NTIA Technical Memorandum 14-505

AGH/NTIA: A Video Quality Subjective Test with Repeated Sequences

Margaret H. Pinson and Lucjan Janowski

technical memorandum

U.S. DEPARTMENT OF COMMERCE • National Telecommunications and Information Administration

NTIA Technical Memorandum 14-505

AGH/NTIA: A Video Quality Subjective Test with Repeated Sequences

Margaret H. Pinson and Lucjan Janowski

U.S. DEPARTMENT OF COMMERCE

June 2014

DISCLAIMER

Certain commercial equipment and materials are identified in this report to specify adequately the technical aspects of the reported results. In no case does such identification imply recommendation or endorsement by the National Telecommunications and Information Administration, nor does it imply that the material or equipment identified is the best available for this purpose.

CONTENTS

Figures vi

Tables vii

1 Introduction 1
1.1 Subject Rating Behavior 1
1.2 Source Video Reuse 2
1.3 Suitability of Subject Screening Methods 3

2 New Terms 4
2.1 Full Matrix of SRCs by HRCs 4
2.2 Related Source Sequences 5
2.3 Coding Difficulty Source Sequences 6

3 Experimental Design 8
3.1 Training 9
3.2 Scene Pool A 10
3.3 Scene Pool B 11
3.4 Scene Pool C 11
3.5 Scene Pool D 17
3.6 Ordering and Software Control 23
3.7 Questionnaire 25

4 References 27

Appendix A :Subject Screening Experiment Instruction Script 28

Appendix B :Session Ordering 30

FIGURES

Figure 1. Sample frame depicts ITU-R Rec. BT.802, 525-line scene no. 30, mobile and calendar. 4

Figure 2. Sample frames from the training clips**.**....9

Figure 3. Sample frames from scene pool **A**: (left) “NTIA Rose Path” (right) “NTIA The Foot music video Segment 1.” Each SRC was impaired with all five HRCs.....10

Figure 4.Sample frames from scene pool **B**: (left) “NTIA Red Kayak” (right) “NTIA Red Parrot.” Each SRC was impaired with all five HRCs. 11

Figure 5. Sample frames from the 15 original SRCs within RSRC **flamenco.** Each segment was taken from CDVL sequence “NTIA Flamenco Dancers”. The name below each frame indicates the PVS. 13

Figure 6. Sample frames from the 15 original SRCs within RSRC **question.** The PVS name is indicated below each frame. 14

Figure 7. Sample frames from the 15 original SRCs within RSRC **walkway.** The PVS name is indicated below each frame. 15

Figure 8. Sample frames from the 15 original clips for CD-SRC low coding difficulty**.**....21

Figure 9. Sample frames from the 15 original clips for CD-SRC medium coding difficulty**.**....22

Figure 10. Sample frames from the 15 original clips for CD-SRC high coding difficulty**.**....23

TABLES

Table 1.(SRC × HRC) Matrix for Scene Pool A 10

Table 2.(SRC × HRC) Matrix for Scene Pool B.....11

Table 3. (RSRC × HRC) Matrix for Scene Pool C.....13

Table 4. RSRC to HRC and Session Mapping 16

Table 5. (CD-SRC × HRC) Matrix for Scene Pool D 18

Table 6. Low Coding Difficulty CD-SRC.....18

Table 7. Medium Coding Difficulty CD-SRC.....19

Table 8. High Coding Difficulty CD-SRC 20

AGH/NTIA: A VIDEO QUALITY SUBJECTIVE TEST WITH REPEATED SEQUENCES

Margaret H Pinson[1] and Lucjan Janowski[2]

This report provides full technical details for the video quality subjective test AGH/NTIA. Analyses of this dataset appear in separate publications. The purpose of this document is to provide design details that are beyond the scope of a conference paper or journal article.

Subjective experiment AGH/NTIA includes multiple instances of the same stimuli rated three or six times by the same subject. The goal is to provide insights into the suitability of subject screening methods, the impact of source video reuse on subjective data, and the behavior of subjects when repeatedly rating the same stimuli.

Key words: subject screening, subjective testing, video quality

1 INTRODUCTION

Subjective video quality test AGH/NTIA was designed to examine three issues:

- The behavior of subjects
- The impact of source video reuse on subjective data
- The suitability of subject screening methods

Each question examines the fundamental nature of subjective video quality testing. Consequently, the experimental design became highly complex. This report is intended to provide a full and complete description of the subjective test design and implementation. Analyses of the subjective data will appear in separate publications, with only a brief summary of the experimental design.

1.1 Subject Rating Behavior

The primary issue is subject behavior generally and the accuracy of a subject's ratings in particular. That is, if the same stimulus is shown to one subject many times, what is the distribution of that subject's scores? Are people capable of perfectly repeating the same

[1] The author is with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80305.
[2] The author is with the Department of Telecommunications, AGH University of Science and Technology, Krakow, Poland.

subjective score? If scores change, is this due to an opinion change over time or is there error term associated with the scoring task?

The purpose of this experiment is to identify the important variables and levels of variables that impact subject scoring behaviors. Other datasets will be used to confirm these variables. Additional subjective tests may be required.

A problem arises in that each subject will view an identical sequence many times. Rote memorization could teach the subject the purpose of the experiment, thus invalidating our results. Thus, we need our experiment to include an interesting variety of stimuli that disguise the main purpose of the experiment. The second issue provides that distraction.

1.2 Source Video Reuse

Many video quality subjective experiments are designed to analyze quality differences among hypothetical reference circuits (HRC). The experiment design is usually a full matrix of scenes (SRC) and HRCs. Fundamentally, the test measures whether or not subjects can perceive a difference between two versions of the same stimuli. Those conclusions are assumed to generalize to the entire HRC.

One consequence of this experiment design is that each subject will view and rate the same SRC multiple times, occasionally as many as 25 times. We know from subject interviews that some subjects change their scoring behavior in response to these repeated viewings—focusing on one portion of the sequence instead of paying attention to the whole.

We would like to explore the hypothesis that reuse of source video influences the resulting mean opinion score (MOS) ratings. That is, source video reuse fundamentally changes MOS, thus yielding a less accurate estimate of subject opinion. The corollary is that alternative subjective test designs (that do not reuse SRC) will more accurately estimate MOS. The novel stimuli inherent to these alternative experiment designs will help distract subjects from the presence of the perfectly repeated stimuli mentioned in Section 1.1.

Dataset AGH/NTIA proposes two new experiment designs that depend upon novel stimuli:

- Related Source Sequences (RSRC, defined in Section 2.2)
- Coding Difficult Source Sequences (CD-SRC, defined in Section 2.3)

where each new design replaces each SRC with a set of SRC.

Section 3 describes how these two new experiment designs are combined into a single test. Briefly put, dataset AGH/NTIA will have three viewing sessions. Each session will have three full matrices: (SRC × HRC), (RSRC × HRC) and (CDSRC × HRC). This will allow comparisons between the three techniques within each session. The (SRC × HRC) matrix will be identical for all three sessions, to help identify any time dependent changes in a subject's treatment of SRC.

This experiment will evaluate the (RSRC × HRC) and (CDSRC × HRC) experiment designs. These designs may have advantages over the traditional (SRC × HRC) design when analyzing HRC quality. Additional experiments may be needed if either technique appears promising.

1.3 Suitability of Subject Screening Methods

The third issue addressed by dataset AGH/NTIA is an investigation into subject screening methods. This is a tertiary use of the data rather than a primary design goal. Consequently, it has minimal impact on the experiment design.

There are two philosophies on subject screening for video quality subjective tests. The first philosophy, preferred by behavior scientists, is that all subjective data should be retained unless there is a highly compelling reason to discard that subject's data. A clear and compelling case must be made individually for each subject discarded, and this should be a rare event. We do not know why subjects respond differently to stimuli, so all subject data should be retained. Subject screening is done manually.

The second philosophy is to ensure that flawed subjective data is discarded. We know that problems occur: inattention, a misunderstanding of the task, a flaw in the subjective test playback system, or score entry error. Flawed subjective data corrupts the data analysis, typically making it more difficult to differentiate between stimuli. An assumption is made that there is an underlying true quality value for each stimuli, which all subjects agree upon. Therefore, the penalty for discarding a good subject is small, while the error introduced by retaining a bad subject is high. Subject screening is done automatically (see for example ITU-T Rec. P.913).

The third use of dataset AGH/NTIA is to study existing subject screening methods and allow for research into a potentially improved subject screening method. Such an improved method might eliminate the assumption that subjects must agree, and allow there to be genuine differences of opinion among subjects. This research requires multiple scores by a single subject for a single stimulus so that an improved method can be compared to an existing screening technique that uses repeated stimuli.

2 NEW TERMS

Video quality subjective tests are designed around the choice of SRCs and HRCs. One SRC is an exact segment of video, for example the first 10 seconds of the renowned “calendar and mobile” sequence from ITU-R Rec. BT.802 (see Figure 1).

Figure 1. Sample frame depicts ITU-R Rec. BT.802, 525-line scene no. 30, mobile and calendar.

2.1 Full Matrix of SRCs by HRCs

Subjective tests are typically designed to include one or more full matrixes of (SRC × HRC). The list of HRCs might be MPEG-2 at five different bit-rates. This design makes two hypotheses:

Hypothesis 1a: There is no impact on subjective ratings when a subject sees the same SRC multiple times using different HRCs.

Hypothesis 1b: Two different HRCs can be most accurately compared using one identical SRC.

Hypothesis 1b is credible. By pairing one SRC with multiple HRCs, the source content variable can be eliminated from the data analysis.

Hypothesis 1a is more tenuous. We know that subjects will memorize the SRC over the course of an experiment. This memorization can influence their scoring behavior. Exit interviews with subjects indicate that some subjects will focus exclusively on one portion of a sequence to make

their rating, for example the swirling black lines of the sheep in Figure 1. This probably makes their ratings more consistent—yet perhaps provides a less accurate representation of the subject's true opinion of the quality of each processed video sequence (PVS).

As mentioned in Section 1.2, one goal of dataset AGH/NTIA is to investigate two fundamentally different approaches for source content. The goal of each approach was to maintain the traditional subjective test design (SRC × HRC), while eliminating SRC re-use. To do this, we explored two different mechanisms for replacing each SRC with a set of SRCs, within the experimental design.

2.2 Related Source Sequences

Let us define a set of related source sequences (RSRC) to be a set of SRCs that have visually similar content. We replace each SRC in the subjective test design with this set of sequences. Thus, the test design changes from (SRC × HRC) to (RSRC × HRC). During data analysis, the SRCs in each RSRC set are treated as if they were identical.

The RSRC idea depends upon the availability of multiple SRCs that have visually similar content. One example is videoconferencing footage: head and shoulder shots of several different people shot with a still camera and similar backdrops. A second example is a variety of pans across a crowded stadium, where each segment has similar pan speed, zoom, and lighting conditions. A third is different segments of a single football game. Each pool of footage depicts one idea, produced using similar filming and editing techniques.

The RSRC design makes two hypotheses:

Hypothesis 2a: There is an impact on subjective ratings when a subject sees the same SRC multiple times (e.g., for different HRCs).

Hypothesis 2b: SRCs with very similar subject matter and video production will have comparable perceptual quality when impaired.

Let us assume a subjective test with a full matrix design of (RSRC × HRC). If there are ***m*** HRCs, then each RSRC would be a set of ***m*** SRCs depicting similar content. This is how we expect an RSRC would normally be used.

Dataset AGH/NTIA has a full matrix of (RSRC × HRC) for each session, so as a consequence each RSRC will be a set of 3 × ***m*** SRCs. These extra matrices should yield information on whether or not an RSRC yields stable ratings.

There are three advantages to the RSRC:

- SRC memorization is avoided, so the ratings should be a more accurate reflection of our subjects' true opinions.

- Boredom is reduced. Boredom causes inattention and is a key reason for limiting the duration of viewing sessions.

- Each PVS is scored more independently, because the subject cannot memorize the SRC.

There are three disadvantages to the RSRC:

- The data is confounded by the differences among the set of SRCs associated with a RSRC. This adds an uncontrolled variable to the experimental design. We hope to minimize the impact of this unexplained variable by choosing SRCs with similar content, editing, camera work, camera settings, aesthetics, etc.

- The similarity of content within a RSRC set may allow subjects to memorize some aspects of the content that reappear. For example, consider an example RSRC of crowd surveillance footage depicting different people in the same hallway. The subject might focus on a painting that is always visible in the hallway, instead of paying attention to the changing crowd of people.

- The SRCs within one RSRC may have very different coding difficulties (i.e., encoding at the same bitrate with the same codec may produce very different perceived quality). This disadvantage inspires our second proposal.

2.3 Coding Difficulty Source Sequences

Let us define a coding difficulty source sequence (CD-SRC) as a set of SRC that have a similar coding difficulty. For example, one CD-SRC set might include sequences with low coding difficulty, while another might include sequences with high coding difficulty. Since the decision is made by an automated algorithm, these SRCs may have very different visual characteristics (e.g., different content types, camerawork, editing, and aesthetics).

The CD-SRC idea depends upon the availability of an automated algorithm to calculate the coding difficulty of an SRC. The Appendix of Fenimore et al. [3] provides a suitable algorithm: scene criticality. We hope that this scene criticality algorithm will allow us to estimate the coding difficulty of an SRC: that is, which SRC will yield a similar bit-rate when encoded with a constant QP (Quantization Parameter) level. If the CD-SRC idea is valid but the scene criticality metric is not accurate enough, it could be replaced with a better one.

The RSRC design makes two hypotheses:

Hypothesis 3a: There is an impact on subjective ratings when a subject sees the same SRC multiple times.

Hypothesis 3b: SRCs with similar coding difficulties will have comparable perceptual quality when impaired.

Let us assume a subjective test with a full matrix design of (CD-SRC × HRC). If there are ***m*** HRCs, then each CD-SRC would be a set of ***m*** SRCs with similar coding difficulties. This is how we expect a CD-SRC would normally be used.

Dataset AGH/NTIA has a full matrix of (CD-SRC × HRC) for each session, so as a consequence each CD-SRC will be a set of 3 × ***m*** SRCs. These extra matrices should yield information on whether or not a CD-SRC yields stable ratings.

There are three advantages to the CD-SRC:

- SRC memorization is avoided, so the ratings may be a more accurate reflection of our subjects' true opinions.

- Boredom is minimized. Each SRC may depict completely different content, thus maintaining the subject's interest.

- In the case of ACR (Absolute Category Rating), a subject rate would be closer to the typical situation where a subject sees an impaired content live but cannot compare it with identical (SRC) or even similar (RSRC) content under different conditions (HRC).

There are three disadvantages to the CD-SRC:

- The variety of SRCs within each CD-SRC adds an uncontrolled variable to the experiment design. We hope to minimize the impact of this unexplained variable by choosing SRCs with similar coding difficulty, so that the artifacts produced by the codec are minimized.

- The SRCs within a CD-SRC set will have very different aesthetic characteristics. It is thus possible that subjects' interest in the content matter will be another variable that strongly influences the HRC variable.

- The number of SRCs needed for an experiment increases as a factor of the number of HRCs.

3 EXPERIMENTAL DESIGN

Dataset AGH/NTIA adheres to ITU-T Rec. P.910. The absolute category rating (ACR) scale was used. The subjective test was run automatically by a preliminary version of subjective test software. Since then, this subjective test software was finalized and made available by Catellier and Connors [1]. The subjective test was performed on a 17" laptop and automated software used by Catellier et al. [2]. The experiment sessions were conducted in a simulated living room.

Subject instructions were read from the script that appears in Appendix A. Data were gathered from a total of 28 subjects. All but one of the subjects were recruited from a temporary hiring agency. The remaining subject was a video quality expert. Three of the 28 subjects were given special instructions:

- The expert viewer knew the purpose of the experiment (from Section 1.1) and tried to perfectly repeat each PVSs prior score.

- Two subjects were intentionally given incorrect instructions, in an attempt to ensure the presence of outliers. These two subjects (#18 and #19) were told to simultaneously rate both the quality of the video and their opinion of the content. The goal was to ensure two subjects whose ratings would reflect a misunderstanding of the rating task. We assume that these subjects should be removed from the dataset by an automated subject screening algorithm (see Section 1.3).

The test was divided into three viewing sessions. Taken together, all three sessions took approximately one hour. Two subject orderings were created (red and blue). The ordering of sessions was held constant, as was the ordering of video sequences within each session. This was done to eliminate the influence of random chance on the subjects' ability to replicate a prior score. Our analysis of subject rating behavior (see Section 1.1) may require time series analyses, which will be impossible if the presentation order is randomized.

Five HRC levels were selected from a pool of variable bit-rate AVC/H.264 encodings that ranged from 400 kbps to 8 Mbps. The video encoding bit-rates were manually chosen so that each video depicted the desired quality level, as judged by the authors. The five quality levels were:

- Original

- Goodplus (i.e., slightly above "good" on the ACR scale)

- Goodminus (i.e., slightly below "good" on the ACR scale)

- Poorplus (i.e., slightly above "poor" on the ACR scale)

- Poorminus (i.e., slightly below "poor" on the ACR scale)

Manually selected quality levels were used in an attempt to minimize the influence of source content on HRC quality level. This same design choice was made in [2] and the resulting ratings indicated reasonable accuracy in the experimenter's ability to manually select consistent quality

levels. The minimal influence of source content was intended to allow stronger conclusions to be reached when comparing different scene pools.

All of the source scenes are available on the Consumer Digital Video Library (CDVL, www.cdvl.org). All scenes are 8 seconds long and converted to 720p 59.94fps. The scenes were originally filmed in a variety for formats (e.g., 1080i 29.97fps, 1080p 29.97fps and 720p 29.97fps).

The source scenes (SRC) were divided into four scene pools: A, B, C, and D. Each scene pool was treated differently with regards to the choice of HRCs and the repeated viewings (or lack thereof).

Each of the following sub-sections describes the content for each scene pool. The following naming convention was used for each processed video sequence (PVS): an abbreviated scene name, an underscore, and the quality level. Each sub-section will identify the name of each SRC as used by CDVL (in quotes), to make it easier for the reader to locate that SRC. In some cases the CDVL SRC is longer than 8 seconds, and a sub-section of that longer sequence was used.

3.1 Training

Three SRC were used for the training session. Each SRC depicted a simulated news clip, as shown in Figure 2. The training session contained four PVS:

PVS Name	CDVL Name
newsbottledwater_original	"NTIA simulated news bottled water" (1080i 29.97fps)
newsbottledwater_poorminus	"NTIA simulated news bottled water" (1080i 29.97fps)
newsdogpark_goodminus	"NTIA simulated news dog park" (1080i 29.97fps)
newsdriving1_poorplus	"NTIA simulated news driving version 1" (1080i 29.97fps)

newsbottledwater newsdogpark newsdriving1

Figure 2. Sample frames from the training clips**.**

3.2 Scene Pool A

Scene pool **A** contains two SRCs:

SRC Name	CDVL Name
thefoot01	from "NTIA The Foot music video Segment 1" (1080i 29.97fps)
rosepath	from "NTIA Rose Path" (1080i 29.97fps)

See Figure 3 for sample frames. These two SRCs were paired with all five HRCs (i.e., original, goodplus, goodminus, poorplus, and poorminus). All ten PVSs were each included twice in each of the three sessions. Therefore, each subject rated each PVS six times. The total ratings available for scene pool **A** is (2 SRC × 5 HRC × 6 repeats) = 60 ratings per subject.

Table 1 shows the (SRC × HRC) matrix. The number in each cell identifies the number of ratings made by a single subject. Here "6" indicates that each subject rated the same PVS six times.

Table 1. (SRC × HRC) Matrix for Scene Pool A

SRC	Original	Goodplus	Goodminus	Poorplus	Poorminus
thefoot01	6	6	6	6	6
rosepath	6	6	6	6	6

The six ratings allow for maximum information on each person's scoring error. The other motivation was to provide information about within-session trends as well as between-session trends. The size of this scene pool was intentionally limited, to alleviate subject boredom and inattention.

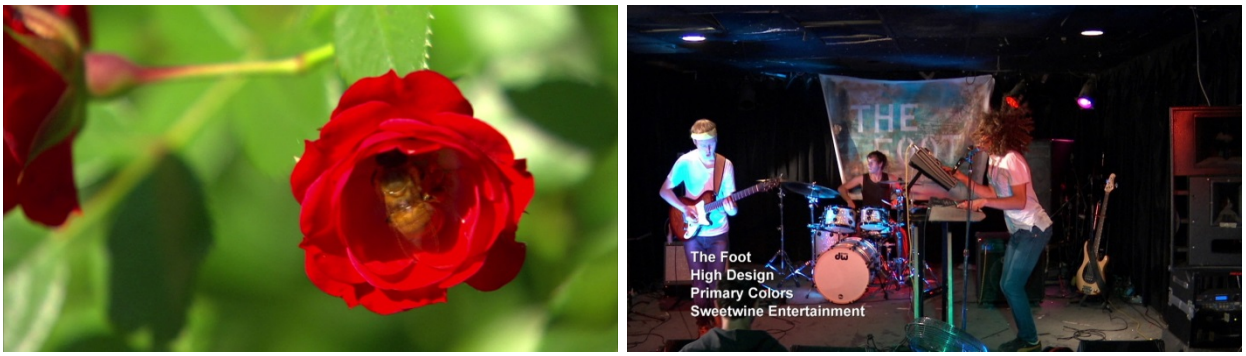


Figure 3. Sample frames from scene pool **A**: (left) "NTIA Rose Path" (right) "NTIA The Foot music video Segment 1." Each SRC was impaired with all five HRCs.

3.3 Scene Pool B

Scene pool **B** contains two SRC:

SRC Name	CDVL Name
redkayak	from “NTIA Red Kayak” (1080p 29.97fps)
parrot	from “NTIA Red Parrot” (1080i 25fps, played faster)

See Figure 4 for sample frames. These two SRCs were paired with all five HRCs. All ten PVSs were each included once in each of the three sessions. Therefore, each subject rated each PVS three times. The total ratings available for scene pool **B** is: (2 SRC × 5 HRC × 3 repeats) = 30 ratings per subject.

Table 2 shows the (SRC × HRC) matrix. The number in each cell identifies the number of ratings made by a single subject. Here “3” indicates that each subject rated the same PVS three times.

Table 2. (SRC × HRC) Matrix for Scene Pool B

SRC	Original	Goodplus	Goodminus	Poorplus	Poorminus
redkayak	3	3	3	3	3
parrot	3	3	3	3	3

Scene pool **B** provides information about between-session trends.

Figure 4. Sample frames from scene pool **B**: (left) “NTIA Red Kayak” (right) “NTIA Red Parrot.” Each SRC was impaired with all five HRCs.

3.4 Scene Pool C

Scene pools **A** and **B** adhere to the traditional philosophy in subjective testing of creating a full (SRC × HRC) matrix. Scene pool **C** investigates the proposed alternative subjective experiment design theories outlined in section 2.2. Scene pool **C** includes a full matrix of (RSRC × HRC). Thus, many different instantiations of the same scene idea are paired with different HRCs.

Scene pool **C** contains three RSRC:

RSRC Name	CDVL Name
flamenco	edited from the 7 minute sequence "NTIA Flamenco Dancers" (1080i 29.97fps)
question	edited from "PSCR Mock Police Interview" (1080i 29.97fps) sequences
walkway	edited from "Angled stadium walkway, set B" version 1 to 8 and "Angled stadium walkway, set E" version 3 to 9 (1080i 29.97fps)

These three sets of video were chosen for their homogeneous appearance.

Figure 5 shows sample frames for the 15 SRCs within RSRC **flamenco**. Each frame is labeled below by the name of this SRC. The RSRC flamenco is taken from pieces of a 7 minute fully edited audiovisual sequence depicting two flamenco dancers with live music. Some of the SRCs in the flamenco RSRC set overlap by up to 0.5 seconds. The video was edited to avoid scene cuts and fades appearing close to the beginning or end of any 8 second sequence. Fades are visible in some of the sample frames.

Figure 6 shows sample frames for the 15 SRCs within RSRC **question**. The raw footage is taken from simulated police interviews between a police detective (with his back to the camera) and an actor. The actors were asked to enact a variety of visibly obvious emotional responses. Each SRC is taken from a different edited segment of this footage from CDVL. The sequence names within CDVL are listed below each SRC. For brevity, "PSCR Mock Police" is replaced by "…".

Figure 7 shows sample frames for the 15 SRCs within RSRC **walkway**. The raw footage is taken from security footage of an indoor stadium walkway during the break of a sporting event. Each SRC is taken from a different edited segment of this footage from CDVL. The sequence names within CDVL are listed below each SRC. For brevity, "Angled stadium walkway" is replaced by"…". RSRC question and walkway contain no scene cuts.

These three RSRCs were paired with all five HRCs. The experiment design for scene pool **C** is: (3 RSRC × 5 HRC). Put another way, each RSRC contains fifteen different SRCs, which were used to create 15 PVSs (i.e., one each). These PVSs were distributed evenly among the five HRCs as follows: 3 original, 3 goodplus, 3 goodminus, 3 poorplus, and 3 poorminus.

In addition, the RSRC PVSs that appeared in the first session were repeated identically in the 2nd and 3rd sessions. This provides extra data to analyze subject rating behavior (see Section 1.1).

Table 3 shows the (RSRC × HRC) matrix. The number in each cell identifies the number of ratings made by a single subject. This number is ambiguous, because it ignores whether the rated PVS was novel or repeated. Table 3 is gives a big picture; Table 4 contains the missing details. This mapping of RSRC, SRC and HRC is shown in Table 4. Each row is one PVS. The last column ("Sessions") identifies the session or sessions in which this PVS appears. Note that five of the 15 PVSs for each RSRC are assigned to each session, such that all five quality levels were represented in each session and each SRC appeared once. Then the five PVSs from the 1st session are repeated in the 2nd and 3rd session, to yield two additional data points in Table 3.

Table 3. (RSRC × HRC) Matrix for Scene Pool C

RSRC	Original	Goodplus	Goodminus	Poorplus	Poorminus
flamenco	5	5	5	5	5
question	5	5	5	5	5
walkway	5	5	5	5	5


flamenco1


flamenco2


flamenco3


flamenco4


flamenco5


flamenco6


flamenco7

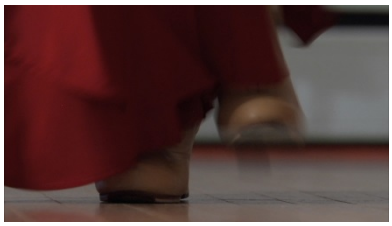

flamenco8


flamenco9


flamenco10


flamenco11


flamenco12


flamenco13


flamenco14

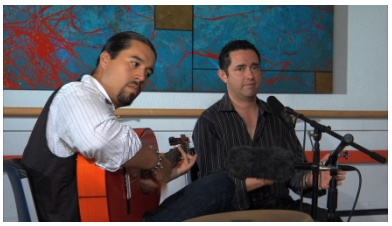

flamenco15

Figure 5. Sample frames from the 15 original SRCs within RSRC **flamenco.** Each segment was taken from CDVL sequence "NTIA Flamenco Dancers". The name below each frame indicates the PVS.

question assaultf2v1 from "PSCR Mock Police Interview of Assault, Female 2 version 1"

question assaultf2v3 from "… Interview of Assault, Female 2 version 3"

question assaultf2v5 from "… Interview of Assault, Female 2 version 5"

question assaultf3v3 from "… Interview of Assault, Female 3 version 3"

question assaultf3v4 from "… Interview of Assault, Female 3 version 4"

question assaultf3v6 from "… Interview of Assault, Female 3 version 6"

question assaultm3v1 from "… Interview of Assault, Male 3 version 1"

question assaultm3v4 from "… Interview of Assault, Male 3 version 4"

question carf1v1 from "… Interview of Car Accident, Female 1 version 1"

question carf1v5 from "… Interview of Car Accident, Female 1 version 5"

question carf2v2 from "… Interview of Car Accident, Female 2 version 2"

question carf2v3 from "… Interview of Car Accident, Female 2 version 3"

question carm1v1 from "… Interview of Car Accident, Male 1 version 1"

question carm1v3 from "… Interview of Car Accident, Male 1 version 3"

question storem2v2 from "… Interview of Store Robbery, Male 2 version 2"

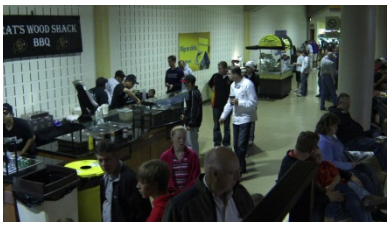
Figure 6. Sample frames from the 15 original SRCs within RSRC **question.** The PVS name is indicated below each frame.

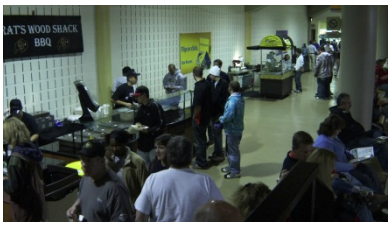

walkway bver1 from "… set B (some people) version 1"


walkway bver2 from "… set B (some people) version 2"


walkway bver3 from "… set B (some people) version 3"


walkway bver4 from "… set B (some people) version 4"

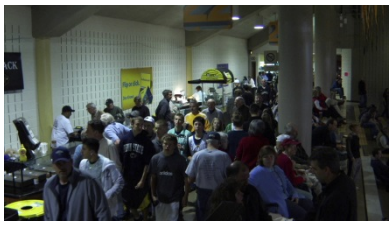

walkway bver5 from "… set B (some people) version 5"


walkway bver6 from "… set B (some people) version 6"


walkway bver7 from "… set B (some people) version 7"


walkway bver8 from "… set B (some people) version 8"

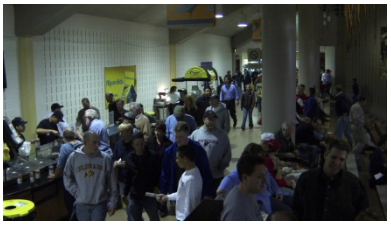

walkway ever3 from "… set E (many people) version 3"


walkway ever4 from "… set E (many people) version 4"


walkway ever5 from "… set E (many people) version 5"


walkway ever6 from "… set E (many people) version 6"


walkway Ever7 from "… set E (many people) version 7"


walkway Ever8 from "… set E (many people) version 8"

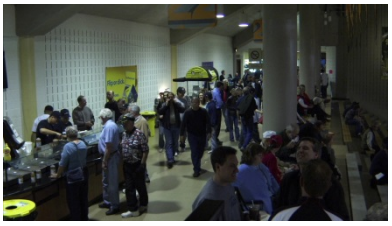

walkway ever9 from "… set E (many people) version 9"

Figure 7. Sample frames from the 15 original SRCs within RSRC **walkway.** The PVS name is indicated below each frame.

Table 4. RSRC to HRC and Session Mapping

RSRC	SRC	HRC	Sessions
flamenco	flamenco1	original	1, 2, 3
flamenco	flamenco2	original	2
flamenco	flamenco3	original	3
flamenco	flamenco4	goodplus	1, 2, 3
flamenco	flamenco5	goodplus	2
flamenco	flamenco6	goodplus	3
flamenco	flamenco7	goodminus	1, 2, 3
flamenco	flamenco8	goodminus	2
flamenco	flamenco9	goodminus	3
flamenco	flamenco10	poorplus	1, 2, 3
flamenco	flamenco11	poorplus	2
flamenco	flamenco12	poorplus	3
flamenco	flamenco13	poorminus	1, 2, 3
flamenco	flamenco14	poorminus	2
flamenco	flamenco15	poorminus	3
question	assaultf2v1	original	1, 2, 3
question	assaultf2v3	original	2
question	assaultf2v5	original	3
question	assaultf3v3	goodplus	1, 2, 3
question	assaultf3v4	goodplus	2
question	assaultf3v6	goodplus	3
question	assaultm3v1	goodminus	1, 2, 3
question	assaultv3v4	goodminus	2
question	carf1v1	goodminus	3
question	carf1v5	poorplus	1, 2, 3
question	carf2v2	poorplus	2
question	carf2v3	poorplus	3
question	carm1v1	poorminus	1, 2, 3
question	carm1v3	poorminus	2
question	storem2v2	poorminus	3
walkway	Bver1	original	1, 2, 3
walkway	Bver2	original	2
walkway	Bver3	original	3
walkway	Bver4	goodplus	1, 2, 3
walkway	Bver5	goodplus	2
walkway	Bver6	goodplus	3
walkway	Bver7	goodminus	1, 2, 3
walkway	Bver8	goodminus	2
walkway	Ever3	goodminus	3
walkway	Ever4	poorplus	1, 2, 3
walkway	Ever5	poorplus	2
walkway	Ever6	poorplus	3
walkway	Ever7	poorminus	1, 2, 3
walkway	Ever8	poorminus	2
walkway	Ever9	poorminus	3

3.5 Scene Pool D

Scene pool **D** investigates the proposed alternative subjective experiment design theories outlined in section 2.3. Scene pool **D** includes a full matrix of (CD-SRC × HRC). Where scene pool **C** treats groups of visually similar content as a single RSRC, scene pool **D** treats groups of similarly complex content as a single SRC. Part of the purpose of scene pool **D** is to disguise the actual purpose of the experiment as a whole.

There are three CD-SRCs: high coding complexity, medium coding complexity, and low coding complexity. Each CD-SRC contains 15 SRCs, for a total of 45 SRCs. The number of required SRCs was high enough that obtaining a sufficient number of high quality source HDTV sequences was a problem. Consequently, two or three visually similar SRCs were occasionally included in scene pool D (e.g., different segments of the "Mr. Fins" animation). The coding complexity was measured using the scene criticality algorithm (see Fenimore et al. [3]). Scene criticality was measured before scenes were edited (i.e., using the whole video sequence as provided on CDVL).

The **low** CD-SRC set spans scene criticality values [2.1, 2.6]. Figure 8 shows a sample frames for each SRC within CD-SRC low. Table 6 describes each PVS for CD-SRC low. The information displayed identifies the SRC to HRC mapping, the sequence name within CDVL, the original format, and the sessions where that PVS appears (column Sn).

The **medium** CD-SRC set spans scene criticality values of [2.76, 3.0]. Figure 9 shows sample frames for each SRC within CD-SRC medium. Table 7 describes each PVS.

The **high** CD-SRC set spans scene criticality values of [3.26, 3.63]. Figure 10 shows sample frames for each SRC within CD-SRC high. Table 8 describes each PVS.

Within each CD-SRC, scenes were ordered by scene criticality and the five impairment levels created in order (1=original, 2=goodplus, 3=goodminus, 4=poorplus, 5=poorminus, 6=original, 7=goodplus, etc.) For SRC "crane", it was difficult to create a coding impairment with the desired quality level, therefore the HRC association was modified by swapping two adjacent video clips (CD-SRC medium, SRC "lion" and "crane".

These three CD-SRC were paired with all five HRCs. The experiment design for scene pool **D** is: (3 CD-SRC × 5 HRC). Put another way, each CD-SRC contains fifteen different SRCs, which were used to create 15 PVSs (i.e., one each). These PVSs were distributed evenly among the five HRCs as follows: 3 original, 3 goodplus, 3 goodminus, 3 poorplus, and 3 poorminus.

In addition, the CD-SRC PVSs that appeared in the first session were repeated identically in the 2nd and 3rd sessions. This provides extra data to analyze subject rating behavior (see Section 1.1). Table 5 shows the (CD-SRC × HRC) matrix. The number in each cell identifies the number of ratings made by a single subject. This number is ambiguous, because it ignores whether the rated PVS was novel or repeated. Table 5 is gives a big picture; Tables 6, 7, and 8 contain the missing details.

This mapping of CD-SRC, SRC and HRC is shown in Tables 6, 7, and 8. Note that five of the 15 PVSs for each CD-SRC are assigned to each session, such that all five quality levels were

represented in each session and each SRC appeared once. Then the five PVSs from the 1st session are repeated in the 2nd and 3rd session, to yield two additional data points in Table 5.

Table 5. (CD-SRC × HRC) Matrix for Scene Pool D

CD-SRC	Original	Goodplus	Goodminus	Poorplus	Poorminus
low	5	5	5	5	5
medium	5	5	5	5	5
high	5	5	5	5	5

Table 6. Low Coding Difficulty CD-SRC

CD-SSRC	SRC	CDVL Sequence Name	Format	HRC	Sn
low	denver	NTIA Denver Skyscrapers	1080p 29.97fps	original	1,2,3
low	horses	NTIA Horses in a Fall Field	1080p 29.97fps	goodplus	1,2,3
low	nightlights1	NTIA Night Lights in the City	1080p 29.97fps	goodminus	1,2,3
low	nightlights2	NTIA Night Lights in the City	1080p 29.97fps	poorplus	1,2,3
low	radiointro1	NTIA Radio Intro Part	1080i 29.97fps	poorminus	1,2,3
low	radiointro2	NTIA Radio Intro Part	1080i 29.97fps	original	2
low	mrfins1	NTIA Mr. Fins	1080p 23.98fps	goodplus	2
low	mrfins2	NTIA Mr. Fins	1080p 23.98fps	goodminus	2
low	mrfins3	NTIA Mr. Fins	1080p 23.98fps	poorplus	2
low	mrfins4	NTIA Mr. Fins	1080p 23.98fps	poorminus	2
low	thecloud	NTIA The Cloud	1080p 29.97fps	original	3
low	cropdusters1	NTIA cropduster at sunrise	1080i 29.97fps	goodplus	3
low	cropdusters2	NTIA cropduster at sunrise	1080i 29.97fps	goodminus	3
low	parksquirrel	NTIA squirrel at a park	1080p 29.97fps	poorplus	3
low	controlledburn	NTIA Controlled Burn	1080i 29.97fps	poorminus	3

Table 7. Medium Coding Difficulty CD-SRC

CD-SSRC	SRC	CDVL Sequence Name	Format	HRC	Sn
medium	snowmnt	NTIA Snow Mountain	1080p 29.97fps	original	1,2,3
medium	tulip1e	NTIA outdoor mall with tulips (1e)	1080p 29.97fps	goodplus	1,2,3
medium	aspen	NTIA Aspen Trees in Fall Color, Rapid Scene Cuts	1080p 29.97fps	goodminus	1,2,3
medium	mallfountains	NTIA simulated news, mall fountains	1080i 29.97fps	poorplus	1,2,3
medium	zoocuts	NTIA Zoocuts	1080p 29.97fps	poorminus	1,2,3
medium	bikepark	NTIA Bike Path Park	1080i 29.97fps	original	2
medium	halftimemusic	NTIA Halftime Music at Football Game	1080p 29.97fps	goodplus	2
medium	aspenwalk	NTIA Aspen Trees in Fall Color, Slow Scene Cuts	1080p 29.97fps	goodminus	2
medium	lion	NTIA Lion	1080p 29.97fps	poorminus	2
medium	crane	NTIA Elephant Crane	1080p 29.97fps	poorplus	2
medium	speedbag	NTIA Speed Bag	1080p 29.97fps	original	3
medium	foxandbird2	NTIA Fox and Bird, All Available Footage	1080p 23.98fps	goodplus	3
medium	foxandbird3	NTIA Fox and Bird, All Available Footage	1080p 23.98fps	goodminus	3
medium	foxandbird4	NTIA Fox and Bird, All Available Footage	1080p 23.98fps	poorplus	3
medium	mallmusic	NTIA simulated news, mall music	1080i 29.97fps	poorminus	3

Table 8. High Coding Difficulty CD-SRC

CD-SSRC	SRC	CDVL Sequence Name	Format	HRC	Sn
high	twostreamseasy	NTIA Two Streams Easy	1080p 29.97fps	original	1,2,3
high	duckfeed	NTIA Duck Feed	1080i 25fps	goodplus	1,2,3
high	stream	NTIA Stream	1080p 29.97fps	goodminus	1,2,3
high	bluespruce2a	NTIA Colorado blue spruce (2a)	1080i 29.97fps	poorplus	1,2,3
high	westwindeasy	NTIA Ode to the West Wind Easy	1080p 29.97fps	poorminus	1,2,3
high	snowfall5a	NTIA snowy day in the city (5a)	1080p 29.97fps	original	2
high	rainbo2a	NTIA Rainbow Collage Wide View: format 59.94fps interlace	1080i 29.97fps	goodplus	2
high	westwind	NTIA Ode to the West Wind	1080p 29.97fps	goodminus	2
high	bluespruce1e	NTIA Colorado blue spruce (1e)	1080p 25fps	poorplus	2
high	bluespruce3a	NTIA Colorado blue spruce (3a)	1080i 29.97fps	poorminus	2
high	purple1a	NTIA Purple Collage Narrow View: format 59.94fps interlaced	1080i 29.97fps	original	3
high	scarletoakeasy	NTIA Scarlet Oak Easy	1080p 29.97fps	goodplus	3
high	snowfall3a	NTIA snowy day in the city (3a)	1080i 29.97fps	goodminus	3
high	stadpancheer	NTIA stadium cheer and pan	1080i 29.97fps	poorplus	3
high	tulip4e	NTIA outdoor mall with tulips (4e)	1080p 25fps	poorminus	3

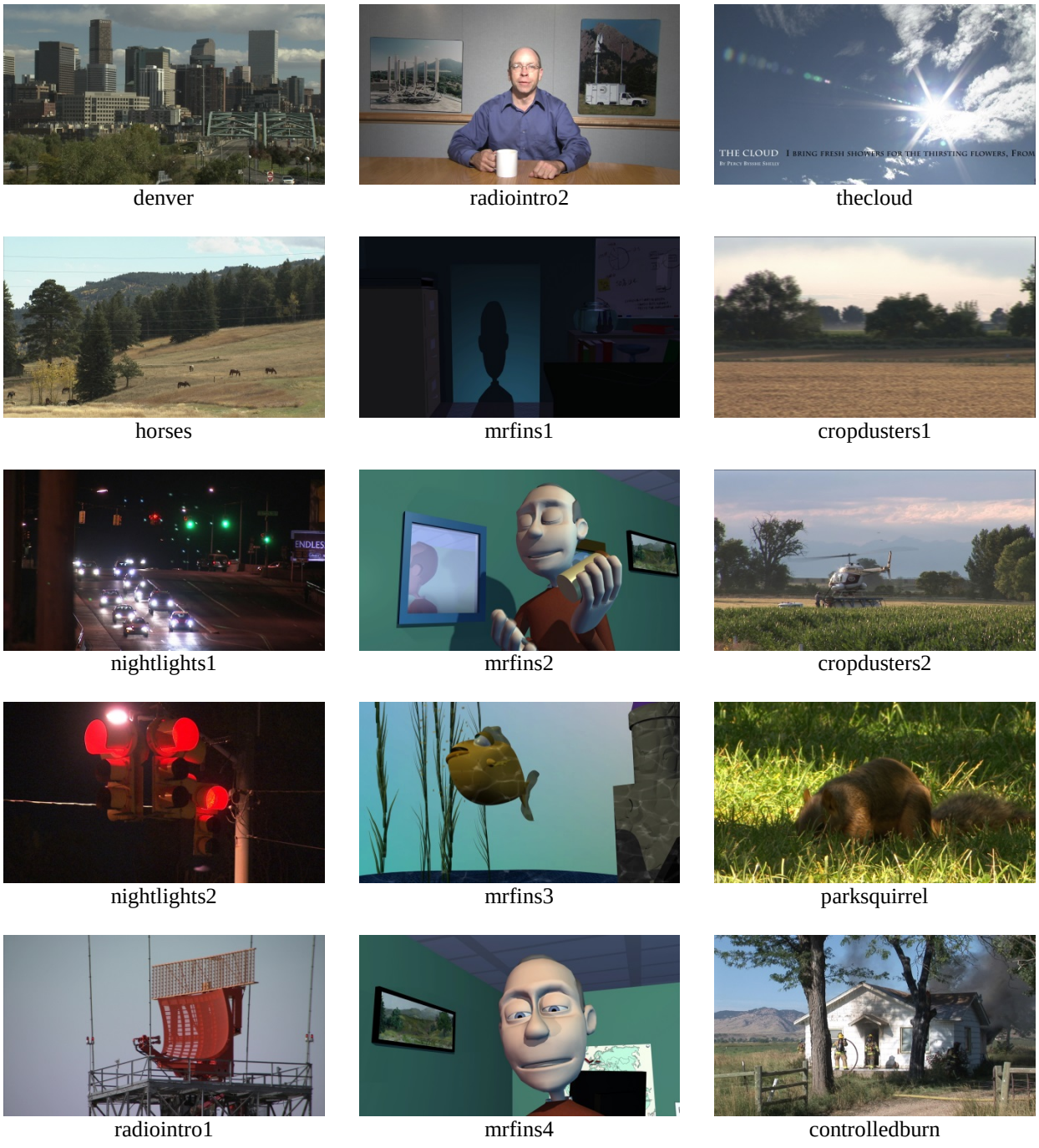
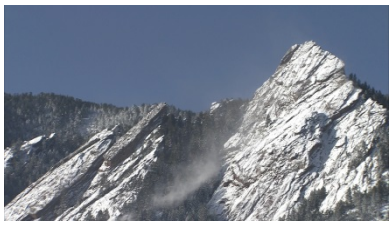


Figure 8. Sample frames from the 15 original clips for CD-SRC low coding difficulty.


snowmnt


bikepark

speedbag


tulip1e

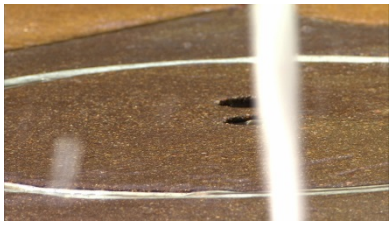

halftimemusic


foxandbird2


aspen


aspenwalk


foxandbird3


mallfountains


lion

foxandbird4


zoocuts


crane

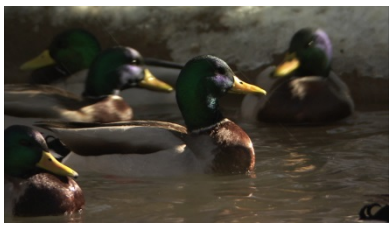

mallmusic

Figure 9. Sample frames from the 15 original clips for CD-SRC medium coding difficulty.

twostreamseasy


snowfall5a


purple1a


duckfeed


rainbo2a

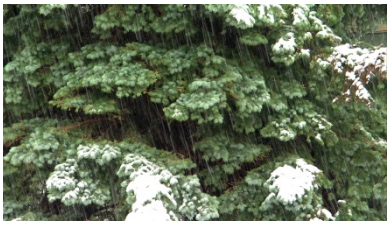
scarletoakeasy


stream

westwind


snowfall3a


bluespruce2a


bluespruce1e


stadpancheer

westwindeasy


bluespruce3a


tulip4e

Figure 10. Sample frames from the 15 original clips for CD-SRC high coding difficulty**.**

3.6 Ordering and Software Control

ITU-R Rec. BT.500, ITU-T Rec. P.910 and ITU-T Rec. P.913 recommend that different groups of subjects view and rate sequences in different orders. The assumption is that the order of PVSs within the subjective test sessions influences subjective ratings. The session ordering issue

differs from “recency,” otherwise known as the “forgiveness effect,” in which the quality of sequences with time varying quality is influenced by the temporal ordering of impairment levels.

For the AGH/NTIA experiment, we would like to see the influence of PVS order on subjective scores. If the PVSs were displayed to subjects in a random order, then it might be difficult to look for patterns. Therefore, we decided to use a fixed ordering of PVSs in each session, and a fixed ordering of sessions to subjects.

Two fixed orders were chosen (red and blue). They were intentionally designed to be different and to contain some non-random elements. The four scene pools (A, B, C, and D) indicate the content for each of the three sessions.

Within each session, the PVSs of the red ordering were chosen as follows:

- Four PVSs with quality goodplus
- Four PVSs with quality poorplus
- The remaining sequences, ordered randomly
- Four PVSs with quality goodminus
- Four PVSs with quality poorminus (these are the last four PVSs in the session)

Within each session, the PVSs of the blue ordering were chosen as follows:

- Four PVSs with quality poorplus
- Four PVSs with quality goodplus
- The remaining sequences, ordered randomly
- Four PVSs with quality poorminus
- Four PVSs with quality goodminus (these are the last four PVSs in the session)

All subjects saw these same sequence presentations. See Appendix B for the presentation order of each session.

This choice is diametrically opposed to a random ordering. If the hypothesis is true (that prior scores have a major influence on current opinion), then we should see a difference among subjects based on the ordering they saw (red or blue). If we see a negligible difference among subjects based on their ordering, then the hypothesis is false. Either way, these ordering differences were intended to help us determine whether people rate each SRC independently throughout the test, or whether people modify their ratings based on what they have already seen in the test.

3.7 Questionnaire

Subjects were given a written questionnaire after each session. The questionnaire after session 1 and 2 had the following two questions:

- What did you like about this session?
- What did you not like about this session?

The questionnaire after session three had the following 15 questions:

1. What did you like about this session?
2. What did you not like about this session?
3. How well did the instructions and training session prepare you for this session? Was there anything missing? Was there extra information you would have appreciated?
4. How did you decide on the quality of video sequences?
5. You saw some video sequences many times. How did this impact the way you decided on quality ratings?
6. You saw some video sequences only once or twice. How did this impact the way you decided on quality ratings?
7. If you saw new content: How much did your the interest in the content impact your rating?
8. You were given an option to replay each sequence.
    a. How frequently did you use this option?
    b. Why did you use this option?
    c. How did the replay option impact your quality judgments?
9. Generally about the experiment:
    a. Was the experiment interesting or boring?
    b. When did you have an easy time staying alert and paying attention?
    c. When was it difficult to pay attention?
10. You are sitting in a simulated living room for this experiment. What is your opinion of this viewing environment?
11. Have you participated in another audio or video subjective experiment, at ITS or elsewhere? [If yes]

a. What was that experiment like? Did you rate quality in that previous experiment, or were you identifying objects?

b. How did your experience today compare with those previous experiments? What did you like better about each one?

c. What do you like or dislike about this simulated living room, in comparison the environments from previous experiments?

12. Did you try to give the same answer to a sequence just because it was similar to a previously seen sequence?

4 REFERENCES

[1] A. Catellier and L. Connors, “Web-enabled subjective test (WEST) research tools manual,” NTIA Handbook HB-14-501, Jan. 2014.

[2] A. Catellier, M. Pinson, W. Ingram and A. Webster, “Impact of Mobile Devices and Usage Location on Perceived Multimedia Quality,” *Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*, Jul. 2012.

[3] C. Fenimore, J. Libert and S. Wolf, “Perceptual effects of noise in digital video compression,” *140th SMPTE Technical Conference*, Pasadena, CA, Oct. 28-31, 1998.

APPENDIX A: SUBJECT SCREENING EXPERIMENT INSTRUCTION SCRIPT

Thanks for coming in today to participate in our study. In this study, we will ask for your opinion of the quality of video sequences.

This experiment is mostly self-paced. There will be three sessions, each from 10 to 20 minutes long. Some sessions are longer than others. During each session you will watch a series of short videos and rate them on an evenly spaced scale from excellent to poor. After each session, you may take a break.

Do you have any questions?

[Note: do eye test]

Welcome to the living room. We want this room to feel like home. We can adjust the lighting level to your comfort. We can turn the ceiling lights, the dragonfly lights, or the fluorescent ceiling lights on or off. We can leave the curtains open or close them.

You may choose to either relax in the easy chair or sit at the table using the office chair. You may change where you are sitting during the test if you wish.

Observe and listen carefully to the entire clip before making your judgment. Keep in mind that you are rating the combined visual quality of the clip rather than the content of the clip. If the subject of the clip is pretty or boring or annoying, for example, please do not consider that when evaluating the overall quality of the clip. Simply ask yourself what you think about the quality of clip.

Remember, there is no right or wrong answer. Everyone’s opinion will be different. We simply want to record your opinion.

In order to familiarize yourself with the device and voting process, we will begin a short practice session. Please ask any questions you have as we go.

When you’re comfortable, play the video shown on the screen by tapping the “Play" button. After the clip is finished playing, a ratings screen will appear. Feel free to replay the video as many times as is necessary. When ready, tap the rating that best describes your opinion of the clip, then tap the vote button. Another video will automatically be loaded. Continue this process until the device says you are finished.

[practice session]

At the end of the test session, I should be waiting in the lab. If I’ve been called away, have a seat. I will return shortly, but if you need to get a hold of me, call an extension listed on the phone.

After the test session I will measure the distance between your eyes and the laptop. I will ask you a few questions, and then you will have a chance to take a break. After that, you will begin the next session.

APPENDIX B: SESSION ORDERING

The PVS order for blue group was as follows. PVSs with a manually chosen order are marked in bold.

Training

1 newsdriving1_poorplus.mp4
2 newsbottledwater_original.mp4
3 newsdogpark_goodminus.mp4
4 newsbottledwater_poorminus.mp4

Session 1

1 **A_thefoot01_poorplus.mp4**
2 **D_nightlights2_poorplus.mp4**
3 **C_walkwayever4_poorplus.mp4**
4 **B_parrot_poorplus.mp4**
5 **C_walkwaybver4_goodplus.mp4**
6 **B_parrot_goodplus.mp4**
7 **D_horses_goodplus.mp4**
8 **A_thefoot01_goodplus.mp4**
9 C_flamenco4_goodplus.mp4
10 B_redkayak_original.mp4
11 D_bluespruce2a_poorplus.mp4
12 D_denver_original.mp4
13 A_thefoot01_poorminus.mp4
14 D_zoocuts_poorminus.mp4
15 D_aspen_goodminus.mp4
16 A_rosepath_goodminus.mp4
17 B_redkayak_goodminus.mp4
18 A_rosepath_original.mp4
19 A_thefoot01_original.mp4
20 A_rosepath_goodminus.mp4
21 A_thefoot01_original.mp4
22 C_questioncarm1ver1_poorminus.mp4
23 D_twostreamseasy_original.mp4
24 A_rosepath_goodplus.mp4
25 A_thefoot01_poorminus.mp4
26 C_flamenco7_goodminus.mp4
27 C_questionassaultf2ver1_original.mp4
28 C_flamenco10_poorplus.mp4
29 C_questioncarf1ver5_poorplus.mp4
30 B_parrot_original.mp4
31 C_flamenco13_poorminus.mp4
32 A_rosepath_poorminus.mp4

33 D_duckfeed_goodplus.mp4
34 B_redkayak_goodplus.mp4
35 A_rosepath_poorplus.mp4
36 B_redkayak_poorplus.mp4
37 D_mallfountains_poorplus.mp4
38 A_rosepath_goodplus.mp4
39 D_westwindeasy_poorminus.mp4
40 A_rosepath_poorminus.mp4
41 D_stream_goodminus.mp4
42 A_rosepath_poorplus.mp4
43 C_walkwaybver1_original.mp4
44 A_thefoot01_goodminus.mp4
45 A_rosepath_original.mp4
46 D_tulip1e_goodplus.mp4
47 D_snowmnt_original.mp4
48 A_thefoot01_goodminus.mp4
49 C_questionassaultf3ver3_goodplus.mp4
50 B_redkayak_poorminus.mp4
51 C_questionassaultm3ver1_goodminus.mp4
52 C_flamenco1_original.mp4
53 **D_radiointro1_poorminus.mp4**
54 **B_parrot_poorminus.mp4**
55 **C_walkwayever7_poorminus.mp4**
56 **A_thefoot01_poorplus.mp4**
57 **B_parrot_goodminus.mp4**
58 **C_walkwaybver7_goodminus.mp4**
59 **A_thefoot01_goodplus.mp4**
60 **D_nightlights1_goodminus.mp4**

Session 2

1 **B_parrot_poorplus.mp4**
2 **A_thefoot01_poorplus.mp4**
3 **D_nightlights2_poorplus.mp4**
4 **C_walkwayever4_poorplus.mp4**
5 **A_thefoot01_goodplus.mp4**
6 **D_horses_goodplus.mp4**
7 **B_parrot_goodplus.mp4**
8 **C_walkwaybver4_goodplus.mp4**
9 D_westwind_goodminus.mp4
10 C_questionassaultf3ver4_goodplus.mp4
11 A_rosepath_goodminus.mp4
12 D_zoocuts_poorminus.mp4
13 A_thefoot01_original.mp4
14 A_rosepath_poorminus.mp4
15 D_mallfountains_poorplus.mp4

16 C_questionassaultm3ver1_goodminus.mp4
17 C_flamenco1_original.mp4
18 D_snowfall5a_original.mp4
19 D_rainbo2a_goodplus.mp4
20 D_duckfeed_goodplus.mp4
21 B_redkayak_goodplus.mp4
22 D_halftimemusic_goodplus.mp4
23 C_walkwaybver1_original.mp4
24 C_flamenco14_poorminus.mp4
25 A_rosepath_goodminus.mp4
26 C_flamenco11_poorplus.mp4
27 A_rosepath_poorminus.mp4
28 B_redkayak_poorminus.mp4
29 D_mrfins2_goodminus.mp4
30 C_flamenco5_goodplus.mp4
31 D_mrfins3_poorplus.mp4
32 C_questionassaultf2ver1_original.mp4
33 D_twostreamseasy_original.mp4
34 B_redkayak_original.mp4
35 C_questioncarm1ver3_poorminus.mp4
36 C_walkwaybver2_original.mp4
37 A_thefoot01_poorminus.mp4
38 C_questioncarf2ver2_poorplus.mp4
39 D_westwindeasy_poorminus.mp4
40 C_questioncarf1ver5_poorplus.mp4
41 D_mrfins1_goodplus.mp4
42 C_walkwaybver8_goodminus.mp4
43 B_redkayak_poorplus.mp4
44 D_aspen_goodminus.mp4
45 C_flamenco7_goodminus.mp4
46 A_rosepath_poorplus.mp4
47 C_walkwaybver5_goodplus.mp4
48 B_redkayak_goodminus.mp4
49 A_rosepath_poorplus.mp4
50 D_aspenwalk_goodminus.mp4
51 D_stream_goodminus.mp4
52 C_walkwayever5_poorplus.mp4
53 D_radiointro2_original.mp4
54 C_flamenco2_original.mp4
55 D_mrfins4_poorminus.mp4
56 A_rosepath_original.mp4
57 C_walkwayever8_poorminus.mp4
58 D_bikepark_original.mp4
59 A_rosepath_goodplus.mp4
60 D_bluespruce3a_poorminus.mp4

61 A_rosepath_original.mp4
62 C_questionassaultf3ver3_goodplus.mp4
63 C_flamenco13_poorminus.mp4
64 C_flamenco10_poorplus.mp4
65 D_denver_original.mp4
66 A_rosepath_goodplus.mp4
67 D_snowmnt_original.mp4
68 D_lion_poorminus.mp4
69 D_tulip1e_goodplus.mp4
70 C_questionassaultm3ver4_goodminus.mp4
71 A_thefoot01_original.mp4
72 C_flamenco4_goodplus.mp4
73 C_questionassaultf2ver3_original.mp4
74 A_thefoot01_poorminus.mp4
75 D_bluespruce2a_poorplus.mp4
76 B_parrot_original.mp4
77 A_thefoot01_goodminus.mp4
78 D_bluespruce1e_poorplus.mp4
79 C_flamenco8_goodminus.mp4
80 D_crane_poorplus.mp4
81 A_thefoot01_goodminus.mp4
82 C_questioncarm1ver1_poorminus.mp4
83 **C_walkwayever7_poorminus.mp4**
84 **B_parrot_poorminus.mp4**
85 **D_radiointro1_poorminus.mp4**
86 **A_thefoot01_poorplus.mp4**
87 **B_parrot_goodminus.mp4**
88 **C_walkwaybver7_goodminus.mp4**
89 **A_thefoot01_goodplus.mp4**
90 **D_nightlights1_goodminus.mp4**

Session 3

1 **B_parrot_poorplus.mp4**
2 **D_nightlights2_poorplus.mp4**
3 **C_walkwayever4_poorplus.mp4**
4 **A_thefoot01_poorplus.mp4**
5 **C_walkwaybver4_goodplus.mp4**
6 **A_thefoot01_goodplus.mp4**
7 **D_horses_goodplus.mp4**
8 **B_parrot_goodplus.mp4**
9 A_thefoot01_goodminus.mp4
10 C_walkwaybver3_original.mp4
11 D_zoocuts_poorminus.mp4
12 C_flamenco15_poorminus.mp4
13 D_controlledburn_poorminus.mp4

14 D_snowmnt_original.mp4
15 D_scarletoakeasy_goodplus.mp4
16 A_thefoot01_original.mp4
17 C_questioncarf2ver3_poorplus.mp4
18 D_snowfall3a_goodminus.mp4
19 C_questionstorem2ver2_poorminus.mp4
20 D_cropdusters1_goodplus.mp4
21 D_foxandbird3_goodminus.mp4
22 D_cropdusters2_goodminus.mp4
23 B_redkayak_goodplus.mp4
24 D_duckfeed_goodplus.mp4
25 C_questionassaultf2ver1_original.mp4
26 C_flamenco12_poorplus.mp4
27 D_speedbag_original.mp4
28 B_parrot_original.mp4
29 A_rosepath_poorminus.mp4
30 D_stadpancheer_poorplus.mp4
31 C_flamenco1_original.mp4
32 C_walkwayever3_goodminus.mp4
33 A_rosepath_poorminus.mp4
34 C_questionassaultf3ver3_goodplus.mp4
35 A_rosepath_goodplus.mp4
36 C_walkwayever9_poorminus.mp4
37 C_questionassaultf2ver5_original.mp4
38 A_thefoot01_poorminus.mp4
39 D_foxandbird4_poorplus.mp4
40 A_thefoot01_poorminus.mp4
41 C_questionassaultf3ver6_goodplus.mp4
42 D_bluespruce2a_poorplus.mp4
43 C_flamenco9_goodminus.mp4
44 A_rosepath_goodplus.mp4
45 D_mallmusic_poorminus.mp4
46 D_foxandbird2_goodplus.mp4
47 C_flamenco10_poorplus.mp4
48 C_questionassaultm3ver1_goodminus.mp4
49 A_thefoot01_original.mp4
50 C_flamenco6_goodplus.mp4
51 A_rosepath_original.mp4
52 C_questioncarm1ver1_poorminus.mp4
53 D_aspen_goodminus.mp4
54 C_flamenco3_original.mp4
55 A_rosepath_poorplus.mp4
56 D_westwindeasy_poorminus.mp4
57 A_rosepath_goodminus.mp4
58 D_tulip4e_poorminus.mp4

59 B_redkayak_original.mp4
60 B_redkayak_poorplus.mp4
61 C_walkwaybver1_original.mp4
62 D_thecloud_original.mp4
63 C_flamenco7_goodminus.mp4
64 D_tulip1e_goodplus.mp4
65 D_stream_goodminus.mp4
66 C_walkwaybver6_goodplus.mp4
67 A_rosepath_poorplus.mp4
68 C_walkwayever6_poorplus.mp4
69 A_rosepath_goodminus.mp4
70 D_parksquirrel_poorplus.mp4
71 D_mallfountains_poorplus.mp4
72 D_purple1a_original.mp4
73 C_questioncarf1ver1_goodminus.mp4
74 C_questioncarf1ver5_poorplus.mp4
75 D_twostreamseasy_original.mp4
76 D_denver_original.mp4
77 C_flamenco4_goodplus.mp4
78 A_thefoot01_goodminus.mp4
79 C_flamenco13_poorminus.mp4
80 B_redkayak_poorminus.mp4
81 A_rosepath_original.mp4
82 B_redkayak_goodminus.mp4
83 **C_walkwayever7_poorminus.mp4**
84 **A_thefoot01_poorplus.mp4**
85 **B_parrot_poorminus.mp4**
86 **D_radiointro1_poorminus.mp4**
87 **A_thefoot01_goodplus.mp4**
88 **C_walkwaybver7_goodminus.mp4**
89 **B_parrot_goodminus.mp4**
90 **D_nightlights1_goodminus.mp4**

The PVS order for red group was as follows. PVSs with a manually chosen order are marked in bold.

Training
1 newsbottledwater_original.mp4
2 newsdriving1_poorplus.mp4
3 newsdogpark_goodminus.mp4
4 newsbottledwater_poorminus.mp4

Session 1
1 **D_horses_goodplus.mp4**

A_thefoot01_goodplus.mp4
C_walkwaybver4_goodplus.mp4
B_parrot_goodplus.mp4
D_nightlights2_poorplus.mp4
A_thefoot01_poorplus.mp4
B_parrot_poorplus.mp4
C_walkwayever4_poorplus.mp4
A_thefoot01_original.mp4
C_questioncarf1ver5_poorplus.mp4
D_zoocuts_poorminus.mp4
C_flamenco10_poorplus.mp4
D_bluespruce2a_poorplus.mp4
A_rosepath_poorplus.mp4
D_aspen_goodminus.mp4
C_questioncarm1ver1_poorminus.mp4
D_twostreamseasy_original.mp4
A_thefoot01_goodminus.mp4
A_rosepath_poorminus.mp4
B_redkayak_goodplus.mp4
A_thefoot01_poorminus.mp4
B_parrot_original.mp4
A_rosepath_poorplus.mp4
C_walkwaybver1_original.mp4
C_questionassaultm3ver1_goodminus.mp4
C_flamenco7_goodminus.mp4
B_redkayak_poorplus.mp4
C_flamenco13_poorminus.mp4
D_denver_original.mp4
A_rosepath_original.mp4
A_thefoot01_poorminus.mp4
D_tulip1e_goodplus.mp4
A_thefoot01_original.mp4
D_duckfeed_goodplus.mp4
A_rosepath_poorminus.mp4
B_redkayak_original.mp4
A_rosepath_goodplus.mp4
C_flamenco4_goodplus.mp4
C_questionassaultf3ver3_goodplus.mp4
D_westwindeasy_poorminus.mp4
C_questionassaultf2ver1_original.mp4
A_thefoot01_goodminus.mp4
A_rosepath_goodminus.mp4
D_snowmnt_original.mp4
A_rosepath_goodminus.mp4
B_redkayak_goodminus.mp4

47 A_rosepath_original.mp4
48 D_mallfountains_poorplus.mp4
49 B_redkayak_poorminus.mp4
50 C_flamenco1_original.mp4
51 D_stream_goodminus.mp4
52 A_rosepath_goodplus.mp4
53 **B_parrot_goodminus.mp4**
54 **C_walkwaybver7_goodminus.mp4**
55 **D_nightlights1_goodminus.mp4**
56 **A_thefoot01_goodplus.mp4**
57 **C_walkwayever7_poorminus.mp4**
58 **B_parrot_poorminus.mp4**
59 **A_thefoot01_poorplus.mp4**
60 **D_radiointro1_poorminus.mp4**

Session 2

1 **D_horses_goodplus.mp4**
2 **B_parrot_goodplus.mp4**
3 **A_thefoot01_goodplus.mp4**
4 **C_walkwaybver4_goodplus.mp4**
5 **A_thefoot01_poorplus.mp4**
6 **C_walkwayever4_poorplus.mp4**
7 **D_nightlights2_poorplus.mp4**
8 **B_parrot_poorplus.mp4**
9 C_flamenco11_poorplus.mp4
10 C_questionassaultm3ver4_goodminus.mp4
11 D_mrfins1_goodplus.mp4
12 C_walkwaybver2_original.mp4
13 D_mrfins4_poorminus.mp4
14 C_flamenco7_goodminus.mp4
15 A_rosepath_original.mp4
16 D_bluespruce2a_poorplus.mp4
17 A_rosepath_goodminus.mp4
18 D_mallfountains_poorplus.mp4
19 D_aspen_goodminus.mp4
20 C_walkwayever8_poorminus.mp4
21 C_flamenco10_poorplus.mp4
22 C_walkwayever5_poorplus.mp4
23 D_lion_poorminus.mp4
24 B_redkayak_goodplus.mp4
25 A_thefoot01_poorminus.mp4
26 C_questionassaultf2ver1_original.mp4
27 D_snowfall5a_original.mp4
28 D_twostreamseasy_original.mp4
29 B_parrot_original.mp4

30 C_flamenco5_goodplus.mp4
31 D_radiointro2_original.mp4
32 D_duckfeed_goodplus.mp4
33 C_questionassaultf2ver3_original.mp4
34 B_redkayak_original.mp4
35 A_thefoot01_poorminus.mp4
36 A_rosepath_goodminus.mp4
37 C_flamenco1_original.mp4
38 A_rosepath_poorminus.mp4
39 C_flamenco4_goodplus.mp4
40 C_questioncarm1ver1_poorminus.mp4
41 D_westwindeasy_poorminus.mp4
42 C_walkwaybver8_goodminus.mp4
43 C_questioncarf2ver2_poorplus.mp4
44 D_bluespruce1e_poorplus.mp4
45 B_redkayak_poorplus.mp4
46 C_flamenco14_poorminus.mp4
47 A_rosepath_goodplus.mp4
48 C_flamenco2_original.mp4
49 D_bikepark_original.mp4
50 A_thefoot01_goodminus.mp4
51 C_questionassaultf3ver3_goodplus.mp4
52 D_zoocuts_poorminus.mp4
53 C_flamenco13_poorminus.mp4
54 D_bluespruce3a_poorminus.mp4
55 D_aspenwalk_goodminus.mp4
56 C_questioncarm1ver3_poorminus.mp4
57 D_westwind_goodminus.mp4
58 D_rainbo2a_goodplus.mp4
59 B_redkayak_poorminus.mp4
60 A_rosepath_poorminus.mp4
61 B_redkayak_goodminus.mp4
62 C_questionassaultf3ver4_goodplus.mp4
63 C_walkwaybver5_goodplus.mp4
64 A_thefoot01_original.mp4
65 A_thefoot01_original.mp4
66 C_questioncarf1ver5_poorplus.mp4
67 D_denver_original.mp4
68 A_rosepath_poorplus.mp4
69 C_flamenco8_goodminus.mp4
70 C_questionassaultm3ver1_goodminus.mp4
71 A_rosepath_goodplus.mp4
72 D_mrfins3_poorplus.mp4
73 D_tulip1e_goodplus.mp4
74 A_rosepath_original.mp4

75 D_stream_goodminus.mp4
76 D_snowmnt_original.mp4
77 A_thefoot01_goodminus.mp4
78 D_mrfins2_goodminus.mp4
79 D_crane_poorplus.mp4
80 D_halftimemusic_goodplus.mp4
81 C_walkwaybver1_original.mp4
82 A_rosepath_poorplus.mp4
83 **D_nightlights1_goodminus.mp4**
84 **C_walkwaybver7_goodminus.mp4**
85 **B_parrot_goodminus.mp4**
86 **A_thefoot01_goodplus.mp4**
87 **D_radiointro1_poorminus.mp4**
88 **A_thefoot01_poorplus.mp4**
89 **C_walkwayever7_poorminus.mp4**
90 **B_parrot_poorminus.mp4**

Session 3

1 **B_parrot_goodplus.mp4**
2 **D_horses_goodplus.mp4**
3 **A_thefoot01_goodplus.mp4**
4 **C_walkwaybver4_goodplus.mp4**
5 **D_nightlights2_poorplus.mp4**
6 **B_parrot_poorplus.mp4**
7 **C_walkwayever4_poorplus.mp4**
8 **A_thefoot01_poorplus.mp4**
9 C_questioncarf1ver1_goodminus.mp4
10 C_flamenco9_goodminus.mp4
11 D_stadpancheer_poorplus.mp4
12 D_parksquirrel_poorplus.mp4
13 C_walkwaybver1_original.mp4
14 D_speedbag_original.mp4
15 C_flamenco7_goodminus.mp4
16 D_cropdusters2_goodminus.mp4
17 C_flamenco13_poorminus.mp4
18 D_snowfall3a_goodminus.mp4
19 B_redkayak_poorminus.mp4
20 D_stream_goodminus.mp4
21 D_mallfountains_poorplus.mp4
22 D_duckfeed_goodplus.mp4
23 A_rosepath_goodplus.mp4
24 B_redkayak_poorplus.mp4
25 A_rosepath_original.mp4
26 C_walkwaybver6_goodplus.mp4
27 D_cropdusters1_goodplus.mp4

28 C_questionassaultf3ver3_goodplus.mp4
29 A_thefoot01_poorminus.mp4
30 D_mallmusic_poorminus.mp4
31 A_thefoot01_original.mp4
32 C_questionassaultf2ver1_original.mp4
33 D_foxandbird3_goodminus.mp4
34 D_thecloud_original.mp4
35 C_questionassaultf2ver5_original.mp4
36 C_questioncarm1ver1_poorminus.mp4
37 D_aspen_goodminus.mp4
38 A_thefoot01_goodminus.mp4
39 C_walkwayever6_poorplus.mp4
40 C_flamenco15_poorminus.mp4
41 C_walkwayever3_goodminus.mp4
42 A_rosepath_goodplus.mp4
43 D_snowmnt_original.mp4
44 D_scarletoakeasy_goodplus.mp4
45 A_thefoot01_original.mp4
46 A_rosepath_poorplus.mp4
47 B_parrot_original.mp4
48 C_walkwayever9_poorminus.mp4
49 A_rosepath_poorplus.mp4
50 D_tulip4e_poorminus.mp4
51 D_twostreamseasy_original.mp4
52 C_questionassaultf3ver6_goodplus.mp4
53 A_rosepath_poorminus.mp4
54 D_foxandbird4_poorplus.mp4
55 C_questionassaultm3ver1_goodminus.mp4
56 C_flamenco6_goodplus.mp4
57 B_redkayak_goodplus.mp4
58 A_rosepath_goodminus.mp4
59 C_flamenco4_goodplus.mp4
60 D_denver_original.mp4
61 C_flamenco12_poorplus.mp4
62 D_purple1a_original.mp4
63 D_westwindeasy_poorminus.mp4
64 A_thefoot01_poorminus.mp4
65 D_foxandbird2_goodplus.mp4
66 D_bluespruce2a_poorplus.mp4
67 C_flamenco1_original.mp4
68 C_questionstorem2ver2_poorminus.mp4
69 C_flamenco10_poorplus.mp4
70 B_redkayak_original.mp4
71 A_rosepath_goodminus.mp4
72 D_zoocuts_poorminus.mp4

C_walkwaybver3_original.mp4
A_thefoot01_goodminus.mp4
C_questioncarf1ver5_poorplus.mp4
D_controlledburn_poorminus.mp4
B_redkayak_goodminus.mp4
C_flamenco3_original.mp4
A_rosepath_poorminus.mp4
D_tulip1e_goodplus.mp4
A_rosepath_original.mp4
C_questioncarf2ver3_poorplus.mp4
A_thefoot01_goodplus.mp4
D_nightlights1_goodminus.mp4
B_parrot_goodminus.mp4
C_walkwaybver7_goodminus.mp4
D_radiointro1_poorminus.mp4
A_thefoot01_poorplus.mp4
B_parrot_poorminus.mp4
C_walkwayever7_poorminus.mp4

FORM **NTIA-29**
(4-80)

U.S. DEPARTMENT OF COMMERCE
NAT'L. TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION

BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION NO.	2. Government Accession No.	3. Recipient's Accession No.
TR-14-505		

4. TITLE AND SUBTITLE	5. Publication Date
AGH/NTIA: A Video Quality Subjective Test With Repeated Sequences	June 2014
	6. Performing Organization Code NTIA/ITS.T

7. AUTHOR(S)	9. Project/Task/Work Unit No.
Margaret H Pinson and Lucjan Janowski	3141012-300
8. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Telecommunication Sciences National Telecommunications & Information Administration U.S. Department of Commerce 325 Broadway Boulder, CO 80305	10. Contract/Grant Number.

11. Sponsoring Organization Name and Address	12. Type of Report and Period Covered
National Telecommunications & Information Administration Herbert C. Hoover Building 14th & Constitution Ave., NW Washington, DC 20230	

14. SUPPLEMENTARY NOTES

15. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)

This report provides full technical details for the video quality subjective test AGH/NTIA. Analyses of this dataset appear in separate publications. The purpose of this document is to provide design details that are beyond the scope of a conference paper or journal article.

Subjective experiment AGH/NTIA includes multiple instances of the same stimuli rated three or six times by the same subject. The goal is to provide insights into the suitability of subject screening methods, the impact of source video reuse on subjective data, and the repeatability of subjects.

16. Key Words (Alphabetical order, separated by semicolons)

subject screening, subjective testing, video quality

17. AVAILABILITY STATEMENT	18. Security Class. (This report)	20. Number of pages
☒ UNLIMITED.	Unclassified	54
☐ FOR OFFICIAL DISTRIBUTION.	19. Security Class. (This page) Unclassified	21. Price: N/A

NTIA FORMAL PUBLICATION SERIES

NTIA MONOGRAPH (MG)

A scholarly, professionally oriented publication dealing with state-of-the-art research or an authoritative treatment of a broad area. Expected to have long-lasting value.

NTIA SPECIAL PUBLICATION (SP)

Conference proceedings, bibliographies, selected speeches, course and instructional materials, directories, and major studies mandated by Congress.

NTIA REPORT (TR)

Important contributions to existing knowledge of less breadth than a monograph, such as results of completed projects and major activities. Subsets of this series include:

JOINT NTIA/OTHER-AGENCY REPORT (JR)

This report receives both local NTIA and other agency review. Both agencies' logos and report series numbering appear on the cover.

NTIA SOFTWARE & DATA PRODUCTS (SD)

Software such as programs, test data, and sound/video files. This series can be used to transfer technology to U.S. industry.

NTIA HANDBOOK (HB)

Information pertaining to technical procedures, reference and data guides, and formal user's manuals that are expected to be pertinent for a long time.

NTIA TECHNICAL MEMORANDUM (TM)

Technical information typically of less breadth than an NTIA Report. The series includes data, preliminary project results, and information for a specific, limited audience.

For information about NTIA publications, contact the NTIA/ITS Technical Publications Office at 325 Broadway, Boulder, CO, 80305 Tel. (303) 497-3572 or e-mail info@its.bldrdoc.gov.