# Towards Video Quality of Experience and Selective Attention: A Subtitle-Based Measurement Study

Lingyan Zhang, Qibo Sun, Shangguang Wang, Sen Su, and Fangchun Yang

State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Haidian, Beijing, 100876 China
{zhanglingyan; qbsun; sgwang; susen; fcyang}@bupt.edu.cn

***Abstract*—Selective attention is the idea that visual attention is not granted equally to all regions of the visual field, but rather focuses on high-attention regions. In this paper, we propose that selectively improving the performance of video in regions that are given high attention by users can significantly improve the quality of experience (QoE) of video services. We investigate users' attention regions, and choose the subtitle region as a representative high-attention region to present a subtitle-based measurement study. This study concentrates on how and how much the subtitle region affects the QoE of video. First of all, we verify that changes in performance of the subtitle region can cause an illusion about users' judgment on video bitrate. Secondly, we present a logarithmic relationship between video bitrate and the QoE of video. Thirdly, we measure the improved QoE by improving the performance of the subtitle region. Finally, we demonstrate that selectively improving the performance of the subtitle region can significantly improve the QoE of video at little cost.**

***Keywords-component; selective attention; high-attention region; quality of experience (QoE); video bitrate; subtitle performance;***

## I. INTRODUCTION

The Internet is now firmly established as a source of video services, and the explosion in ownership of connected devices is driving users' insatiable desire to use video services. As users spend increasing amounts of time on video services, they are no longer satisfied with the mere availability of video content, but increasingly expect a certain quality level in their viewing experience. In contrast to QoS, such as video bitrate, quality of experience (QoE) attempts to take into account non-technical parameters [1]. In practice, this is achieved by considering the characteristics of the human perceptual system.

Selective attention [2] is an important concept in the human perceptual system. In normal situations, visual attention is not distributed equally to all regions of the visual field, but is instead focused on some high-attention regions [3]. Under limited network resources, we believe that preferentially improving the performance of some high-attention regions (e.g., the subtitle region) can significantly improve the QoE of video services, with no need to improve the performance of all video regions.

In this paper, we go one step further and propose to apply the concept of selective attention to QoE improvement of video services. Specifically, since video displaying is the main form of various video services, we try to figure out the high-attention region of video, and improve the performance of the high-attention region to obtain a higher QoE at little cost. Through a preliminary experiment, we discover that the subtitle region is one of users' high-attention regions and users consistently maintain a high standard of subtitle performance. Therefore, we choose the subtitle region as a representative of high-attention regions to present a subtitle-based measurement study.

In this study, we aim to figure out how and how much the subtitle region affects the QoE of video through a series of subjective experiments. First of all, we verify that changes in performance of the subtitle region can cause an illusion of users' judgment on video bitrate. Secondly, we demonstrate a logarithmic relationship between video bitrate and the QoE of video through subjective experiments, which does correlate well with the logarithmic model for VoIP and web browsing obtained from objective results presented in [6, 7]. Third, we measure the improved QoE by improving the performance of the subtitle region. Videos with clear subtitles are scored an average of approximately one point higher than that with blurring subtitles. Considering the improved QoE value, the two percent increase of the peak bitrate is well worth than it conveys. Finally, the results show that selectively improving the performance of the subtitle region can significantly improve the QoE of video at little cost. And the subtitle-based measurement study provides a basis for a more detailed understanding and improvement of video QoE by combining with the concept of selective attention.

## II. SELECTIVE ATTENTION

Selective attention [2] is an important concept in the human perceptual system. In normal situations, the human perceptual system always selects the most important objects in the visual field and processes them first. Visual attention is not distributed equally to all regions of the visual field, but is instead focused on certain regions known as high-attention regions [3]. In this paper, our goal is to investigate the applicability of selective attention for video, and then to focus on applying the concept of selective attention to QoE improvement of video.

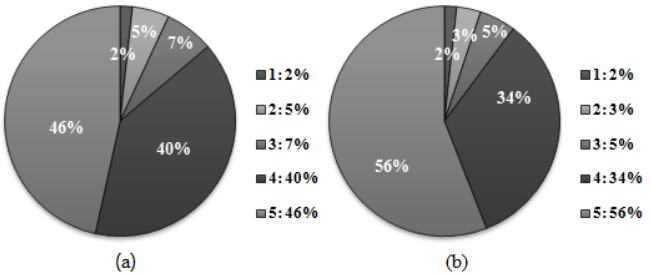


Fig. 1. The pie chart of the percentages of user attention levels on the subtitle region which range from 1 ("Lowest attention") to 5 ("Highest attention"). (a) The pie chart when it is a native language video. (b) The pie chart when it is a foreign language video.

We conduct a preliminary experiment to figure out users' attention regions when watching a video. In the experiment, the video screen is divided into several regions based on position (e.g., upper left region, center left region) and on the information contained therein (e.g., face region, subtitle region). We make a list of the video regions, video contents (e.g., speech, TV show, sports, documentaries) and languages (e.g., native language or foreign language).

A total of 116 subjects, 60 males and 56 females, participate in the experiment. The subjects rate their attention levels to different regions at a given instance. We adopt a five-point Likert scale to measure the attention level which ranges from 1 ("Lowest attention") to 5 ("Highest attention"). The results show that 88% of the subjects favor only those regions that are appealing or providing useful information rather than attending to the entire video screen. As *Fig. 1* shows, 86% of the subjects pay high attention to the subtitle region regardless of video content, and 56% of the subjects pay the highest attention to the subtitle region when watching a foreign language video. These results prove the existence of high-attention regions of video, confirming the applicability of selective attention to the video QoE.

As the subtitle region is clearly a high-attention region and plays a vital role during users' viewing process, we are motivated to give a subtitle-based measurement study. We believe that selectively improving the performance of the subtitle region can result in a QoE improvement.

## III. Experiment and Results

In this section, we aim to figure out how and how much the subtitle region affects the QoE of video through a series of subjective experiments. First of all, we verify whether changes in performance of the subtitle region could cause an illusion of users' judgment on video bitrate. Secondly, we discuss the relationship between video bitrate and QoE. Thirdly, we measure the improved QoE by improving the performance of the subtitle region.

### A. Experiment Configuration

In this experiment, we adopt the MSU Perceptual Video Quality tool [4] and the Stimulus Comparison Adjectival Categorical Judgment (SCACJ) method [5] to evaluate the QoE of video, as shown in *Fig. 2*.

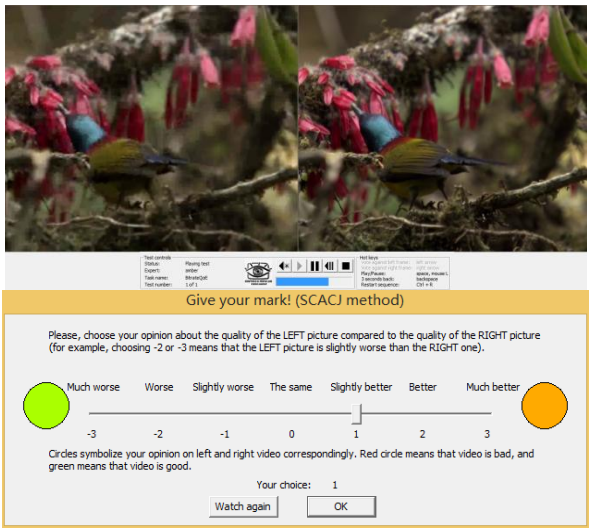


Fig. 2. The experiment tool and the seven-point scale in SCACJ.

The SCACJ methodology uses a "just-notice-able difference" approach, in which the viewer is presented with a pair of video clips side by side and directly rates the difference between the pair. We adopt a seven-point scale of mean opinion score (MOS) as the main measurement as shown in *Fig. 2*; and we call it the DMOS which ranges from -3 ("Much worse") to 3 ("Much better") and obtains higher granularity. The SCACJ method is used in our test because this method assesses the relationships among conditions more fully and it helps report the existence of perceptible differences (e.g. same, different), the existence and direction of perceptible differences (e.g. less, same, more), or judgments of extent and direction.

We recruit 40 participants (20 males and 20 females; age range 18–28, including a variety of college students). Through the training process, each subject becomes familiar with the testing environment and the manual scoring function. To mitigate the order effects, the display order of video pairs and the side for each video clip are randomized. After viewing each video pair, the subject immediately rates the QoE of video or gives the judgment on video bitrate.

### B. Video Bitrate Illusion Caused by Subtitle

We propose a hypothesis that changes in performance of the subtitle region could cause an illusion about users' judgment on video bitrate, thus affect the QoE. To demonstrate this, we carry out an experiment.

Video clips used in the experiment are distinguished based on the subtitle type and the video language type. The subtitle type includes no subtitle (hereafter, referred to as NS), blurring subtitle (BS), and clear subtitle (CS). The video language type includes native language and foreign language. We use four

TABLE I. THE ERROR RATES FOR FOUR CATEGORIES OF VIDEO PAIRS

| Categories of video pairs | 240p | 360p | 480p | 720p | 1080p |
|---|---|---|---|---|---|
| NS*native language | 5% | 7.5% | 5% | 2.5% | 2.5% |
| NS*foreign language | 5% | 7.5% | 5% | 2.5% | 2.5% |
| BS*CS*native language | 22.5% | 22.5% | 12.5% | 5% | 2.5% |
| BS*CS*foreign language | 27.5% | 22.5% | 12.5% | 15% | 5% |

categories of videos pairs: 1) the native language video pairs in which the video clips are with NS; 2) the foreign language video pairs in which the video clips are with NS; 3) the native language video pairs in which one video clip is with BS and the one is with CS; 4) the foreign language video pairs in which one video clip is with BS and the one is with CS. Each category includes video clips which are encoded in five different video bitrate (240p, 360p, 480p, 720p and 1080p) on YouTube[1].

After watching a video pair, the subjects judge which video is encoded in higher bitrate. *Table I* shows the error rates of users' judgments on different categories of videos. As shown in *Table I*, there are no differences in error rate between native language and foreign language video pairs with NS. In contrast to NS video pairs, users have a much higher error rate when watching a BS video clip and a CS video clip at the same time. It is because that subtitles deliver so much information important in aiding the users' understanding of video that they always hold users' attention. The high attention on subtitles causes an illusion that the video with CS has higher bitrate than the video with BS, despite their video bitrate is the same.

### C. Relationship between QoE and Video Bitrate

As changes in the performance of the subtitle region can cause an illusion about video bitrate, we further our study of figuring out the relationship between the video bitrate and QoE for videos with NS.

*Fig. 3* shows the QoE value obtained for videos with NS as a function of video bitrate $x$; the applied logarithmic model $f(x)$ is shown as well. The x-value is the video bitrate and the y-value is the corresponding QoE value measured as DMOS. Each dot represents a single DMOS for the given bitrate $x$. The model function is obtained by means of nonlinear regression. The function is defined as follows:

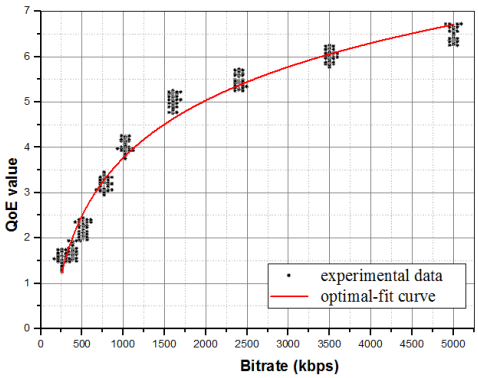


Fig. 3. Optimal-fit curve of QoE value as function of bitrate for NS video clips which specified in equation (3)

[1] https://support.google.com/youtube/answer/2853702?hl=en

$$f(x) = \alpha - \beta \bullet \ln(x + \chi) \quad (1)$$

We use the optimization toolbox of MATLAB to find an optimal fitting function (for the unknown parameters $\alpha$, $\beta$ and $\chi$ in (1)) such that the normal error $E$ is minimized. The normal error is defined as the sum of residuals $r_i$ for $n$ measurements $(x_i, y_i)$ with an average bitrate $x_i$ and a measured DMOS $y_i$:

$$E = n\sum_{i=1}^{n} r_i, r_j = f(x_i) - y_i \quad (2)$$

The mean squared error $MSE = \frac{1}{n}\sum_{i=1}^{n} r_i^2$ or the normalized mean square error $NMSE = MSE / (Max[y_i] - Min[y_i])$ should approach zero where $Max[y_i]$ is the maximum of $y_i$ and $Min[y_i]$ is the minimum of $y_i$. In addition, the correlation coefficient $R$ between the model function and the measured data (or the coefficient of determination $R^2$) should approach one for a close-to-perfect match. The optimal fitting function obtained, i.e., $f(x)$, is as follows:

$$f(x) = 1.817 \bullet In(x - 8.721) - 8.766 \quad (3)$$

In the derivation, (3) yields $R$=0.987, $R^2$ =0.975, *MSE*=0.172, and *NMSE=0.0372*; all metrics indicate a very good match between the measurement data and the applied logarithmic models. The measurement results and their respective logarithmic regression function are presented in *Fig. 3*. And the logarithmic relationship from subjective result presented here does correlate well with the logarithmic model for VoIP and web browsing obtained from objective results presented in [6, 7].

### D. Improved QoE of Video

Since changes in performance of the subtitle region affects users' judgment on video bitrate, thus affect the QoE, we need further measure the improved QoE by improving the performance of the subtitle region.

According to the statistic [8], the video bitrate of the video from YouTube has a clear peak at around 320kbps. Therefore, the video clips used in this experiment are encoded in 320 kbps. And there are six types of subtitles used in the experiment. It ranges from 1 ("Much clear") to 6 ("Much blurring"), as shown


Fig. 4. Performance levels of subtitles which range from 1 ("Much clear" to 6 ("Much blurring")

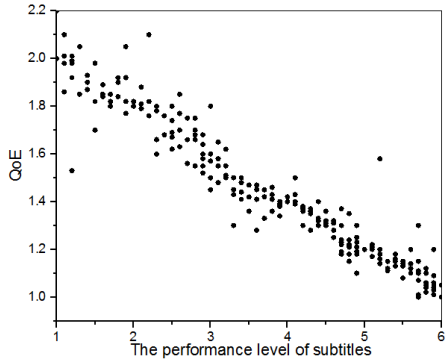


Fig. 5. The QoE value of subtitles in different performance levels which range from 1 ("Much clear") to 6 ("Much blurring")

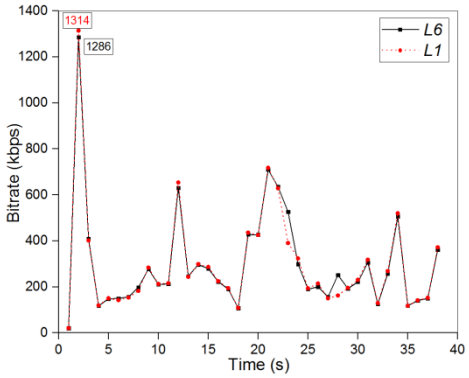


Fig. 6. Average bitrate comparison between *L6* subtitles and *L1* subtitles over time

in *Fig. 4*. They are referred to as *L1*, *L2*, *L3*, *L4*, *L5*, and *L6*.

*Fig. 5* shows the QoE value of subtitles in different performance levels. Because subtitles always hold users' attention, the QoE value of video decreases as subtitles become more and more blurring. In contrast to video clips with *L6* subtitles, video clips with *L1* subtitles are scored an average of approximately one point higher, which is a significant improvement in the five-point Likert scale. *Fig. 6* shows the average bitrate comparison between *L6* subtitles and *L1* subtitles over time. Compared to *L6* subtitles, the bitrate of *L1* subtitles is increased from a peak of 1286 kbps to 1314 kbps. Considering the improved QoE, the two percent increase of the peak bitrate is well worth than it conveys. Therefore, selectively improving the performance of the subtitle region can significantly improve the QoE of video at little cost.

## IV. Conclusion

In this paper, we apply an important concept in the human perceptual system, that of selective attention, to the QoE improvement of video.

In analyzing the results of a preliminary experiment, we find that the subtitle region is a high-attention region and users consistently maintain high quality standard for the performance of the subtitle region. Therefore, we take the subtitle region as a representative high-attention region in carrying out a series of carefully designed experiments, and present a subtitle-based measurement study. The results show that changes in the performance of the subtitle region can cause an illusion about users' judgment on video bitrate, thus affect QoE. Videos with clear subtitles are judged to have higher video bitrate than videos with blurring subtitle, despite using the same video. Meanwhile, we obtain a logarithmic relationship between the video bitrate and QoE, which does correlate well with the logarithmic model for VoIP and web browsing presented in [6, 7]. On this basis, we further measure the improved QoE by improving the performance of the subtitle region. Videos with clear subtitles are scored an average of approximately one point higher than that with blurring subtitles. Considering the improved QoE value, the two percent increase of the peak bitrate is well worth than it conveys. Finally, we demonstrate that selectively improving the performance of the subtitle region can significantly improve QoE at little cost.

This paper is a significant first step in combining the QoE of video with the high-level concept of selective attention. We carry out a number of subjective experiments, and present a measurement study. The results of subjective measurement in this paper can serve to verify or to question the results of objective measurement. We believe that this work provides an important basis for a more detailed understanding and improvement of video QoE by combining with the concept of selective attention. Quantifying and correlating the cognitive factors to form a valid general methodology requires further research and experimentation.

## V. Achknowledgment

This work was supported by the National Natural Science Foundation of China under grants no. 61571066 and 61472047.

## References

[1] P. Juluri, V. Tamarapalli, and D. Medhi. "Measurement of Quality of Experience of Video-on-Demand Services: A Survey." *IEEE communication surveys & tutorials*, vol. 18, No. 1, pp. 401-418, 2015.

[2] Itti, L. and C. Koch, "Computational modelling of visual attention". *Nature reviews neuroscience*, vol.2, no. 3, p. 194-203, 2001.

[3] U. Ansorge, S. Buchinger, C. Valuch, A. R. Patrone, and O. Scherzer, "Visual attention in edited dynamical images," *Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications (SIGMAP-2014)*, pp. 198-205, 2014

[4] D. Vatolin and O. Petrov, "MSU perceptual video quality tool," 2012. Available.

[5] TU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," *International Telecommunication Union*, 2002.

[6] P. Reichl, S. Egger, R. Schatz, and A. D'A M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36-41, 2010.

[7] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36-41, 2010.

[8] X. Cheng, J. Liu, and C. Dale. "Understanding the characteristics of internet short video sharing: A YouTube-based measurement study," *IEEE Transactions on Multimedia*, vol.15, no.5, pp. 1184-1194, 2013.