

Optimizing the Perceptual Quality of Real-Time Multimedia Applications

Jingxi Xu and Benjamin W. Wah
The Chinese University of Hong Kong

Designing a multimedia system to achieve high perceptual quality requires a black-box approach to tune control inputs according to preferences from subjective tests. The authors present OptPQ, a systematic method for optimizing perceptual quality using just-noticeable difference (JND) profiles.

The goal in designing a real-time multimedia system is to achieve high perceptual quality under every possible operating condition, where perceptual quality is the result of a subjective assessment by system users. Achieving high perceptual quality requires making tradeoffs between the multiple quality metrics observed by users, and translating those tradeoffs into analytical data that can be used to tune control inputs at run time. This translation is difficult because the relation among control inputs, quality metrics, and perceptual quality is complex and unknown, especially under resource constraints such as limited network bandwidth.

Here, we present our method, called OptPQ, for optimizing perceptual quality. Our systematic method for finding operating points achieves good perceptual quality using an off-line measured just-noticeable difference (JND) profile that captures human awareness of adjustments made to control inputs. Further-

more, an online method combines multiple independent JND profiles to find the best control inputs.

Understanding the Tradeoffs

Figure 1 depicts an application with n control inputs and m quality metrics. The underlying runtime condition, such as network bandwidth, might impose constraints on the control inputs and introduce tradeoffs. These tradeoffs require that developers consider all quality metrics and constraints together to achieve high perceptual quality.

We first define the perceptual quality of a multimedia system as follows:

Definition 1. A simplex quality metric of a multimedia system has a corresponding control input in which the perceptual quality of the metric is monotonic with respect to the control input.

Definition 2. The perceptual quality of a multimedia system with multiple quality metrics is the combined quality perceived by subjects when using the system's user interface.

As a running example, Figure 2 illustrates a voice-over-IP (VoIP) multimedia application. The system has complex interacting quality metrics with two control inputs: mouth-to-ear delay (MED) and audio quality parameter (AQP). Both inputs affect the audio signal quality (ASQ) and the interactivity, which in turn determine the user's perceptual quality.

ASQ is a combination of source and channel qualities. Source quality is the quality after compression, which is controlled by AQP (with higher AQP resulting in better source quality). Channel quality reflects the reliability in transmitting audio packets, including the concealment of delay jitters and lost packets. Source and channel qualities are determined by the network condition and controlled by MED and AQP.

In contrast, interactivity is controlled by MED and affects the efficiency and symmetry of a conversation. Efficiency measures the fraction of time extended between a conversation with and without network delay; symmetry measures the ratio between the average times experienced by a speaker when the conversation switches direction from local to remote versus remote to local. Longer MEDs lead to worse efficiency and symmetry.

Other real-time multimedia systems also involve tradeoffs. For instance, Figure 3 depicts a real-time multiplayer online game. It has a control input for setting the buffer size for

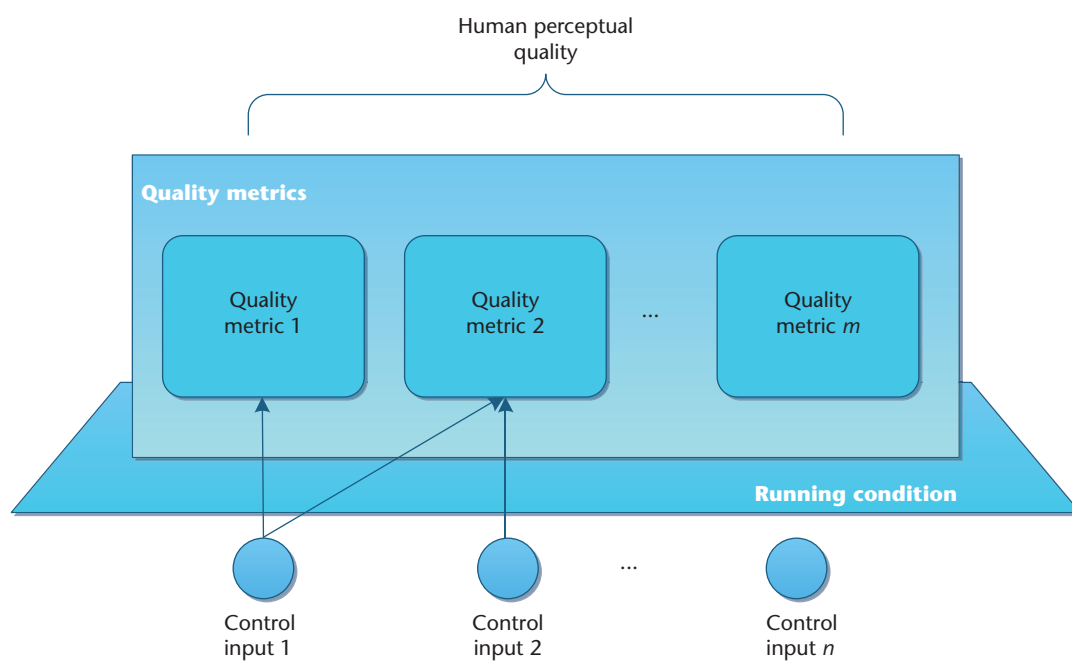


Figure 1. The relation among control inputs, quality metrics, and perceptual quality. A multimedia application is perceived by users as a black box as far as perceptual quality is concerned. Under resource constraints, tradeoffs must be made among the quality metrics to achieve high perceptual quality.

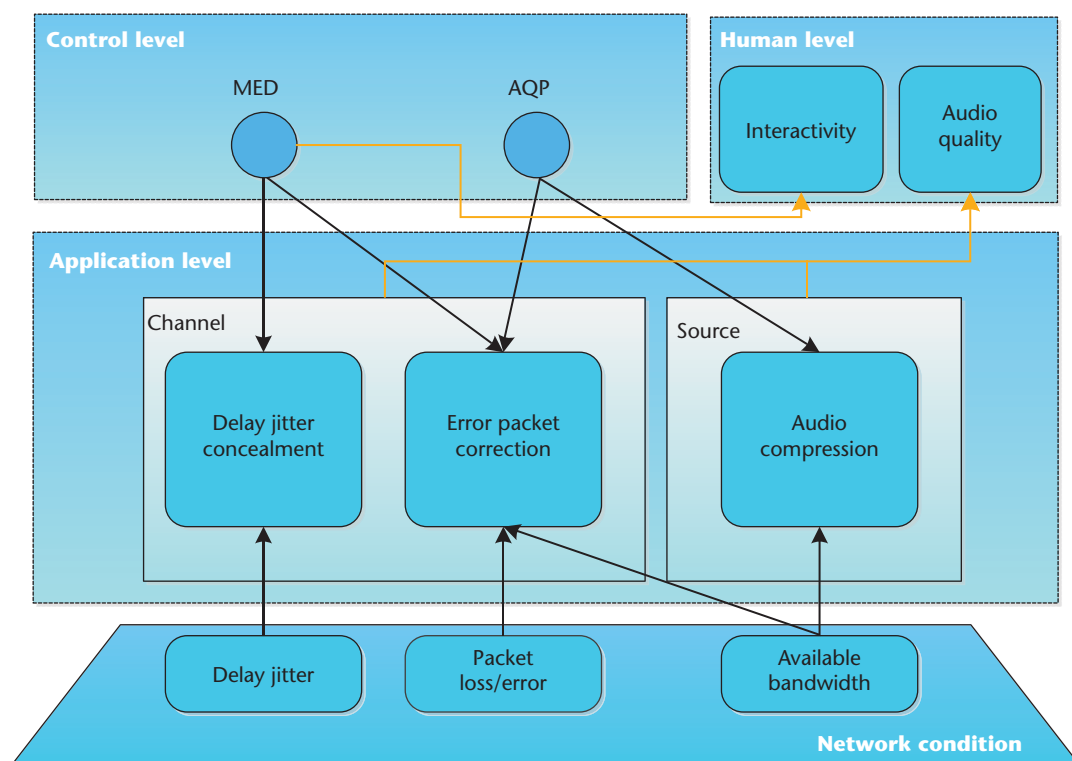
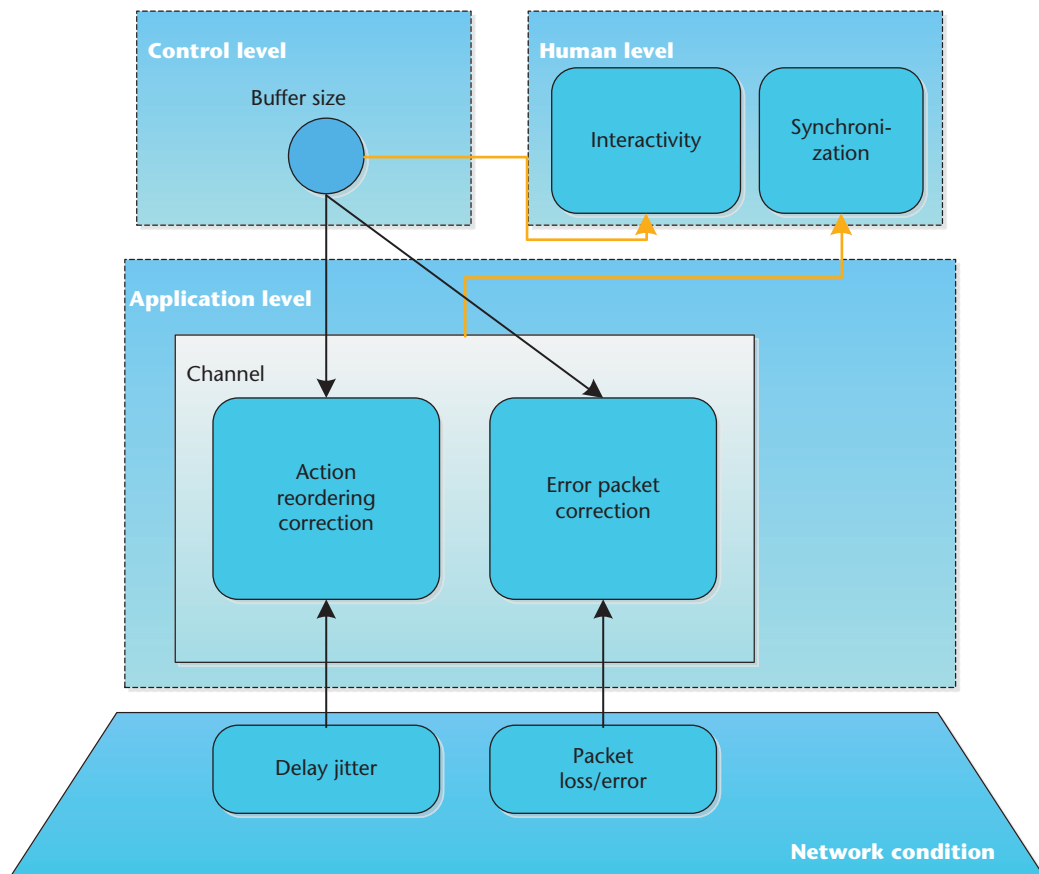


Figure 2. A VoIP system as a real-time multimedia application with two control inputs. The mouth-to-ear delay (MED) determines the interactivity of a conversation and the jitter-buffer size (which affects jitters and loss concealment under given network conditions). The audio quality parameter (AQP) controls the source quality after compression and the channel quality when some packets are lost. Both metrics are then captured by the audio signal quality (ASQ) of a conversation. Perceptual quality depends on both ASQ and interactivity.

Figure 3. A multiplayer online game has buffer size as the major control input, which acts like MED in VoIP, balancing interactivity and synchronization.



accommodating late packets, which affects the perceptual quality of playing the game. A larger buffer can delay the acknowledgments of actions, but smooth delay jitters and maintain correct synchronization among local and remote clients.¹

As another example, in a virtual reality system with a head-mounted display, the processing time before displaying a virtual scene is a control input that affects perceptual quality. A longer processing time will lead to smoother graphics when rendering changes in a scene but will incur longer delays before the changes are perceived.

Challenges and Approaches

Here, we describe the issues involved in subjective evaluations, as well as previous work for solving these problems.

Difficulties in Evaluating Perceptual Quality

Figure 4 illustrates the two difficulties involved when optimizing the control of a real-time multimedia system to achieve high perceptual quality.

First, runtime conditions (like network bandwidth and losses) or information needed by the control inputs might be outdated.² Such information might be too large to be collected,

or there might be a delay in its collection. Optimization using outdated information might not lead to high perceptual quality.

Second, human perception is not well understood, and there is no well-defined relation between perceptual quality and control inputs.³ Some previous approaches have developed approximate analytic models for measuring perceptual quality. For instance, quality metrics can be simplified using linear relations or heuristics^{4,5} to allow online measurements and closed-form mappings from control inputs to perceptual quality. However, such models usually cannot capture the complex relations between control inputs and perceptual quality. Other analytic models might be more accurate for evaluating the perceptual quality of a given application but not applicable in online measurements because the process is computationally expensive. The perceptual evaluation of speech quality (PESQ)⁶ used in VoIP is one such example.

In short, the perceptual quality of real-time multimedia systems is hard to optimize because it lacks well-defined analytic models for relating perceptual quality to control inputs and to the available quality metrics.

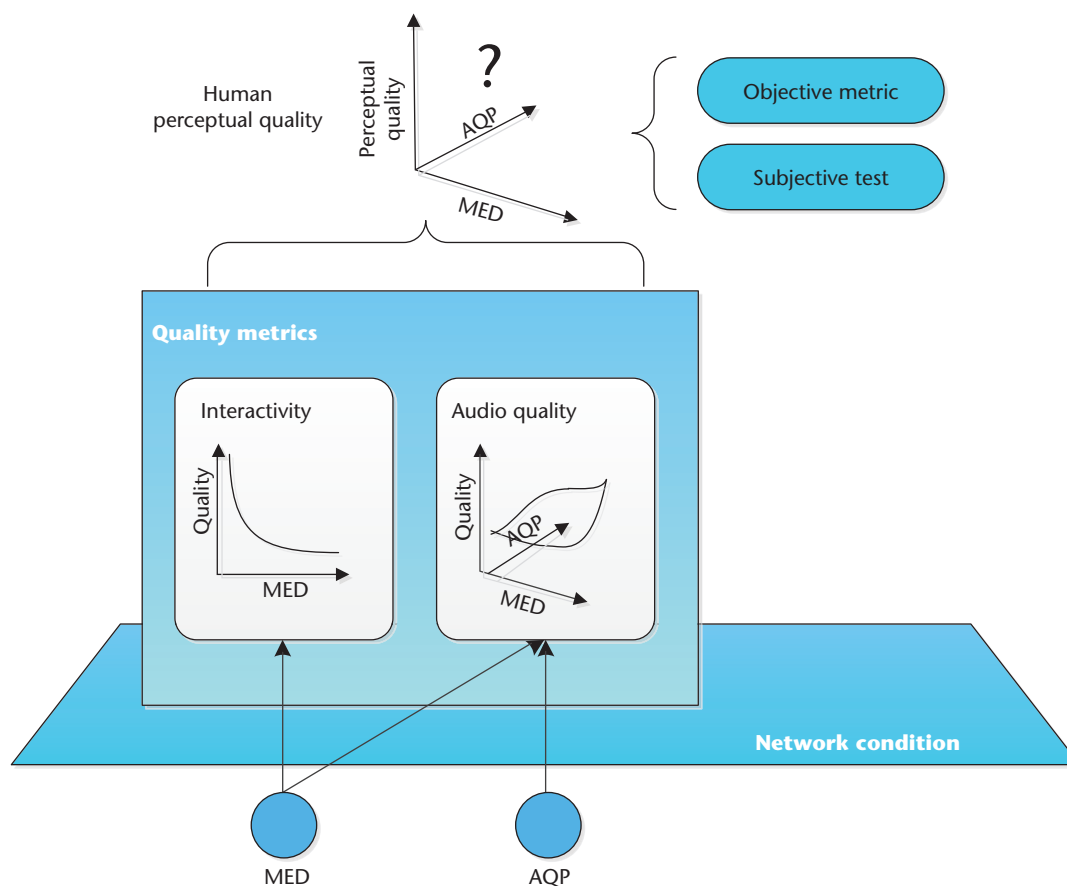


Figure 4. Two difficulties involved when optimizing the perceptual quality of VoIP. a) Proper tradeoffs cannot be made because runtime network information is outdated. b) The perceptual quality is not well modeled and might need to be measured by expensive subjective tests. Specific to the VoIP application, the complexity of optimizing perceptual quality is high because the dependent control inputs (MED and AQP) that affect ASQ cannot be independently considered.

Issues Related to Subjective Tests

Without well-defined and reliable quality metrics for optimizing perceptual quality, previous approaches^{7,8} resorted to offline subjective tests for measuring perceptual quality. A subjective test involves a number of subjects reporting their perceptual opinions of a system. The results collected offline are then used to guide the setting of control inputs at run time.

Subjective tests can be based on either absolute or relative assessments. In absolute assessments, a standard method is to evaluate a mean opinion score (MOS), which is obtained by asking subjects to report their opinion using an absolute category rating and by taking an algebraic mean of their opinion when evaluating the same output.⁹ Using absolute assessments imposes a total order on perceptual quality, which might differ from reality. For instance, two scenarios in VoIP under different MEDs could lead to different

interactivity and ASQ, but it doesn't mean one will necessarily be perceptually better than the other in subjective tests.

In contrast, in relative assessments, subjects are asked to do pairwise comparisons of two observed outputs under different control input settings and to choose the setting leading to the output with better perceptual quality (similar to that in Annex E of ITU P.800⁹). Figure 5 illustrates such an application in VoIP.

Relative assessments have higher complexity than absolute assessments because the Cartesian product of combinations of control-input settings must be tested. In practice, relative assessments have difficulty in handling more than one control input and more than two quality metrics. Heuristics must be employed to limit the search space. For example, simplifying assumptions about the relation among subjects' opinions were made in a previous study⁷ to

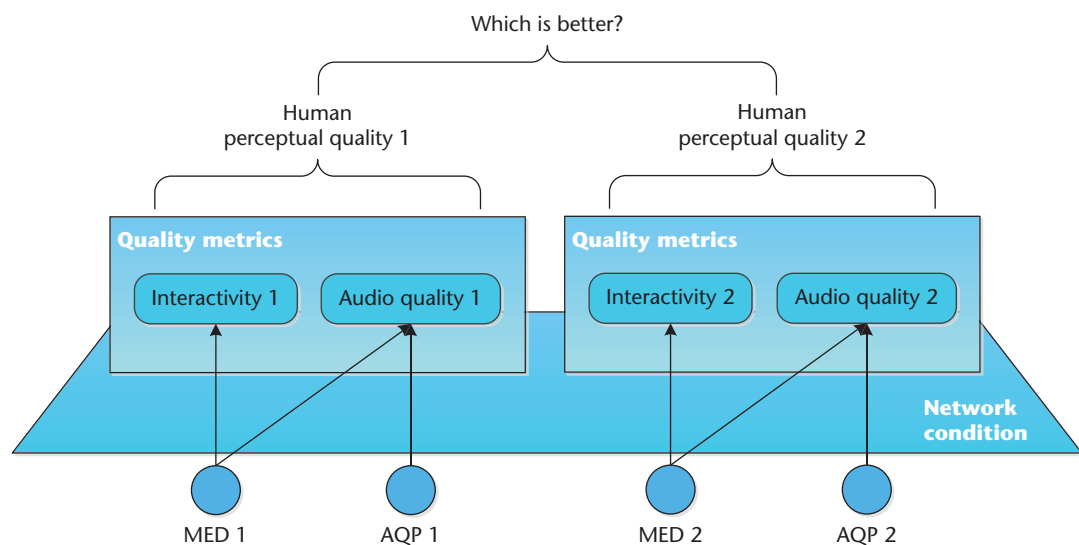


Figure 5. Pairwise comparisons are suitable for discovering a better alternative in perceptual quality when complex tradeoffs are involved. Users are asked to choose between two scenarios with better perceptual quality, each under a different setting of control inputs—in this case, (MED_1, AQP_1) versus (MED_2, AQP_2) .

limit the search space. Note that while absolute assessments give a total order of perceptual quality, relative assessments only give a partial order. Only when all the alternatives are comparable can the two be equivalent.

One of the main issues with subjective tests is that they are expensive to conduct. This is especially true when the number of quality metrics is large, leading to a prohibitively large number of combinations to be tested.

Optimizing Perceptual Quality

There have been several previous studies on developing a general method for optimizing the perceptual quality of multimedia systems. Researchers conducted studies to combine existing quality metrics with metric selection using offline psychophysical measurements^{10,11} or using a heuristic method, such as evolutionary algorithms.¹² However, these approaches are limited in that they depend on existing quality metrics and just provide a framework for combining them.

There have also been approaches that use a black-box method to optimize multimedia systems without well-modeled metrics. Batu Sat, along with one of us (Benjamin Wah), proposed a statistical method for performing an offline search of optimal controls of a VoIP system and then generalizing them online using a learned SVM.⁷ The limitation of this method is that it cannot decompose the combined perceptual quality into simplex quality metrics, leading to

a large number of subjective tests when there are multiple quality metrics. For example, optimal controls must be found in a discrete manner under low, medium, and high network latencies. At run time, one of these optimal controls is used according to the classification result of the current network latency.

Zixia Huang proposed a modified version of Sat and Wah's method for a video-conferencing system.⁸ It has similar limitations and requires a large number of subjective tests to acquire the optimal controls.

Our Approach

To address the issue of the large combination of quality metrics to be tested, we must analyze the metrics' dependencies and their effects on perceptual quality.

In the simplest case, when the quality metrics independently affect perceptual quality, then the perceptual quality corresponding to each metric can be optimized independently. In more general cases, the metrics can be treated as independent, because their dependence is not perceptible in a small region of the control-input value.

By exploiting this independence, we propose conducting subjective tests for each quality metric separately to determine how the metric affects perceptual quality. Without enumerating the combinations of quality metrics and their effect on perceptual quality, our approach

aims to greatly reduce the number of subjective tests needed.

To address the issue of the high cost of subjective tests, we next present an effective way of sampling the space of all possible subjective tests. By exploiting the inability of humans to perceive small changes in the control inputs, it is not necessary to conduct subjective tests whose control inputs have changed slightly from those of tests already conducted. Furthermore, by exploiting the continuity of the results of subjective tests, a majority of the results on subjective tests not conducted can be interpolated from those already conducted.

Exploiting Just-Noticeable Differences (JND)

JND is the minimal change of an input (or multiple inputs) that can be perceived by humans in output. It has been employed to study human perception in numerous applications, including light intensity, brightness, loudness of sound, and various multimedia applications.¹³

JND is a statistical concept defined with respect to a given *awareness*, which indicates the fraction of a sufficient number of human subjects who can correctly identify the output caused by the changed input, when the original input (reference) and the changed input are presented one after another in a random order. A 75 percent awareness level is generally used in psychophysics studies. Under this awareness level, JND is the amount of change of reference input *ref* in order to achieve 75 percent awareness: $JND(ref|awareness = 75 \text{ percent})$. The following definition relates awareness to relative perceptual quality.

Definition 3. Let p be the awareness of a pairwise comparison of two outputs of a multimedia system, one due to a reference input (*ref*) and other due to a modification of the reference ($ref + mod$). Then, $p > 0.5$ indicates that $ref + mod$ has better relative perceptual quality than *ref*, $p < 0.5$ indicates that the quality is worse, and $p = 0.5$ indicates the same perceptual quality.

If we are interested in the awareness of a changed input $ref + mod$ from the reference *ref*, we can define JND as an awareness function: $p(ref, mod)$, where p is the fraction of humans perceiving the change from *ref*.

Previous studies^{14,15} on JND generally rely on Weber's law, which states that the reference and the JND are related by a linear relation. Although this property works well in simple applications, such as determining changes in

the length of a line, it does not apply in many real-time multimedia applications. In these applications, there are complex tradeoffs among changes in control inputs and the resulting perceptual quality. These tradeoffs are demonstrated later in the nonlinear relation between interactivity/AQP and JND in VoIP.

Past works also assume that awareness fits into analytical models like cumulative normal or logistic.¹³ Again, subjective tests on actual multimedia applications show that these models are not always applicable and that general analytical models are hard to find.¹

A JND Profile

With one control input, a JND profile plots $p(ref, mod)$, where p is the awareness that measures the probability of subjects who can identify the output due to the modified reference $ref + mod$ from that of *ref* when the outputs are presented in a random order.

As an illustration, consider the generation of the JND profile of the VoIP application using interactivity as the quality metric. We first recruit a group of N subjects and present them with two system outputs (each in the form of an interactive conversation) in a random order: one generated using MED_{ref} as the control input, and the second using $MED_{ref} + MED_{mod}$. We then ask the subjects to choose the conversation they feel is less interactive (caused by a longer *MED*). We then calculate the fraction of subjects who can identify the right conversation. When N is sufficiently large, the fraction is a good approximation of p . To make awareness meaningful with respect to changes of perceptual quality, we use the absolute value to represent the fraction, and a positive (or negative) value to represent an improvement (or degradation) in quality after the change.

We have discovered two properties of human perception in a JND profile¹ that can significantly reduce the number of subjective tests required. First, we know that the same change made to a larger reference is less significant because the effect of the change is smaller. Second, a larger change to the same reference has a greater effect because it is more noticeable. These properties are stated in the following axiom:

Axiom 1. (a) Awareness p is monotonically nonincreasing with respect to *ref* because a given *mod* is less noticeable with a larger *ref*. (b) p is monotonically nondecreasing with respect to *mod* because a larger *mod* is more noticeable for a given *ref*.

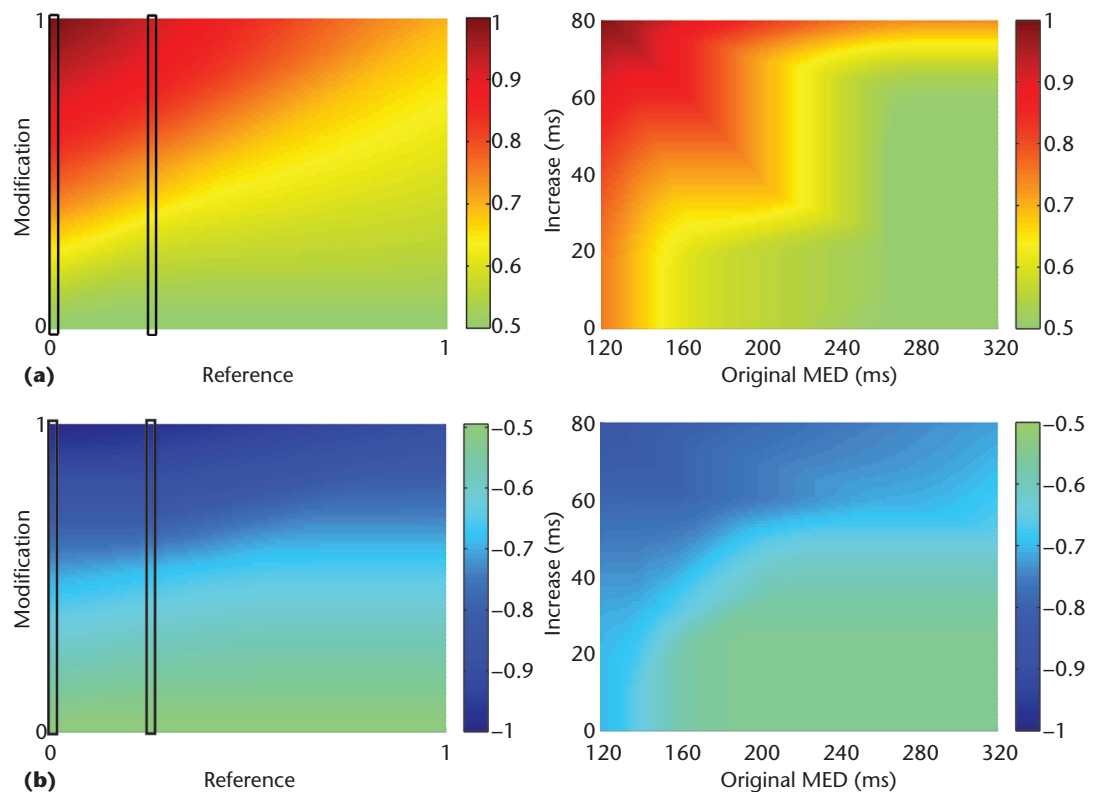


Figure 6. JND profiles when the simplex quality metric is (a) improving or (b) degrading with respect to the control input. Left panels: synthetic profiles illustrating an ideal condition (both axes normalized to [0, 1]); the bar shows the increasing absolute awareness of the change with a larger modification. Right panels: real profiles in VoIP in an error-prone network showing the fraction of subjects who can correctly identify the output with better ASQ (or poorer interactivity) caused by an increased MED. An absolute value of awareness indicates the fraction, whereas a negative value indicates a degradation in perceptual quality.

Note that the order of ref and $ref + mod$ can be reversed when evaluating a quality metric. We should first test whether subjects are less sensitive to the same change when ref is larger than $ref + mod$. If not, then we reverse their order.

Figure 6 illustrates two JND profiles when the simplex quality metric either improves or degrades monotonically after the control input is changed with respect to ref . We show the ideal JND profiles in the left panels, the real profiles for VoIP in the right panels, the upper ones with respect to ASQ, and the lower ones with respect to interactivity. Although the ideal profiles are smooth, the real profiles might not be. For example, the step pattern in Figure 6a is due to the discrete time interval for receiving parity packets in an error-prone network. Only when MED is increased to the deadline of an additional parity packet can ASQ be improved.

Relation between Perceptual Quality and JND

To show the relation between awareness and the corresponding perceptual quality, Figure 7 illustrates the relative perceptual quality of the ideal profiles in the left panels of Figure 6. Consider two references ref_1 and ref_2 . Starting from ref_1 in Figure 7a, the absolute value of awareness increases monotonically (because perceptual quality is better) when the modified control input $ref_1 + mod$ is larger. This also applies to ref_2 , but the increase is slower, because perceptual quality increases with a slower trend. Figure 7b shows a similar behavior, where negative awareness indicates a degraded quality after the change.

This observation illustrates an important relation between awareness and relative perceptual quality—namely, when compared to the same reference, a higher awareness indicates better perceptual quality. This is stated formally as follows:

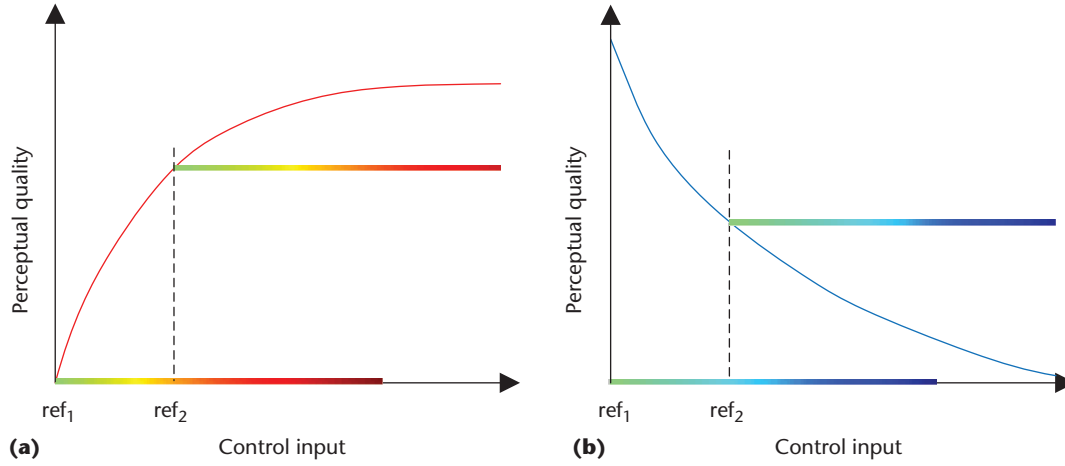


Figure 7. Relation between awareness and relative perceptual quality. The horizontal color bars (obtained from the left panels in Figure 6) show how the absolute value of awareness increases as the modification is increased. (a) The awareness of the improved quality becomes larger as the control input increases with respect to a fixed reference ref_1 . The increase is slower when the fixed reference is ref_2 . (b) The awareness of the degraded quality becomes larger as the modification is increased. Note that the y-axis shows only the relative magnitude of perceptual quality.

Lemma 1. Let $Q(x)$ be the perceptual quality of a simplex quality metric controlled by input x . With given reference ref ,

$$\begin{aligned} p(ref, mod_1) &> p(ref, mod_2) \\ \Rightarrow Q(ref + mod_1) &> Q(ref + mod_2) \end{aligned} \quad (1)$$

$$\begin{aligned} Q(ref + mod_1) &> Q(ref + mod_2) \\ \Rightarrow p(ref, mod_1) &\geq p(ref, mod_2) \end{aligned} \quad (2)$$

Proof: (1) can be proved by using Axiom 1, namely, $p(ref, mod_1) > p(ref, mod_2) \Rightarrow mod_1 > mod_2$, as well as Definition 1—namely, $mod_1 > mod_2 \Rightarrow Q(ref + mod_1) > Q(ref + mod_2)$.

(2) can be proved by using Definition 1—namely, $Q(ref + mod_1) > Q(ref + mod_2) \Rightarrow mod_1 > mod_2$, as well as Axiom 1—namely, $mod_1 > mod_2 \Rightarrow p(ref, mod_1) \geq p(ref, mod_2)$.

With Lemma 1, the (relative) perceptual quality of a simplex quality metric can be fully evaluated by the corresponding JND profile. Next, we present an efficient algorithm for finding the JND profile.

Measuring a JND Profile

The cost of getting a complete JND profile is prohibitive if we need to conduct subjective tests for all combinations of ref and $ref + mod$. Instead, we can sample combinations of (ref, mod) and interpolate those unmeasured combinations. For this approach to work, we need to determine which pairs to measure, and

how to assure that the errors of those unmeasured combinations are within a tolerable threshold.

With Axiom 1, the awareness of any test point in a JND profile can be proved to be bounded by that of the two diagonal points of a rectangular region containing the test point. This property is stated as follows:

Theorem 1. Let two test points in a JND profile be, respectively, (ref_1, mod_1) with awareness p_1 and (ref_2, mod_2) with awareness p_2 . If $ref_1 < ref_2$ and $mod_1 > mod_2$, then for any other unmeasured test point $p_3(ref_3, mod_3)$ where $ref_1 \leq ref_3 \leq ref_2$ and $mod_1 \geq mod_3 \geq mod_2$, we have $p_1 \geq p_3 \geq p_2$.

The theorem can be proved by using the continuity property of awareness in a JND profile. Because the axes are defined with respect to ref and mod , it follows that a profile can be divided into rectangular regions in testing, and that interpolations can be performed in each region.

Corollary 1. A JND profile can be divided into rectangular regions with subjective tests conducted on its diagonal corners, and the awareness of any point inside the region is bounded by that of the corner points.

Based on this result, we propose a greedy algorithm to plan the subjective tests. We first identify the region with a looser


```

Require:  $\hat{p}(\text{ref}, \text{mod})$  or fraction of subjects who correctly identify the modified
control input  $\text{ref} + \text{mod}$ ;  $\delta$  or required error threshold.

Ensure: JND profile  $p(\text{ref}, \text{mod})$ .

1. Measure  $\hat{p}(0,1)$  and  $\hat{p}(1,0)$ , add them to  $P_{\text{tested}}$ .

while  $\max |\hat{p}_i - \hat{p}_j| > \delta$  where  $i, j \in P_{\text{tested}}$  and no mid-point in between tested do

    Perform subjective tests to measure  $\hat{p}_m$ , the mid-point of  $\hat{p}_i$  and  $\hat{p}_j$ .

    Add  $\hat{p}_m$  to  $P_{\text{tested}}$ .

end while

Fix  $\forall p \in P_{\text{tested}}$  that do not satisfy monotonicity.

2. Interpolate  $p(\text{ref}, \text{mod})$  with the fixed  $P_{\text{tested}}$ .

```

Figure 8. Algorithm 1—the greedy algorithm for finding the just-noticeable difference (JND) profile.

bound (uncertainty) on awareness, conduct subjective tests at the center of this region, divide the region into four subregions, and repeat the process until the uncertainty is satisfactorily tight. Here, uncertainty in awareness of a region is defined to be the difference between its largest and smallest awareness, which are actually located at the diagonal corners. As we measure the awareness from large to small regions, the uncertainty is always available for the next step. The greedy algorithm is outlined in Figure 8.

Figure 9 further illustrates application of the greedy algorithm on the VoIP application.

Combining Multiple JND Profiles

Here, we present methods for combining independent JND profiles, each developed for a simplex quality metric, when evaluating the overall perceptual quality of a multimedia application. Because responses from subjects leading to the multiple profiles are independent, our approach is based on computing the combined awareness using probability theory.

Combining Independent JND Profiles

We define the combined awareness to measure the result of a pairwise comparison when evaluating the relative perceptual quality Q of an application (see Definition 3).

Definition 4. The combined awareness of multiple independent JND profiles is the fraction of sufficiently many subjects who can notice the output caused by a changed input to have better

perceptual quality than the output caused by the original input.

Recall that awareness of a simplex quality metric represents the fraction of subjects who can correctly identify the output caused by a modified reference from the original when they are shown in a random order. Although this is a probabilistic concept, the combined awareness of multiple independent simplex quality metrics is not simply a product of their awareness because awareness is not independent, even when the corresponding metrics are independent. For example, when there is no modification, the awareness of ASQ and interactivity is 50 percent, but the combined awareness is still 50 percent, not $1 - (1 - 0.5)(1 - 0.5) = 0.75$.

To allow awareness of multiple independent simplex metrics to be interpreted probabilistically, we convert the awareness into a new measure called *significance* defined below. To unify our definition, we use in this section the absolute value of p as awareness.

Definition 5. Significance $\mu(\text{ref}, \text{mod})$ is the fraction of N subjects who can correctly perceive the change after ref is changed by mod when N is sufficiently large.

For the same subjective test, awareness p and significance μ are related to each other as follows:

$$\mu = 2p - 1. \quad (3)$$

This can be verified by considering it in subjective tests with given μ and N . Here, μN

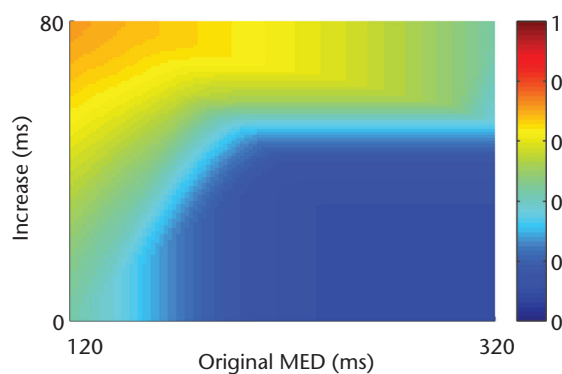
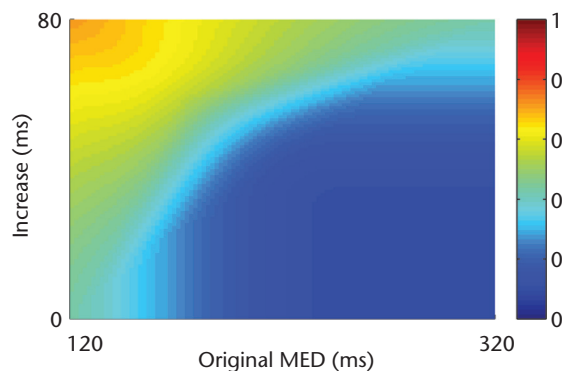
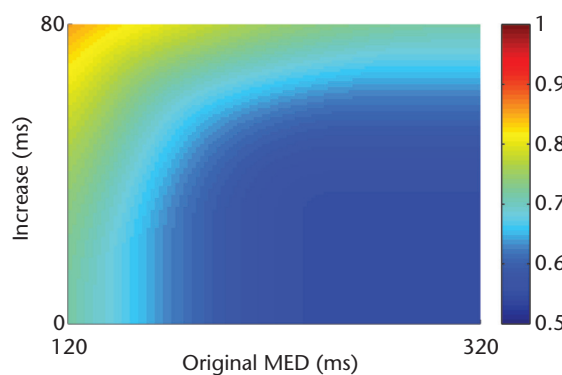
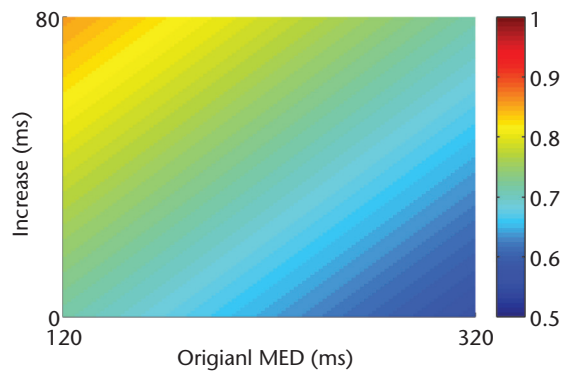
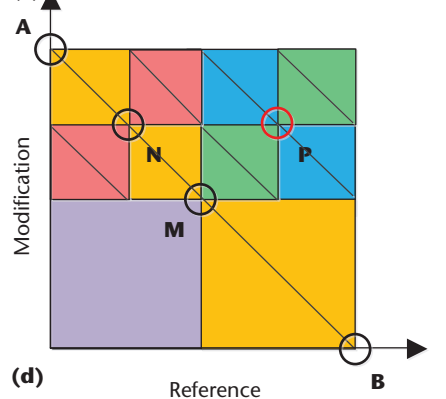
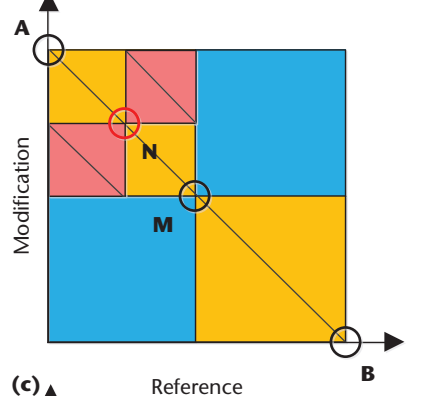
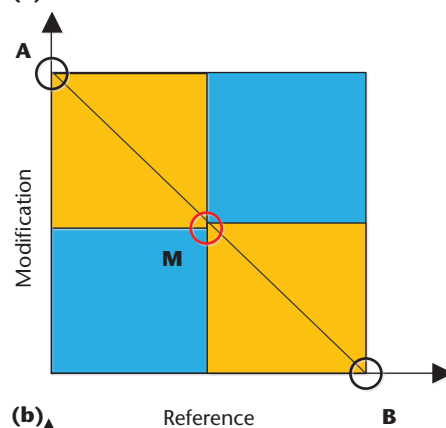
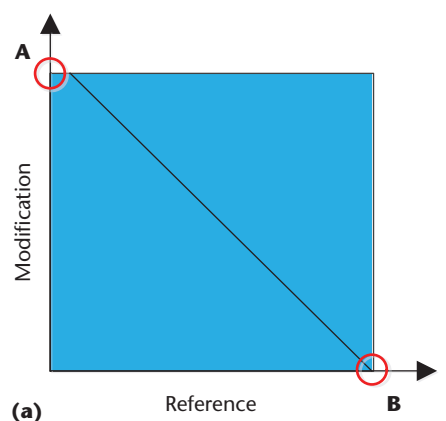


Figure 9. The application of a greedy algorithm to successively divide a JND profile into finer regions for subjective tests. The color is rescaled for clarity. (a) The direction of monotonicity is indicated by the diagonals. We start by testing the upper-left and lower-right boundary points of the profile. (b) We conduct subjective tests at the center point of the profile. (c) We further divide the profile into four subregions, measure the region with the largest uncertainty (in the upper-left end of the figure) and then further subdivide it for testing. (d) We measure the remaining regions using the same method. The awareness in the resulting JND profile is an interpolation of those points verified by subjective tests. Note that this figure shows only the first five test points.

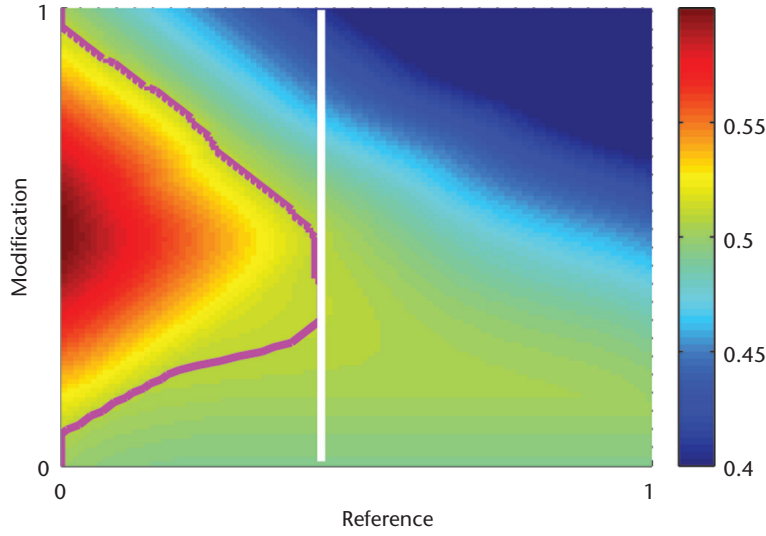


Figure 10. A simplified, combined JND profile from the two synthetic profiles in the left panels of Figure 6. Awareness $p > 50$ percent indicates subjects prefer an increased reference, whereas $p < 50$ percent indicates subjects prefer the original reference. The pink curve bounds the region where awareness is larger than 50 percent. The solid line indicates the local optimum in the reference.

subjects will be able to identify the change in ref , whereas the remaining $(1 - \mu)N$ subjects cannot identify the change and answer only with a random guess. Therefore,

$$\mu N \times 1 + (1 - \mu)N \times 0.5 = pN \Rightarrow \mu = 2p - 1.$$

Before we can calculate the combined awareness, we need the following axiom to relate the perceptual quality of multiple simplex metrics and the combined relative perceptual quality. For simplicity, we present the results only for cases with two simplex quality metrics. A general case with more than two quality metrics can be similarly derived.

Axiom 2. Let Q_1 and Q_2 (Q'_1 and Q'_2 , respectively) be the relative perceptual quality of two simplex quality metrics with respect to the original (modified) control inputs. Further, let Q_{comb} and Q'_{comb} be the corresponding combined relative perceptual quality. Then,

- a) $Q'_1 > Q_1$ and $Q'_2 \geq Q_2 \Rightarrow Q'_{comb} > Q_{comb}$
- b) $Q'_1 \leq Q_1$ and $Q'_2 < Q_2 \Rightarrow Q'_{comb} < Q_{comb}$
- c) $Q'_1 = Q_1$ and $Q'_2 = Q_2 \Rightarrow Q'_{comb} = Q_{comb}$
- d) $Q'_1 > Q_1$ and $Q'_2 < Q_2 \Rightarrow Q'_{comb} ? Q_{comb}$
- e) $Q'_1 < Q_1$ and $Q'_2 > Q_2 \Rightarrow Q'_{comb} ? Q_{comb}$

The first two conditions in the axiom can be explained as follows. When we present two alternative outputs for a system with two sim-

plex quality metrics, a subject will always be able to identify the alternative with better perceptual quality if the perceptual quality of one quality metric is improved while the other is not degraded.

The third condition corresponds to the case in which subjects cannot identify a change in perceptual quality for both quality metrics. As a result, subjects will respond with a random guess to the combined case.

The last two conditions correspond to cases in which one metric is improved and the other is degraded. Depending on the amount of modification with respect to the reference, the subjects might notice better, the same, or worse overall perceptual quality between the outputs corresponding to the reference and the modified reference. The outcome will not be known until actual subjective tests are performed.

For the last two cases, we assume that their probability is low for the majority of references and modifications (to be verified at the end of this section). Under this assumption, we simplify their combined awareness to 0.5; that is, subjects will respond with random guesses. This simplification might generate some small errors in awareness. However, these errors are only large in extreme regions in the combined JND profile; they are not significant near the regions of interest.

To compute p_{comb} , the combined awareness of two simplex quality metrics when the control input changes from ref to $ref + mod$, let μ_1 (or μ_2) be the fraction of subjects who notice an improvement (or degradation) for the two quality metrics. Then the fraction of subjects is

$$\begin{cases} \mu_1(1 - \mu_2) & \text{who prefer modification} \\ \mu_2(1 - \mu_1) & \text{who do not prefer modification} \\ (1 - \mu_1)(1 - \mu_2) & \text{who find no difference} \\ \mu_1\mu_2 & \text{who find improvement in the first and degradation in second} \end{cases} \quad (4)$$

The combined awareness is then calculated as

$$\begin{aligned} p_{comb} &= \mu_1(1 - \mu_2) \times 1 + \mu_2(1 - \mu_1) \times 0 \\ &\quad + (1 - \mu_1)(1 - \mu_2) \times 0.5 + \mu_1\mu_2 \times 0.5 \\ &= \frac{1 + \mu_1 - \mu_2}{2} \\ &= 0.5 + p_1 - p_2. \end{aligned} \quad (5)$$

According to Definition 4, $p_{comb} = 1$ is for subjects who choose the modified input

(corresponding to the first case in Equation 4); $p_{\text{comb}} = 0$ is for subjects who choose the original input (corresponding to the second case in Equation 4); and $p_{\text{comb}} = 0.5$ is for subjects who make a random guess (corresponding to the last two cases in Equation 4).

Figure 10 depicts the resulting JND profile derived using Equation 5 when combining the two synthetic JND profiles in the left panels of Figure 6.

Note that in computing p_{comb} , the contribution of those subjects who find improvement in one metric but degradation in the other in (5) is $0.5\mu_1\mu_2$. Figure 11 illustrates the value of this term for every point in the combined JND profile in Figure 10. The result shows that the amount is small throughout the bottom and the middle parts of the profile, which are the regions of interest and contain the local maxima in relative perceptual quality.

Best Operating Points Using the Combined Profile

Given the JND profile of the combined awareness, we can derive the resulting relative perceptual quality in a way similar to that presented earlier in the “Relation between Perceptual Quality and JND” section. The result will let us find the best control input that gives the best relative perceptual quality. The following corollary is used to search for the local maxima in perceptual quality:

Corollary 2. For any given δ ,

- $Q(x)$ is the local maximum in $[x, x + \delta]$ if $p(x, y) \leq 0.5$ for all $0 \leq y \leq \delta$;
- $Q(x)$ is not the local maximum in $[x, x + \delta]$ if $p(x, y) > 0.5$ for any $0 \leq y \leq \delta$.

Proof: Based on Definition 1, the first part follows from the fact that $p(x, y) \leq 0.5$ indicates $Q(x) \geq Q(x + y)$. Similarly, the second part follows from the fact that $p(x, y) > 0.5$ indicates $Q(x) < Q(x + y)$.

In the corollary, we define the local maximum only from x to $x + \delta$ because, in a JND profile, we can measure awareness only when the control input is increased. However, if we check for optimality only on one side, it is possible that $x - \delta \leq x' < x$ exists such that $Q(x') > Q(x) > Q(x + y)$, where $0 \leq y \leq \delta$. To assure that the result is also the maximum on the other side of the region, we look for the first ref in the combined JND profile that satisfies

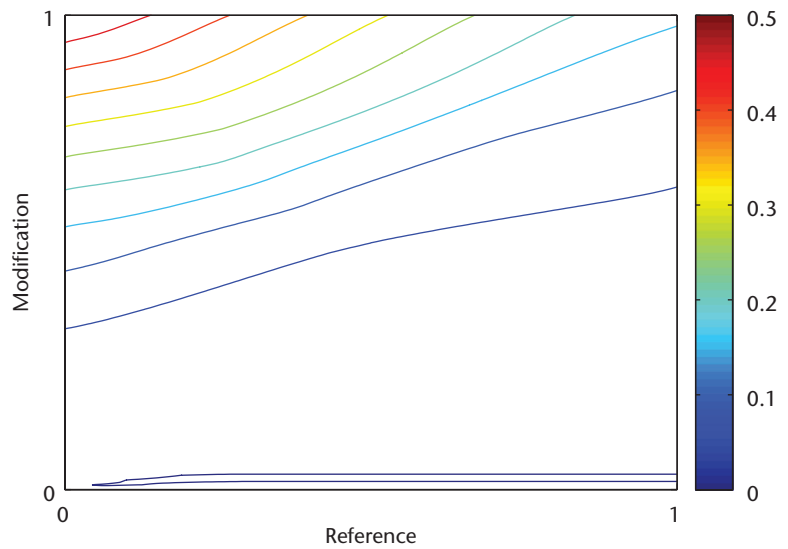


Figure 11. The contour shows the value of $0.5\mu_1\mu_2$ in Equation 5 (see the main text). The values at the bottom and middle parts are small. These regions are of interest and contain the local maxima in relative perceptual quality. Its small value means that this term will not significantly affect the search result.

$p(ref, y) \leq 0.5$ for all $0 \leq y \leq \delta$. Then, $Q(ref - \delta) > Q(ref)$ should not exist; otherwise, $ref - \delta$ can be found to satisfy the condition, considering that δ is sufficiently small.

In practice, we like to find the local maximum within the largest possible range, and the largest δ we can identify in the JND profile is the range of the increase y_{max} . Therefore, we discard any ref that does not satisfy $p(ref, y) \leq 0.5$, where $0 \leq y \leq y_{\text{max}}$, until we find the first ref that satisfies the condition. Figure 10 illustrates the region where we have discarded the ref (bounded by the pink curve), as well as the first ref that satisfies the condition (indicated by the solid white line). Figure 12 further shows the resulting relative perceptual quality derived from the combined JND profile.

Because there might be multiple local maxima in perceptual quality depending on δ , we can reduce δ from y_{max} to find other $refs$ if necessary.

Using the JND profiles for the VoIP application in the right panels of Figure 6, the left panel in Figure 13a shows the corresponding combined JND profile. The local maxima is identified by the right solid lines in the left panel, which corresponds to the best MED for achieving high perceptual quality. The left solid line illustrates another local maximum that we can find when δ is reduced. The right panel in Figure 13a illustrates the relative perceptual

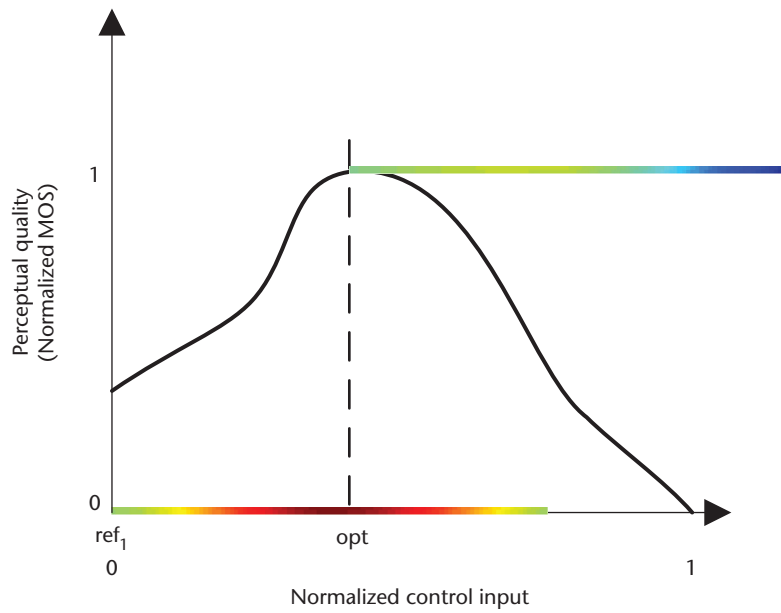


Figure 12. The resulting overall relative perceptual quality derived from the combined JND profile in Figure 10. The profile indicates a better change with $p > 0.5$, a poorer change with $p < 0.5$, and no change with $p = 0.5$. The awareness bar starting from ref_1 shows the awareness increasing until the local maximum opt is reached. The awareness bar starting from opt shows the awareness decreasing from 0.5 as the reference increases.

quality derived from the JND profile. The profile in a fast-conversation scenario shows the optimal MED $ref = 240$ and the second optimal MED $ref = 210$. The first optimal MED is better because $p(210, 30) > 0.5$.

Generalization of the Results in VoIP

With the combined JND profile found by offline subjective tests, identifying the best MED is now transformed into a simple runtime search. Specifically, as the network condition changes, the JND profile of ASQ can change with respect to the network delay and the buffer size. For example, when the average network delay increases, we can simply increase the starting MED in the JND profile of ASQ (by including the extra buffering delay for concealing lost packets) without changing the profile itself. We then combine this new profile with the original interactivity profile and find the optimal MED to be used at runtime.

We can modify the JND profile of ASQ in this fashion because a small change of network latency generally does not affect the generally low network loss rate. We do not need to modify the interactivity profile because the new network condition should not change human

sensitivity on interactivity. In short, we can do online modification of the offline-measured JND profile, combine the profiles at run time, and perform a runtime search of the control input leading to the best perceptual quality. This approach lets us generalize our method for different network conditions, without measuring new profiles each time.

Another generalization is needed when interactivity changes. In VoIP, the conversational behavior might change when a different topic is discussed. A business conversation might have a fast turn frequency, whereas a casual conversation might instead be slow. These two conversational conditions will lead to different JND profiles on interactivity. Figures 13a and 13b illustrate the JND profile based on the interactivity profile measured for a fast and slow conversation, respectively, and the common ASQ profile in Figure 6a. We find that the two profiles are not very different, considering the distribution of the improved (red) and degraded (blue) regions. Consequently, interpolations can be made between the two JND profiles for interactivities in between. All these computations can be done at run time.

Comparison with the Related Method

Here, we compare our method to a peer method⁷ using the VoIP application.

Solution Quality

Figure 13 shows the optimal controls found by our method and the peer method. The solid white lines in Figure 13a show the two best MEDs found by our method, whereas the dashed line shows the best MED ($ref = 245$) found by the peer method. The resulting optimal MEDs are very similar with comparable performance.

To further demonstrate the advantage of the proposed method, in Figure 13b we present the profile in a slow-conversation scenario, which shows a much larger optimal MED (400 ms) found by our method. The optimal MED found by the peer method is $ref = 270$, which has poorer perceptual quality than ours: $p(270, 80) > 0.9$ and $p(270, 130) = 1.0$. That is, nearly 100 percent of the subjects will prefer the increase, which shows that our MED provides better perceptual quality.

Cost of Subjective Tests

As discussed earlier, because we have decomposed the measure of awareness on interactivity

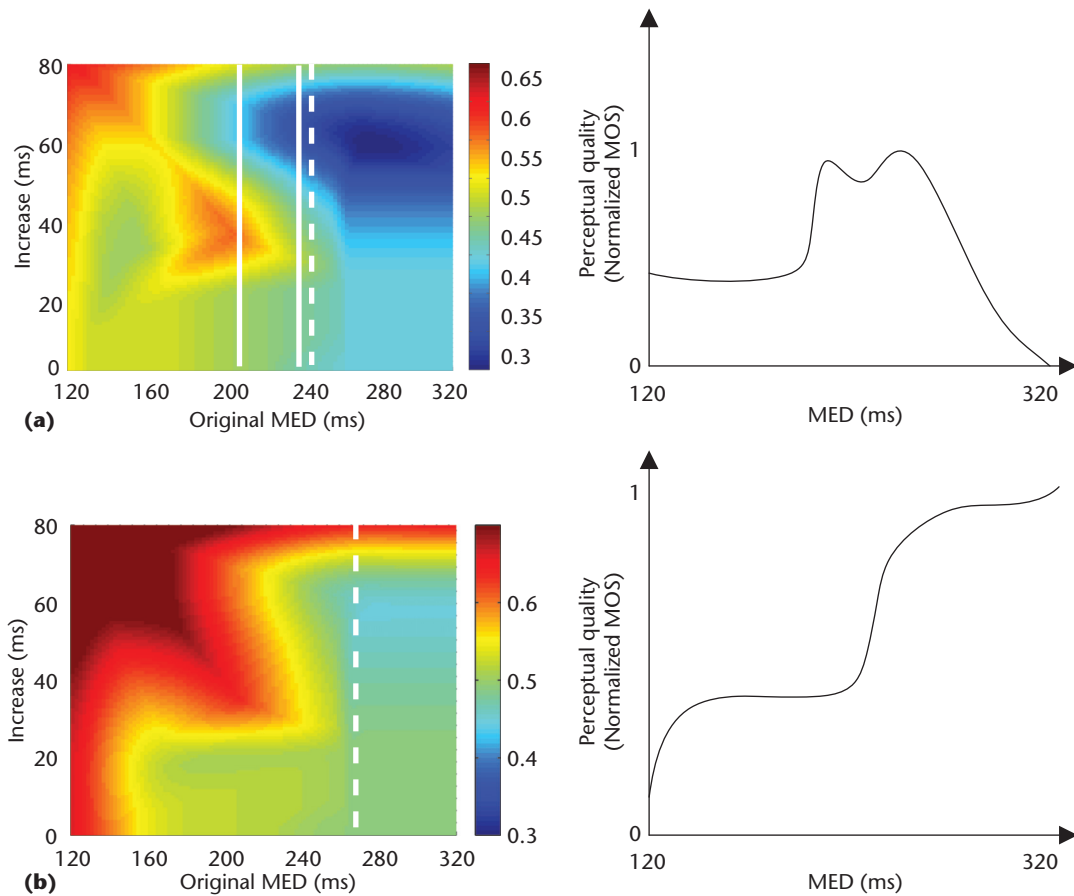


Figure 13. The combined JND profile and the corresponding relative perceptual quality measured for (a) a fast conversation and (b) a slow conversation for the VoIP application using the profiles on ASQ and interactivity in Figure 6. The optimal controls found by our method are shown with the solid white lines; optimal controls found by the previous method⁷ are shown with the dashed white line.

and signal quality, we do not need to perform subjective tests for finding the optimal MED when network latency changes. In comparison, the peer method needs new subjective tests for each condition of network latency. Depending on the discretization of latency, the subjective tests will need to be repeated multiple times.

Moreover, when the conversational scenario changes, the peer method must redo the subjective tests for each network latency. For example, if network latency is discretized to three levels, and the conversation has four scenarios, the peer method will need to perform subjective tests in 3×4 cases (combinatorially increasing), while our method needs to measure only $1 + 4$ cases (linearly increasing). Obviously, our method will need fewer subjective tests when the combination of running conditions is larger.

Our approach can be generalized to applications with multiple quality metrics and control inputs.

For applications with multiple independent JND profiles, our approach can be directly applied by deriving subjects' awareness using a voting strategy and probabilistic arguments when there are more than two changes. In this case, the combined JND profile might display a more complex multimodal behavior.

For applications with multiple independent control inputs, we can compute their combined awareness using probabilistic arguments developed here. In this case, the combined JND profile is multidimensional, because the control inputs might vary within the constraints. Under a given set of constraints, the optimal combination of control inputs will need to be searched to attain the optimal combined awareness.

For applications with multiple dependent control inputs, it will not be possible to combine the resulting profiles probabilistically. In this case, we must study the sensitivity of one input with respect to others. For example, in the VoIP application, AQP and MED are dependent control inputs because they both

affect ASQ. In most cases, it might be possible to fix the less dominant control inputs and to focus on only one. In the VoIP application, AQP will most likely be fixed and changed only under extreme conditions.

MM

Acknowledgments


This special issue is a collaboration between the 2014 IEEE International Symposium on Multimedia (ISM 2014) and *IEEE MultiMedia*. This article is an extended version of the keynote, "Just Noticeable Differences—Optimizing the Perceptual Quality of Real-Time Multimedia Systems over the Internet," presented at ISM 2014.

References

1. J. Xu and B.W. Wah, "Concealing Network Delays in Delay-Sensitive Online Interactive Games Based on Just-Noticeable Differences," *Proc. Int'l Conf. Multimedia and Expo (ICME)*, 2013, pp. 1–6.
2. J. Xu and B.W. Wah, "Exploiting Just-Noticeable Difference of Delays for Improving Quality of Experience in Video Conferencing," *Proc. Multimedia Systems Conf.*, 2013, pp. 238–248.
3. B. Sat and B.W. Wah, "Playout Scheduling and Loss-Concealments in VoIP for Optimizing Conversational Voice Communication Quality," *Proc. 15th Int'l Conf. Multimedia*, 2007, pp. 137–146.
4. *The E-Model: A Computational Model for Use in Transmission Planning*, Int'l Telecommunications Union, ITU-T G.107, 2014; <https://www.itu.int/rec/T-REC-G.107>.
5. *Opinion Model For Video-Telephony Applications*, Int'l Telecommunications Union recommendation, ITU-T G.1070 Ed.2, July 2012; <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=9050>.
6. A.W. Rix et al., "Perceptual Evaluation of Speech Quality (PESQ): A New Method for Speech Quality Assessment of Telephone Networks and Codecs," *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2001, pp. 749–752.
7. B. Sat and B. Wah, "Statistical Scheduling of Offline Comparative Subjective Evaluations for Real-Time Multimedia," *IEEE Trans. on Multimedia*, vol. 11, no. 6, 2009, pp. 1114–1130.
8. Z. Huang and K. Nahrstedt, "Perception-Based Playout Scheduling for High-Quality Real-Time Interactive Multimedia," *Proc. 31st Int'l Conf. Computer Communications*, 2012, pp. 2786–2790.
9. J.G. Beerends et al., "Perceptual Evaluation of Speech Quality (PESQ) the New ITU Standard for End-To-End Speech Quality Assessment Part II: Psychoacoustic Model," *J. Audio Eng. Society*, vol. 50, no. 10, 2002, pp. 765–778.
10. W. Wu et al., "Quality of Experience in Distributed Interactive Multimedia Environments: Toward a Theoretical Framework," *Proc. 17th Int'l Conf. Multimedia*, 2009, pp. 481–490.
11. W. Wu et al., "CZLoD: A Psychophysical Approach for 3D Tele-Immersive Video," *Trans. Multimedia Computing, Communications, and Applications*, vol. 8, no. 3s, 2012, p. 39.
12. A. Arefin, R. Rivas, and K. Nahrstedt, "OSM: Prioritized Evolutionary QoS Optimization for Interactive 3D Teleimmersion," *ACM Trans. Multimedia Computing, Communications, and Applications (TOMCAP)*, vol. 10, no. 1s, 2014, p. 12.
13. J.A. Ferwerda, "Psychophysics 101: How to Run Perception Experiments in Computer Graphics," *ACM SIGGRAPH 2008 Classes*, 2008, p. 87.
14. I. Cheng and P. Boulanger, "Feature Extraction on 3-D Texmesh Using Scale-Space Analysis and Perceptual Evaluation," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 10, 2005, pp. 1234–1244.
15. P. Hinterseer et al., "Perception-Based Data Reduction and Transmission of Haptic Data in Telepresence and Teleaction Systems," *IEEE Trans. Signal Processing*, vol. 56, no. 2, 2008, pp. 588–597.

Jingxi Xu is a PhD candidate in the Department of Computer Science and Engineering, The Chinese University of Hong Kong. His research interests include strategies for the optimization of multimedia systems, especially the video conferencing and online interactive games. Xu received his bachelor's degree from Sun Yat-Sen University. He is a student member of IEEE. Contact him at jxxu@cse.cuhk.edu.hk.

Benjamin Wah is the Provost of The Chinese University of Hong Kong and the Wei Lun Professor of Computer Science and Engineering, as well as Professor Emeritus, University of Illinois, Urbana-Champaign. His research interests include big data analytics, nonlinear optimization, and multimedia signal processing. Wah received his PhD in engineering from UC Berkeley. He is a fellow of IEEE, the ACM, and the American Association for the Advancement of Science. Contact him at bwah@cuhk.edu.hk.

 Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.