Perception-based Playout Scheduling for High-Quality Real-time Interactive Multimedia

Zixia Huang, Klara Nahrstedt
University of Illinois at Urbana-Champaign
E-mail: {zhuang21, klara}@illinois.edu

Abstract—**Existing media playout scheduling (MPS) schemes usually focus on selecting and scheduling packets according to optimized Internet media metrics, which are only partially relevant to the subjective human perception in the interactive system. The MPS design challenges are two-fold. First, human *preferences* are concurrently dominated by multiple quality attributes of the streaming media whose perceptual tradeoffs were not well understood, so they were not used as an integral part of an efficient MPS design. Second, people's perceptions can be impacted by the *flicker effect* caused by Internet dynamics and the resulting MPS adaptations. In this paper, we propose a new and adaptive perception-based MPS scheme to deliver high-quality real-time interactive multimedia. We first investigate the perceptual tradeoffs among the multi-modal *bundle streaming qualities* in a real Internet environment. We then present our MPS design that finds the bundle quality tradeoffs, while minimizing flicker degradations. Evaluation results show the performance of our MPS scheme.**

I. Introduction

Modern interactive multimedia technologies can allow distributed users to conduct shared and realistic activities. Tele-immersive (TI) systems [1], [2], for example, enable such collaborations by producing and rendering audio and multi-view video streams (called *multi-stream bundle*) at the (local) sender and (remote) receiver sites in real time (Fig. 1).

The **bundle streaming quality** of a real-time interactive multimedia system is usually decided by multiple quality attributes: (a) *media signal quality*, the clarity of media playout; (b) *interactivity*, the one-way end-to-end delay (EED) of media packets between the sender input devices (microphone/camera) and receiver outputs (speaker/display); (c) *synchronization*, the in-pace playout of audio-visual signals. Because of these coexisting attributes, the overall quality can be represented as a multidimensional quality *point* with each orthogonal dimension representing an individual attribute [3] (Section 2).

Media playout scheduling (MPS) has been heavily studied to reduce the negative impacts of Internet imperfections on the system quality. Unfortunately, existing MPS algorithms create **tradeoffs** among the bundle streaming quality attributes. For example, we have to sacrifice the interactivity quality by increasing the video or audio latencies, so that larger receiver buffers can be employed to equalize greater Internet jitter, and thus, improve the media signal quality. On the other hand, the synchronization skew is determined by the one-way latency difference of the time-correlated audio and videos. We can, of course, properly set the receiver buffer size in order to reduce the skew to minimum, but this approach may affect both media signal quality and interactivity. Because of the quality tradeoffs, it is difficult for MPS to decide a best operating quality point with the optimal human perception during the presence of Internet degradations. Only a set of local *non-dominated* Pareto optimal points can be accessed (i.e., in each such point, it is not possible to improve one quality dimension without worsening the other dimensions).

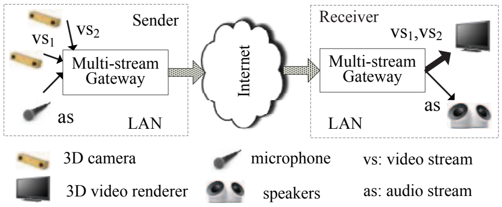

Fig. 1. System architecture for two-site tele-immersion.

Another factor impacting the user perception is the **flicker effect** (i.e., the perceptible change of bundle streaming qualities), which is caused by the network condition variations and system adaptations. Because different operating quality points may not expect the same level of flickers under Internet dynamics and playout controls (Section 3), it is important to locate a point which can minimize the flicker degradation.

Previous work. The multi-modality of the bundle streaming quality creates difficulties in presenting their combined effects on overall **human perception** in a closed form [1], [4]. Hence, offline subjective evaluations have been conducted in literature to guide the online MPS adaptation. However, existing subjective metrics (e.g., MOS and CMOS from ITU-R BT.500 [5]) have been proven unsuccessful in describing the multi-dimensional quality tradeoffs [3], [4]. Hence, the perception-based MPS studies built upon these metrics [6], [7] cannot capture the resulting diversity of user opinions (Section 2). Our previous VoIP studies [4], [8] addressed the MPS tradeoff between the audio signal quality and interactivity, but they did not study the video impacts in the media system and the flicker degradations on human perception.

Our contributions. The goal of this paper is to propose a MPS scheme for real-time interactive multimedia systems to deliver time-correlated multi-sensory multi-stream bundle with high streaming quality and minimal flicker effects. To achieve this, we propose a new subjective metric: *preference*, which is able to identify the *dominant* opinion among diverse user votes based on the hypothesis testing. Using the offline subjective preference results and a novel perception-based genetic approach, we design our online MPS algorithm which

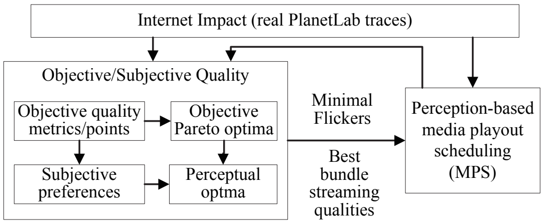
Internet Impact (real PlanetLab traces)
Objective/Subjective Quality
Objective quality metrics/points
Objective Pareto optima
Subjective preferences
Perceptual optma
Minimal Flickers
Best bundle streaming qualities
Perception-based media playout scheduling (MPS)

Fig. 2. Overall framework of the paper

adapts the objective bundle streaming metrics and searches for the locally perceptual optimal quality points. We conduct statistical flicker estimations among these perceptual quality optima under Internet dynamics. Fig. 2 shows the overall framework presented in the paper.

II. Interactive Multimedia Quality Metrics

We investigate both objective and subjective metrics, and discuss their joint implications on the MPS design. For illustration purposes, we will use the TI system as the example in the rest of the paper. WLOG, we assume each TI site is configured with one audio and multiple video streams. We use the 3D TI video codec developed by UC Berkley [9] where there is no coding dependency across the video frames (i.e., video frames are equally important). We define the term *macroframe* as a set of correlated multi-view video frames, which are synchronously produced at the input devices and presented at the media outputs.

A. *Objective Metrics of Bundle Streaming Quality*

Many objective metrics have been identified to describe different aspects of the media bundle streaming quality. To reduce the objective quality space, we only study four user-observable metrics in this paper.

(1) Synchronization quality: audio-visual skew (x_S). x_S is defined as the EED difference of audio frames (EED$_A$) and video macroframes (EED$_V$), i.e., $x_S = \text{EED}_V - \text{EED}_A$. Note that EED includes the media codec latency, the receiver buffer size, the application *transmission delay* (the time to push the media data from the operating system onto a physical link), and the combined Internet propagation and switch/router delay (abbr. *Internet delay*) over the physical links.

(2) Interactivity: one-way EED (x_D). Because EED$_V$ and EED$_A$ can be different, we follow ITU-T G.1070 [7] and give them an equal weight, i.e., $x_D = (\text{EED}_V + \text{EED}_A)/2$.

(3) Audio (x_A) and video (x_V) signal quality. We approximate x_A with a range from 0 to 4.5 based on the distribution of audio frame unavailability at the scheduled playout [8]. Because video frames are equally important, we simply pick the multi-view video macroframe rate to represent x_V.

We use *frames per second* (fps) as the unit of x_V, *milliseconds* (ms) for x_D and x_S, and [0–4.5] for x_A. Hence, the overall system objective quality can be described as a 4-dimensional space with each quality *point* $\vec{x}$ in the space:

$$\vec{x} = \{x_V,\ x_A,\ x_D,\ x_S\} \tag{1}$$

System control. Both x_V and x_A depend on the receiver video and audio buffer sizes, which are decided by the corresponding MPS-controllable EED$_V$ and EED$_A$ values. Because larger delay fluctuations can be smoothed as the result of a prolonged receiver buffer size, x_V and x_A are improved as EED$_V$ and EED$_A$ increase. But this will increase (degrade) x_D and impact x_S (Fig. 3). A mapping can be found between EED$_V$ and EED$_A$, and the objective quality space, given a network condition NC:

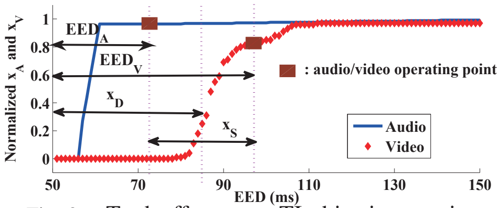

Fig. 3. Tradeoffs among TI objective metrics.

$$\text{Given NC}:\ \{\text{EED}_V,\ \text{EED}_A\} \mapsto \vec{x} = \{x_V,\ x_A,\ x_D,\ x_S\} \tag{2}$$

Because the mapping from the MPS control variables EED$_V$ and EED$_A$ to the objective quality space $\vec{x}$ can be accessed online, objective quality metrics are usually employed for real-time MPS control. Due to a lack of closed form describing their tradeoffs and combined impacts, a MPS scheme is unable to find a best operating (quality) point leading to the optimal overall perceptual quality without the aid of offline subjective evaluations.

B. *Subjective Metrics of Bundle Streaming Quality*

We conduct subjective tests to find a mapping from the objective quality space to the overall human satisfaction QS.

$$\vec{x} = \{x_V,\ x_A,\ x_D,\ x_S\} \mapsto QS(\vec{x}) \tag{3}$$

In our study, we ask people to compare two samples (i.e., TI objective quality points) featuring different quadruple values, and give a rating score 1, 0, or -1 representing the quality of the first sample is {*better, undistinguishable, worse*} compared to the second sample. To capture the diversity of user votes, we compute the percentage of votes for each score: $(pp^s_1,\ pp^s_0,\ pp^s_{-1})$.

Example 2.1: We have shown in [1] the complete subjective evaluation results for the TI application where two people (one person in each site) are conducting a social conversation. An example is the comparison between $\vec{x}^1 = \{20, 4.0, 600, 0\}$ and $\vec{x}^2 = \{5, 4.0, 200, 0\}$. The voting result of 19 subjects is $(pp^s_1,\ pp^s_0,\ pp^s_{-1}) = \{47.4\%,\ 5.2\%,\ 47.4\%\}$.

We follow our previous study [8] to decide the *dominant* opinion from $(pp^s_1,\ pp^s_0,\ pp^s_{-1})$ with $> 50\%$ probability ($p > 0.5$) and a certain level of statistical significance. We model the subjective opinion by a multinomial distribution with 3 outcomes. We then conduct hypothesis testing by selectively combining two options within the 3 outcomes, and describing the *for* or *against* probabilities of the third opinion using an equivalent binomial distribution. [8] prescribes that an option i (i can be 1, 0, -1) is dominant if the following hypothesis is accepted (N is the total number of votes):

$$H_0 : \left(pp^s_i, \sum_{j \neq i} pp^s_j\right) \text{ is drawn from binomial}(N, p \geq 0.5) \tag{4}$$

Definition 2.1: Two objective quality points are "*inconclusive*" if there is no dominant opinion in $(pp^s_1,\ pp^s_0,\ pp^s_{-1})$.

Definition 2.2: A user *preference* is either aligned with the dominant opinion, or with the "inconclusive" comparison.

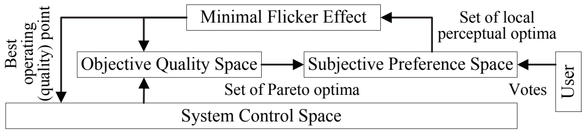

Fig. 4. Interaction of quality metrics with control variables

Example 2.2: For 90% significance, we know from Eqn. 4 that at least 12 out of 19 votes (i.e., at least 63.2% of votes) need to agree on an opinion. Because none of the opinions in Example 2.1 exceeds this number, the comparison is "inconclusive".

The reason leading to an "inconclusive" comparison is that people's perceptions are dominated by different quality attributes of the two points. We have observed from the user study [1] that as two objective points are moving apart on the tradeoff plane, and one dimension is gradually improving as another dimension is worsening, the likelihood of outputting an "inconclusive" comparison is increasing.

C. Metrics for Flicker Effects

An objective quality point can change in two cases: (1) the network condition changes; and (2) the system adaptation moves its operating (quality) points. The flicker effect is incurred as the result of the perceptible change of a quality point. We formulate the flicker to capture this system behavior.

Definition 2.3: The flicker $fd(\vec{x}^1, \vec{x}^2)$ between two points $\vec{x}^1 = \{x_V^1,\ x_A^1,\ x_D^1,\ x_S^1\}$ and $\vec{x}^2 = \{x_V^2,\ x_A^2,\ x_D^2,\ x_S^2\}$ is:

$$\begin{aligned} fd(\vec{x}^1, \vec{x}^2) = w_V \cdot |x_V^1 - x_V^2| + w_A \cdot |x_A^1 - x_A^2| \\ + w_D \cdot |x_D^1 - x_D^2| + w_S \cdot |x_S^1 - x_S^2| \end{aligned} \tag{5}$$

where w_V, w_A, w_D and w_S are the normalized weights.

Subjective evaluations show that a larger flicker generally creates a greater human discomfort [1].

D. Implications on MPS Design

Fig. 4 shows the interaction of MPS control variables and the resulting subjective/objective metrics. The control variables can only achieve the MPS adaptations by employing objective metrics serving as direct online performance indicators of different quality attributes. But due to the tradeoffs among the objective metrics, only a set of non-dominated Pareto optimal points can be obtained. Because these objective points cannot describe the real user satisfactions, offline subjective evaluation results are needed for selecting the perceptual optimal points among the Pareto optima online. We employ our proposed subjective preference metric in the offline user study due to its capability to capture the user opinion diversity. As discussed in Section 2.B, two objective quality points can be mutually "inconclusive". Thus, a total quality order of these points cannot be obtained, and only a partial order can be accessed. In other words, we are unable to obtain a globally perceptual optimal point, but rather several points which are locally perceptual optimal, and thus, are perceptually no "worse" than each other (i.e., mutually "undistinguishable" or "inconclusive"). We then select the quality point among these local optima with the minimal flicker estimation, as the best online operating point. We will present the detailed design in Section 3.

III. Design of Perception-based MPS

In this section, we present a new perception-based MPS scheme for real-time interactive multimedia applications that adapt the objective quality operating points under Internet dynamics, and consistently deliver high subjective media bundle perception with minimal flicker effects.

At each MPS control update, the selection of a new operating (objective quality) point expects the following three **goals**. First, the candidate point must be decided dynamically based upon the recent (**short-term**) Internet conditions. Second, because the network condition may change between two consecutive control updates, the candidate should also be robust to Internet dynamics, in the sense that we prefer a candidate (and its corresponding control values $\{\text{EED}_V,\ \text{EED}_A\}$) that will introduce a smaller flicker due to any potential Internet condition variations (based on **long-term** Internet observations). Third, the receiver control prefers a new operating point closer to the point immediately before the control update so that the flicker effect due to the system adaptations can be reduced. Hence, the overall MPS scheme can be divided into three steps. We will use the TI setting in Section 2 for illustrations.

A. ***Searching Local Perceptual Optima***

The playout control updates the local perceptual optima according to the most recent short-term Internet conditions, by finding the tradeoffs among the quality points and employing human preference results to select among the resulting Pareto optima. The difficulty, however, is that there are infinite number of objective quality points on the operating plane, so searching over the entire plane can be costly. Here, we propose a perception-based genetic approach to address the issue.

The tradeoffs among multiple objective quality metrics in the continuous space can be formulated as a multi-objective optimization problem. The genetic algorithm has been proven successful in locating Pareto optimal points efficiently. In our study, we modify the original genetic algorithm and incorporate the subjective preference metric, so that the new algorithm is able to search the local perceptual optima.

In the TI application, each objective quality metric within $\vec{x} = \{x_V,\ x_A,\ x_D,\ x_S\}$ can be regarded as an objective function in the multi-objective optimization formulation, and $\vec{u} = \{\text{EED}_V,\ \text{EED}_A\}$ is the input control vector. Based on the Internet statistics within the most recent (short-term) timing window, we are able to find the mapping from the control vector to the objective metrics. We utilize the genetic NSGA-II implementation [10] for its computation efficiency. The key idea of the genetic approach is that in each iteration, it generates new *populations* (i.e., a set of $\vec{u} = \{\text{EED}_V,\ \text{EED}_A\}$ at different values) using the standard crossover/mutation methods. It then prioritizes the solutions based on their *fitness* values (defined below), and selects those with better fitness as the populations in the next iteration. We make modifications to NSGA-II to incorporate the human subjective preferences.

First, we redesign the fitness value of the solutions based on two factors. (1) Original Pareto ranking in NSGA-II [10]. For each solution $\vec{u}^i$ in the current iteration, we compute the number of other solutions in the populations that dominate this solution, and denote it as $h(\vec{u}^i)$. A solution $\vec{u}^1$ whose $h(\vec{u}^1) = 0$ means a non-dominated solution. (2) Subjective preference status. A non-dominated solution $\vec{u}^2$ that is perceptually no "worse" (in terms of the preference metric) than any other solution up till the current iteration is given the top priority. Here, subjective comparisons are conducted between the corresponding objective quality points of the solutions. We set the fitness value of $\vec{u}^2$ to $h(\vec{u}^2) = -1$. We call $\vec{u}^2$ as a *perceptual elitist*. Hence, in our algorithm, a solution with a smaller fitness value h represents a better solution.

Second, we maintain the complete list of perceptual elitists, which is always used as a part of the populations of the next iteration. In each iteration, a solution in the elitist list is removed from the list if (a) it is dominated by other solutions in the population, or (b) it is perceptually "worse" than a new non-dominated solution. Note that multiple mutually "undistinguishable" solutions can be crowded together. To maintain the diversity within the elitist list, we prescribe that for any two solutions in the list, their Euclidean distance should not be less than a minimal threshold δdu (= 50 ms in our evaluations).

Our online perception-based genetic algorithm terminates beyond certain preset timing constraints. It will return solutions most approximate to the actual local perceptual optima that are mutually "undistinguishable" or "inconclusive".

B. Statistical Flicker Estimation

Among the local perceptual optima obtained above, we estimate their flickers under potential Internet changes. We conduct the statistical estimation based on the previous long-term records of Internet statistics of the same connection.

We describe the network condition by three parameters: the Internet delay average d_0, the jitter j and the bandwidth availability r. The Internet random losses are negligible, and we assume they can be concealed by the media codec. Except d_0, whose variation is very small, the other two parameters can be described as random variables J and R. We assume their mutual independence. Their distributions can be estimated based on long-term record statistics. We also describe the video macroframe size sv as a known distribution SV, and the codec latency for each video macroframe is a constant d_c.

Given a fixed bandwidth r, the video macroframe delay is represented as a function r:

$$D_V(r) = SV/r + d_0 + J + d_c \tag{6}$$

As discussed in Section 2.A, the video delay includes the codec delay d_c, the Internet delay (d_0 and J), and the transmission delay SV/r. $D_V(r)$ is decided by the distributions of both J and SV. On the other hand, the audio frame delay is mainly decided by the Internet delay d_0 and J. The audio codec and transmission delays are negligible.

$$D_A = d_0 + J \tag{7}$$

Suppose the candidate local perceptual optimum is $\vec{x}^c = \{x_V^c,\ x_A^c,\ x_D^c,\ x_S^c\}$ with the corresponding control values EED_V^c and EED_A^c. We compute the percentage of video macroframes or audio frames (denoted as g_V and g_A) that arrive within EED_V^c and EED_A^c, given the Internet jitter and transmission delay variations based on S and J distributions.

$$g_A = P(D_A \leq \text{EED}_A^c), \quad g_V(r) = P(D_V(r) \leq \text{EED}_V^c) \tag{8}$$

Here, $P(\cdot)$ represents the probability of the input expression. Because $(1 - g_A)$ and $(1 - g_B)$ are actually the unavailable percentage of audio frames and video macroframes, which directly decide x_A and x_V (discussed in Section 2.A), we know the mapping F_V/F_A from g_V/g_A to x_V/x_A:

$$F_V \ : \ g_V \mapsto x_V \quad F_A \ : \ g_A \mapsto x_A \tag{9}$$

Therefore, $|F_A(g_A) - x_A^c|$ and $|F_V(g_V) - x_V^c|$ can be used to approximate the change of the audio and video signal quality from $\vec{x}_A^c$ and $\vec{x}_V^c$ due to the Internet condition changes.

As the bandwidth r varies according its distribution R, the expected difference of the media signal quality (denoted as δx_V and δx_A) over R at $\vec{x}^c$ becomes:

$$\delta x_A|_{x_A^c} = |F_A(g_A) - x_A^c| \tag{10}$$

$$\delta x_V|_{x_V^c} = \int_r |F_V(g_V) - x_V^c| f_R(r)\ dr \tag{11}$$

By plugging Eqn. 10 and 11 into Eqn. 5, we finally come up with the expected flicker fda at the candidate $\vec{x}^c$, assuming δx_A and δx_V are mutually independent.

$$fda|_{\vec{x}^c} = w_A \cdot \delta x_A|_{x_A^c} + w_V \cdot \delta x_V|_{x_V^c} \tag{12}$$

C. Online Adaptation

The receiver MPS updates the operating point periodically based on the results from Section 3.A and 3.B. It also takes into account the system adaptation flickers.

We let the set of local perceptual optima be $\vec{\mathcal{X}}^C = \{\vec{x}^1, \ldots \vec{x}^c, \ldots\}$, and their corresponding statistical flicker estimations be $fda|_{\vec{x}^c}$ (Eqn. 12). We assume the operating point before the MPS update be $\vec{x}^1$. We compute the flicker between $\vec{x}^1$ and $\vec{x}^c$ incurred as the result of the system adaptation, and denote it as $fdb|_{\vec{x}^c} = fd(\vec{x}^c, \vec{x}^1)$ (Eqn. 5).

The flickers are introduced jointly by both Internet variations and system adaptations. We use a heuristic weighted linear function: $w_1 \cdot fda + w_2 \cdot fdb$ to estimate their combined impact. The reason we study fda and fdb separately is that various multimedia applications usually attach different importances to the two factors, so we are able to assign them different weights in each application. Hence the receiver selects the best operating point among $\vec{\mathcal{X}}^C$ that can minimize the combined flicker impact.

$$\vec{x}^{opt} = \arg \min_{\vec{x}^c \in \vec{\mathcal{X}}^C} \{w_1 \cdot fda|_{\vec{x}^c} + w_2 \cdot fdb|_{\vec{x}^c}\} \tag{13}$$

By computing EED_V and EED_A values from $\vec{x}^{opt}$, the receiver then sets its buffer sizes and schedules the media packets to be sent to the output devices accordingly.

IV. Performance Evaluation

We develop a real TI testbed to evaluate our perception-based MPS scheme. Each TI site is configured with 1 audio and 3 video streams. To make our results repeatable, we prerecord at the sender the size of audio frames and video macroframes, so that our testbed is able to transmit the same media data in each experiment. These video and audio data are also transmitted between PlanetLab nodes to collect the real

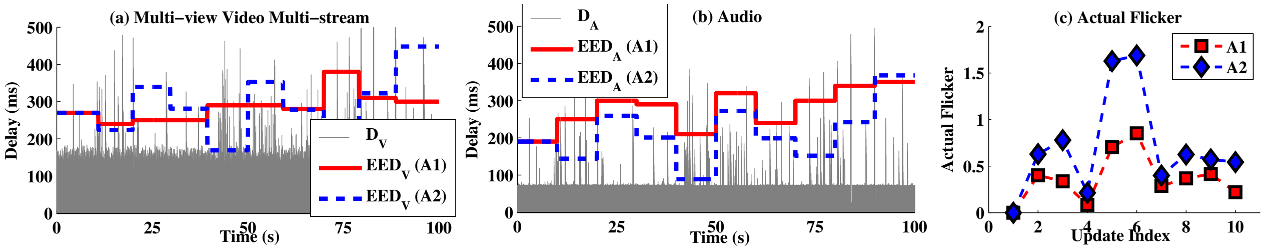

Fig. 5. Comparison between perception-based MPS (A1) and non-perception-based algorithm (A2) [2]. Internet traces are between IL,USA to the Netherlands. D_V and D_A are defined in Eqn. 6 and 7. $d_c = 70$ ms. Flicker values are shown for each MPS update.

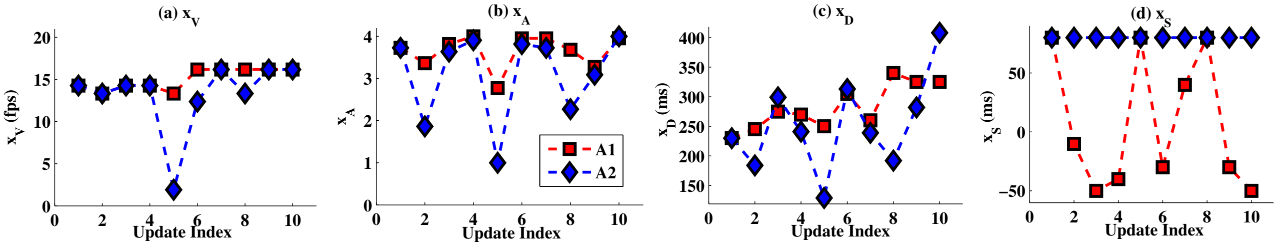

Fig. 6. Objective quality metric values for each 10-sec duration. A1: perception-based MPS. A2: non-perception-based algorithm.

Internet traffic for evaluation purposes. A network emulator is then implemented between the TI sender and receiver gateways to replay the PlanetLab delay and loss distributions. Due to space limit, our paper will only show the experiment results between IL,USA and the Netherlands with a fixed bandwidth of 15 Mbps over 100-second duration. We use the subjective user study in [1] to guide our MPS adaptation. In [1], two people (one in each TI site) are conducting social conversation.

As a comparison, we also evaluate the non-perception-based MPS algorithm previous studied [2], [8]. In [2], [8], EED$_A$ is decided based on the α-percentile of the cumulative distribution function (CDF) of D_A (Eqn. 7), i.e., CDF$_{D_A}$(EED$_A$) $= \alpha$, using the most recent delay statistics. We then set EED$_V$ $=$ EED$_A + 80$, where 80 ms is the maximum audio-visual skew that cannot be noticed [1], [2].

Fig. 5 and Fig. 6 present the comparison results. In perception-based MPS, we set the normalized weights in the flicker formulations (Eqn. 5) to be the inverse of the maximal possible value of each objective quality metric in the TI system (i.e., $w_V = 1/20$, $w_A = 1/4$, $w_D = 1/400$ and $w_S = 1/400$). We also heuristically set $w_1 = 0.2$ and $w_2 = 0.8$ in Eqn. 13. In non-perception-based algorithm, we set $\alpha = 0.95$. For both algorithms, we update the MPS control every 10 seconds. The most recent 10-second duration is used for characterizing the short-term network conditions by both algorithms.

Fig. 5(a) and (b) present the variations of EED$_V$ and EED$_A$ in response to Internet dynamics. They show that both algorithms are able to adapt the video and audio receiver buffer sizes based on the Internet jitter. However, our perception-based MPS outputs far smaller fluctuation magnitudes. This can be explained by two reasons. First, a local optimal objective quality point may not have the best media signal quality, so both EED$_V$ and EED$_A$ do not have to accommodate all delay spikes. Second, our perception-based MPS algorithm takes into account the flicker effects caused by the system control, so it purposefully reduces EED$_V$ and EED$_A$ fluctuations. In addition, we find that the robustness of the non-perception-based algorithm to a sudden Internet change depends on the short-term timing window size, while our perception-based MPS sets a more reasonable value for both EED$_V$ and EED$_A$ because of the internal statistical flicker estimation mechanism.

Within each 10-second duration, we calculate the operating points of both algorithms (Fig. 6) and the resulting flickers experienced by users under the combined impacts of both receiver control updates the real Internet dynamics (Fig. 5(c)). We show that our perception-based MPS scheme outputs a smaller flicker compared to the non-perception-based version, and that it offers less fluctuations in the audio and video signal quality, and the interactivity. These results prove the credit of our perception-based MPS algorithm.

Acknowledgement. This study is supported by NSF CNS 0834480, 0964081, 1012194, IIP 1110178, by UPCRC grant from Intel and Microsoft, and by Grainger Grant.

References

[1] Z. Huang, A. Arefin, P. Agarwal, K. Nahrstedt, and W. Wu, "Understanding the human perceptions in tele-immersive shared activity," University of Illinois Technical Report, Urbana, IL, 2011.

[2] Z. Huang, W. Wu, K. Nahrstedt, A. Arefin, and R. Rivas, "Tsync: A new synchronization framework for multi-site 3D tele-immersion," in *Proc. ACM NOSSDAV*, Jun. 2010.

[3] B. Sat and B. W. Wah, "Statistical scheduling of offline comparative subjective evaluations for real-time multimedia," *IEEE Transaction on Multimedia*, vol. 11, no. 6, pp. 1114–1130, Oct. 2009.

[4] B. Sat and B. Wah, "Playout scheduling and loss-concealments in VoIP for optimizing conversational voice communication quality," in *Proceedings of ACM MM*, Sep. 2007, pp. 137–146.

[5] ITU-BT.500, "Methodology for the subjective assessment of the quality of television pictures," 2002.

[6] A. Meddahi, H. Afifi, and G. Vanwormhoudt, ""MOSQoS": Subjective VoIP Quality for Feedback Control and Dynamic QoS," *IEEE ICC*, 2006.

[7] ITU-G.1070, "Opinion model for video-telephony applications," 2007. [Online]. Available: http://www.itu.int/rec/T-REC-G.1070/en/

[8] Z. Huang, "The design of a multi-party VoIP conferencing system," M.S. thesis, University of Illinois at Urbana-Champaign, Urbana, IL, 2009.

[9] R. Vasudevan and et al, "A methodology for remote virtual interaction in teleimmersive environments," in *ACM MMSYS*, 2010.

[10] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, pp. 182–197, 2002.