



The NappingPlayer — Projective Mapping Experiments on Android Tablets

Werner Robitza, Yohann Pitrey, Benjamin Kitzinger, Helmut Hlavacs

Entertainment Computing Research Group, University of Vienna, Austria

firstname.lastname@univie.ac.at

Abstract

Projective Mapping opens up new ways of interpreting experienced quality from a bottom-up perspective and identifies the main dimensions along which it is classified. This methodology comes from the sensory sciences, and asks users to place stimuli on a tablecloth in front of them, according to their perceived similarity. They may then be freely labeled and grouped in order to explore the semantics of the users' perception. We apply Projective Mapping in the domain of Quality of Experience assessment, and implement a novel application for tablet PCs, which allows for a more efficient and accurate test procedure. This paper presents the application and results from a preliminary study.

Index Terms: Quality of Experience, Experiment Methodologies

1. Introduction

In the domain of Quality of Experience (QoE), subjective quality evaluation methods typically follow a standardized approach. A sequence of stimuli is shown to a number of observers, and after each, they are asked to give their quality rating on a predefined scale. Popular amongst those methodologies is Absolute Category Rating (ACR) and Pair Comparison (PC) [1] or Double Stimulus Impairment Scale (DSIS) [2]. The advantage of protocols like these is that a large number of stimuli can be tested at once, even with many test subjects, and the interpretation and analysis of results has been well established.

Researchers may use the ratings obtained from such experiments as ground truth data for modeling the impact of distortions or degradations on the quality as perceived by the user. Due to the complexity and variety of possible degradations and their subjective interpretation in different viewing contexts, it is not always easy to pinpoint the reason for certain ratings. This is exacerbated by the fact that often only one single rating is taken, without any way for the observer to provide more (qualitative) feedback. In viewing contexts that may be new to the user—or even entirely new research domains such as 3DTV—the interpretation of ratings on a simple quality scale might be different from previous recommendations. For example, it is not given that for a 3DTV experiment, participants will use a Mean Opinion Scale [2] in the same fashion

as for “regular” TV. Also, the factors leading to the user judgements may generally not be known in advance.

The problems with classical methods can be alleviated when looking at “quality” from a bottom-up perspective: users should be able to express the perceived quality in specific attributes that directly pertain to their senses. “Napping” (Projective Mapping) is a testing procedure that allows researchers to ask for those attributes. In this paper we present our implementation of the Napping methodology, the NappingPlayer, which can be directly used on a tablet device. By recording the test results digitally, our application makes it possible to significantly reduce the amount of time needed to prepare, conduct, and evaluate Napping experiments. In the following, we will explain the procedure (Section 2) as well as the application (Section 3) and show preliminary results from two experiments conducted with the NappingPlayer in Section 4.

2. Projective Mapping

In the field of sensory sciences, traditionally food products are evaluated based on sensoric attributes (mostly olfactory, gustatory and visually). These constitute opinions about the presented material, much like in the domain of multimedia quality testing. Projective Mapping [3] is a testing procedure that stems from the sensory sciences. Here, a number of non-expert participants (called “panelists”, usually ten or less) is chosen. Their task is to sort a set of stimuli on a tablecloth placed in front of them. This procedure is called “Napping”, derived from the French word “nappe” for “tablecloth”. The only criterion for sorting these stimuli is their perceived difference: similar stimuli are placed closer, and different ones are positioned apart from each other. The participants are given no other instructions for placement. Depending on the test procedure, the stimuli may then be additionally categorized into groups, to each of which the user can give multiple labels. This extension is called “Sorted Napping” [4], and was already applied in the domain of video quality [5]. Depending on the use case, there exist variations of this procedure, where for example the labeling can be restricted to a predefined vocabulary [6].

Traditionally, conducting such an experiment with

video content is time-consuming. To illustrate this, we describe a process that is usually followed. To familiarize ourselves with the classical procedure, we also internally tested it at our department. As participants cannot “move” the video stimuli *per se*, each video is given a number. When the panelist wants to see a video, they must ask the experimenter to play it for them, e.g. on a video projector or screen. Paper cards with the numbers are handed to the user as proxies to put on the tablecloth. The experimenter will first play all videos in sequence. In order to save time, the videos can be shown to multiple panelists simultaneously. After this step, each video can be separately requested by a single panelist. At the same time, the panelists may sort the cards on the tablecloth in front of them. They may also use a sheet of paper for taking notes.

After the user has finished the placement, the positions of the cards on the nappe have to be drawn on a small sheet of paper, where the participant may create groups (e.g. by drawing circles around the numbers) and label the videos. Each video can belong to several groups, and every group can have overlapping labels. When the grouping is finished, a post-test interview may be conducted. It involves asking the participant about the main attributes they used, explaining their nappe configuration.

To finally analyze the data, the positions then have to be converted to coordinates, which are entered on a computer. Typically, this would involve measuring the sheet with a ruler. The labels for the groups are entered in a similar fashion (e.g., a spreadsheet program). All of these post-experiment steps may take as long (or longer) than the actual test itself and thus impose huge efforts on research personnel. They are also prone to errors (in the sense of inaccuracies) during the coordinate translation steps.

3. The NappingPlayer

To overcome the drawbacks of a manual Napping procedure with video content, we implemented the procedure for tablet devices based on the *Android* mobile OS.¹ The application is called “NappingPlayer”. It was made completely open source and available to the public under <http://github.com/slhck/napping-player>. In order to set up the test, the experimenter will transfer the stimuli as video files onto the tablet. During the test, the panelist may enter their name on the device. Much like in a real-life experiment, the NappingPlayer presents the user a landscape view of a virtual tablecloth. Its size depends on the screen dimensions of the tablet itself, and its background is medium grey so as not to introduce color bias during the testing. The user may hold the tablet freely during the test (see Figure 1).

¹<http://www.android.com/>



Figure 1: Napping experiments with the NappingPlayer can be conducted in various contexts without the need for a table or projection equipment.

The test procedure with the NappingPlayer is as follows: after entering their name, each video is played to the user in a sequential fashion, directly on the tablet itself. Via in-app preferences, the presentation can be zoomed to full screen, but videos may also be shown in their native resolution to avoid the scaling to have an impact on the experienced quality. After seeing a video, it will be represented as a small “card” (a button) on the virtual tablecloth. The user may sort the cards by dragging them with a finger, or view a video again by tapping its card. Sorting multiple cards at the same time by using more than one finger is also supported. The user may sort and watch the videos as long as they like to, and then press “Finish”. Once the positioning is done, groups can be created by tapping (and thereby selecting) cards that should belong to the group. Every group can also be tagged with free-choice keywords (see Figure 2) by means of an input field. The labeling step—just like in the regular Napping—is optional and can be disabled for a test session through the app’s preferences.

Comparing this procedure with the classical method shows that using the NappingPlayer allows for much quicker execution. Experiments can be easily set up by just placing the video files onto a folder on the tablet’s file system. The NappingPlayer will pick up the files and construct the test set automatically. The experiment data—including the positions and labels of the cards—are stored on the device. When the observer is finished, they may also send their results directly via e-mail to a preconfigured address. This allows the experimenter to gather the data immediately after a test, and visualize the outcome of the analysis (see Section 4).

The NappingPlayer source code will be accompanied with a script for the open source statistics software

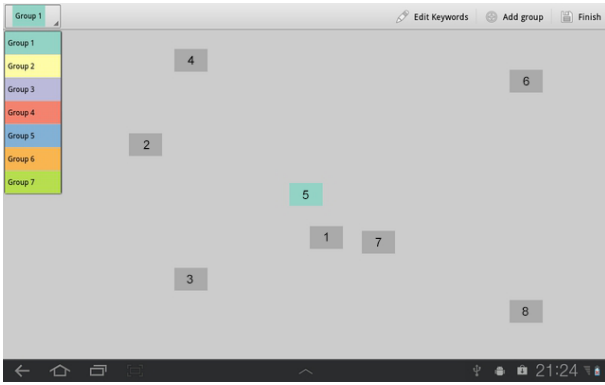


Figure 2: Virtual tabletop with eight videos and optional grouping (top left).

R .² This script allows to transform the experiment data into formats compatible with the *FactoMineR* and *SensoryMineR* packages for *R*, which were (partially) written by the authors of [4, 6] to facilitate the data analysis for experiments in sensory sciences. It also includes examples of how to generate plots typically used for Projective Mapping analysis. The script therefore makes it possible to analyze an entire folder of NappingPlayer data within seconds.

4. Preliminary Tests and Results

To test the NappingPlayer, we conducted two rounds of experiments. Both were carried out in the scope of the QUASSUMM project.³ As stimuli, we downloaded 20 popular clips from YouTube in their original quality. All clips could be classified as user-generated content (i.e. no clip was taken from a movie or television show). The clips featured various contents: from home videos showing animals to semi-professionally produced clips depicting tips and tricks. To prepare them, we scaled them to 480 by 320 pixels and cut them to 20 seconds length. The sources were selected to span a broad range of video and audio quality: some showed heavy compression artifacts and stuttering motion, while the impairment of others was barely visible.

In our tests, we wanted to find the answers to the following questions regarding the app itself as well as the presented content:

- Can the NappingPlayer be used to replace the traditional Napping method for video content?
- Does it result in a significant improvement of time needed to conduct a whole experiment—from the setup to the data analysis and presentation of results?

²<http://www.r-project.org/>

³<http://univie.ac.at/quassumm>

- According to which perceptual dimensions do users classify video content? Which quality attributes are used, or does content play an important role as well?

4.1. Experiment 1

15 students participated in the first experiment. Here, we did not lay focus on a specific task, meaning that—contrary to typical Projective Mapping experiments—observers were not explicitly briefed to evaluate stimuli on their quality attributes. While it can be imagined that food products, without any further instruction, will be sorted by their gustatory, olfactoric (and perhaps visual) characteristics, videos carry different dimensions: 1) the quality of the video and audio streams, respectively, and 2) the content of the sequence shown. In the first experiment, users were simply told to place the videos according to their similarity, and then label them with whatever attribute they wanted.

To combine the results of the observers into one “average” tabletop, we performed a Hierarchical Multi Factor Analysis (HMFA) on the individual configurations (i.e. placement of the cards and their labels). The HMFA is an extension of a regular Multi Factor Analysis (MFA), in which the goal is to scale individual configurations describing the same stimuli to a common configuration. [4] The HMFA, in addition to the coordinates of the stimuli, takes into account the categorization from the labeling task.

The HMFA results let us plot the stimuli placement on the two main principal components (Figure 3). The first two dimensions of the analysis combined explain 35.5% of the variance in the data. We can observe that users exclusively classified videos according to their content, with videos of animals grouped on the left, tips and tricks as well as funny and entertaining videos at the bottom / right, and humans, nature and serious topics grouped at the top / right. With respect to the stimuli in this experiment, we can thus see that for user-generated content the dimensions can be loosely interpreted as *animal* – *human* and *funny* – *serious*.

The focus on content attributes can be confirmed through the keywords assigned to the video sequences. Combining synonymous words, 38 different labels were used, however no labels were related to the visual or auditory quality of the content: users appeared to disregard this factor entirely. Most attributes described the actual content of the video (e.g., “tips and tricks”), while several meta-attributes were also used (e.g., “uninteresting”, “professional”, “pointless”).

4.2. Experiment 2

We observed that in experiment 1, panelists did not sort the stimuli according to their visual and auditory quality,

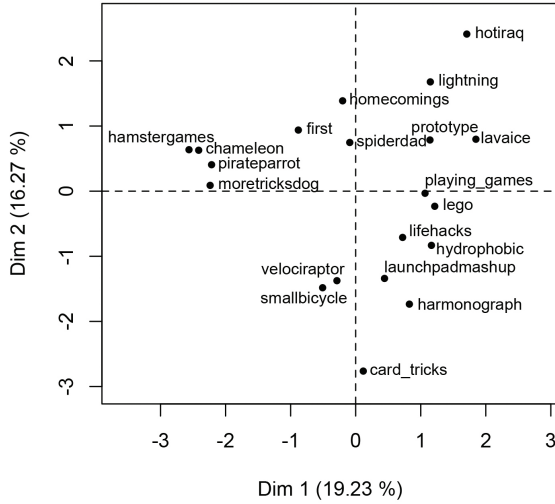


Figure 3: Factor map from Hierarchical Multi Factor Analysis for experiment 1.

but according to content instead. We therefore repeated the sessions with different instructions to see whether this task would evoke quality attributes only. In the second experiment, 5 panelists took part. To elicit quality-only attributes during the main test session, every panelist discussed with the experimenter about what quality meant for them when looking at video, trying to describe what they had previously experienced (e.g. on television or when watching YouTube). For simplicity reasons, audio was not included in this experiment. The remainder of the experiment was however conducted in a similar fashion as experiment 1.

Again, we performed an HMFA on the individual configurations. Figure 4 shows the two main principal components, which explain 61.37% of the variance (compared against 35.5% in experiment 1). From left to right we see the quality of videos decreasing. The cluster of the videos in the top right constitutes the videos with the lowest visual quality (as confirmed by expert screening). The meaning of the second dimension however can not be clearly interpreted without looking at the attributes assigned to the groups. Figure 5 shows the correlation circle for all attributes and the two main dimensions. We can observe attributes pertaining to spatial degradations (e.g. from “sharp”, “clear” to “blocky” and “pixelated”) and temporal degradations (e.g. from “fluid” to “stuttering” and “jerky”). With regard to the second dimension we now see that blurry videos and washed out colors are perceived differently—panelists seemed to classify them neither as particularly good or bad in terms of quality. In fact, the clips found in this region of the common configuration do not appear jerky and do not show visual artifacts.

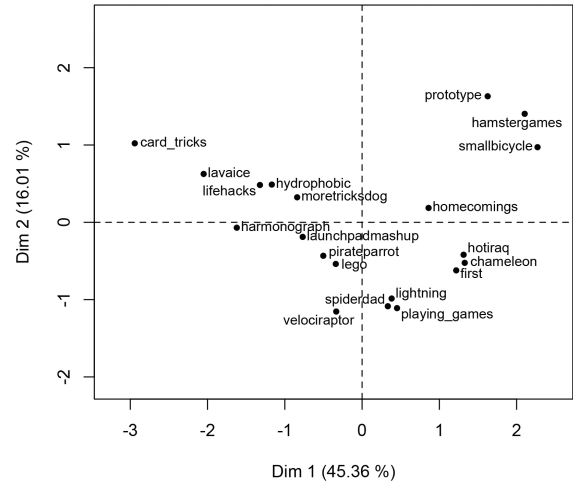


Figure 4: Factor map from Hierarchical Multi Factor Analysis for experiment 2.

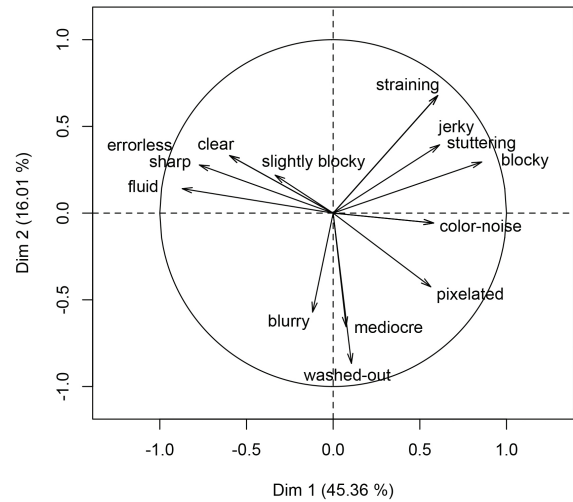


Figure 5: Correlation circle for attributes used in experiment 2. For readability, the individual panelist’s dimensions have been removed and the labels were translated from German.

5. Discussion

5.1. Methodology

Projective Mapping and its variants have a long history as methodologies in the sensory sciences. They only have been recently adapted for the use in multimedia quality. The advantages of the Napping procedure can be found in the richness of the results—instead of a collection of single numerical or categorical ratings, researchers get insight into all the dimensions along which quality is experienced. Especially for uncharted research areas, conducting Napping experiments can help in directing future research. When we developed the NappingPlayer, we wanted to facilitate the conduction of Napping experiments, which would otherwise be time-consuming and resource-intensive to perform.

The results of experiment 1 raise interesting questions. While we had assumed that participants would instinctively sort video according to mixed dimensions (content *and* quality), they turned out to be neglecting experienced quality factors entirely. Why would this happen? We explain this result based on several factors: First of all, the videos—most of which showed a considerable amount of visual degradation due to generation loss, improper editing or recording—were viewed on a mobile screen: In internal tests with video quality experts from our research group, we were not able to perceive the same degree of artifacts on the mobile when compared to a reference computer display. In addition to that, differing consumer expectations regarding mobile and “regular” multimedia consumption might have also biased their reception. Lastly, the use of real-life material instead of content specifically produced for video quality testing resulted in a realistic watching scenario (e.g., watching YouTube video on a tablet at home). It appears that in such a scenario, consumers do not pay attention to the visual quality, since the entertainment value of the videos is of more importance.

Experiment 2 gave us more insight into the quality attributes used for classifying our set of user generated videos. The main dimension according to which video quality was perceived includes both spatial and temporal factors such as the fidelity of the image as well as fluid playback. Here, panelists were able to identify mostly degradations due to low bitrate, wrong frame-rate settings or generation loss. The second dimension however describes the overall sharpness and color vividness, which might be a property of the video source itself, or even an aesthetic (and deliberate) factor, such as desaturated colors. While the labels help in understanding the individual configurations, their full meaning can only be clarified during post-test surveys and discussions.

5.2. Usability

As observed during the test and confirmed during the post-test interviews, participants did not express any major usability problems with the software itself and generally enjoyed participating. They liked being able to conduct the experiment at their own pace, watching videos again if they felt the need to do so. Most observers enjoyed “playing” with the buttons and dragging them around.

Minor improvements could go into allowing for a quicker labeling method, where circles can be drawn around the video’s cards instead of manually selecting every single card. Over all, participants however had no issues creating labels with the current implementation.

5.3. Time Factors

In comparison to manually conducting the Napping procedure, it was possible to save several hours of time, especially in the following tasks:

- *Preparation:* No cards have to be printed, cut out, and distributed to the observers. A room with large enough tables for the potential number of panelists simultaneously participating is not needed when using a tablet device.
- *Conduction:* Giving each panelist the opportunity to watch videos at their own rate means that the overall time for the experiment is lowered, since no panelist has to wait for the other. In practice, performing the experiment with only one video projector proved to be inefficient, since observers might not want to view the same videos as others again. We estimate that, depending on the number of participants, about 30 minutes can be saved during this step.
- *Data analysis:* With traditional Projective Mapping, analysis time increases linearly with the number of observers. Transcribing coordinates and labels is not needed with the NappingPlayer and cut down roughly three hours of two researchers entering the data to less than a few minutes.

6. Conclusion

Projective Mapping as a test method is becoming more and more popular in the field of QoE. It provides meaningful data about the user’s sensoric experiences, helping researchers understand the aspects of quality from a bottom-up perspective. The experiments are time-consuming though, but can be conducted much faster on mobile devices.

We developed a prototype of an Android application that allows to conduct Napping experiments in a mobile environment. Preliminary experiments have shown that

the application is a suitable replacement for the traditional pen-and-paper method, while providing for a much more time-efficient test procedure and data analysis, with new possibilities for mobile testing. Our experiments have also highlighted the importance of content and context during quality tests, to the extent of severe quality impairments not being factored into the sorting process by users, when not explicitly asked to.

Future research will go into performing more specific quality-focused experiments and extending the application to provide for variations of the common procedure. In order to elicit a higher number of sensoric attributes, familiarizing the subjects with the topic of visual quality could also be encouraged, e.g. by collecting a set of quality attributes *before* the actual test, using training videos. In a similar fashion, this was proposed in [5]. The labeling task can then further be restricted to the attribute set defined before the test, much like a panelists of wine tasting experts would use a predefined (and well-described) taxonomy of attributes instead of choosing their own.

Our NappingPlayer is freely available to the research community as a tool to facilitate Napping experiments and provide further insight into what quality means to the user.

7. References

- [1] “Recommendation ITU-T P.910: Subjective video quality assessment methods for multimedia applications,” 2008.
- [2] “Recommendation ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures,” 2012.
- [3] E. Risvik, J. A. McEwan, J. S. Colwill, R. Rogers, and D. H. Lyon, “Projective mapping: A tool for sensory analysis and consumer research,” *Food quality and preference*, vol. 5, no. 4, pp. 263–269, 1994.
- [4] M. Cadoret, S. Lê *et al.*, “The sorted napping: A new holistic approach in sensory evaluation,” *Journal of Sensory Studies*, vol. 25, no. 5, pp. 637–658, 2010.
- [5] D. Strohmeier, “Open profiling of quality: A Mixed Methods Research Approach for Audiovisual Quality Evaluations,” Ph.D. dissertation, Technische Universität Ilmenau, 2011.
- [6] L. Perrin, R. Symoneaux, I. Maître, C. Asselin, F. Jourjon, and J. Pagès, “Comparison of three sensory methods for use with the Napping procedure: Case of ten wines from Loire valley,” *Food Quality and Preference*, vol. 19, no. 1, pp. 1–11, 2008.