Z24: Usability of Multimedia

Mark Handley

(based on work by M. Angela Sasse)

Overview

- Background: Internet Videoconferencing
- Quality Assessment
  - Task performance and subjective assessment
  - Physiological responses
- Users and QoS parameters

Background

- Internet research at UCL Computer Science since 1973
- IP Videoconferencing since 1992 (MICE, ReLaTe)
  - Suffering from broken-up audio, low framerates
- Much research work on conferencing tool development:
  - rat, vic, nte, sdr
- Humans don't always respond in predictable ways:
  - Need evaluation of which things *actually work.*

A/V Quality Assessment

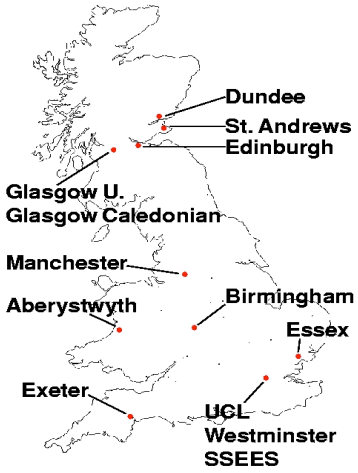
- Telecommunications
  - ITU 5-point scales
  - Mean Opinion Score (MOS)
  - Short clips, no relation to task
  - "spot the degradation"

Quality of the speech/ Score	
Excellent	**5**
Good	**4**
Fair	**3**
Poor	**2**
Bad	**1**

- HCI-based assessment approach
  - Real users in context
  - Real tasks/materials
  - Repeated trials
  - Is quality adequate/comfortable for the task?

PIPVIC-2 Trials

- Piloting IP Videoconferencing over UK Academic Network
- 13 partner sites
- Jan. - Sept. 1999
- multicast (rat, vic, wb, nte)
- teaching, research, admin
- lectures and tutorials :
  - languages
  - medical
  - sociology
  - business studies
  - history
  - computer science



Trial results

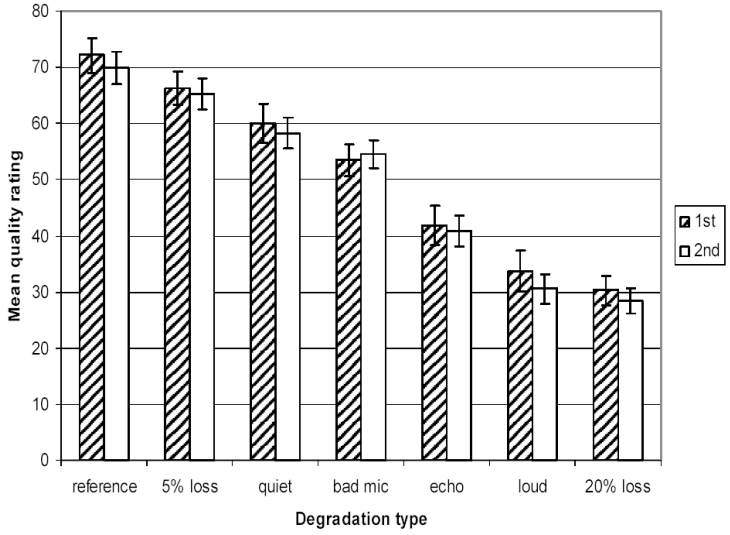
- Objective and subjective quality were collected for all the multimedia sessions
  - 1/3 of session reports rated that the audio quality was less than adequate
  - but *objective* quality (network and tool stats) good (< 5% packet loss) throughout sessions
- Qualitative data and recordings indicate that the session setup and usage are the main cause of dissatisfaction.

Experimental study

- Compare network with other packet loss (repaired with packet repetition)
  - *Reference (no loss), 5% loss, 20% loss*
- Volume differences
  - *too loud, too quiet*
- Acoustic problems
  - *echo, bad mike*

Watson & Sasse: Procs ACM Multimedia 2000

Results: Subjective



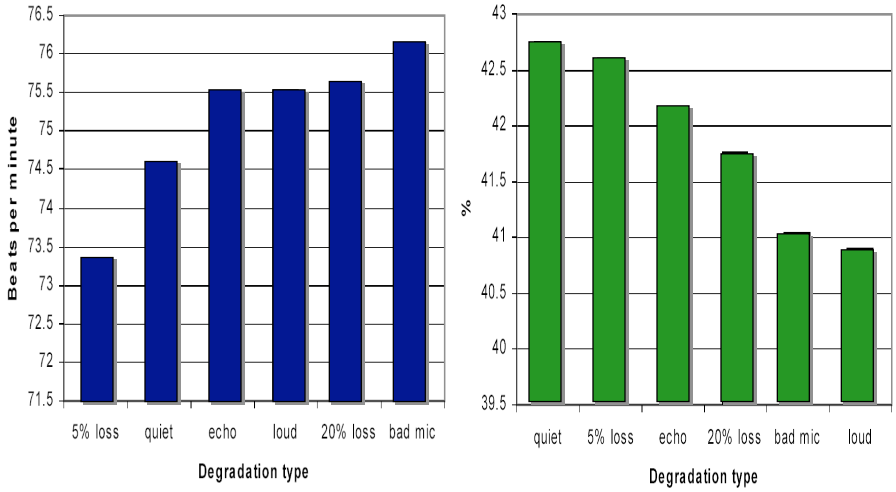
Results: *subjective* ratings

- Ratings were very consistent across repeated measures.
- No significant difference between the reference condition, *5% packet loss,* and *quiet.*
- *Bad microphone* significantly worse.
- No significant difference *between echo*, *loud* and *20% packet loss.*

Capturing users' physiological responses

- Procomp measurement device on left hand
  - *Blood Volume Pulse*
  - *Heart Rate*
- Under stress, BVP decreases and HR increases.
- Objective measure of the "cost" to the user.

Results: Objective



Results: physiological responses

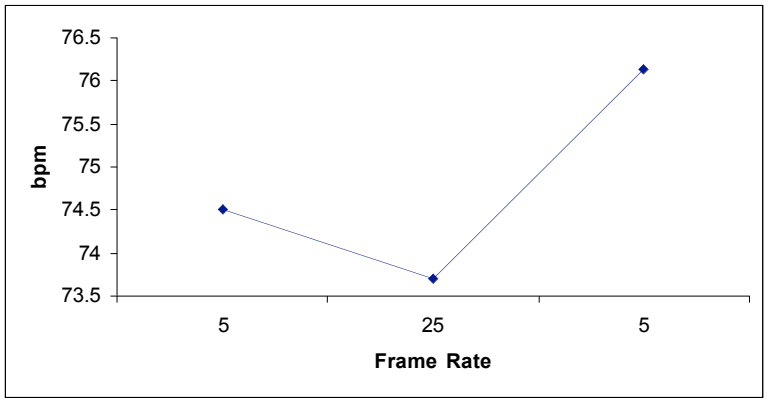
- *bad microphone, loud* and *20% loss* were significantly more stressful than *quiet* and *5% loss*
- *echo* was more stressful than *quiet* and *5% loss*
- *bad microphone* and *loud* more stressful than *20% loss*

Video Experiment 1

- Effects of 5fps and 25fps.
- Participants watched recorded University admission interviews and answered questions.
- Physiological & subjective assessment.

Wilson & Sasse Procs HCI 2000

Mean Heart Rate



Results

- Significant increase in galvanic skin response (GSR) and heart rate (HR) at 5 fps for 78% of participants
- Subjectively, 16% noticed change in frame rate
- 18% felt under stress due to quality
- Subjective responses are cognitively mediated, subject to memory biases

Physiology implications

- Differences between subjective ratings and physiology
- Indication of strain, fatigue in longer-term use (stress, RSI)
- Danger of mis-attribution (e.g. in interviews)

Summary (1)

- Users evaluate quality in terms of perceived *utility* for a *task* performed (*value*)
- QoS parameters vary across tasks
  - Speed, resolution,
  - security, allowing focus on task
- *Predictable*, *consistent* quality valued more highly than variable, “best available”.
- Each task has critical thresholds and critical period that determines perception of quality per session

Bouch et al., Procs IWQoS 2000

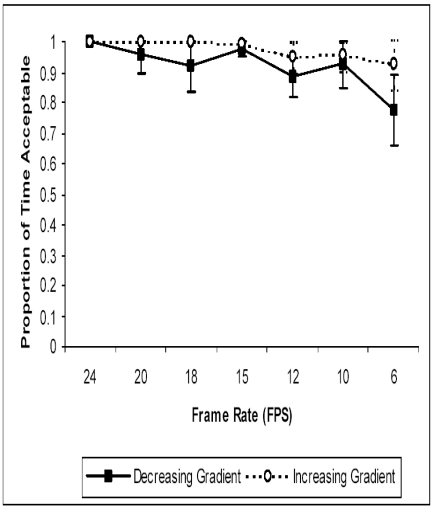
Video Experiment 2

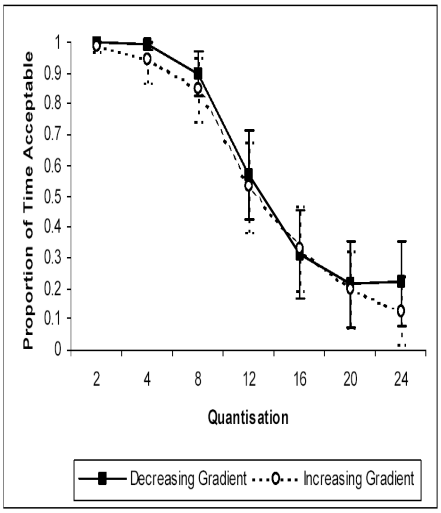
- Watching football clips
  - particular emphasis on video on handheld devices, such as 3G phones, etc.
- Goal was to evaluate the tradeoff between frame rate and quantization.
  - Conventional guidelines recommend high frame rates for high-motion video such as sports events.

Details

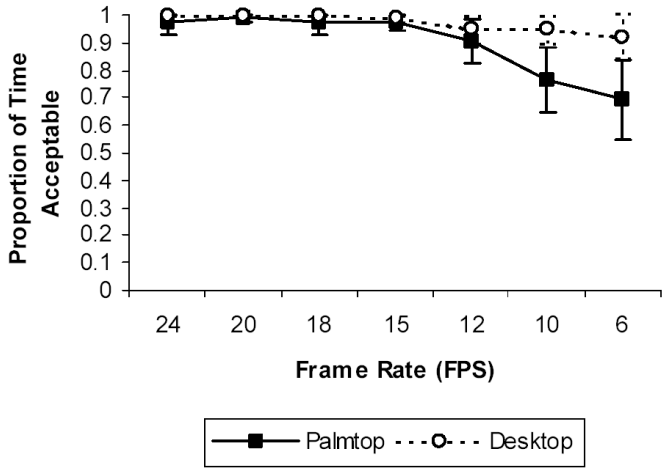
- Two video streams:
  - CIF (352x288 pixels) on a LCD monitor.
  - QCIF (176x144 pixels) on an iPAQ palmtop.
- Vary framerate between 6fps and 25fps
- Vary quantization parameters.
- Verbal feedback from subjects
- Use eye-tracker
- Use heart measurements.

Results: CIF size

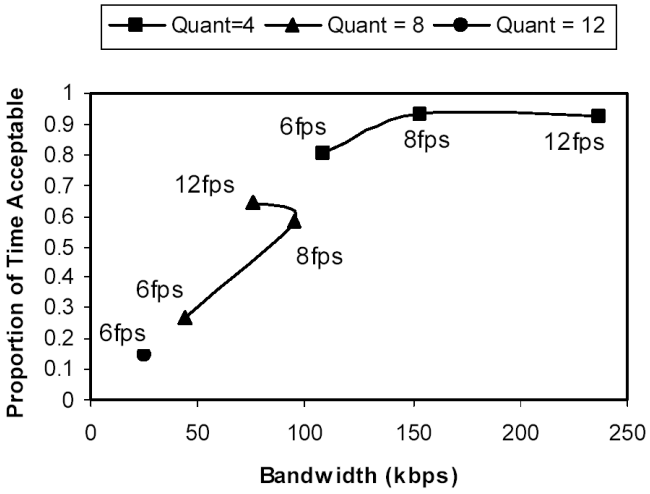




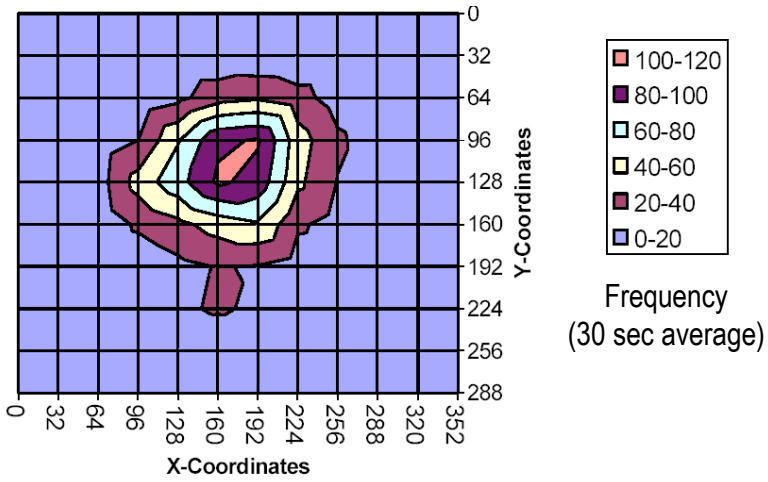
Results: CIF vs QCIF



Bandwidth vs Acceptability



Eye Tracking



Summary

- Audio is difficult
  - External factors (echo, microphone position, volume) can seriously degrade quality.
- Video is a little more forgiving.
  - No single right answer for how to tradeoff frame size, frame rate, quantization.
  - Depends very much on the task in question.
- Cost of service influences expectations and how quality is evaluated.
- User evaluation is an important part of design
  - Results are not always what you expect!