Duration Neglect in Multi-episodic Perceived Quality

Conference Paper · June 2016

READS

4

4 authors, including:

Dennis Guse

Technische Universität Berlin

18 PUBLICATIONS **30** CITATIONS

SEE PROFILE

Anna Wunderlich

Technische Universität Berlin

2 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Duration Neglect in Multi-episodic Perceived Quality

Dennis Guse, Anna Wunderlich, Benjamin Weiss, Sebastian Möller
Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin
{dennis.guse, benjamin.weiss, sebastian.moeller}@tu-berlin.de
a.wunderlich@campus.tu-berlin.de

***Abstract*—In this paper, the occurrence of duration neglect for multi-episodic perceived quality is investigated. Such an effect has been observed for retrospective judgments of individual experiences, showing that the actual duration of exceptional parts of an experience are not reflected in retrospective judgments. Duration neglect has, so far, only received limited attention for perceived quality, but represents an important aspect especially for multi-episodic usage. Here, it is not known if a duration neglect occurs and thus the actual duration of an individual usage episode affects judgments on multi-episodic perceived quality. The results of a subjective experiment are reported that investigated the occurrence of duration neglect in one session of multi-episodic use. In this experiment, the duration of a single degraded episode was varied between two multi-episodic conditions. The results show that a duration neglect occurs for the multi-episodic judgments as well as the episodic judgments of this usage episode.**

***Keywords*—Quality of Experience; Telecommunication Services**

I. Introduction

Quality of Experience (QoE) of telecommunication services represents an essential subjective evaluation from a user's and also customer's point of view. Recently, the investigation of the formation process of perceived quality over multiple usage episodes has been undertaken [1]. A usage episode is a task-driven interaction with a telecommunication service. It is self-contained and meaningful to a user, such as watching a movie or a telephone call. For the so-called multi-episodic perceived quality, it is of interest how the perceived quality of previous episodes (experiences) determines multi-episodic judgments. With regard to retrospective judgments of individual experiences, several effects or biases have been observed, such as a recency effect and a neglect of duration [2]. These effects describe that not all parts of an experience are considered equally important for a retrospective judgment. A duration neglect has been observed for retrospective judgments of pain as well as for perceived quality. There, the actual duration of an outstanding part is not reflected in a retrospective judgment, i. e., an experience with shorter duration of a comparable outstanding part is judged very similar to a comparable experience with a longer outstanding part. Duration neglect has so far only received limited attention for the investigation of perceived quality. In this paper, one experiment is presented, investigating the existence of a duration neglect for multi-episodic use in one session in which one single usage episode is degraded.

II. Related Work

Duration neglect has been first observed for retrospective judgments of pain [2]. There, it has been observed that momentary judgments (taken while experiencing) reflect the duration properly while retrospective judgments are not affected by differences in duration. With regard to perceived quality, this has also been reported by Hands and Avons for video quality [3]. It could be observed that for 30 s video sequences, the retrospective judgments were not different if either 5 s or 10 s of degradations were presented. However, participants were able to assess the actual duration of the degraded part correctly. This indicates that while the duration could be memorized and recalled, this information was not taken into account for the retrospective judgments of perceived quality. First investigations on multi-episodic judgments were conducted by Möller et al. [4] and Guse et al. [5]. They investigated the formation process for repeated, similar usage episodes over a usage period of 12 and 14 days, respectively. To derive a *Mean Opinion Score* (MOS), multiple participants were exposed to the same multi-episodic condition by defining how, when, and for what the provided service should be used as well as the performance for each usage episode. After each usage episode, participants judged the perceived quality of this usage episode (*episodic judgment*). *Multi-episodic judgments*, which requested to judge the perceived quality of all usage episodes *so far*, were assessed repeatedly after several usage episodes. In fact, the results of both experiments were limited, as *a)* Möller et al. failed to create measurable effects of degraded episodes on multi-episodic judgments, and *b)* Guse et al. could only show that the adaptation speed of multi-episodic perceived quality to performance changes is relatively slow. In the latter case, even a presentation of three consecutive days without degradations did not produce a significant increase of multi-episodic judgments. In both experiments, the performance was only varied between episodes, i. e., a uniform performance was provided within each episode. With regard to multi-episodic perceived quality, duration neglect was not yet investigated.

III. Experiment

An experiment was conducted to investigate if duration neglect can be observed in a multi-episodic use case. In contrast to previous work on multi-episodic perceived quality, which covered usage periods of several days, this experiment was conducted for multi-episodic use in one session. Here, one session consists of multiple episodes that are performed sequentially by a participant. This complements prior work and allows to investigate multi-episodic perceived quality under standardized laboratory conditions. For the investigation of duration neglect, a passive listening-only situation was selected. This allows to define the duration of each episode precisely, as varying user behavior cannot influence the duration. Here,

978-1-5090-0354-9/16/$31.00 ©2016 IEEE

Fig. 1. 7-point continuous scale with German labels; labels from left-to-right: extremely bad (0), bad (1), poor (2), fair (3), good (4), excellent (5) and ideal (6) [6] (own illustration).

a speech-only audio book[1] was used, which was cut into individual episodes. Six usage episodes were cut with a similar duration: 174 s to 199 s (μ = 184 s). An additional usage episode was cut with approximately the doubled duration (362 s). Two multi-episodic conditions were investigated, varying the duration of one single degraded usage episode. Both presented six usage episodes while episode 1 to 4 and 6 were presented in the best performance. The 5th episode was presented with uniform degradations. In condition A, this episode was presented in normal duration while in condition B this episode was presented in doubled duration. After finishing a usage episode, the episodic judgment was taken on the 7-point continuous scale (Figure 1). Multi-episodic judgments were taken after the 3rd and 6th episode on the same scale. For the best performance, the PCM-coded content was not further processed and presented in stereo with 44.1 kHz. The uniform degradations of the 5th episode were created by compressing the speech signal with LPC-10 (mono-only). This codec has been developed for low bandwidth radio transmission while providing understandability rather than speech quality. For this experiment, a pair of *Sennheiser HD25-1* connected to the internal sound card of a *Microsoft Surface* was used. The sound pressure level was calibrated to 75 dB 20 μPa (babble noise). This experiment was conducted with 20 female and 16 male participants (normal hearing) aging from 18 to 32 years ($\mu = 25.0$). A between-subject design was applied.

IV. Data Analysis

Statistical evaluation is conducted by applying the Wilcoxon rank-sum test and quality judgments are reported as *Mean Opinion Score* (MOS) with standard deviation in brackets. Between the two conditions neither the episodic judgments for the non-degraded episodes ($W = 4338.50$, $p = 0.325$) nor the multi-episodic judgments after the 3rd usage episode ($W = 188.50$, $p = 0.368$) are significantly different, i. e., the between-subject design did not produce a difference between the two conditions. These usage episodes resulted in an episodic MOS of 4.7 (0.6). The multi-episodic judgment after the 3rd usage episode resulted in a MOS of 4.6 (0.6). Table I shows the episodic judgments of the 5th usage episode and the final multi-episodic judgments for the two conditions. With regard to episodic judgments of the 5th usage episode, no significant difference between episodic judgments is found ($W = 111.50$, $p = 0.123$), resulting in an episodic MOS of 1.0 (0.6). The episodic judgments for best performance episodes versus degraded episodes are significantly different ($W = 2.50$, $p < 0.001$). Moreover, the multi-episodic judgments after the 6th usage episode are not significantly different between the two conditions ($W = 181.00$, $p = 0.507$).

[1] As audio book Isabelle Allende's *City of the Beasts* (German, ISBN: 978-3867171915) was used.

TABLE I. Episodic MOS and multi-episodic MOS. Standard deviation in brackets.

Condition	5th usage episode		Multi-episodic Judgment
	Duration	Episodic Judgment	
A	similar	0.8 (0.6)	3.9 (0.6)
B	double	1.2 (0.5)	3.7 (0.6)

V. Discussion & Conclusion

The results show that a neglect of duration could be observed twice: for episodic judgments and the final multi-episodic judgments. For both judgments, no significant differences between the two conditions could be observed, although the 5th usage episode was either presented with the same duration or in doubled duration. On the contrary, increasing the number of degraded usage episode (with a similar duration) and thus a longer *overall* duration of experienced degradation is expected to result in a reduction of multi-episodic judgments. The results indicate that first the episodic perceived quality is derived and memorized, which is then used to derive the multi-episodic perceived quality. Moreover, this leads to the conclusion that the episodic judgments reflect the impact of each usage episode on multi-episodic judgments sufficiently, i. e., no further information about the duration seems to be required. This is an important aspect for the prediction of multi-episodic judgments, as prediction models were so far based on episodic judgments alone [4], [5]. The underlying reason for the observed duration neglect could not be determined in this experiment. It must be noted that in this experiment, the performance within a usage episode remained nearly constant. This leaves open if a duration neglect also occurs in the presence of noticeable varying performance within an episode. In addition to this, also taking episodic judgments, which were used for verification of the between-subject design, might affect the formation process of multi-episodic judgments. Moreover, a duration neglect was in this experiment only investigated for one session of multi-episodic use with a simulated audio-on-demand service by presenting an audio book. This leaves open if a neglect of duration can also be observed in usage periods spanning several days as well as for different types of telecommunication services.

References

[1] B. Weiss, D. Guse, S. Möller, A. Raake, A. Borowiak, and U. Reiter, "Temporal development of quality of experience," in *Quality of Experience*. Springer International Publishing, 2014, pp. 133–147.

[2] B. L. Fredrickson and D. Kahneman, "Duration neglect in retrospective evaluations of affective episodes," *Journal of personality and social psychology*, vol. 65, no. 1, pp. 45–55, 1993.

[3] D. S. Hands and S. E. Avons, "Recency and duration neglect in subjective assessment of television picture quality," *Applied Cognitive Psychology*, vol. 15, no. 6, pp. 639–657, Nov. 2001.

[4] S. Möller, C. Bang, T. Tamme, M. Vaalgamaa, and B. Weiss, "From single-call to multi-call quality: a study on long-term quality integration in audio-visual speech communication," in *12th Annual Conference of the International Speech Communication Association*. Florence, Italy: ISCA, 2011, pp. 1485–1488.

[5] D. Guse, B. Weiss, and S. Möller, "Modelling multi-episodic quality perception for different telecommunication services: first insights," in *Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, Sep. 2014, pp. 105–110.

[6] ITU-T Recommendation P.832, *Subjective performance evaluation of hands-free terminals (05/2000)*. International Telecommunication Union, May 2000.