The Trouble with Choice: An Empirical Study to Investigate the Influence of Charging Strategies and Content Selection on QoE

Andreas Sackl, Patrick Zwickl
FTW Telecommunications Research Center Vienna
Donau-City-Straße 1/3rd floor, 1220 Vienna, Austria
Email: {sackl|zwickl}@ftw.at

Peter Reichl
Université Européenne de Bretagne/Télécom Bretagne
2, rue de la Châtaigneraie, 35576 Cesson Sévigné, France
Email: peter.reichl@telecom-bretagne.eu

***Abstract*—Despite the fact that novel QoE based charging mechanism are vitally needed, the complex interrelation of payment and quality perception has been examined only marginally so far. In this paper we want to describe a comprehensive experiment which investigates the intricate interplay of content selection, quality decisions & evaluation and payment strategies in the context of a video on demand scenario. Beside depicting methodological challenges and providing recommendations for further empirical work, we also compare current findings with our previous work to reveal new sights and research attempts.**

I. Introduction

Telecommunication providers are in an inauspicious position: While service and content providers offering profitable products like cloud based storage solutions or video streaming platforms, the network service providers have to deal with rising traffic volume and decreasing revenues. Due to the market success of flat rate pricing over usage-based alternatives, the best effort Internet access has taken a predominant position. Therefore, beyond technical and economic feasibility of new charging regimes, the attractivity for customers is of high importance in order to motivate a transition. Thus, Quality of Experience (QoE) may serve as principal tool for investigating the customers service satisfaction, which may on economic terms essentially relate to customers loyalty and their willingness to purchase network products. However, the negative influences of pricing on the perceived quality, as described in [1], has insufficiently been empirically discussed until now.

Thus, in this paper we want to discuss challenges of empirical, laboratory based experiments which investigate charging mechanism via the users Willingness-To-Pay (WTP) for enhanced network quality and also its impact on user evaluation, e.g. how does spending money influence subjective quality perception. After a short overview regarding related work and our previous trials in the next section, an updated test setup which enables comprehensive analysis possibilities is presented in section three. In the subsequent section our findings & difficulties resulting from our trial are presented and recommendations for further economic related user studies are given. The last section is reserved for conclusions and acknowledgments.

II. Related Work & Research Questions

Traditionally, WTP is defined as the highest price an individual is willing to accept to pay for some good or service, see e.g. [2]. However, this definition does not take service quality into account, therefore we propose to extend it and define WTP as the highest price an individual user is accepting to pay for a service which is delivered to her at a specific minimum QoE. It can be investigated either via surveys or via observations, i.e., laboratory and field experiments. Surveys are easy to conduct but lead to unreliable results whereas experiments are more complex but also provide more valid findings. Both attempts have been executed in the field of network service quality and charging, e.g. [3] simply asked test participants if they would be willing to pay for the presented video qualities and in the M3I project [4], studies have been conducted in which participants received money in advance which could be spent to increase the video quality of a consumed movie (the remaining deposit was paid out afterwards).

In 2011, we conducted an experimental willingness-to-pay study based on the previous mentioned M3I experiment, in which we improved shortcomings like tariffing complexity of the original experiment, see [5] for more details. Our results revealed some interesting interrelations between QoE and WTP, as well as illustrated that both can be examined within the same trial and both are related to each other: The subjective perception of quality is influenced by previous purchasing decisions, i.e., it seems that quality is evaluated by the customers *more positively* when preceded by a *monetary decision*. In [6] we explained this effect with the socio-psychology theory of cognitive dissonance, which provides a suitable justification of the gained results.

However, these findings are based on two studies which were combined to investigate the effect of charging on QoE. For increasing the consistency of testing (e.g., due to test groups), the current paper will present an improved setup, which investigates QoE ratings with and without user decisions (regarding quality levels, purchases, and content) in the same setting.

Quality Class	Q0	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
VBR [kBit/s]	128	181	256	362	512	724	1024	1448	2048	2896
Priceplan A [€]	0	0.105	0.211	0.316	0.421	0.526	0.632	0.737	0.842	0.947
Priceplan B [€]	0	0.158	0.316	0.474	0.632	0.789	0.947	1.105	1.263	1.421
Priceplan C [€]	0	0.211	0.421	0.632	0.842	1.053	1.263	1.474	1.684	1.895
Quality Class	**Q10**	**Q11**	**Q12**	**Q13**	**Q14**	**Q15**	**Q16**	**Q17**	**Q18**	**Q19**
VBR [kBit/s]	4096	5793	8192	11585	16384	23170	32768	32768	32768	32768
Priceplan A [€]	1.053	1.158	1.263	1.368	1.474	1.579	1.684	1.789	1.895	2
Priceplan B [€]	1.579	1.737	1.895	2.053	2.211	2.368	2.526	2.684	2.842	3
Priceplan C [€]	2.105	2.316	2.526	2.737	2.947	3.158	3.368	3.579	3.789	4

TABLE I
QUALITY CLASSES WITH VIDEO BITRATE AND FEES

Therefore, the following research question is defined in order to act as an audit and extension of our previous results in [5], [6]:

RQ1: What is the influence of quality, content and charging decisions on quality perception?

In our previous work we did not change the prices for quality enhancement during the trial, e.g. in [5] all users were *charged equally* for choosing a certain quality level. Especially regarding *price discrimination* considerations, the influence of varying fees for quality enhancement has to be considered in empirical experiments. Therefore, we extended our setup with the possibility to apply three charging mechanism, named *price plan A, B and C*. As depicted in Table I, each available video quality class *Q0 - Q19* is charged differently depending on the set price plan. The lowest quality class Q0 is always for free, wheres the additionally classes are linearly getting more expensive up to the highest quality class Q19 which is charged with €2 - 4. The influence of varying charing fees for quality enhancement is investigated via the second research question:

RQ2: What is the influence of price plans on quality selection and perception?

In our trial each user has to make overall three quality decisions and for each decision one of the three previously mentioned price plans can be applied, e.g. user 18 has to make her first quality decision under price plan C, her second quality decision under price plan B and her last quality selection under price plan A. Therefore, several *price plan patterns* can be set for each user. Table II depicts the three relevant patterns we used in our trial: we are able to investigate the influence of *increasing prices* (pattern I), the influence of *decreasing prices* (pattern II) and the influence of *constant prices* (pattern III), which is recapped in the third research question:

RQ3: What is the influence of price plan patterns on quality selection and perception?

For a more detailed investigation of influencing factors explaining the customer's selection behavior regarding price and quality, we implemented the quality classes Q16 - Q19: whereas the video bitrate remains *constant* at the highest available value, the prices *increases* up to the maximum charged fee, see Table I. Therefore, it is not rational to chose a higher quality class than Q16, but due to individual preferences a certain market demand may exist. Thus, the last research question can be formulated as follows:

RQ4: What is the influence of unsubstantiated expensive quality classes on quality selection and perception?

Pattern	Description	Iteration 1	Iteration 2	Iteration 3
I	*Increasing prices*	A	B	C
II	*Decreasing princes*	C	B	A
III	*Constant Prices*	B	B	(C/A)

TABLE II
PRICEPLAN PATTERNS

III. STUDY SETUP

Due to space limitations, only a brief overview of our study's complex technical setup, i.e., test components and procedures, is given in the following subsection. There are three QoE ratings for each user-initiated quality/content decision, which provide the basis for our analysis. Therefore, this section is aligned regarding these subjective quality evaluations.

A. Rating I: without user decisions

At the beginning of the trial, our test participants have to evaluate 17 video clips (1080p, h.264 codec, duration of 10 seconds) encoded with 17 videobitrates ranging from 128 kBit/s to 32768 kBit/s, see quality classes Q0 to Q16 in Table I. Each video bitrate is tested once per user. According to [7], our participants use a standard, continuous 5-point ACR-scale ranging from "excellent" to "bad" to evaluate the automatically presented videos and with the yes/no question "Would you consume this video in the presented quality at home?" the so called *acceptance* is determined.

This first round is used to investigate the subjective quality perception *without any user decisions*, i.e., the participants do not make any decision regarding the consumed content, quality or expenditure. In this paper we refer to these measurements as *Rating 1*, see Table III.

B. Rating II: Content and payment decisions

Similar to [5] and [4], our test users received cash in advance of €10, which could be used to increase the presented video quality while consuming a selected movie, i.e. the participants have to pay with *real money* to increase the quality. At the end of the test, the remaining deposit is paid out to the users. Reminder: The lowest available quality was always for free and for the highest quality the user had to pay

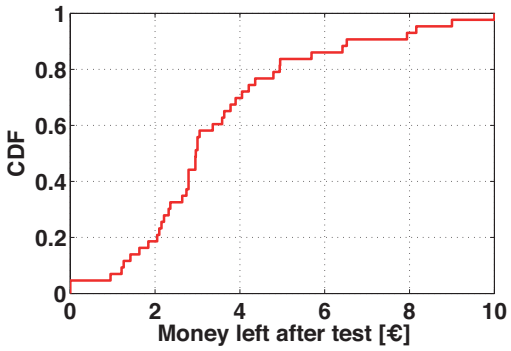

Fig. 1. Paid out money at the end of the test. Only 4.6% of our users did not spent any money on quality enhancement and therefore received the maximum paid out deposit of €10

	Description	Content selected?	Monetary / Quality decision?
Rating 1	begin of study	no	no
Rating 2	after video consumption	yes	yes
Rating 3	end of study	yes	no

TABLE III
QoE Measurements

between €2 and €4 per selected movie, depending on the current price plan, see Table I.

We use an adaptive streaming system, based on Apple's HLS (HTTP Live Streaming) to enable the functionality of adapting the video quality during consumption. Our users are able to chose via an iPad-Interface their preferred videos from a set of 20 action movies, each with a duration of 20 minutes. After a movie is chosen, the test user is able to try out all video qualities by interaction with a so called jog wheel [1], i.e. turning the wheel clockwise increases the quality and vice versa. According to Table I, the user is able to switch between 20 quality classes. The current deposit and the current charge, e.g. €0.211 for quality level Q2 in price plan A, are displayed via a small screen located near the TV set. The description of the effective quality class is *not* displayed, i.e., only monetary aspects are directly visible. Therefore, the user is forced to make a decision based on current *charge* and *perceived quality / QoE* instead of considering *objective, QoS-related* aspects like the displayed name of the currently applied quality class. Missing quality labels also lead to the effect that our users are not able to remember and automatically apply the previously set quality and therefore the actual decision is based on current quality perception and charges.

The user is able to try out all quality classes for free during the first 5 minutes of the video. Thereafter, the last selection is irreversibly set and the related fee is withdrawn from the deposit. After the video has ended, the participant rates the perceived video quality in the same way as described in the previous subsection. We refer to this measurement as *Rating 2*, see Table III.

[1] http://retail.contourdesign.com/?/products/22, last accessed: May 21, 2013

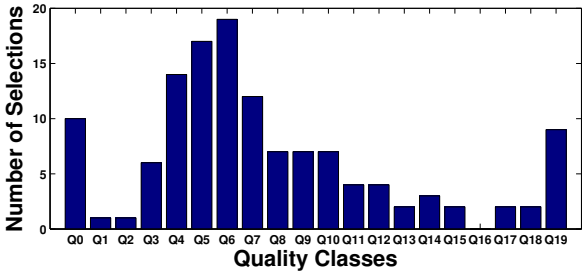

Fig. 2. Number of selections for each quality class for all users and all iterations

Overall, each user selects three videos (iteration 1,2 and 3) and each participant is randomly assigned to a certain price plan pattern, see Table II. For example, pattern II means that price plan C is applied to the first selected movie (iteration 1), price plan B is applied to the second movie (iteration 2) and price plan C is applied to the third selected movie (iteration 3). Overall, 14 users were assigned to patter I, 15 users to pattern II and 14 users to pattern III.

C. Rating III: without payment decisions

At the end of the complete test, i.e., after three movies have been selected, consumed & evaluated and some additional questionnaires have been filled out, the user has to rate again her selected three videos with the selected quality, but of course with no additional payment. We refer to this evaluation as *Rating 3*, see Table III.

IV. Results

Overall, 43 test users from Austria participated in our study (31 female, 12 male). 26 of them were between 18 and 30 years, 10 were between 31 and 45 years old and 7 were older than 45 years. Most of them (16) were employed or students (16), only 8 participants were in a relationship or married and only 3 users were experienced with charged Video on Demand services.

A. Methodological challenges

Empirical user studies including decision processes (regarding content, quality, payment,...etc.) are always critical in terms of *practicability*, i.e., one can only be sure that the experimental design successfully worked *after the study* has been conducted. In our case, only two users did not spend any amount of their deposit to increase the presented video quality, i.e., a large amount of our participants actually spent money and therefore interacted with the system and made decisions regarding enhancement, see Figure 1. This overall behavior shows that our choice regarding deposit, charging, available content and video duration was suitable for the test purpose. For more details, please see [8].

In contrast to usual QoE user studies, e.g. [9], in which every tested condition appears aligned with a test plan, in the current experiment test subjects individually select their

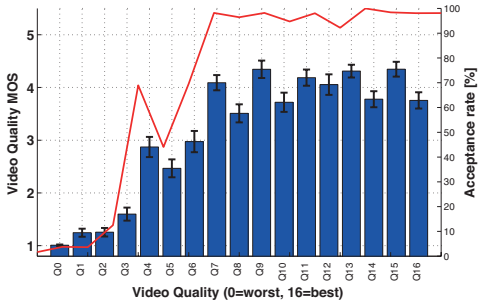

Fig. 3. Quality perception without any influence by charging, quality and content selection. Bars and 95%-CI intervals show video quality MOS ratings whereas the red line indicates acceptance rate

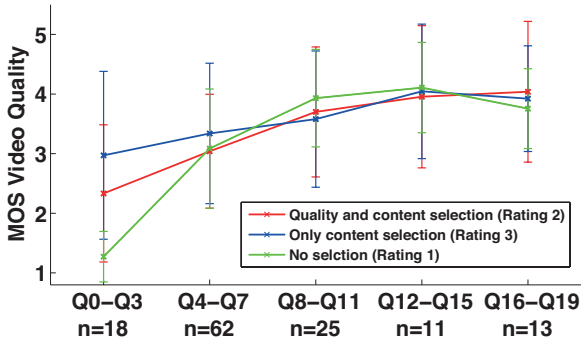

Fig. 4. Original 20 quality classes were aggregated to 5 classes, MOS video quality ratings: 5=best and 1=worst, confidence intervals are given for a level of 0.05, Ratings 1,2 and 3 refer to Table III

condition, i.e., some quality levels may be more popular than others. For example, quality level Q5 was selected 19 times whereas quality class Q17 was chosen only twice. Thus, there may not be meaningful results for each individual quality class as some only have a sample size of two. Therefore, we decided to aggregate our 20 quality classes. After some attempts we decided to combine 4 quality classes to one superclass resulting in 5 superclasses (Q0-Q3, Q4-Q7...etc.) in order to increase the sample sizes. On the one hand we unfortunately lose granularity regarding quality classes, but on the other hand more robust analysis can be conducted.

In the following subsections IV-D and IV-E we compared MOS ratings with other aspects. To guarantee comparability with the MOS (Mean Opinion Score) ratings, we decided to compare *mean* values of the selected quality and the spent money instead of using other metrics, e.g. median values.

B. Quality Evaluation without any user decisions

Figure 3 shows the results regarding subjective quality perception of videos which were not influenced by any user decisions regarding payment, quality or content, see subsection III-A. We observe that over the entire range of quality classes (Q0 to Q16), there is a significant correlation between quality perception (blue bars, including confidence intervals) and acceptance rate (red line). Hence, this illustrates that our

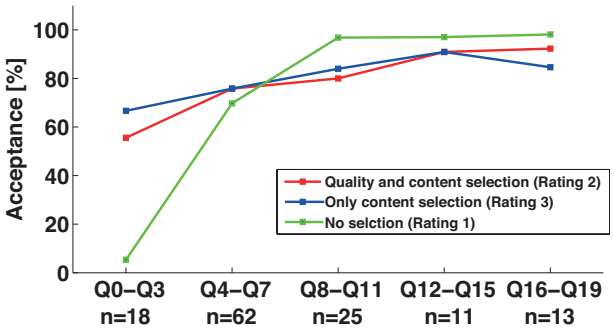

Fig. 5. Original 20 quality classes were aggregated to 5 classes, acceptance is the percentage amount of users who positively answered the question "Would consume this video in this quality at home?", Ratings 1,2 and 3 refer to Table III

choice regarding video bitrates were appropriate, i.e. over all quality classes ratings increases when video bitrate rises. According to Table III, this measurement represents *Rating 1*.

C. Influence of user decisions on quality perception

In [6], we discussed to which extent the social-psychological effect of *cognitive dissonance* leads to different quality perceptions regarding monetary decisions. According to this hypothesis, paying for an enhanced quality level leads to a more positive subjective quality evaluation in contrast to an evaluation without any monetary decisions. As far as the current experiment is concerned, our results depicted in Fig. 4 and 5 cannot confirm this hypothesis. Figure 4 and 5 show that there are no significant differences regarding MOS values and acceptance rate between ratings with and without monetary decisions (red vs. blue line). Figure 6 depicts the differences of the Rating 1 and Rating 2 (DiffMOS): For low quality levels Q0-Q3 and Q4-Q7, the difference is negative, which is contrary to our hypotheses that quality and paying decisions *positively* influence quality evaluation, whereas higher quality levels tendentially lead to positive, smaller differences.

Hence, based on the results of the current experiment, we cannot confirm our earlier hypothesis, and the effect of payment on subjective quality perception cannot be clarified unambiguously: Our previous work indicates a significant interrelationship, whereas the current experiment leads to opposite results. Both experiments could be affected by test setup characteristics, e.g., amount of quality classes, content etc., by unidentified user characteristics or other unknown biases. Therefore, research question RQ1 can not definite & satisfactorily be answered, and we have to leave further clarification to future work.

However, it seems that *active content selection* could have an positive impact on quality perception: The green line in Figure 4 and 5 depicts the results of *Rating 1*, i.e. without any user decision regarding content, quality or charging. Especially for low video qualities scenarios (Q0-Q3), the acceptance rate is clearly lower compared to *Rating 2* and *Rating 3*, and there are also tendencies in the MOS quality ratings, even if they are

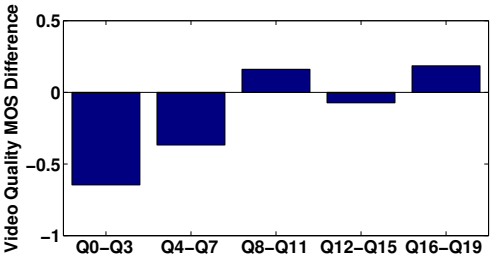

Fig. 6. Difference between ratings with payment and without payment decision (DiffMOS)

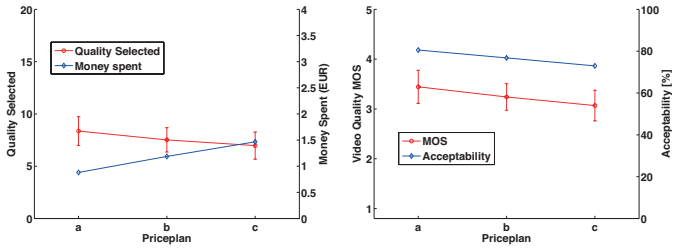

Fig. 8. Influence of price plans on average selected quality (left) and average subjective perceived quality (right)

not significant. *Content selection* as the distinguishing factor between *Rating 1* and *Rating 2/3* could explain the differences in the quality perception.

D. Influence of price plans

As depicted in Table I various price plans were applied, e.g. user 17 was assigned to price plan A for his first movie selection (Iteration 1) and to price plan B for his second movie selection (Iteration 2). Figure 8 (left) shows small changes for both average selected quality (red line) and average spent money (blue line) for different charging: Even if prices increases, on average our participants did not significantly reduce the selected quality (red line), which leads to an increase on average money spent (blue line) when more expensive price plans are applied. At least under the circumstances of our experiment, there is a tendency that users prefer constant quality to constant payments on average.

The MOS ratings in Figure 8 (right, red line) are in line with the average selected quality ratings in Figure 8 (left, red line), i.e., different price plans do not influence the subjective perceived quality.

E. Influence of price plan patterns

As depicted in Table II, various combinations of the price plans A, B and C were applied. In Figure 7 the influence of the patterns on the average selected quality, the average money spent, the video quality MOS ratings and acceptance rate is depicted. Whereas the selected quality for the increasing price plan (A-B-C) did not change during the first two iterations, the subjective ratings regarding video quality MOS and acceptance rate decreased. For decreasing prices (C-B-A), there is hardly any change in selected quality and MOS ratings.

It seems that the subjective quality perception is influenced by previous pricing factors: if a comparison between *current* and *previous* iteration leads to a *negative* evaluation, i.e. the same quality is more expensive now, the resulting subjective perceived quality is lowered. The depicted spent money per price plan pattern is in line with our former findings: whereas quality is held constant, personal payment is adapted to increasing or decreasing prices.

F. Generosity & cognitive dissonance

As shown in Table I, the quality classes Q16, Q17, Q18 and Q19 were identically coded with 32,768 kBit/s but differently charged. From a rational point of view, it does not make any sense to invest in a higher quality class above Q16. Interestingly, nobody chose quality class Q16 whereas higher quality classes have been selected 13 times, see Figure 2. Regarding the three price plans, the distribution of choosing these classes is equally distributed: A=4 (11.1%), B=6 (10.7%) and C=3(8.1%). Therefore, the price plans have no influence on squandering money. We conducted a t-test to compare the DiffMOS values between quality selections above and below Q16, but no differences have been found, i.e. generosity does not effect the quality perception in a negative or positive way.

Thus, *first degree price discrimination* for QoE charging is, beside its legal and regulatory implications, a feasible tool for increasing provider revenues without hampering the product/quality perceptions. In particular, especially the customer segment of quality-seeking customers may be subdivided and targeted by different maximum prices according their willingness-to-pay and price sensitivity of demand (third degree price discrimination).

After the first iteration, our participants filled out a questionary dealing with cognitive dissonance related questions. Among other, they were asked: *How painful was the financial and the quality decision?* As depicted in Figure 9, there is a small tendency that with increasing quality classes and charging fees, the approval of this question increases. But for the quality classes higher than Q16, the approval drops to the lowest value. It seems that spendthrift users have no troubles with paying the highest price even if it is unnecessary from a rational point of view.

There were several other questions handed out after the first iteration which deals with cognitive dissonance related aspects, e.g. *I asked myself if I really needed the selected video quality* or *I asked myself if my purchasing was correct*. There were no significant correlations between these questions and the spent money and also no significant correlations regarding the diffMOS-values. Therefore, these findings do not support our assumption that cognitive dissonance is involved in subjective quality perception and purchasing situations.

V. Conclusion and Outlook

Our experiment clearly shows, that the connection between spending money and quality perception is more intricate than expected. Whereas previous findings indicates that selecting and paying for a certain video quality positively influence

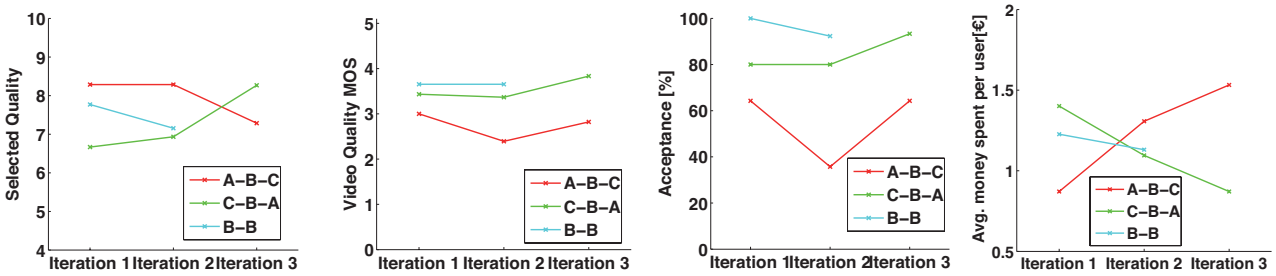

Fig. 7. Influence of price plan patterns on average selected quality, average subjective perceived quality, acceptance rate and average spent money

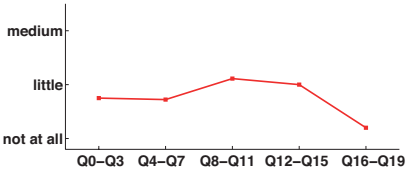

Fig. 9. Ordinate indicates the average approval regarding the question "How painfull was the financial decision and the quality decision?" per quality class

quality perception, the experimental setup presented in this paper leads to inconclusive results. For low quality levels, *quality and payment* decisions have a slightly negative influence on the MOS ratings, whereas *content decisions* positively influence subjective quality evaluation. Further research is needed to investigate the complex interrelationship between these three decision types.

Experimental setups which enable user decision lead to unpredictable outcome. Our trial demonstrated that for the analysis of the gained data, certain conversions are necessary to guarantee valid results, e.g. the combination of selected quality classes. Different metrics could be applied to compare various (user) aspects, e.g. to guarantee comparability we decided to combine MOS ratings with *mean* values of other variable instead of using for example *median values*.

Price plan patterns influence observed user behavior. Whereas *decreasing* prices (C-B) do not affect the selected quality, the subjective MOS and acceptance ratings, *increasing* prices (A-B) negatively affect user ratings, i.e., the selected quality remains constant, but MOS and acceptance ratings decreases. Our experiment was a first attempt to combine pricing strategies with quality perception and especially Figure 7 raises more questions than it answers.

The results presented in this paper and the findings of our previous work [6] indicates that the *acceptance rate* might be a superior indicator for the influences of user decisions than MOS ratings. Further research is needed to investigate this assumption and to progress the underlying methodology.

Until now, in realistic WTP-studies test users received a deposit *in advance* which can be used to enhance the presented quality. To strengthen the claim of validity, a study could be conducted in which participants have to use their own, *brought along money*. Needless to say that this leads to manifold challenges regarding the test design, but the gained results could extend previous findings with new and precious insights.

VI. Acknowledgments

We would like to thank Christoph Piesbergen for providing us with the cognitive dissonance questionnaire. We also like to dedicate special thanks to Stefan Suette for his support during the trial execution. Additional support from the Austrian government, the City of Vienna in the context of the COMET program and from the EU Marie Curie program in the context of the RBUCE WEST International Research Chair NICE is gratefully acknowledged. Parts of this work has been performed in the CELTIC project QuEEN. The research leading to these results received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n ° 248567 for the ETICS project.

References

[1] Peter Reichl, Patrick Maillé, Patrick Zwickl, and Andreas Sackl, "A fixed-point model for qoe-based charging," in *Proc. ACM SIGCOMM 2013 Workshop on Future Human-Centric Multimedia Networking, Hong Kong, China*. 2013, ACM.

[2] Christian Breidert, *Estimation of Willingness-to-pay*, Deutscher Universtitätsverlag, 2005.

[3] M. Ries, O. Nemethova, and M. Rupp, "On the Willingness to Pay in Relation to Delivered Quality of Mobile Video Streaming," in *Proceedings of the International Conference on Consumer Electronics*. 2008, IEEE.

[4] David Hands, Ed., *FP5 Project M3I, IST–1999–11429, Deliverable 15/2 – M3I user experiment results*, 2002.

[5] Andreas Sackl, Sebastian Egger, Patrick Zwickl, and Peter Reichl, "The QoE Alchemy: Turning Quality into Money. Experiences with a Refined Methodology for the Evaluation of Willingness-to-pay for Service Quality," in *Proc. QoMEX (Quality of the Multimedia Experience) 2012, Yarra Valley, Australia*. July 2012, IEEE.

[6] Andreas Sackl, Patrick Zwickl, Sebastian Egger, and Peter Reichl, "The role of cognitive dissonance for qoe evaluation of multimedia services," in *Globecom Workshops (GC Wkshps), 2012 IEEE*, Dec., pp. 1352–1356.

[7] International Telecommunication Untion (ITU), *ITU-R BT.500-7 Methodology for the subjective assessment of the quality of television pictures*, 2002.

[8] Andreas Sackl, Patrick Zwickl, and Peter Reichl, "Qoe alchemy 2.0: An improved test setup for the pecuniary bias of qoe," in *Proc. QoMEX (Quality of the Multimedia Experience) 2013, Klagenfurt, Austria*. July 2013, IEEE.

[9] Sebastian Egger, Peter Reichl, Tobias Hoßfeld, and Raimund Schatz, " 'Time is Bandwidth'? Narrowing the Gap between Subjective Time Perception and Quality of Experience," in *IEEE ICC 2012 - Communication QoS, Reliability and Modeling Symposium (ICC'12 CQRM)*, Ottawa, Ontario, Canada, June 2012.