$$Load\ (\rho) = \frac{No.\ sources \quad \times \quad Sourceon_{prob} \quad \times \quad Rate_{peak}}{Link\ \ capacity}$$

Mixing VoIP and data traffic



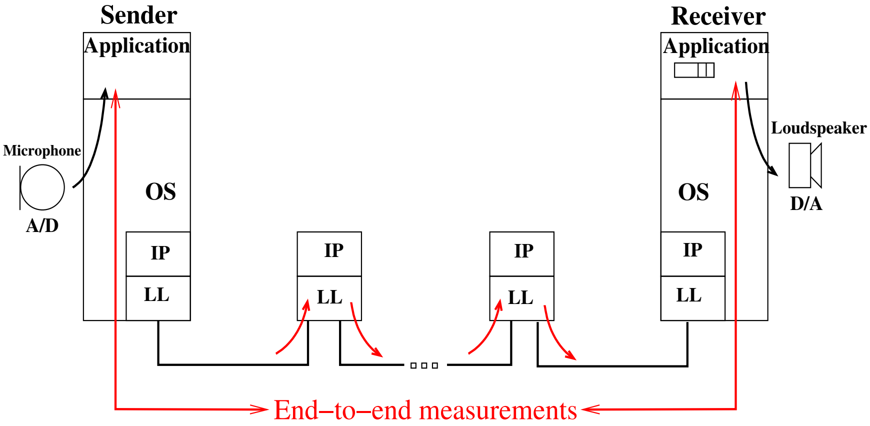
Measured and modelled VoIP interarrival histograms



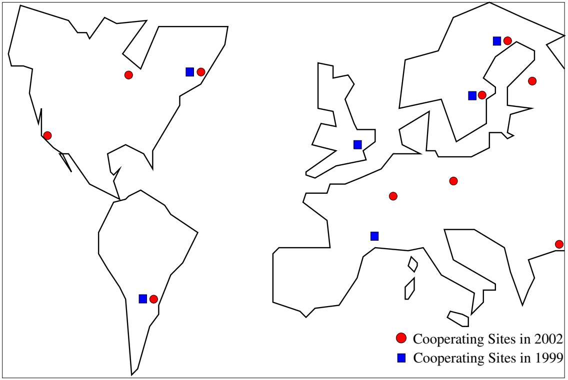


- Left histogram is a (typical) VoIP session from Argentina to Stockholm (packetisation interval 20ms)
- Right plot is 3 choices of random network delays (2 Exponentials, 1 Gaussian)
- $x = 0$ (origin) represents packets back to back
  $x = 1$ (maximum) at the packetisation interval and for $x > 1$ (tail) dispersed

Measuring wide-area VoIP quality



Co-operating test sites used in 1999 and 2002



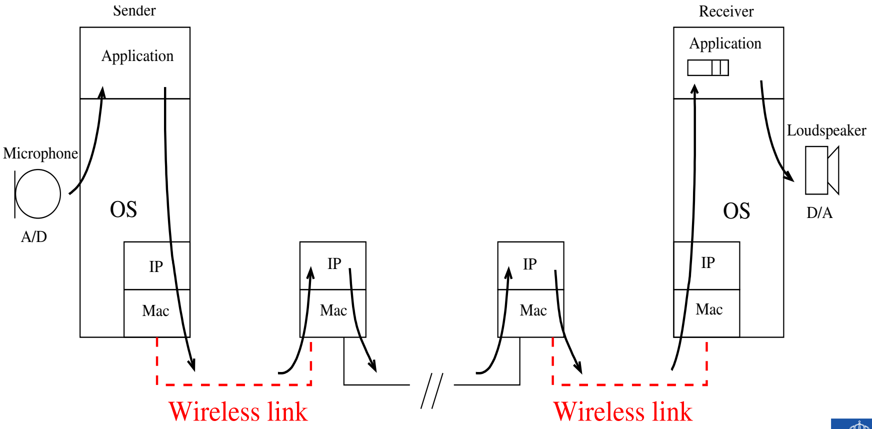
An active measurement method was used, a 70 second pre-recorded conversation between the marked sites was sent once/hour

Measurement goals and brief results

Comprehensive loss, delay and jitter assessment of VoIP in 2002

- Some indication of the trend by comparing the results with those taken in 1999
  - Improvements from 1999's measurements: full-mesh topology, totally automatic invocation and more hosts (next slide)
  - Also investigated some new items: asymmetry issues, time of day effects, different packet sizes and silence suppression
- Collected a repository of VoIP sample sessions (approx. 22,500 calls)
- Most calls $<$ 2% loss & 10 ms jitter (delay is location dependent)
- Quality has slightly improved over the past three years
- VoIP still not usable on a global scale, two sites showed poor quality
- Infra-structure not distance (or the number of hops) is important
- Hundreds of downloads, at least six publications from our data

Moving onto wireless access...



Wireless measurements

Now we look at IEEE 802.11b access, and again use active measurements of the channel. In particular we look at the MAC layer behaviour, environment and the role of competing traffic.

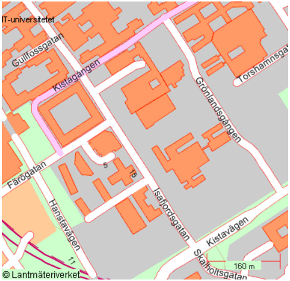
- Pure distance effects using line-of-sight between a single sender and receiver (outside)
- Distance effects with line-of-sight (same premises)
- Distance effects without line-of-sight (same premises)
- Competing traffic effect in ad-hoc mode (same room)
- Competing traffic effect in infra-structure mode (same room)
- IEEE specific bitrate selection, RTS/CTS (same premises)

SWEDISH INSTITUTE OF COMPUTER SCIENCE SICS

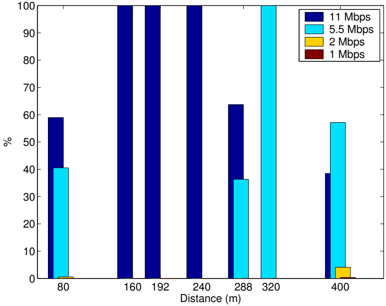
KTH VETENSKAP OCH KONST

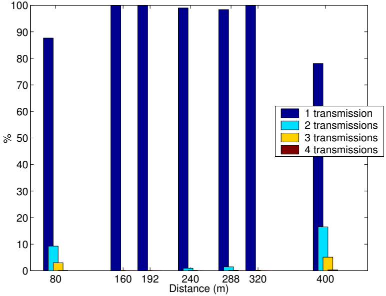
Line-of-sight setup

We took eight measurements at distances from 80 to 400 meters We recorded the loss, delay, jitter, bitrate & no. of MAC transmissions The sender was stationary and the receiver moved



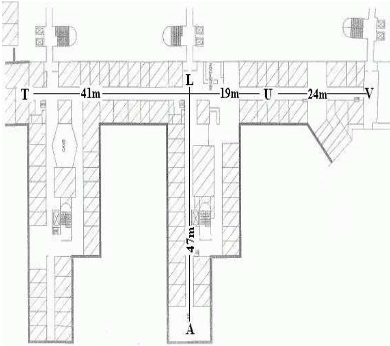
Line-of-sight results





Left histogram shows the rates at which each frame was sent, the right one is the number of retransmissions at each distance. VoIP quality is good.

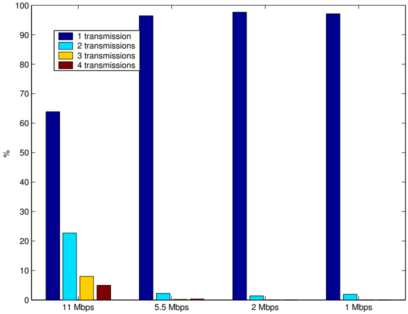
Line-of-sight in the office I



Locations	fraction of losses (%)	round-trip time (ms)	jitter (ms)
$A->L$ (47m)	[0.0, 0.0, 0.0]	[1.9, 2.0, 2.2]	[0.1, 0.1, 0.2]
$T->U$ (60m)	[0.0, 0.2, 0.9]	[1.9, 2.0, 2.9]	[0.2, 0.4, 0.9]
$T->V$ (84m)	[0.0, 0.4, 1.4]	[1.8, 2.2, 3.5]	[0.1, 0.2, 0.5]

Non line-of-sight office results

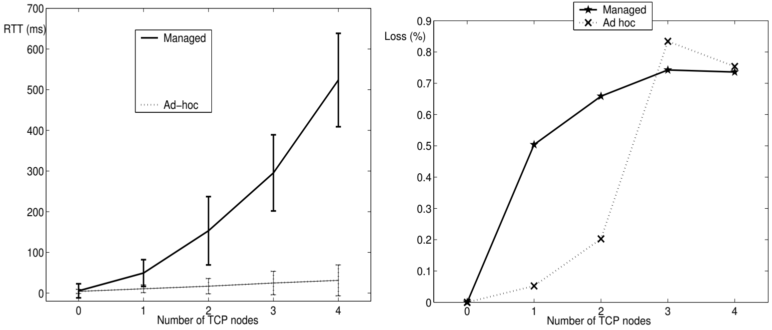
In this setup, the receiver is around a corner. This time we fixed the maximum bitrate and observed the rates selected.



5.5 Mbits in this case is a reasonable choice

Voice quality and competing TCP traffic

If we now look at competing TCP traffic and its implications in both ad-hoc and managed modes

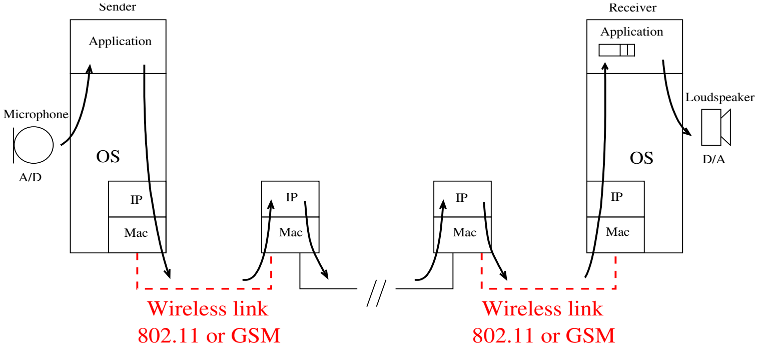


Left plot is the delay & the right plot the loss, the delay is very high

Conclusions for real-time voice and 802.11b access

- Generally good quality
- 802.11b performs better than we expected (or reported)
- However even line-of-sight vulnerable to interference
- Loss is generally caused by intervening obstacles
- Delay and jitter arise from competing traffic
- However, problems remain: low SNRs, obstacles & heavy load are some that we observed
- Access point can "add" high delays (scheduling & queueing)
- Collated approximately 8 Gig of data (application and MAC)
- Used information from both layers wherever possible

802.11-based voice with alternative access



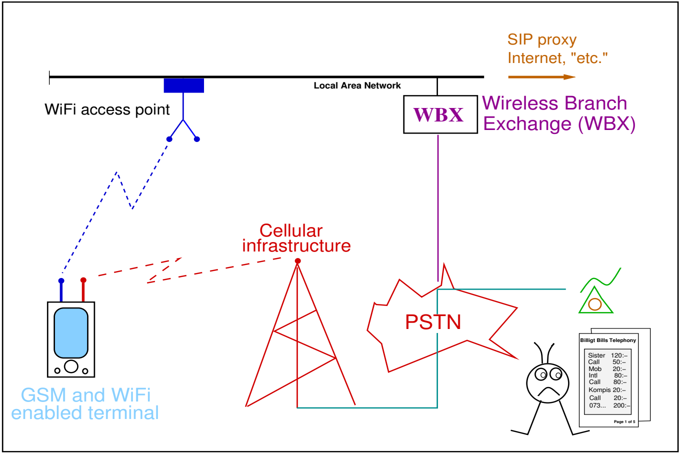
- Use WiFi network for voice whilst in "good" conditions and switch to the cellular network when the quality is not
- Switch with minimum disruption to the user

SWEDISH INSTITUTE OF COMPUTER SCIENCE SICS

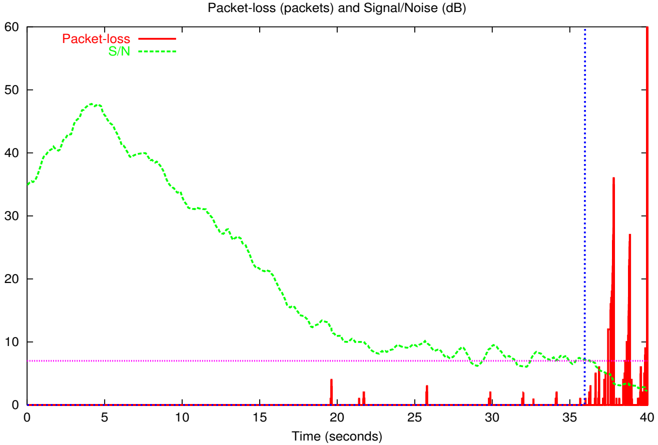
Quality prediction problems

- The problem is the 5 seconds setup to the PSTN cellular network
- Therefore some *prediction* is needed in the WiFi network
- One needs to initiate a handover to the network before poor quality is experienced
- But also not to confuse this with short-term audible (but tolerable) glitches
- Need to handle leaving and entering WiFi network
- Also don't want to flip between the two networks

Handover infra-structure



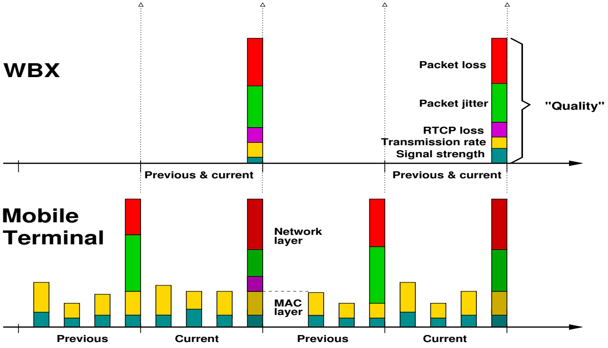
Loss and signal to noise ratio



Explanation

- The magenta horizontal line shows the SNR, it is acceptable but could mislead to waiting too long before a handover
- The blue vertical line shows the start of loss that would be unacceptable, but it is too late to initiate a successful handover
- So one has to combine the parameters and weigh them in a spam filter-type approach
- Also it is clear degrading quality over consecutive time intervals is more serious than ones in a single interval

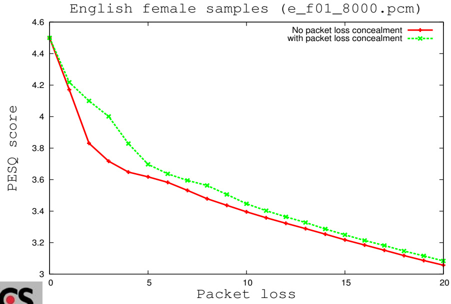
Combining quality and link-layer parameters



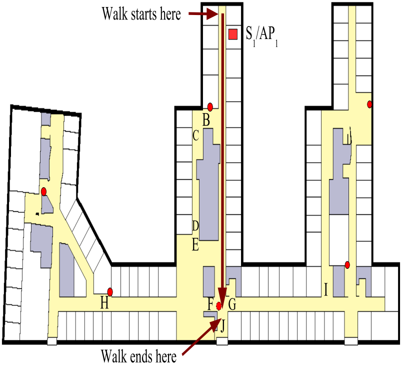
Component	interval 1	interval 2	both intervals
Loss ($> 4\%$)	-1	-1	-1
Jitter ($> 85ms$)	-1	-1	-1
SNR ($> -70db$)	-	+1	-
SNR ($< -90db$) & rate ($< 2Mbits$)	-	-1	-
Rate ($> 3Mbits$)	-	+2	-
2 lost messages	-	-	-1
Total	$\sum X$	$\sum Y$	≥ 0 do handover

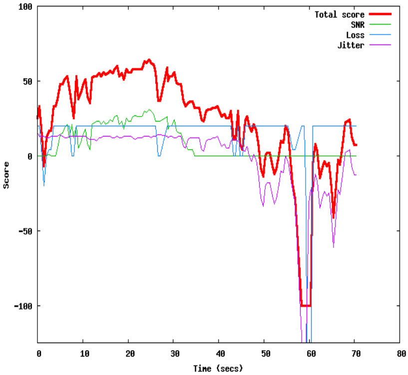
Small diversion - human tolerance to loss

- How much loss do we know is acceptable?
- Well, the ITU have developed a psychoacoustic model for human speech quality evaluation called PESQ
- No need for human listeners, the model estimates the quality close to a humans rating 4.5 (excellent) to 1 (very poor)
- 10 losses seem to be approximately a lowering of one rating



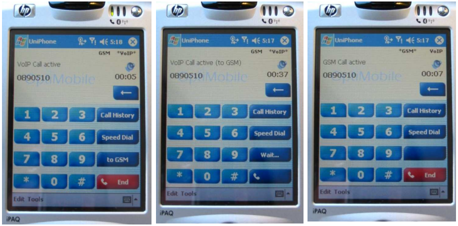
Test example



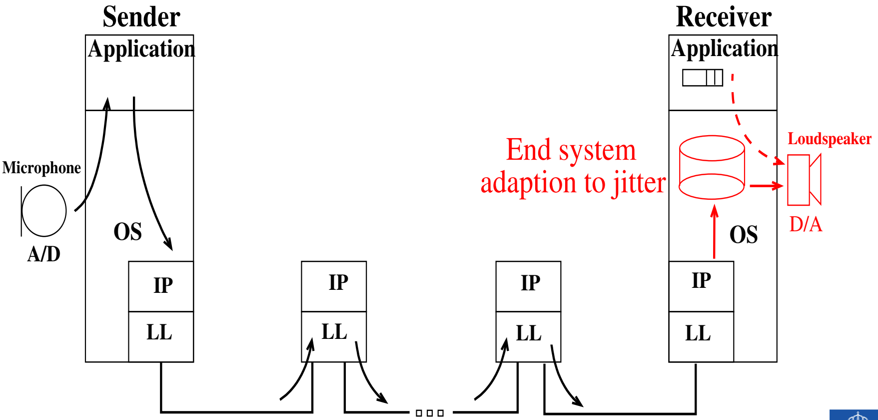


Demonstration

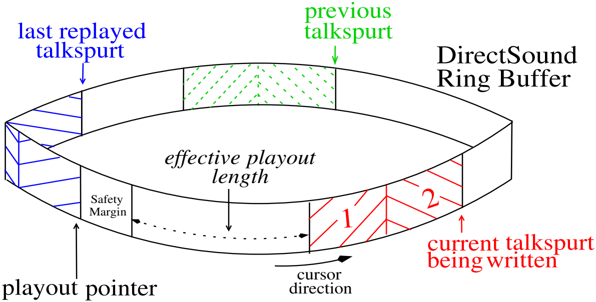
- We have tested with a single cell with one access point (AP)
- By walking away from the AP we decrease the reception signal & and hence the communication quality
- An estimation of the quality is done in the handset and also sent to the WBX
- Using the score system described earlier it will initiate a handover when the quality falls below acceptable levels



End-system adaption to network jitter



Directsound playout support



- Implemented using DirectX interface by Microsoft
- Circular buffer, pointers rotate anti-clockwise
- Talkspurts written contiguously, adapt buffer length during silence periods

Mouth-to-ear delay measurements

Audio Tool	Latency (ms)
Sicsophone prototype	25-100
Vocal Internet Phone 4.5	450-550
NetMeeting 2.1	620
VAT 3.4 (Solaris)	1200
RAT 3 (Solaris)	1500

- Windows (98, NT) operating systems with SoundBlaster audio cards
- Simple square wave used rather than a speech sample for easier triggering and delay calculations
- Point here is to show the end-systems should not be neglected, and one can change them relatively easily (unlike the network!)

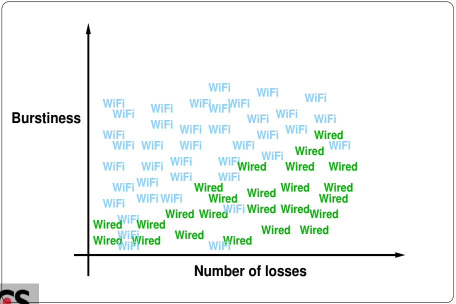
Voice quality estimation

Estimating the quality of telephony streams over IP networks in real-time is useful for:

1. Reporting quality back to provider (application writer too)
2. Calls do not need to be continued and use valuable resources
3. Give the user some feedback what is happening (is it the network, computer or me?)
4. Not easy to deduce the quality when using different codecs
5. Path characteristics during and after the call
6. Skype also allows some human feedback
7. Help determining a handover point (more later)

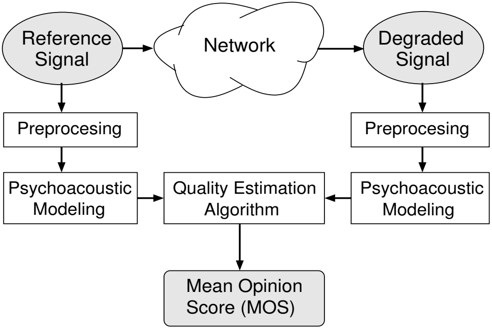
Basic technique: Off line precalculation

- The idea is pre-calculate the most common loss patterns
- Using this loss information we can "calibrate" the receiver
- And interpolate values that we didn't calculate
- Possible to "parameterise" the receiver for different codecs
- The secret is to choose the right combination



How to do it: PESQ

- Perceptual Estimation of Speech Quality (PESQ)
- PESQ compares a reference signal to a degraded one
- Gives a human-understandable rating (MOS)



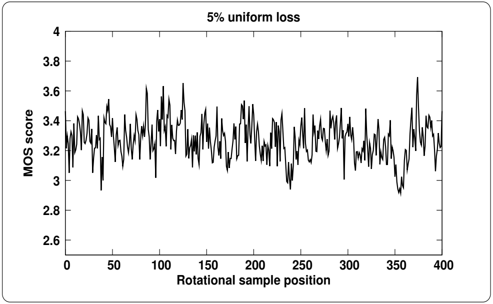
ITU's PESQ standard in words and numbers

PESQ MOS	Linguistic equivalent	Quality degradation
4.5	Excellent	None
4	Good	
3.5	Good/Fair	"Moderate"
3	Fair	
2.5	Fair/Poor	"Severe"
2	Poor	
1	Bad	

- A MOS score of ≥ 0.5 is usually audible
- 0.5 scores are not in the ITU standard (we added them)
- We also added an extra column, quality degradation

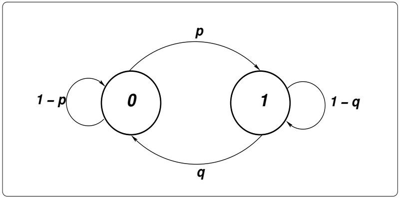
Uniform losses

- Losses are from a uniform distribution from 1% to 50%
- Losses are applied to 8 second ITU standard samples
- We rotated the loss pattern one packet at a time
- With 5% loss, the PESQ scores can vary up to 0.7 MOS points (below)



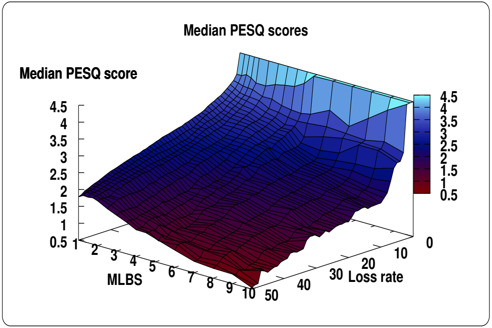
Bursty losses

- We consider losses up to 50% and bursts up to 10 packets
- Basically it covers the wireless network scenario
- Losses are generated according to a Gilbert model
- p is the *uncorrelated* loss and 1-q the *correlated* losses



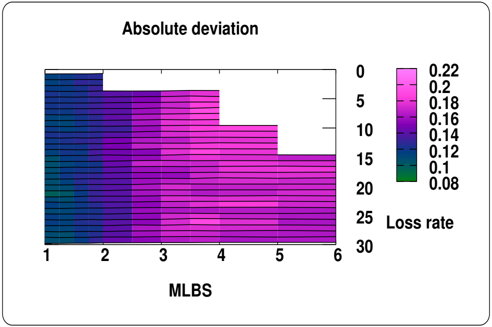
Bursty losses I

- The loss rate (LR) is the major factor in degrading the quality
- Burstiness is also significant (shown as MLBS)
- There is an initial fall in quality then flattens somewhat
- Shown below are the results without packet loss concealment



Absolute errors

If one uses the median, then some errors exist:

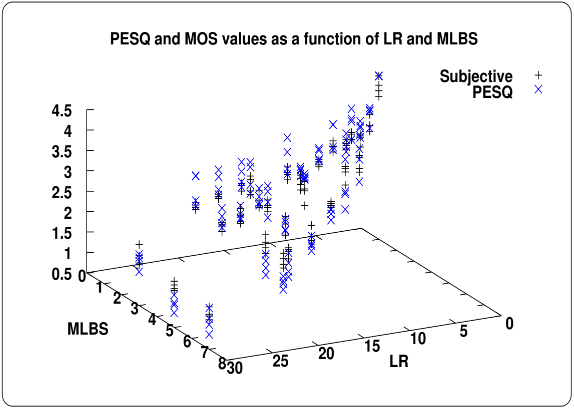


Subjective assessment

A P.800 based test (ITU designed) was carried out with

- 11 subjects
- 168 samples, corresponding to 42 configurations
- We chose a 9–point scale.
- Anchors were used during the warm–up session.
- We compared PESQ to our subjective scores
- Correlation was quite high, given the losses considered (0.86)
- PESQ's estimation did not degrade at the same rate as the network conditions degraded
- It also tends to over–estimate when the burstiness is low, and under–estimate when high
- Our estimation is robust & close to human perception

PESQ and subjective scores



We also compared the scores with P.563 (single sided metric)

Wireless VoIP challenges

- WiFi voice is not yet a true competitor to the cellular network
- Lack of true handsets (currently only PDAs), also people *like* their phones as fashion accessories, cameras etc.
- Security issues exist WiFi, WEP is known to be broken
- Quality issues related too:
  - PDAs are not voice optimised
  - WiFi primarily a data communication technology
  - Quality related to the environment, walls, windows etc.
  - Use of unlicensed spectrum (2.4 Ghz)
  - Other interfering devices: access points, users on the same channel, Bluetooth devices, microwaves ovens and so on
  - User mobility is unpredictable

Voice quality conclusions

- We have looked at the complete system in separate studies
- We do not assume any network QoS support
- Simply "how is the system performing today?"
- Good quality is possible on todays Internet (as you know)
- Operators want isolate VoIP traffic or give it higher priority
- We found well-provisioned links are almost as satisfactory
- Certain infrastructures need upgrading for reasonable VoIP
- End-systems should not be ignored

Short summary of publications

Research & prototypes of VoIP systems over 5 years

- 2001: Started with call quality and link dimensioning
- 2003: Moved onto the end-systems, performance issues
- 2003: Investigated WAN measurements, backbone (1999 too)
- 2003: Some modelling of network delays and VoIP
- 2003: Prediction aspects: can we predict poor quality calls
- 2003: Licentiate thesis
- 2004: Moved to wireless, 802.11 measurements
- 2005: Handovers from 802.11 to cellular networks (ongoing)
- 2006: Estimating subjective voice quality using ITU methods
- 2008: Wrote PhD thesis
- 2009: Defence :-)

SWEDISH INSTITUTE OF COMPUTER SCIENCE SICS

KTH VETENSKAP OCH KONST

Some obvious things...

Research issues exist from both the operator & end-user perspectives:

- Operator viewpoint
  - Multiservice networks will happen, separate networks are costly for operators
  - Voice still is the biggest revenue earner for Telco's, so VoIP needs accommodating in the multi-service the network
  - Some issues exist here
- End-user viewpoint
  - Internet Telephony needs wireless access (802.11x or WiMAX) to seriously compete with the cellular systems
  - More subjective assessment is needed in the VoIP domain
  - Loss, delay and jitter are not really sufficient in order to get a *real* human understanding

Current(ish) research issues

- Measurements on latest technologies (WiMAX)
- PESQ-like assessment on incoming calls
- Improve handover performance
  - Prediction from noisy signals, robustness of parameters
  - Time series analysis, e.g. Kalman filter
- Wide-scale deployment of IP telephony
  - Wired (mass migration to IP)
  - Wireless (fon like, DECT wide-range)
- 3D telephony
  - Better spatial separation
  - Higher quality interactions
  - Tools for separating background noise