Redefining ITU-T P.912 Recommendation Requirements for Subjects of Quality Assessments in Recognition Tasks

Mikołaj I. Leszczuk Joel Dumke Lucjan Janowski Artur Koń

Quality Assessment for Recognition Tasks (QART)
Video Quality Experts Group (VQEG)

December 12, 2011

Table of contents

Introduction

- Problems of quality measurements for task-based video partially addressed in Recommendation ITU-T P.912
  - Title: “Subjective Video Quality Assessment Methods for Recognition Tasks”
  - Published: 2008
  - Introducing:
    - Basic definitions
    - Methods of testing
    - Psycho-physical experiments
- Section 7.3 (“Subjects”): “Subjects who are experts in the application field of the target recognition video should be used.”
- Nevertheless, to best authors’ knowledge, expert viewer issue not well verified in specific academic research
- Consequently, we compared groups of subjects assessing video quality for task-based video

Is **Subjects' Proficiency** Necessary?

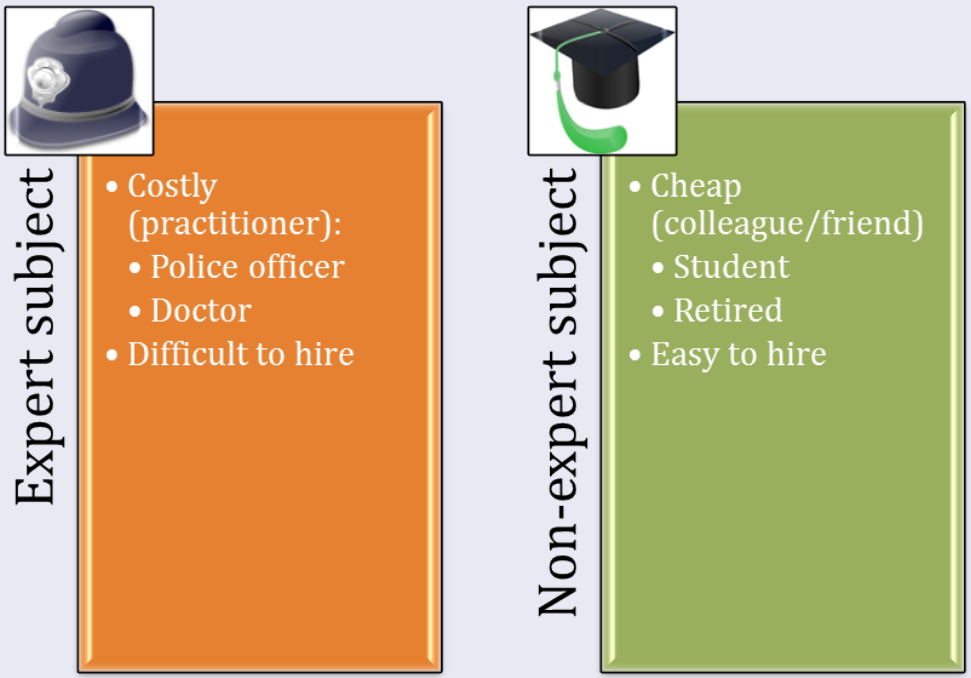

Figure: *Do I really need to be a security officer in order to participate in a test checking my ability to read license plate numbers in compressed video?*

Table of contents

Four Categories of Lighting Condition Scenarios

1. Outdoor, daytime light
2. Indoor, bright with flashing lights
3. Indoor, dim with flashing lights
4. Indoor, dark with flashing lights

Three Different Distances Used for Clips Creating

1. 5.2 meters for indoor scenarios
2. 10.9 meters for outdoor scenarios, objects close
3. 14.6 meters for outdoor scenarios, objects far

Table of contents

Viewing Conditions of Room Where Test Took Place Following ITU-R BT.500-12 and ITU-T P.910

- Ratio of luminance of inactive screen to peak luminance: ≤ 0.02
- Ratio of the luminance of the screen, when displaying only black level in a completely dark room, to that corresponding to peak white: ≈ 0.01
- Maximum observation angle relative to the normal (this number applies to CRT displays, whereas the appropriate numbers for other displays are under study): 30°
- Ratio of luminance of background behind picture monitor to peak luminance of picture: ≈ 0.15
- Other room illumination: *low*

Arranged Viewing Conditions

Figure: Test environment

Example of User Interface

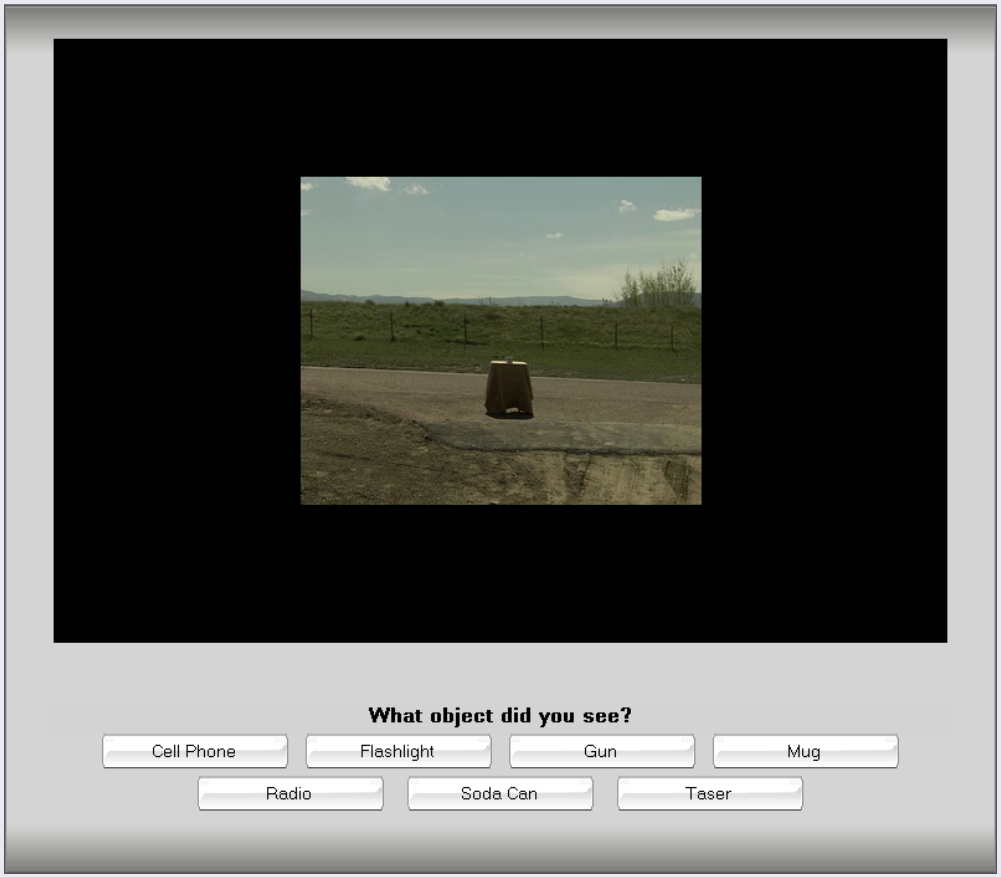

Figure: User interface for subjective target recognition task test performed

NTIA Test-Plan (1/2)

- NTIA performed the object recognition tests with two groups of viewers
- The Practitioner group
  - All subjects were volunteers and weren't paid for the test
  - Most received invitational travel to Boulder, CO
  - All of them had experience in public safety, including:
    - Police
    - Firemen
    - EMS
    - Forensic Video Analysts
  - Very few were outside the range of 30-60 years old
  - Three had minor color vision problems — their results were not significantly different

NTIA Test-Plan (2/2)

- The Non-Practitioner group
  - Subjects having no experience in image recognition
  - All subjects were paid through a temp agency to take the test
  - None of them had experience in public safety
  - Subjects had a wide variety of ages, but skewed young
  - Two had minor color vision problems — their results were not significantly different

AGH Test-Plan

- Subjects having no experience in image recognition
- All subjects volunteers and weren't paid for their job
- None of them with experience in public safety area

- Almost all subjects 20-25 years old
- One of them with color vision problems — did no worse than other viewers so his results included

Introduction to Results

- Test results:
  - For each test: 15540 answers totally given
  - For AGH test: 10096 correct and 5444 wrong — 65% of right answers
- Best and worst conditions
  - Best recognition — outdoor, stationary, close distance scenarios (89.5% for AGH)
  - Worst recognition — indoor, moving, dark light scenarios (25.4% for AGH)

Results on Recognition Rates

- Lighting
  - Significant influence on recognition, best recognition under daylight and bright lighting
  - For 1536 kbit/s bit-rate recognition changes from 92% with bright to 30% with dark lighting conditions
- Motion
  - Influence strongly depending on other conditions
  - For indoor, dim lighting 27% difference between stationary and moving objects recognition
- Distance
  - Influence larger for lower bit-rates
  - 36% difference between far and close outdoor objects

Observations

- Subjects' accuracy growing during the test, suggests that testers learned how objects were carried
- Under some conditions better results for CIF resolution videos (than for higher resolutions)
- Often viewers didn't watch whole clip before giving the answer
- Most of subjects didn't take any brake during the tests

Comparison of Unmotivated with Motivated Subjects

- Significant difference between results, about 17% more right answers in motivated subjects' experiments
- Under good conditions (high bit-rate, enough lighting etc.) results of both groups are quite similar, but difference grows fast while conditions are degrading, than in case of very bed conditions (like far, moving object) difference decreases again
  - 43% difference for CIF, 64 kbit/s bit-rate, outside, far, stationary objects clips
  - 6% difference for VGA, 1536 kbit/s bit-rate, outside, close, stationary objects clips
  - 9% differnce for CIF, 64 kbit/s bit-rate, outside, far, moving objects clips
- Both, motivated and unmotivated subjects, achieved best and worst results for same scenarios

Comparison of Unmotivated with Motivated Subjects

- Light changes
  - For indoor, moving objects scenarios with bright lighting motivated subjects achieved about 7% higher recognition, for dim lighting conditions difference grows to 21%, and then fall to 17% for dark lighting

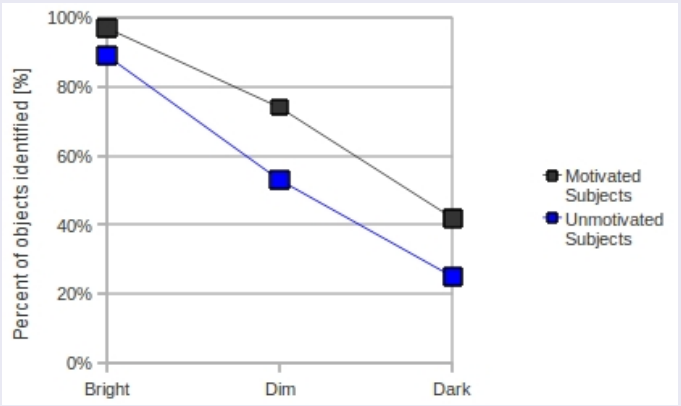

Figure: Indoor, Moving

Comparison of Unmotivated with Motivated Subjects

- Motion changes
  - For inside, stationary, dim lighting conditions scenarios unmotivated subjects achieved 7% lower recognition, for the same lighting conditions, but moving objects, difference changes to 21%

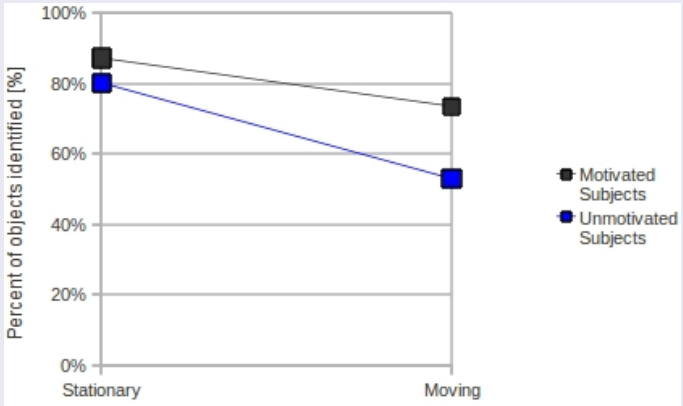

Figure: Indoor, Dim Lighting

Comparison of Unmotivated with Motivated Subjects

- Distance changes
  - For close, outdoor, stationary scenarios difference between motivated and unmotivated subject's recognition is 6%, for far objects motivated subjects achieved just 1% lower result when unmotivated subjects's recognition decreased for 24%

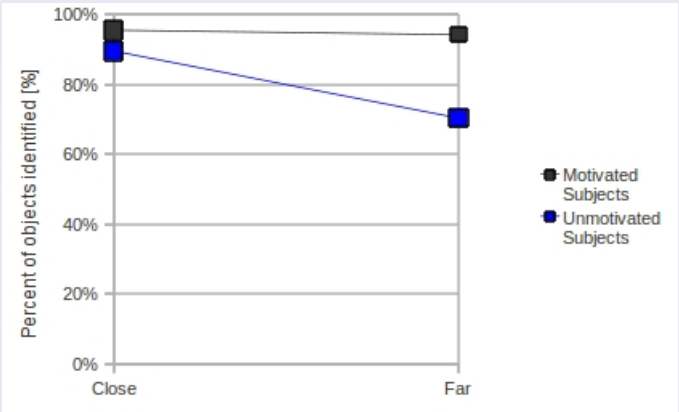

Figure: Outdoor, Stationary

Conclusion on Standardization

- Comparison developed for task-based video
- Specifications amendments for ITU-T P.912 Recommendation developed
- Consequently first sentence of Section 7.3 (“Subjects”) of ITU-T P.912 to get rephrased into: “Subjects who are motivated should be used.”
- Assisting researchers of task-based video quality to identify subjects that will allow them to successfully perform psychophysical experiment required

Acknowledgement

- European Community's Seventh Framework Program (FP7/2007-2013) for received funding for research leading to these results under grant agreement №. 218086 (INDECT)
- U.S. Department of Commerce for received funding for research under Public Safety Communications Research project
- The Department of Homeland Security's Office of Interoperability and Compatibility, and NIST's Office of Law Enforcement Standards