3G Services Present New Challenges For Network Performance Evaluation
Outline

• Synopsis of speech, audio, and video quality evaluation metrics
• Performance evaluation challenges related to 3G multimedia applications
• Models for multimedia quality evaluation metrics
• Solutions to be implemented in 3G field test equipment
• TEMS view on the multimedia quality evaluation metrics in real time and autonomous tools
• Status of the development of multimedia quality evaluation metrics
• Conclusions
Synopsis of speech, audio and video quality evaluation metrics

- Three QoS metrics mainly define subscriber perception of network quality: Speech, Audio and Video
- Quality of speech is perceived as satisfactory in 2G networks, as is the audio-video on TV broadcast (HDTV) networks
- New challenges are emerging from 3G networks
Synopsis (cont.)

New Challenges emerging for 3G networks

• Voice and data convergence
• Coexistence of PS and CS core networks
• Implementation of a large set of applications
  – divided into quality classes and requiring different priorities and quality levels
    • conversational: video conferencing; voice/speech
    • streaming: video and audio
    • interactive: games, e-commerce
    • background: e-mail, ftp
• Eventually, the deployment of the Next Generation Network (NGN)
  • wireless and wireline networks converged in an IP-based multi-service next generation network (NGN)
Types of quality evaluation metrics to be considered

- **Subjective:**
  - subscribers’ opinion on the network delivered services
  - measured by the means of off-line listening tests performed in subjective testing labs and being therefore time and cost consuming

- **Objective:**
  - algorithms that provide an estimate of the subjective opinion
  - accuracy expressed by statistical metrics, such as correlation coefficient and prediction error, and reflecting the objective estimator’s error in regard to the subjective opinion
  - implemented in field test equipment
Objective quality evaluation metrics

- Model (generally defining the metric’s accuracy)
  - Perceptual (based on models of human perception of video/audio/speech)
  - Non-perceptual (based on network’s parameters)

- Method (generally defining the appropriate field tool type for performing the test measurements)
  - Intrusive:
    - uses test stimuli, such as speech samples, video clips,
    - ensures high accuracy (high correlation and low prediction error in regard to the subscriber’s opinion)
    - adequate for off-line processing tools and/or autonomous tools
  - non-intrusive:
    - uses in service signals requiring the estimation of the source (originally sent) signal
    - lower accuracy if compared to the intrusive method
    - adequate for real time monitoring tools
Synopsis (cont.)

Objective quality evaluation metrics

• Performance evaluation methods for the objective quality evaluation metrics
  – statistical metrics such as correlation coefficient and prediction error that compare objective and subjective scores;
  – statistical metrics’ values define the performance of the objective quality evaluation metric
Synopsis (cont.)

**Speech: listening and conversational** (considers double-talk and echo effects)

- **Perceptual metrics**
  - Intrusive: PESQ, ITU-T (Telephony Sector) standard (Feb. 2001)
    - Listening Subjective Opinion Score and speech diagnosis
    - Listening Subjective Opinion Score and speech diagnosis

- **Non-perceptual (network parameter based metrics)**
  - Non-intrusive
    - Speech Quality Index SQI (RF parameter based), Ericsson - TEMS
      - Listening Subjective Opinion Score and network’s air interface diagnosis
    - Call Clarity Index CCI (network parameters based), ITU-T standard (Feb 2002)
      - Conversational Subjective Opinion Score and fixed CS network’s diagnosis
      - Listening Subjective Opinion Score and IP network’s diagnosis
Synopsis (cont.)

Audio-TV broadcast (music, non-speech signals)

- Perceptual
    - tested and validated for applications such as music codecs evaluation, network planning;
    - not validated for transmission errors characteristic to 2.5G and 3G networks
Synopsis (cont.)

**Video-TV broadcast (HDTV)**

- **Perceptual metrics**
    - a set of four algorithms recommended (no statistical differences between the algorithms’ performance)
      - algorithms that measure specific artifacts introduced into the video as perceived by a human viewer
  - Non-intrusive:
    - under development within a joint group based on the Video Quality Expert Group VQEG, ITU-T-SG12 and ITU-R 6Q groups
    - Standard expected beginning of 2005

- **Non-perceptual metrics**
  - Intrusive: ANSI T1.801.03, NTIA (1996)
    - a set of objective metrics that serve as basis for the video quality measurement
Performance evaluation challenges related to 3G multimedia applications

• Multimedia quality evaluation
  – complex subjective environment that require a set of quality evaluation metrics:
    • video, audio and respectively the combined video-audio (or equivalently multimedia) metrics
    • video and audio components contribute with different weights to the total multimedia metric
    • each metric is defined by different subjective aspects and characterized therefore by different accuracies
Performance evaluation challenges (cont.)

Challenges characteristic to the algorithms’ development

- Subjective aspects that define the proper development, testing and evaluation of the objective metrics
  - video formats: the Common Intermediate Format (CIF: 352 lines x 288 pixels) and Quarter CIF (QCIF: 176 lines x 144 pixels) for PC and respectively for mobile phones and PDAs
  - the large variety of displays for PDAs and mobiles and their continuous development
  - the type of subjective tests that are required for multimedia types of video formats, displays (PDA, mobile) and characteristic conditions (packet loss, rebuffering)
- Absolute/ degraded, continuous or fixed subjective scale (ITU-T P.910, P.911, P.920 standards)
Challenges characteristic to the algorithms’ development

- Objective aspects that define the algorithms’ accuracy
  - the low and very low rate video codecs
    - Mobile: 16kb/sec-320kb/s, could be starting as low as 1.5 kb/sec;
    - PC-CIF: 128kb/s-704kb/s;
    - PC- ITU-R Rec.601: 320kb/s-2Mb/s;
  - the large variety of video codecs, such as:
    - H.264 (MPEG4), H.263 (3GPP), Windows Media Video 9, RealPlayer
  - large variety of video contents: sport, news, movies, games
  - the registration procedure (temporal and spatial synchronization) of the video sequences
    - synchronization procedures required to synchronize audio and video (lips sync) for video-audio quality evaluation
Performance evaluation challenges (cont.)

Challenges characteristic to the algorithms’ implementation

- Capturing procedures for video and audio
- CPU capacity and memory, especially for intrusive metrics that require the storage of the reference and the test sample
- Hypothetical reference conditions (HRC) or degradation conditions characteristic to the 3G networks
  - packet loss and delay: variable and with a large range (mainly characteristic to video conference application)
  - frame errors (mainly characteristic to video conference application)
  - frame rates: variable and large range (2.5....30 fps)
  - rebuffering (mainly characteristic to streaming application)
Models for multimedia quality evaluation metrics

- Based on network’s parameters (RF, IP) (falling into the so-called “In service Non-intrusive Measurement Devices” category)
  - non-video based
  - non intrusive
  - straightforward network trouble shooting and network diagnosis
  - limited accuracy
  - arise some difficulties when it comes to mapping to the subjective domain
  - adequate for multimedia quality evaluation in real time field test tools

- Based on objective measures calculated from the video signal features (reference and processed)
  - video signal based
  - intrusive type
  - limited accuracy
  - might be less difficult to map to the subjective domain, but however requiring extended subjective tests
  - adequate for video quality evaluation in autonomous test tools
Models for multimedia quality evaluation metrics (cont.)

- objective metrics/parameters
  - PeakSNR, SNR, RMSE, colour PeakSNR (also known as fidelity metrics)
  - edge energy difference, motion energy differences, spatial frequency difference, repeated frames (also known as ANSI metrics)

- Based on video parameters as perceived by the viewers (called perceptual metrics)
  - video signal based
  - non-intrusive and/or intrusive type
  - more accurate
  - convenient mapping to the MOS domain
  - adequate for video quality evaluation in autonomous test tools
  - video parameters (such as jerkiness, block distortion, blurriness, noise and colourfulness), which measure video artifacts as perceived by the viewer
Examples of artifacts as perceived by the viewer

- **Blurriness**
  - **Defined as:** loss of fine detail and the smearing of edges in the video
  - **Causes:** high-frequency attenuation at some stage of the recording or encoding process, transmission errors, packet loss, low-pass filtering
  - Measured by the Lost Edge Energy Parameter belonging to the Spatial Information (SI) video metric (ANSI standard)
    - for the input (white, bottom left) and output image (white, bottom right)
Examples of artifacts as perceived by the viewer

- **Block Distortion** (blockiness, tiling effects)
  - **Defined as**: distortion of the image characterized by the appearance of an underlying block encoding structure
  - **Causes**: coarse quantization of the spatial frequency components (common to the DCT compression); packet loss, transmission errors, which often can affect entire blocks in the video.
  - The output image (right) contains both tiling and blurring impairments
  - Tiling adds horizontal and vertical spatial information (i.e., the output has more spatial information along the horizontal and vertical SI axis than the input – red arrow)
  - Blurring causes a loss of diagonal spatial information (i.e., the output has less spatial information along a diagonal direction, such as $\alpha = 45$ degrees, than the input – green arrow).
Examples of artifacts as perceived by the viewer

- Another form of block distortion: one or more blocks in the image bear no resemblance to the current or previous scene and often contrast greatly with adjacent blocks.
Models (cont.)

Examples of artifacts as perceived by the viewer

• Noise
  – **Defined as:** the presence of pixels in a processed video, whose content deviates from the original video, most noticeably in smooth regions and around edges (edge noise).
  – **Causes:** noisy recording equipment, compression process or transmission errors
  – The random noise manifests itself as "snow" in the output temporal information TI video metric. The squared values of the TI video metric ($TI^2$) can be used for identifying noise
Models (cont.)

Examples of artifacts as perceived by the viewer

- **Jerkiness**
  - **Defined as**: motion that does not look smooth; the human vision perception of originally continuous motion as a sequence of distinct "snapshots."
  - **Cause**: skipped frames to reduce the amount of video information, packet loss, reduced frame rate, network congestion

- **Colorfulness**
  - **Defined as**: intensity of saturation of colours as well as the spread and distribution of individual colours in the video.
Solutions to be implemented in 3G field test equipment

- Measurements within 3G networks require
  - complex and reliable signaling and synchronization procedures;
  - accurate handling (processing and storing) of large amount of collected data

- Criteria for the implementation solution of quality evaluation metrics into field test tools depends on the
  - equipment’s type (such as real time or autonomous monitoring);
  - metric’s type (speech, video, audio);
    - speech metric is more prone to degradations than audio and video from human perception point of view
  - application type (network development/design, end to end performance evaluation, optimization)
    - the development phase is less demanding on the quality metric’s accuracy than the evaluation and optimization phase
The following table presents some solutions that consider the criteria on the previous slide

<table>
<thead>
<tr>
<th>Field test tool’s type</th>
<th>Tool’s demands</th>
<th>Recommended solution</th>
<th>Recommended application</th>
<th>Quality metric’s type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real time</td>
<td>- CPU capacity and memory</td>
<td>Non-intrusive and non-perceptual</td>
<td>Network development/design</td>
<td>Speech, video, audio</td>
</tr>
<tr>
<td></td>
<td>- Acceptable lower accuracy of the quality metric</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-intrusive and perceptual</td>
<td>E2E performance evaluation, network optimization</td>
<td>Video, audio</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intrusive and perceptual</td>
<td>E2E performance evaluation, network optimization</td>
<td>Speech, audio video</td>
</tr>
<tr>
<td>Autonomous, Off-line</td>
<td>Quality metric’s accuracy</td>
<td>Intrusive and perceptual</td>
<td>E2E performance evaluation, network optimization</td>
<td>Speech, audio, video</td>
</tr>
</tbody>
</table>
TEMs view on the multimedia quality evaluation metrics in real time and autonomous tools

• TEMS products’ approach
  – real time test tools and autonomous monitoring tools
    • TEMS Investigation 2.3 for CDMA2000 (cellular and PCS) and for CDMA450
    • TEMS Automatic 3.2 for CDMA2000 (cellular and PCS)
  – implementation of main quality metrics that reflect subscriber’s opinion on the 3G network’s performance
    • Speech quality evaluation metric and speech diagnosis
      – Non-intrusive and non-perceptual (network’s parameter based) metric: Speech Quality Index
      – Intrusive and perceptual metric: PESQ
      – Implemented in TEMS Investigation and TEMS Automatic
    • Multimedia (audio-video) quality evaluation metric: under development
      – Non-intrusive and non-perceptual metric
    • Video streaming quality evaluation metric: under development (participation into the standardization work going on currently)
      – Intrusive and perceptual metric
TEMS view on the multimedia quality evaluation metrics in real time and autonomous tools (cont.)

TEMS Investigation CDMA

Real-time solution for monitoring, troubleshooting, and optimization of CDMA networks.

• Network roll-out
  – RF planning
  – Installation & Commission
  – Verification
  – Optimization

• Network operations
  – Professional troubleshooting
  – Network tuning
  – Network expansion

• Supports CDMA2000 (cellular and PCS) and for CDMA450
TEMS view on the multimedia quality evaluation metrics in real time and autonomous tools (cont.)

**TEMS Investigation CDMA**

**The Operator’s Challenge**

- Identify disruptions in network quality
- Increase customer satisfaction
- Reduce subscriber churn
- Handle both air interface and data testing efficiently and accurately

**TEMS Investigation is the answer**

- Combines air interface testing with data testing
- Identifies what the network problems are, and why they occur
TEMs view on the multimedia quality evaluation metrics in real time and autonomous tools (cont.)

TEMS Investigation CDMA

Strengths and Benefits

- Advanced troubleshooting tool
- Network supplier independent
- Used throughout the network life cycle
- User-configurable
- Provides simple test and simulation
- Tests from a user’s perspective
Autonomous CDMA network monitoring system

- Mobile Test Units collect data through the whole network
- True Quality of Service information
- Support for CDMA2000 (cellular and PCS)

Use TEMS Automatic to

- Increase revenue
- Increase efficiency
- Optimize investments
TEMS view on the multimedia quality evaluation metrics in real time and autonomous tools (cont.)

TEMS Automatic

• The Market Situation Today
  – Saturated markets
  – High churn
  – Declining revenue
  – Low data usage

TEMS Automatic provides continuous feedback on network QoS to improve customer satisfaction
TEMS view on the multimedia quality evaluation metrics in real time and autonomous tools (cont.)

TEMS Automatic

TEMS Automatic – System Overview

- TEMS Automatic, Fixed MTUs
- TEMS Automatic, Mobile Test Units (MTUs)
- TEMS Automatic, Data Test Node
- TEMS Automatic, Voice Test Node
- TEMS Automatic, Fixed System
- GSM/GPRS, TDMA, CDMA
- PSTN
- Internet

TEMS Automatic, Report and Presentation
TEM5 view on the multimedia quality evaluation metrics in real time and autonomous tools (cont.)

Preliminary results of a multimedia streaming quality evaluation metric - Subjective perspective

• Multimedia quality depends on the video clip content
  – Audio and video quality components contribute with different weights, but follow similar trends.
  • Clips 1-14: movie
  • Clips 15-28: sports
  • Clips 29-42: news
TEMPS view on the multimedia quality evaluation metrics in real time and autonomous tools (cont.)

Preliminary results of a multimedia streaming quality evaluation metric - Subjective perspective

- Multimedia streaming quality, especially the video quality, is strongly dependent on the bandwidth
  - Quality is very sensitivity to bandwidth changes below 300kbps (QCIF format)
  - Quality variation of more than 2MOS units for a constant, but low (below 64kbps) bandwidth value.
TEM5 view on the multimedia quality evaluation metrics in real time and autonomous tools (cont.)

Preliminary results of a multimedia streaming quality evaluation metric - Subjective perspective

- Multimedia quality is dependent on the channel errors
  - Video component is more sensitive than the audio component to channel errors

Channels
- -3 and -2: reference channels
- -1: clean channel at 64kbps
- 0-4: channels at 64kbps with BLER of 2,5,8,10,13%
- 5: 32kbps reference channel
TEMs view on the multimedia quality evaluation metrics in real time and autonomous tools (cont.)

Preliminary results of a multimedia streaming quality evaluation metric - Subjective perspective

- Video and audio quality metrics respond differently to channel errors depending on the video clips’ content
TEMIS view on the multimedia quality evaluation metrics in real time and autonomous tools (cont.)

Preliminary results of a multimedia streaming quality evaluation metric - Objective perspective of a non-intrusive and non-perceptual quality metric

• Two categories of variables
  – Transport layer variables such as throughput, packet size, delay
  – Streaming client statistics such as frame rate, missed frames, stream rate, jitter

• In each category, the variables contribute with different weights to the multimedia quality metric.

• Estimator of the multimedia quality: function of weighted variables belonging to both categories
Status of the development of the multimedia quality evaluation metrics

• Current available standards
  – Intrusive non-perceptual: NTIA ANSI
  – Intrusive perceptual (HDTV): ITU-R J.144 accompanied by J.145 (calibration functions; mappings to the subjective domain)

• Standardization activity (going on)
  – Non-intrusive perceptual for HDTV
  – Intrusive / non-intrusive perceptual for video streaming only
    • Audio streaming and respectively multimedia streaming for later phases of the standard’s development

• Standardization bodies:
  – Video Quality Expert Group
  – ITU-T SG 12 (E2E transmission performance of networks and terminals) and SG9 (Integrated broadband cable networks and TV and sound transmission)
  – ITU-R 6Q (Objective picture quality parameters and associated measurement and monitoring methods for television images)
Conclusions

• 3G roll out keeps the human perceived speech, audio and video quality in the driver seat
• Multimedia quality evaluation arises performance challenges such as: codecs diversity, low and adaptive rates, video formats (mobile and PDA, QCIF format), degradation types (rebuffering)
• Perceptual and network’s parameters based models developed to suit intrusive and non-intrusive measurements methods
• Criteria for the implementation solution of quality evaluation metrics into field test tools depends
  – on the equipment’s type (such as real time or autonomous monitoring)
  – metric’s type (speech, video, audio);
  – application type (network development/design, end to end performance evaluation, optimization)
Conclusions

- TEMS products implement quality metrics that reflect subscriber’s opinion on the 3G network’s performance
  - Speech quality evaluation metric and speech diagnosis
    - Non-intrusive and non-perceptual (network’s parameter based) metric: Speech Quality index
    - Intrusive and perceptual metric: PESQ
    - Implemented in TEMS Investigation and TEMS Automatic
  - Multimedia (audio-video) quality evaluation metric: under development
    - Non-intrusive and non-perceptual metric
      » Metric based on transport parameters and streaming client statistics
  - Video streaming quality evaluation metric: under development
    (participation into the standardization work going on currently)
    - Intrusive and perceptual metric