**Alternate Video Quality Subjective Testing Methodology**

*Jessica Hooker, NTIA/ITS*

INTRODUCTION

The purpose of this paper is to briefly outline an alternative subjective testing methodology for measuring video/audio quality. Originally recommended and tested by Sullivan and colleagues (2008), this methodology seeks to more accurately assess video/audio quality using naturally occurring content. Current subjective video/audio quality testing standards outlined in ITU recommendations (ITU-R BT.500 & ITU-T P.910) have, as Kortum and Sullivan (2004) phrase it, “reduce[d] the human observer to the role of a simple [degradation] detector.” The alternative methodology proposed in this paper and by Sullivan and colleagues (2008) seeks to **avoid** this by altering the fundamental research question while still pursuing the same end goal of this research. Instead of asking, “can participants detect degradations at different levels of video quality?” this methodology strives to obtain the same answer, but by asking, “What is the lowest level of video quality that participants find *acceptable* (or *sufficient*) in meeting their expectations?” The current video/audio quality testing standards, however, are not able to accurately answer this question, necessitating the development of the methodology described by Sullivan et al. (2004, 2008, 2010), and in this paper (Watson & Sasse, 1998).

The primary assertion of this alternative methodology is that in order to accurately answer this question, the subjective testing protocol needs to more closely resembles a “naturalistic video viewing experience” (Kortum & Sullivan, 2004). Because this methodology seeks mimic a more naturalistic experience, it can also be utilized in more specific video/audio quality contexts that have difficulties being accurately tested with current recommended standards (e.g., telemedicine, sign language, first responder training, etc.).

A general description of the alternate methodology currently being tested at ITS is described below. This methodology has been slightly modified from the Sullivan et al. (2008) model.

VIDEO SELECTION

One of the most significant differences between the current proposed methodology and current ITU standards is video content. The term *content* refers to the subject matter of clip and not the physical characteristics of the clip (e.g., frame rate, resolution, color depth). Standardized clips used in current video/audio quality testing are developed to target specific aspects of encoding and generally overlook the content of the clip (Kortum & Sullivan, 2004; 2010; Frohlich et al., 2012). In selecting clips on the basis of physical attributes as opposed to content, viewing experience become unnatural for participants, limiting the ecological validity of the test and limiting the applicability of the results. Additionally, previous research has shown video content to significantly impact video quality ratings (Kortum & Sullivan, 2004, 2010; Evans, 2011).

In using this methodology, it is recommended that video content be natural video content (e.g., sporting events, movies, television programming) or “simulated” natural content that closely resembles naturally occurring content. Using this type of video content in the testing environment more closely approximates a “real-world” video viewing scenario than does the standardized clips developed for video quality research. In addition to using natural/simulated natural content for testing, we recommend that all test clips be novel to participants. By using novel content, the influence of previous viewing experiences cannot confound the results of the study.

The current experiment under development at ITS will be using a sample of clips containing a single individual discussing various topics in TV interview format for the initial testing of this methodology. While it is optimal to use a variety of clip content in order to more accurately portray naturalistic viewing, a homogenous sample of clip content has been selected for initial testing.

Another factor of video selection to be considered is video length. Most of the current ITU standards have deemed test clips of 10 seconds or less in length to be sufficient for obtaining accurate video/audio quality ratings. This creates a problem in attempting to simulate naturalistic viewing because people tend to watch television for longer intervals of time (Adams, 2000; Kortum & Sullivan, 2010; Frohlich et al., 2012). In using longer clip lengths, participants are able to regulate their attention to and become engaged with the content of the clip (Frohlich et al., 2012). Frohlich and colleagues (2012) suggest that using video clips with a length long enough for participants to transition into a “natural viewing attitude” may be necessary for attaining ecologically valid results.

Video clip length for the current study is approximately 30 seconds. Due to the nature of the content, video clips were edited to end at natural stops, resulting clips of variable lengths of 30-50seconds. While maintaining a constant video length would be optimal, we have opted to preserve the naturalness of speech within the video clips.

TESTING

A primary difference between this testing methodology and ITU standards is the use of “distractor” questions to obscure the true nature of the test. These distractor questions serve two purposes. First, it helps to ensure that participants are not focusing solely on the video/audio quality of the clip, but are instead focusing on the clip as a whole, as they would in a natural viewing scenario (Kortum & Sullivan, 2004, 2010). Second, the use of distractor questions helps to ensure that participants are giving their natural reactions to the video/audio quality without knowing that it is video/audio quality test (Sullivan et al., 2008).

Distractor questions should be designed in consideration of the video content to be used in the test. A total of 4-6 distractor questions should be created along with the target video/audio quality question. Due to the nature of the test, there will be separate questions for video and audio quality to be considered on an individual basis. The distractor questions should probe the participant for information on several aspects of the video being observed. Examples of distractor questions from a Kortum and Sullivan (2004) experiment that used video clips from feature films can below:

* 1)  What category is this film?
* 2)  What parental rating would you assign this film?
* 3)  How would you rate the quality of the acting in the film?
* 4)  Do you know anything about the movie (circle all that apply) :
* 5)  I would like to see this movie:
* 6)  How would you rate the overall quality of the sound and picture for this movie?

The directions given to participants at the start of the experiment should not inform them of the true purpose of the test to avoid biased results. Instructions should be either vague enough as to prevent them knowing it is a video/audio quality test or instructions should be appropriately misleading in the sense that there is a related, but false purpose of the test.

Because this testing methodology involves the use of longer video sequences than specified in current ITU standards, the total number of impairment levels in a given test must be reduced to 5-6 impairment levels in order to keep the total test length less than 1.5 hours. Also, due to the nature of this testing methodology, it is recommended that for a given experiment, only one aspect of the clip, video or audio, be impaired. For initial testing, we have chosen to impair the audio of the selected clips at a 5 impairment levels.

The training condition should consist of no more than 2 videos; one of the lowest quality and one of the highest quality to be used in the experimental condition. The other primary purpose of training will be to expose the participants to the particular interface they will be uses for the test. The current test will be using 2 videos for the training condition. The training videos will consist of the same individual speaking on different topics and will not be used in the experimental condition.