

Performance Evaluation Plan

Purpose

To evaluate the performance of the technology (Applications + Network Infrastructure) as it relates to the users' perceived quality. Since the network connection is an essential part of the interaction between Stanford and U. of Michigan, we plan to cross correlate the quality performance information with the network metrics information collected during the session. The expected outcome is to evaluate the impact that the network degradation has on the users' perceived quality.

Methodology

We will divide this evaluation in two parts. The first part will involve the subjective evaluation of the different streams of data (Video, Audio, and RSV Application) that are received at both ends of the transmission. In this case the evaluation will grade the perceived quality in terms of the technology's intrinsic features. For example, in case of the received video, the evaluation will score the quality of the received images. The assigned score will be impacted if the images are blurry, delayed, frozen from time to time, or missing pixels, etc. In case of audio, the score will be impacted if the grader perceives choppiness, delay, echo, and/or noise in the transmission. In case of RSV, the score will be impacted if the images are delayed and/or the rotations are delayed. Eventually, if features like the blue-dot don't work that might impact the score as well. In this case we propose as graders individuals who will not be playing the role of student and/or instructor.

The second part of the evaluation will involve the interaction between the Instructor and the students. In this case the Instructors and the students will be required to grade the video, the audio, and the applications in terms of the usefulness of the technology for the specific task that they want to accomplish. This grading might differ from the one performed in the first part. For example, the delay of the RSV's received images might be perceived as reasonable good for an external observer, however, a student using RSV might consider that the images are not received quick enough to successfully complete a lesson's exercise, because he/she might feel that the application's latency is distracting, etc. Also an instructor might disagree with an external observer on the quality of the video, since the instructor might have the perception that the video prevented him from conveying the lesson's contents efficiently to the students.

Evaluation

The table below shows the different elements that we will evaluate in this event for the first part. In this case we propose two observers at Stanford and U. of Michigan to complete this part.

	Grader at Stanford	Grader at U. of Michigan
Video (*)	Juan Carlos Aragon	Alex Terzian
Audio (*)	Juan Carlos Aragon	Alex Terzian
RSV	Juan Carlos Aragon	Alex Terzian
VNC	(No need to grade at Stanford)	Alex Terzian

(*) In case of Audio and Video we will grade the audio and the video every 5 minutes through the event, so that we will be able to cross correlate the assigned scores to the collection of network metrics (which will be collected every 5 minutes during the event). RSV and VNC sessions are very short; consequently they will be graded just once.

For the second part of the evaluation we will grade the following elements presented in the table below. In this case the graders will be the persons involved in the interaction (a.k.a the instructor and the students). The grading will happen at the end of the event.

	Grader at Stanford	Graders at U. of Michigan
Video	Instructor: Dr. Shakti Srivastava	Instructor: Dr. Ameer Raof And Students
Quicktime Mov. (Movie presented by U. of Michigan)	Instructor: Dr. Shakti Srivastava	(No need to grade at U. of Michigan)
Audio	Instructor: Dr. Shakti Srivastava	Instructor: Dr. Ameer Raof And Students
RSV	Instructor: Dr. Shakti Srivastava	Instructor: Dr. Ameer Raof And Students
VNC (Movie presented by Stanford)	(No need to grade at Stanford)	Instructor: Dr. Ameer Raof And Students

The scores for all the evaluations are presented in the table below. In this case we always assume that the transmission between Stanford and U. of Michigan is not perfect (or ideal), and thus there is some level of degradation in the network. The scores below allow us to rate the level of that degradation.

GRADE	Corresponds to	Which means that the degradation...
5	Excellent	is not noticeable
4	Good	is noticeable, but not annoying
3	Fair	is annoying
2	Poor	has made the application very hard to use
1	Bad	has made the application unusable