Optimizing Multidimensional Perceptual Quality in Online Interactive Multimedia

Benjamin W. Wah, *The Chinese University of Hong Kong, Shatin, Hong Kong, China*
Jingxi X. Xu, *Tencent, Shenzhen, Guangdong, 518054, China*

Network latencies and losses in online interactive multimedia applications may lead to a degraded perception of quality, such as lower interactivity or sluggish responses. We can measure these degradations in perceptual quality by the just-noticeable difference, awareness, or probability of noticeability (p_{note}); the latter measures the likelihood that subjects can notice a change from a reference to a modified reference. In our previous work, we developed an efficient method for finding the perceptual quality for one metric under simplex control. However, integrating the perceptual qualities of several metrics is a heuristic. In this article, we present a formal approach to optimally combine the perceptual quality of multiple metrics into a joint measure that shows their tradeoffs. Our result shows that the optimal balance occurs when the p_{note} of all the component metrics are equal. Furthermore, our approach leads to an algorithm with a linear (instead of combinatorial) complexity of the number of metrics. Finally, we present the application of our method in two case studies, one on VoIP for finding the optimal operating points and the second on fast-action games to hide network delays while maintaining the consistency of action orders.

The high availability of the Internet has led to the pervasiveness of interactive multimedia applications, such as audio/video conferencing, massive multiplayer online games (MMOs), remote surgery,[1] and the metaverse.[2] These applications are built on the high availability and large bandwidth of today's Internet with stationarity in its near-future network behavior. However, end-to-end delay fluctuations may affect the perceptual quality of demanding interactive multimedia applications.

This article focuses on one-to-one and one-to-many conferencing applications and fast-paced games. In the latter, the reaction times are close to the limit of human reaction time and where attack actions are much faster than players' movements in virtual space. We assume the response requirements in conferencing applications to be no more than 400 ms and a human reaction time in MMOs of around 215 ms.

1070-986X © 2023 IEEE
Digital Object Identifier 10.1109/MMUL.2023.3277851
Date of publication 18 May 2023; date of current version 28 September 2023.

A multimedia application generally has multiple controls mapped to quantitative metrics. Modifying the controls will lead to changes in some or all metrics. Examples of quantitative metrics include signal quality and interactivity, whereas examples of controls include buffers to control end-to-end delays and delays to control action initiations.

In this article, we are interested in finding good operating points of interactive multimedia applications that optimize the user-perceived quality at runtime. In the literature, there are two general approaches to define perceptual quality: *just-noticeable distortion* (*JND*) and *awareness*.

In psychophysics, *JND* is the minimal change of the original input [or *reference* (ref)] to its modification (or *distortion*: $ref + mod$), whose effect can be perceived by humans.[3] In contrast, *awareness* p is a probability defined as follows.

Definition 1: Awareness $p(ref, mod)$ is the probability that a subject can identify a changed setting of a multimedia system. In this setup, we ask a subject to make a pairwise comparison of two outputs of an application: one due to a reference input (ref) and the

other a modification (mod) of the reference ($ref+mod$).[3] Given p, ref, and mod, we can relate them in a JND surface $p(ref, mod)$.

As some subjects give the correct response by random guesses, p is larger than the probability of users who notice a change correctly (called the probability of noticeability).

Definition 2: The probability of noticeability p_{note}[4] is the likelihood that subjects can correctly notice a change from ref to $ref+mod$. We can relate $p_{\text{note}}(ref, mod)$ similarly to a JND surface.

To derive p_{note}, we know that, for N independent subjects, Np_{note} of the subjects will notice the change, and $N(1-p_{\text{note}})$ will respond by random guesses. Hence, $Np_{\text{note}} + 0.5N(1-p_{\text{note}}) = pN \Rightarrow p_{\text{note}} = 2p-1$. We are interested in cases with $p_{\text{note}} > 0.5$ ($\Rightarrow p > 0.75$) when more than half of the subjects can correctly identify the change. We generally use this value in psychophysical studies as a threshold for measuring perceptibility. As a well-defined function relates awareness and probability of noticeability, we use them interchangeably.

It is challenging to find the control values to optimize the user-perceived p_{note} because the function relating p_{note} to controls is ill defined. Heuristic methods, such as past studies on quality of experience,[5] do not address the tradeoffs among the nonanalytic perceptual quality metrics. A better approach is learning their relation using offline subjective tests conducted under a broad set of operating conditions and generalizing the property found to runtime. However, subjective tests are expensive to operate even for one scenario, and infinitely many past scenarios grow with the number of control values. Further, there is no single optimal operating point due to the tradeoffs among the multiple metrics.

We have developed, in our past work,[4] a dominance relation that allows us to efficiently find the JND surface of a single metric and one control (see the "Single-Metric Perceptual Quality" section). However, we could only develop a heuristic approach to combine the perceptual quality of multiple metrics.

Our goal in this article is to develop a formal model to optimally combine the perceptual quality of multiple metrics into a joint measure that shows the tradeoffs of the various metrics (see the "Multimetric Perceptual Quality" section). The joint measure leads to an efficient algorithm for finding the optimal tradeoff points under one or more controls and generalizing the results learned offline to runtime operations.

We present the application of our model in two case studies. First, the "Case Study on VoIP" analyzes a VoIP application with multiple metrics under simplex control. Then, the "Case Study on MMOs" shows methods for fast-action games to enforce the consistency of action orders under dependent controls and network delays. Table 1 summarizes the symbols used.

SINGLE-METRIC PERCEPTUAL QUALITY

We first summarize our previous work on evaluating the perceptual quality of a single metric in a JND surface under simplex control.

We have observed, from our experiments on real-time multimedia, the monotonicity of p_{note} to one reference ref and its modification mod. The following hypothesis states this property.

TABLE 1. A summary of the controls and metrics defined.*

Symbol	Meaning
m	$m = (m_1, ..., m_k)$ with k controls for controlling the operation of the application.
q	$q = (q_1, ..., q_n)$ with n quality metrics for measuring perceptual quality, each represented in a ref-mod JND surface.
(ref-mod)	Reference inputs and their modifications in the n JND surfaces, each over a common $2k$-D space, where $ref = (ref_1, ..., ref_k)$, $mod = (mod_1, ..., mod_k)$; in general, when $ref = m$, then $mod = \Delta m$.
$p_{\text{note}}^{q_i}(ref, mod^i)$	Probability of noticeability of metric q_i, $i = 1, ..., n$, where $mod^i = (mod_1, ..., mod_i{}', ..., mod_k)$ is perturbed from mod by only changing the ith component.
$p_{\text{note-f}}^{\text{comb}}(ref, mod)$	Probability of noticeability of the combined metric based on the n quality metrics.
VoIP	$k = 1$ (MED); $n = 2$ with two metrics (ASQ, INT); $ref = \text{MED}$; $mod = \Delta\text{MED}$.
MMO	$k = 1$: the single $m = (m_1)$ is selected from a set of three independent controls applied in isolation or in conjunction and constrained (each leading to a JND surface); $n = 1$ with a metric on the user-perceived delay; ref is the duration of the bullet fired; mod is the one-way network latency.

*ASQ: audio signal quality; MMO: massive multiplayer online game.

Axiom 1: The awareness (or p_{note}) in a JND surface has the monotonicity properties to the reference (ref) and its modification (mod), respectively.[4,6] (a) Awareness is monotonically nonincreasing to ref for given mod. (b) Awareness is monotonically nondecreasing to mod for given ref. (c) For multidimensional (multi-D) modifications, awareness is monotonically nondecreasing when all the component mods are nondecreasing. (d) The boundary case is $p_{\text{note}} = 0$ (or $p = 0.5$) when $mod = 0$.

The hypothesis requires the continuity and smoothness of p_{note} to ref and mod, respectively. These properties are valid because changes to p_{note} in a small region of ref and mod are perceptually indistinguishable.[7] It is a generalization of Weber's law,[3] which hypothesizes a linear relationship between the 1-D ref and 1-D mod.

We illustrate the hypothesis using a VoIP application. The application has two metrics $q =$ [audio signal quality (ASQ), interactivity (INT)] representing signal quality and interactivity in two JND surfaces. The surfaces have the same ref-mod axes, based on the control end-to-end mouth-to-ear delay (MED) and its modification ΔMED. Here, *interactivity* measures the efficiency of a conversation, namely, the fraction of time extended between a conversation with and without a network delay.

Figure 1 plots the color density of the p_{note} of ASQ and INT for each ref-mod combination. For instance, a case with $ref = \text{MED} = 120$ ms and $mod = \Delta\text{MED} = 0$ ms is similar to that of no network delay and no changes to the network delay, and no subject will detect the change ($p_{\text{note}} = 0$). In contrast, when $ref = 120$ ms increases to 200 ms (with $mod = 80$), all subjects will detect the change ($p_{\text{note}} = 1$).

Note that, as mod increases, the p_{note} of ASQ (respectively, INT) will improve (respectively, degrade). To differentiate between metrics that improve versus degrade with increasing mod, we add a negative sign to its p_{note} when the metric degrades with increasing mod, knowing that its absolute value is the actual p_{note}. For example, ASQ (respectively, INT) improves (respectively, degrades) with increasing mod at given ref. Hence, we add a negative sign to the p_{note} of INT. The graphs show that, for a given ref, p_{note} satisfies conditions (a), (b), and (d) in Axiom 1. Further, their absolute p_{note} are monotonically nondecreasing to mod at given ref.

Note that the graph is not symmetric; that is, $p^{\text{A}}_{\text{note}}(ref_A, mod_A) \neq p^{\text{B}}_{\text{note}}(ref_A + mod_A, 0) = 0$ [Axiom 1(d)]. This case happens because the two probabilities are incomparable under different references. However, *A* and *B* have the same *relative perceptual quality* as follows.

Property 1: Relative perceptual quality: Assume two actions A and B with, respectively, $p^{\text{A}}_{\text{note}}(ref_A, mod_A)$ and $p^{\text{B}}_{\text{note}}(ref_B, mod_B)$ on a JND surface. If $ref_B + mod_B > ref_A + mod_A$, and the metric improves (respectively, degrades) with increasing mod, then B has a higher (respectively, lower) or the same relative perceptual quality compared to that of A.

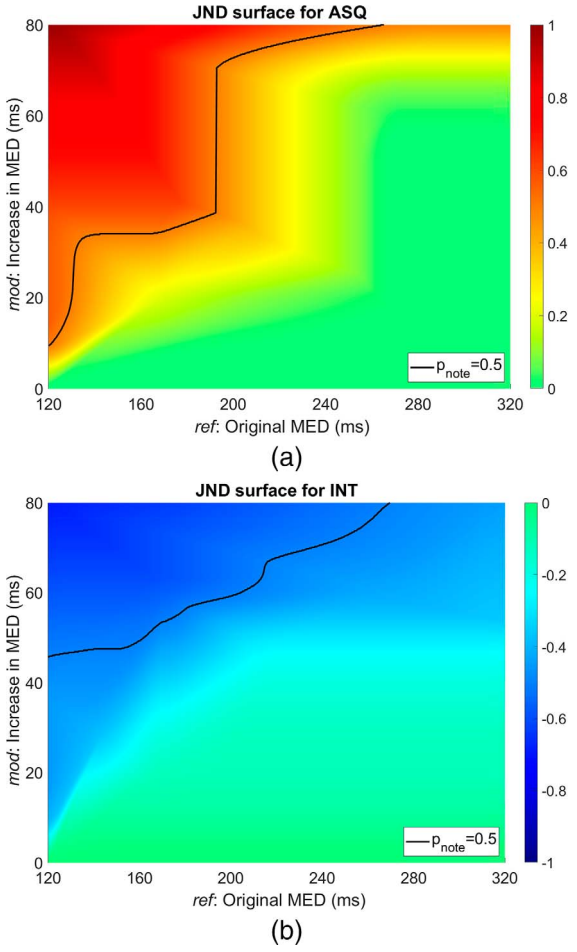

FIGURE 1. (a) The JND surface of the ASQ metric for a VoIP application in an error-prone network relating p_{note} to MED (ref) and ΔMED (mod). The surface is found by interpolating the results of seven subjective tests. (b) The JND surface of INT. The color bar on the right shows the p_{note} levels: an absolute value indicates the fraction of subjects detecting a change; a negative value shows that perceptual quality degrades as mod increases. INT: interactivity; JND: just-noticeable distortion; MED: moth-to-ear delay.

To prove the property, we first transform A and B to a common reference ref_{cm}, as we cannot compare their p_{note} at different references. Without loss of generality, let $ref_{\text{cm}} \leq \min(ref_A, ref_B)$. After the transformation, A

is changed from (ref_A, mod_A) to $(ref_{\text{cm}}, ref_A - ref_{\text{cm}} + mod_A)$, and B is changed from (ref_B, mod_B) to $(ref_{\text{cm}}, ref_B - ref_{\text{cm}} + mod_B)$. At ref_{cm}, $ref_B - ref_{\text{cm}} + mod_B > ref_A - ref_{\text{cm}} + mod_A$ according to our assumption. Hence, the property follows from the monotonicity of p_{note} at ref_{cm} in Axiom 1. □

Axiom 1 allows us to deduce the following error bounds on the p_{note} of points in a rectangular JND region. Consider two points $p_A(ref_A, mod_A)$ and $p_B(ref_B, mod_B)$, and the surface is monotonically nondecreasing (respectively, nonincreasing) with mod (respectively, ref). If $ref_A < ref_B$ and $mod_A > mod_B$, then, for any point $p_M(ref_M, mod_M)$ where $ref_A \leq ref_M \leq ref_B$ and $mod_A \geq mod_M \geq mod_B$, we have $p_A \geq p_M \geq p_B$. Hence, we can bound the p_{note} of any point in the region by those of its diagonal corner points A and B.

This observation leads to pruning the evaluation of those points on a JND surface whose p_{note} is within a threshold δ from its actual value.

Property 2: Dominance pruning on p_{note} to within an error threshold δ. Consider two diagonal corner points A and B of a rectangular block on a JND surface with p_{note} $p_A \geq p_B$, respectively. If $p_A - p_B \leq \delta$, then any point C in the block has p_{note} p_C where $p_A - p_C \leq \delta$ and $p_C - p_B \leq \delta$. Hence, we can approximate p_C by $p_C' = p_A$ or p_B, and $|p_C' - p_C| \leq \delta$.

Based on Property 2, we have developed an efficient binary-divide algorithm to find an approximate JND surface with p_{note} within a threshold δ on the uncertainty of p_{note}. We start by testing the upper-left (*A*) and lower-right (*B*) diagonal points of a rectangular block of the surface (Figure 2). We stop the process if $p_A - p_B \leq \delta$. Otherwise, we conduct subjective tests at the diagonal's center point (*M*) and divide the surface into four regions. As the two butter-colored blocks have p_{note} bounded by their two corner points, we can approximate the top-left (respectively, bottom-right) block when $p_A - p_M \leq \delta$ (respectively, $p_M - p_B \leq \delta$). Since we cannot evaluate the two azure-colored regions by p_M, we can apply the same process to each.

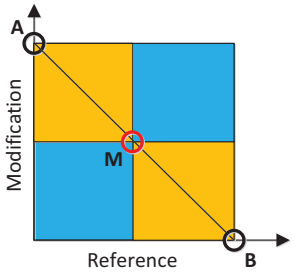

FIGURE 2. A rectangular block on the JND surface. The diagonal indicates the direction of monotonicity, and points *A*, *B*, and *M* have subjective tests conducted.

By subdividing a region into smaller blocks, we can measure the p_{note} at their diagonal points and midpoints. We then approximate the p_{note} of a block when its uncertainty is within δ. Finally, we interpolate the p_{note} in the surface using those points verified by subjective tests. For example, we can interpolate the surface in Figure 1(a) by subjective tests on seven points using $\delta = 0.01$.[4]

This section presents a dominance relation on p_{note}, based on its monotonicity to ref and mod. The property can approximate all the points on a JND surface because they are within an uncertainty threshold compared to the p_{note} of those points already evaluated. The result leads to an efficient realization of a JND surface.

MULTIMETRIC PERCEPTUAL QUALITY

In this section, we extend the earlier result to an application with *n* perceptual quality metrics $q = (q_1, ..., q_n)$ coupled through k controls $m = (m_1, ..., m_k)$. The modification of some or all of the controls leads to changes in the metrics. For each of the n metrics, we represent its p_{note} in a JND surface, defined over a common $2k$-dimensional ref-mod space.

The joint optimization to find a composite p_{note} based on the n JND surfaces is complex because the composite p_{note} does not satisfy Axiom 1 (discussed later). Hence, we cannot apply Property 2 and our binary-divide algorithm to find an approximate JND surface. This condition leads to a combinatorial number of subjective tests to map the joint surface. This section aims to develop the theory for finding the optimal joint JND surface without any new subjective tests.

Numerous past studies exist on combining the JND surfaces due to multiple metrics. Examples include taking the maximum JND[8]; taking square roots to calculate the combined JND[9]; computing an integration function on an explicit JND[10]; evaluating a linear or heuristic function of quality metrics in perceptual evaluation of speech quality (International Telecommunication Union Recommendation G.107) to map control inputs to perceptual quality; and computing the probabilities of cases with consistent awareness but assuming random guesses in cases with inconsistent awareness.[4] These methods are deficient because they do not consider the objective of optimizing p_{note}, namely, the probability that a subject can detect a change.

In the following, we present the theory for optimizing the aggregate p_{note} based on the p_{note} of the individual JND surfaces.

Assumption 1: At a common (ref, mod) point across the n JND surfaces (each representing a metric), users will notice a metric when its corresponding p_{note} increases in subjective tests.

The assumption is reasonable because the p_{note} of a metric represents the probability that subjects would identify a change in this metric.

Definition 3: At a given (ref, mod) across the n surfaces, let $p^{\text{comb}}_{\text{note-f}}(ref, mod)$ be the p_{note} of the combined metric, and $mod = (mod_1, ..., mod_k)$. We assume that $p^{\text{comb}}_{\text{note-f}}(ref, mod)$ is an unknown function f of $p^{q_i}_{\text{note}}(ref, mod^i)$ of metric q_i:

$$p^{\text{comb}}_{\text{note-f}}(ref, mod) = f(p^{q_i}_{\text{note}}(ref, mod^i) | i = 1...n). \quad (1)$$

Here, mod in $p^{q_i}_{\text{note}}(ref, mod)$ is perturbed to mod^i in $p^{q_i}_{\text{note}}(ref, mod^i)$, where the difference between mod and mod^i is in their ith component with mod_i changed to $mod_i^{'}$.

According to Assumption 1, users will notice q_i when perturbing mod to mod^i with mod_i changed to $mod_i{}^{'}$ and $mod_j,\ j \neq i$, unchanged.

Lemma 1: Assume (ref, mod) under the following conditions: (a) at any time, only one metric q_i is perturbed by changing mod to mod^i, and (b) users can perceive the metric with a larger p_{note} at (ref, mod). Then, users can detect the change in metric j with $mod_j{}^{'} > mod_j$ when

$$p_{\max} = \max_{i=1,...,n} p^{q_i}_{\text{note}}(ref, mod^i) = p^{q_j}_{\text{note}}(ref, mod^j). \quad (2)$$

This follows from Axiom 1(c) on the nondecreasing property of p_{note} for the multiple metrics when mod of q_i is perturbed to mod^i. Moreover, by enforcing only one component in mod to be perturbed at any time, users can detect a change in the metric with the largest p_{note}. We do not allow changes in multiple components, as simultaneous changes can reinforce each other, leading to $p^{\text{comb}}_{\text{note-f}}(ref, mod) > p_{\max}$.

Lemma 2: $p^{\text{comb}}_{\text{note-f}}(ref, mod)$ is bounded by $p_{\max}$,

$$p^{\text{comb}}_{\text{note-f}}(ref, mod) \leq p_{\max}. \quad (3)$$

The proof is straightforward by combining Definition 3 and Lemma 1. Equality in (3) occurs when metric q_j in Lemma 1 has mod perturbed to mod^j, and an inequality happens when metric $q_k,\ k \neq j$ is perturbed.

Assumption 2: Given (ref, mod), assume that

$$p^{\text{comb}}_{\text{note-f}}(ref, mod) = \max_{i=1,...,n} p^{q_i}_{\text{note}}(ref, mod^i). \quad (4)$$

This is intuitively correct due to Lemma 2: subjects tend to notice the metric with the largest p_{note} at (ref, mod). Based on the continuity and smoothness of p_{note}, subjects will continue to notice the metric with the largest p_{note} when mod of $p^{q_j}_{\text{note}}(ref, mod)$ is perturbed to mod^j while keeping the other metrics fixed. This case is true even though $p^{\text{comb}}_{\text{note-f}}$ is an unknown function.

To derive the proper tradeoffs among the n metrics, the p_{note} of some metrics must improve while others degrade when increasing mod at given ref. This condition can happen naturally in the application or be enforced by constraining the controls. In the latter case, when the p_{note} of some metrics improves, the p_{note} of others degrades. Note that the tradeoffs of metrics are not possible when the p_{note} of all of the metrics are improving or degrading simultaneously with increasing mod.

In the tradeoffs performed, we avoid one metric having improvements leading to a larger $p_{\max}$ and at the expense of one or more other metrics. Our goal is to solve the following optimization problem $\mathcal{P}$ by choosing mod to minimize $p^{\text{comb}}_{\text{note-f}}(ref, mod)$, namely, the probability of subjects noticing a change:

$$\mathcal{P} = \min_{mod=(mod_1,...,mod_k)} p^{\text{comb}}_{\text{note-f}}(ref, mod). \quad (5)$$

Although we do not know f in (5) where f is defined in (1) in the term minimized, we can apply Lemma 2 and replace $p^{\text{comb}}_{\text{note-f}}(ref, mod)$ by its upper bound $p_{\max}$, assuming only one component in mod is perturbed at any time:

$$\overline{\mathcal{P}} = \min_{mod=(mod_1,...,mod_k)} p_{\max}. \quad (6)$$

We can now derive the optimal solution to (6).

Theorem 1: The optimal solution to (6) is $\overline{mod} = (\overline{mod}_1, ..., \overline{mod}_k)$ across the n metrics, where

$$p^{q_1}_{\text{note}}(ref, \overline{mod}) = ... = p^{q_n}_{\text{note}}(ref, \overline{mod}). \quad (7)$$

Proof: The proof is done by contradiction.

If (7) were false, then, without loss of generality, assume $p^{q_i}_{\text{note}}(ref, \overline{mod}) > p^{q_j}_{\text{note}}(ref, \overline{mod}), i \neq j$. We first note Axiom 1(c) and our requirement that metrics with tradeoffs cannot be monotonically increasing or decreasing simultaneously. We further assume that $p^{q_i}_{\text{note}}$ (respectively, $p^{q_j}_{\text{note}}$) is monotonically nondecreasing (respectively, nonincreasing) with mod. To get $\min_{mod} p_{\max}$, we should reduce mod as much as possible. However, when reducing $mod,\ p^{q_i}_{\text{note}}$ will decrease, but $p^{q_j}_{\text{note}}$ will increase, leading to a possibly larger $p_{\max}$.

Assume that $\overline{mod}$ happens when $p_{\text{note}}^{q_i}(ref, \overline{mod}) = p_{\text{note}}^{q_j}(ref, \overline{mod} + \delta_1),\ \delta_1 \geq 0$. However, we can always find δ_1', where $p_{\text{note}}^{q_i}(ref, \overline{mod}) = p_{\text{note}}^{q_j}(ref, \overline{mod} + \delta_1') < p_{\text{note}}^{q_j}(ref, \overline{mod} + \delta_1)$ for $\delta_1 > \delta_1' \geq 0$ (based on monotonicity). This case will lead to a smaller or equal $p_{\max}$. Hence, it contradicts the assumption that $p_{\text{note}}^{q_j}(ref, \overline{mod} + \delta_1)$ is optimal. Thus, following a similar argument, we can show that δ_1 is zero at the optimal $\overline{mod}$.

We can apply this argument to every pair of counteracting metrics to arrive at some equal p_{note}. By transitivity, it is implied that $p_{\text{note}}^{q_i}$ in (7) must be equal for all metric $i = 1, \ldots, n$. □

Using Theorem 1, we present the following algorithm to find a composite JND surface by searching the n JND surfaces and by identifying the set of (ref, mod) points that satisfy (7).

As ref can be multi-D, we discretize it into k_1 points in the ref space. Similarly, we discretize the multi-D mod into k_2 points. At a given (ref, mod) point (out of $k_1 k_2$ points), we approximate each of the n p_{note} by k_3 decimal places (using the floor function) to avoid the difficulty of matching them in real space. A discretization level of two decimal places would be adequate. The time complexity is $O(n)$ for each point.

Over each of the $k_1 k_2$ points, we match the discretized p_{note} across all the n surfaces. This step has a time complexity of $O(n)$ for each point. Next, we report the optimal points across the n surfaces. Those points with equal p_{note} are the optimal solution to (7). For points not satisfying (7), the p_{note} of their surfaces change at different magnitudes. The implication is that subjects will identify the metric(s) with the largest p_{note}. The total complexity is, therefore, $O(k_1 k_2 n)$.

We can also derive the composite JND surface as a byproduct of the algorithm. It is a multi-D contour graph with ref and mod as its axes. The p_{note} at each of the $k_1 k_2$ discretized points denote the maximum p_{note} across the n surfaces (Assumption 2). The trajectory with all of the p_{note} equal is the optimal solution to (7).

In this section, we have extended our previous result on finding the p_{note} of a single metric to the composite p_{note} of multiple metrics. By assuming that users will notice the metric with a higher p_{note} and that the composite p_{note} is an unknown function driven by the maximum p_{note} across the metrics, we prove that the optimal control to minimize the probability of subjects noticing changes occurs when all the component p_{note} are equal. Further, the results allow us to find the composite p_{note} with complexity based on the discretization levels of the individual JND surfaces. In short, our approach has a linear (instead of combinatorial) complexity of the n metrics.

In the following sections, we illustrate two case studies that use the composite surface to generalize to the optimal runtime operating points.

CASE STUDY ON VoIP

In this section, we apply the result in the last section on the VoIP application described earlier. The application has two metrics with $q = (q_1, q_2) = (\text{ASQ},\ \text{INT})$ and one control with $m = (m_1) = (\text{MED})$. We choose MED as ref because it is an end-to-end delay reference. Figure 1 shows the JND surfaces of the two counteracting metrics ASQ and INT.

The following result identifies the optimal operating points of (7).

Corollary 1: From Theorem 1, we conclude that the optimal operating points $\overline{mod}$ in the composite ASQ–INT JND surface would satisfy

$$p_{\text{note}}^{\text{ASQ}}(ref, \overline{mod}) = p_{\text{note}}^{\text{INT}}(ref, \overline{mod}). \tag{8}$$

Note that there may not be a solution on $\overline{mod}$ to (8) at some ref. This case happens when the surfaces of ASQ and INT do not intersect.

Based on the algorithm in the last section, we search over the two JND surfaces to find $\overline{mod}$ that satisfies (8). The algorithm has $O(k_1 k_2)$ complexity, where $k_1 = k_2 = 500$ are, respectively, the discretization levels of ref and mod used.

Figure 3 shows the composite JND surface found by our search method using the surfaces in Figure 1. Because the negative sign of INT only indicates that the metric degrades with increasing MED, we remap it to an absolute [0,1] scale without changing its meaning. We then find the p_{note} at a given mod across all ref by taking the maximum of the corresponding p_{note} of ASQ and INT using (4). The p_{note} of ASQ and INT are equal at the trajectory and satisfy (8) to within a discretization threshold of 0.005.

The p_{note} at each point in the composite JND surface represents the p_{note} that users would experience in subjective tests. Note that these points do not satisfy the monotonicity property in Axiom 1. For example, the p_{note} to the trajectory's left (respectively, right) corresponds to that of ASQ (respectively, INT) with a larger p_{note}. On the trajectory, we have the minimum p_{note} across all ref (although not visible due to the minor differences). Hence, the composite JND surface cannot be found by the binary-divide method in the "Single-Metric Perceptual Quality" section because Axiom 1 is not satisfied.

Note that the p_{note} on the trajectory are monotonically nondecreasing with increasing mod to within our discretization threshold of 0.005. (Each is computed

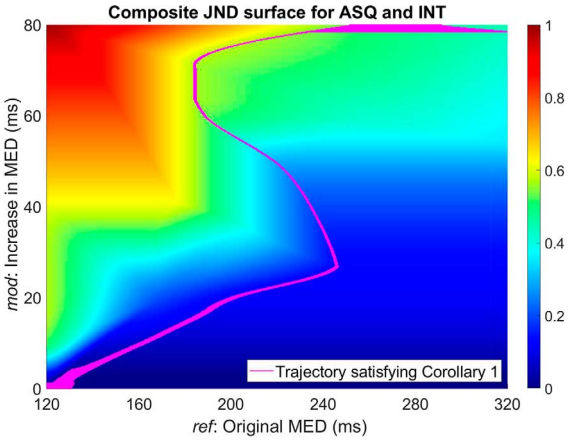

FIGURE 3. Using the ASQ and INT surfaces in Figure 1, the magenta trajectory depicts points satisfying (8). The p_{note} at (ref, mod) on the trajectory is the probability that subjects would not be able to perceptually detect whether ASQ or INT is better compared to ref. On the left (respectively, right) of the trajectory, ASQ (respectively, INT) is more noticeable.

based on a different reference but can be compared after mapping them to a common reference using Property 1.)

The points on the trajectory allow us to choose the optimal operating MED of the application. Note that there may be multiple optimal mod along the trajectory for each MED. In each case, we pick the minimum mod among the candidates because that point will have the smallest p_{note}.

In Figure 3, let the current MED be 120, 200, 246, 280, and 320 ms, respectively. Then, the optimal $\overline{mod}$ would be 0, 19.4, 26.4, 80.0, and 78.5 ms, respectively, by finding the minimum mod on the trajectory (and not by visual inspection of Figure 3). These lead to the optimal MED of 120, 219.4, 272.4, 360, and 398.5 ms, respectively.

The last part of the study is to generalize the offline results to runtime operations. The JND surfaces obtained offline may differ at runtime when the operating condition changes.

Under the same type of conversation, when the measured average delay or the extra buffers for concealing lost packets increases at runtime, the net effect is an increase in the one-way MED. In this case, we assume that the increased MED is to maintain the same level of ASQ as before. The result is to shift the ASQ surface by the extra delay to the left without changing the surface itself. On the other hand, the INT surface remains unchanged because the new network condition would not change the human sensitivity to interactivity. We can then find the best operating points at runtime with the adjusted surfaces.

When the type of conversation changes (from one with a slow to a fast turn frequency or from a casual conversation to a business one), the corresponding JND surface for INT will be different and need to be collected offline.[4] This situation may infrequently happen, as the various INT surfaces do not differ much. In this situation, we can generate the new runtime surfaces of INT by interpolations.

CASE STUDY ON MMOs

In this section, we summarize our previous work[6,11] on MMO and develop a new approach to find the composite JND surface. The application has one control ($k = 1$) based on three independent controls [local lag (LL), local perception filter (LPF1), and LPF2] applied in isolation or conjunction and constrained. Since there are four ways of applying the controls, there are, altogether, four JND surfaces (LL, LPF1, LPF2, and combined). In addition, it has one metric ($n = 1$) on the delay perceived by users, with its ref, the duration of the bullet fired.

We define an *action* as a translation or a movement of a virtual object triggered by a player. A *reference order* (respectively, *actual order*) is an order of actions completed in virtual space without (respectively, with) network delays. Reordering happens when the reference order is different from the actual order. Figure 4(a) illustrates the reordering of two actions between *A*'s and *B*'s views when carried out under network delays.

To maintain consistent outcomes under delays, we like to have *strong consistency*, which requires identical actual and reference orders (thus satisfying the consistency and correctness requirements of Mauve et al.[12]). This condition ensures the consistent outcomes of actions. It implies that overlapping moves from different players with delays must complete in the same order as the original reference order without delays.

We focus, in this section, on the simple case of two players and MMO games with predictable target responses. In this case, a player needs not wait for the target's reaction before proceeding. We do not address cases with unpredictable target responses, as the solution approach is similar. We consider single precise weapons that can attack one target at a time and realize their effects when the action completes.

We aim to reduce the perception of delays between an online MMO with network delays and the same version without delays while satisfying the strong consistency of action orders. The problem is crucial because

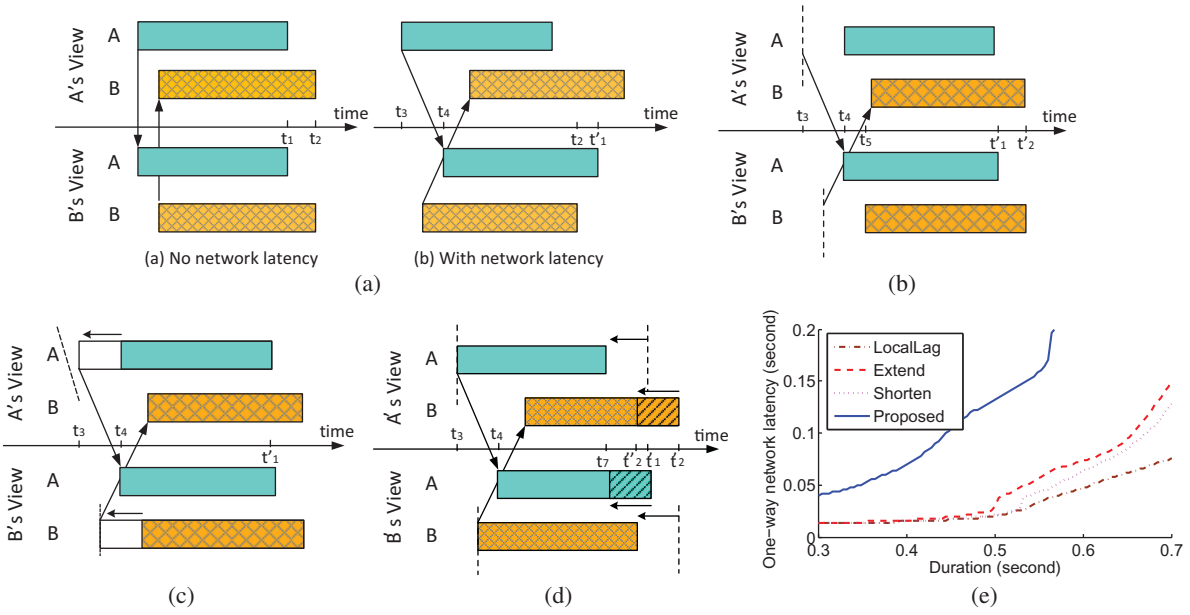

FIGURE 4. Solving the (a) reordering problem[6] using the (b) local lag and the (c) and (d) local perception filters. (e) The $p_{\text{note}} = 0.5$ contours on the JND surfaces of the three primitive and our proposed strategies. (a) Reordering of actions under network delays. (b) Local lag delayed by the most prolonged latency. (c) Extending the durations of local actions. (d) Shortening the durations of remote actions. (e) Evaluation results with $p_{\text{note}} = 0.5$ on BZFlag.

it reduces the perception of delays when running the game online.

To evaluate the result, we have developed an evaluation platform by modifying the code of an open source online tank0battle game, BZFlag.[13] The game has rapid actions as well as fast interactions. To better represent fast-paced games, we modified the base speed of a bullet from the original 100 units/s to 300 units/s. We also made some reasonable assumptions. 1) We estimated the durations of actions in each player based on the speed of the shot and the i to j virtual distance and made them available to all the players. 2) The maximum measured network delay does not change rapidly over time. To conduct subjective tests, we hired 14 students. We showed each subject two cases, one in a reference network without latency and another with the control and latency included. We then asked each to choose the scenario with a slower response. We report p_{note} on whether the subjects could correctly identify the second scenario.

PRIMITIVE STRATEGIES FOR STRONG CONSISTENCY

We examine three primitive strategies in the literature for adjusting the starting times and the durations of actions to ensure strong consistency.

Using LL to Delay the Start of Actions for Each Local Player

Figure 4(b) shows the adjustment by estimating the most prolonged network latency $t_4 - t_3$ and delaying the local action by this amount but without the remote actions. Since the completion times are not changed, strong consistency is maintained.

In evaluating $p^{\text{LL}}_{\text{note}}(ref, mod^{\text{LL}})$, we note that mod^{LL} modifies the action duration governed by the one-way latency. Hence, the JND surface is a 2-D graph with the x-axis as the action duration common in all of the virtual spaces, and the y-axis mod^{LL} equals the one-way latency.

Figure 4(e) shows the $p_{\text{note}} = 0.5$ contour for players to notice a difference from the reference without latency correctly. The LL strategy is subpar because 50% of the subjects can identify the scenario with a short 20-ms one-way latency.

Using LPFs to Extend the Local Action

Starting from the LL setting, the algorithm extends the action in each local player's view to cover the empty period before it starts while keeping the completion time unchanged [Figure 4(c)]. The optimal extension is the one-way latency $(t_4 - t_3)$ from the player who started the action to the receiver: a more extended

period will lead to a more prominent and noticeable change, whereas a shorter extension is infeasible due to the definition of strong consistency. Again, mod^{LPF1} is the one-way latency. Figure 4(e) shows that the result is less noticeable than the LL strategy when applied to actions of the same duration under the same latency.

Using LPFs to Shorten the Remote-Action Duration

Based on the LL state, the algorithm starts the local action in each player's view earlier. It then terminates the remote player's action earlier while keeping its starting time unchanged (and communicating the change to the local player) [Figure 4(d)]. The optimal reduction is the one-way latency $(t_4 - t_3)$ from the player who started the action to the receiver. Figure 4(e) shows less noticeable results than the LL case with significant action durations and one-way delay of mod^{LPF2}.

PROPOSED CONTROL STRATEGY

Figure 5 shows our proposed approach that makes minor adjustments after combining the three primitive strategies.[6] We have observed that minor adjustments on the action duration would be less perceivable as long as the total changes are within the worst-case one-way delay.

The following theorem shows the constraint on the adjustments of the durations in LL, LPF1, and LPF2 to maintain strong consistency.

Theorem 2: Necessary and sufficient condition for strong consistency[6]: Strong consistency is maintained when the total adjustments made by the LL (mod^{LL}) and the LPFs (mod^{LPF1} and mod^{LPF2}) equals $D_{\max}$:

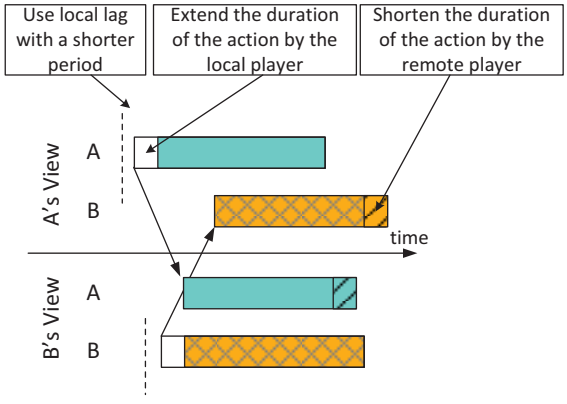

FIGURE 5. The combined strategy comb that makes small adjustments to local lag and local perception filters 1 and 2.

$$mod^{\mathrm{LL}} + mod^{\mathrm{LPF1}} + mod^{\mathrm{LPF2}} = D_{\max}. \tag{9}$$

Readers can find the proof in the article by Xu and Wah.[6]

Theorem 2 provides the condition to allow tradeoffs among the controls. For example, when increasing one value, one or both of the other values must decrease to satisfy (9).

To derive $p^{\mathrm{comb}}_{\text{note-f}}(ref, mod)$, we take the maximum p_{note} of its components based on (4):

$$p^{\mathrm{comb}}_{\text{note-f}}(ref, mod) = \max_{x \in \{\mathrm{LL}, \mathrm{LPF1}, \mathrm{LPF2}\}} (p^{x}_{\mathrm{note}}(ref, mod^{x})). \tag{10}$$

We want to improve the case when applying the three primitive strategies in isolation. Hence, the maximization in (10) is taken over the p_{note} of those three strategies when applied individually.

Similar to (5), we solve $\mathcal{P}$ by choosing $mod = (mod^{\mathrm{LL}}, mod^{\mathrm{LPF1}}, mod^{\mathrm{LPF2}})$ at given ref to minimize $p^{\mathrm{comb}}_{\text{note-f}}(ref, mod)$ subject to (9):

$$\mathcal{P} = \min_{x \in \{\mathrm{LL}, \mathrm{LPF1}, \mathrm{LPF2}\}} p^{\mathrm{comb}}_{\text{note-f}}(ref, mod^{x}). \tag{11}$$

By applying Theorem 1, we can solve (11) by minimizing the perceptual effect of the most dominant artifact.

Corollary 2: The optimal durations of the three adjustments in mod to solve (11) happens at $\overline{mod} = (\overline{mod}^{\mathrm{LL}}, \overline{mod}^{\mathrm{LPF1}}, \overline{mod}^{\mathrm{LPF2}})$ when

$$\begin{aligned} p^{\mathrm{LL}}_{\mathrm{note}}(ref, \overline{mod}^{\mathrm{LL}}) &= p^{\mathrm{LPF1}}_{\mathrm{note}}(ref, \overline{mod}^{\mathrm{LPF1}}) \\ &= p^{\mathrm{LPF2}}_{\mathrm{note}}(ref, \overline{mod}^{\mathrm{LPF2}}). \end{aligned} \tag{12}$$

We can apply the algorithm in the "Multimetric Perceptual Quality" section to solve (11) and (12) under the constraint defined in (9).

In evaluating the combined JND surface $p^{\mathrm{comb}}_{\text{note-f}}$, we note at given (*duration*, *one-way latency*) that the mod^{LL}, mod^{LPF1}, and mod^{LPF2} from the individual surfaces must satisfy (9) to guarantee strong consistency. However, we may not find a set that satisfies (12). In that case, we use (10) to take the maximum p_{note} and solve (11) to minimize the p_{note} of the dominant attribute. Because (11) is nonlinear, the combined surface does not satisfy Axiom 1, and we cannot find the surface using subjective tests as proposed by Xu and Wah.[6]

Figure 4(e) depicts the contour when $p_{\mathrm{note}} = 0.5$. It shows along the y-direction that the combined strategy can maintain strong consistency while concealing the delay effects at much higher network latencies when compared to the three primitive strategies. Similarly, along the x-direction, the combined strategy has

a much higher tolerance to action durations and can provide better loss concealments and jitter removals.

CONCLUSION

In this article, we have developed a unifying approach for deriving the composite perceptual quality of an interactive multimedia application with multiple quality metrics and controls. Our approach evaluates the multiple metrics' joint perceptual qualities by balancing their noticeability probabilities without requiring new subjective tests. The integrated approach is a complete and efficient solution for evaluating the perceptual qualities of interactive multimedia applications.

REFERENCES

1. K. Chan, J. Kwan, and V. Shelat, "Awareness, perception, knowledge, and attitude toward robotic surgery in a general surgical outpatient clinic in Singapore, Asia," *J. Clin. Transl. Res.*, vol. 8, no. 3, pp. 224–233, May 2022.
2. S. Porcu, A. Floris, and L. Atzori, "Quality of experience in the metaverse: An initial analysis on quality dimensions and assessment," in *Proc. 14th Int. Conf. Qual. Multimedia Experience (QoMEX)*, 2022, pp. 1–4.
3. G. T. Fechner, E. G. Boring, and D. H. Howes, *Elements of Psychophysics*. New York, NY, USA: Holt, Rinehart and Winston, 1966 [First published 1860].
4. J. Xu and B. Wah, "Optimizing the perceptual quality of real-time multimedia applications," *IEEE Multimedia Mag.*, vol. 22, no. 4, pp. 14–28, Oct. 2015, doi: 10.1109/MMUL.2015.70.
5. P. Chen and M. E. Zarki, "Perceptual view inconsistency: An objective evaluation framework for online game quality of experience (QoE)," in *Proc. 10th ACM SIGCOMM Workshop Netw. Syst. Support Games*, 2011, pp. 1–6, doi: 10.1109/NetGames.2011.6080978.
6. J. Xu and B. W. Wah, "Consistent synchronization of action order with least noticeable delays in fast-paced multiplayer online games," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 13, no. 1, pp. 1–25, Jan. 2017, doi: 10.1145/3003727.
7. J. Xu and B. W. Wah, "Optimality of greedy algorithm for generating just-noticeable difference surfaces," *IEEE Trans. Multimedia*, vol. 18, no. 7, pp. 1330–1337, Jul. 2016, doi: 10.1109/TMM.2016.2557728.
8. C.-H. Chou and Y.-C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467–476, Dec. 1995, doi: 10.1109/76.475889.
9. H. von Helmholtz, *Kürzeste Linien im Farbensystem: Auszug Aus Einer Abhandlung Gleichen Titels in Sitzgsber*. Berlin, Germany: der Akademie zu Berlin, Dec. 1891.
10. E. N. Dzhafarov and H. Colonius, "Fechnerian metrics in unidimensional and multidimensional stimulus spaces," *Psychonomic Bull. Rev.*, vol. 6, no. 2, pp. 239–268, Jun. 1999, doi: 10.3758/BF03212329.
11. J. Xu and B. W. Wah, "Concealing network delays in delay-sensitive online interactive games based on just-noticeable differences," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2013, pp. 1–6, doi: 10.1109/ICME.2013.6607526.
12. M. Mauve, J. Vogel, V. Hilt, and W. Effelsberg, "Local-lag and timewarp: Providing consistency for replicated continuous applications," *IEEE Trans. Multimedia*, vol. 6, no. 1, pp. 47–57, Feb. 2004, doi: 10.1109/TMM.2003.819751.
13. J. Myers et al. *BZFlag 2.4.2*. (2012). BZFlag. [Online]. Available: http://bzflag.org/

BENJAMIN W. WAH is a research professor at the Chinese University of Hong Kong (CUHK), Shatin, Hong Kong, China and the Franklin W. Woeltge Emeritus Professor of Electrical and Computer Engineering at the University of Illinois, Urbana-Champaign, USA. His research interests include big data and multimedia systems. Wah received his Ph.D. degree in computer science from the University of California, Berkeley, CA, USA, in 1979. He is a Fellow of the American Association for the Advancement of Science, Association for Computing Machinery, and IEEE. Contact him at wah@illinois.edu.

JINGXI XU is a team leader in the Intelligent Algorithm Team 2, WeMeet Product Center, Tencent, Shenzhen, Guangdong, 518054, China. His research interests include improving the quality of videoconferencing in resource-restricted network conditions. Xu received his Ph.D. degree in computer science and engineering from the Chinese University of Hong Kong. Contact him at jingxixuxjx@gmail.com.