# Training Session for Task Recognition

*Lucjan Janowski and Mikołaj Leszczuk*

## Task Recognition Specificity

In many applications the video quality is not as important as the ability to accomplish a specific task for which the video was created. Typical examples are: surveillance systems; a camera installed in a car helping to park; or a remote medical consultation system. A general idea behind the quality tests for task recognition is to find a threshold at which the task can be achieved with a certain probability or accuracy. Therefore, instead of the quality evaluation, the subjective experiment is focused on a task performance measurement. For example, the test might measure the probability that a license plate number is accurately recognized, a car is parked correctly, or a correct diagnosis is made. Therefore, the training session is focused on clearly describing the task description and familiarizing a subject with the test's interface. One can think that explaining a task is especially easy. A task can be described as simply as: "Please type the license number," "Please park a car," or "Please recognize if an organ is healthy," following the previous examples.

Problems of quality measurements for task-based video are partially addressed in an ITU-T Recommendation (P.912, "Subjective Video Quality Assessment Methods for Recognition Tasks," 2008) that mainly introduces basic definitions, methods of testing and psycho-physical experiments. Section 7.4 of ITU-T P.912 ("Instructions to subjects and training session") says that "The subject should be given the context of the task before the video clip is played, and told what they are looking for or trying to accomplish. If questions are to be answered about the content of the video,

the questions should be posed before the video is shown, so that the viewer knows what the task is." We followed such simple training guidance, but still some problems were found. Here we point to some errors we made running subjective task recognition experiments. Possible solutions are proposed.

## Examples

The first example is a license plate recognition task. The task was: "Please write all characters which you are able to read in the text box below" [1]. Analyzing results we found that this description was not precise enough. Some subjects understood that if a character is difficult to read it cannot be read, others try hard to read all characters. As a consequence some subjects recognized just the most obvious characters and others many more of them. Of course, we cannot be sure if better training would change the results much, since we are used to observe different subjects engagement. Nevertheless, a clear training session containing a video with difficult, but possible to read, characters would make it clear for subjects that if they are not sure they should still try their best. More details about errors made by users in this and another recognition tasks experiments can be found in [2].

The second example is an experiment in which goal was object recognition. NTIA ITS performed the object recognition tests with different groups and interfaces [4] [5]. The same experiment was repeated by AGH [3]. For all those experiments, a training session showed a short video with each object and the object name. The experiments' results demonstrate that only one subject misunderstood the training and marked a radio as a mobile phone. With a large number of subjects (164 in total) a training session cannot be blamed. Therefore for a simple object recognition experiment, a simple training session seems to be enough.

The third example is a subjective test of remote ultrasonography conducted by project [6]. The task was to recognize an organ and decide if there were any problems with it. Since the quality of ultrasonography is strongly dependent on the person who is conducting the examination, the test had to be interactive. We explained why the provided quality was low and where the system was to be used (remote places with limited Internet access like Mali in Africa). Nevertheless, in a typical examination, additional information about a patient is available. In cases where there were problems conducting the examination, many other ideas about how to proceed in a real life situation were proposed by doctors. It made the experiment very chaotic. We also noticed that an examination cannot be too long and the tasks (what should be investigated) cannot be very similar, or a doctor will likely lose interest in the experiment. The training session has to include a very clear and detailed explanation of the experiment. The best would be to consult with a doctor to frame the explanation. Also, we should be prepared to give additional information as to why all scenarios had to be conducted to obtain the results needed by the project or algorithm development. Motivation is one of the most crucial parts.

Each task recognition experiment is different. Even if a task description looks easy, we advise that a small preliminary test be run. Not only the results, but also interviews with the subjects taking part in a preliminary test, help to prepare an experiment description and training set that not only explains what to do but also motivates the subjects to perform the task correctly.

## References

[1] Mikołaj Leszczuk, Lucjan Janowski, Piotr Romaniak, Andrzej Głowacz, and Ryszard Mirek. Quality assessment for a licence plate recognition task based on

a video streamed in limited networking conditions. In *Multimedia Communications, Services and Security*, pages 10–18. Springer Berlin Heidelberg, 2011.

[2] Task-based subject validation: reliability metrics, L Janowski Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop ...

[3] Mikołaj I Leszczuk, Artur Kon, Joel Dumke, and Lucjan Janowski. Redefining ITU-T P. 912 recommendation requirements for subjects of quality assessments in recognition tasks. In *Multimedia Communications, Services and Security*, pages 188–199. Springer Berlin Heidelberg, 2012.

[4] VQiPS: Video quality tests for object recognition applications. Public Safety Communications DHS-TR-PSC-10-09, U.S. Department of Homeland Security's Office for Interoperability and Compatibility (June 2010)

[5] VQiPS: Recorded-video quality tests for object recognition tasks. Public Safety Communications DHS-TR-PSC-11-01, U.S. Department of Homeland Security's Office for Interoperability and Compatibility (June 2011)

[6] http://www.qol.unige.ch/research/TeleUSG.html

*Lucjan Janowski is an assistant professor at the Department of Telecommunications (AGH University of Science and Technology). He worked on malicious traffic analysis in CNRS-LAAS in France at a post-doc position in 2007. In 2010/2011 he spent half a year on a post-doc position in University of Geneva working on QoE for health applications. His main interests are statistics and probabilistic modelling of subjects and subjective rates used in QoE evaluation.*

*Mikołaj Leszczuk, PhD, is an assistant professor at the Department of Telecommunications, AGH University of Science and Technology (AGH-UST), Krakow, Poland). He has participated actively as a steering committee member or researcher in several national and European projects, including: INDECT, BRONCHOVID, GAMA, e-Health ERA, PRO-ACCESS, Krakow Centre of Telemedicine, CONTENT, E-NEXT, OASIS Archive, and BTI. He is a member of the VQEG Board and a co-chair of VQEG QART (Quality Assessment for Recognition Tasks) and MOAVI (Monitoring of Audio Visual Quality by Key Indicators) Group.*