Presence in a Three-Dimensional Test Environment: Benefit or Threat to Market Research?

Alma Berneburg *

*Department of Business Studies
University of Applied Sciences Merseburg
Geusaer Straße, 06217 Merseburg, Germany
phone, email: +49 (03461) 46 - 24 32, alma.berneburg@hs-merseburg.de
www: www.hs-merseburg.de/~wohlalma

Abstract

In market research, the adoption of interactive virtual reality-techniques could be expected to contain many advantages: artificial lab environments could be designed in a more realistic manner and the consideration of "time to the market"-factors could be improved. On the other hand, with an increasing degree of presence and the notional attendance in a simulated test environment, the market research task could fall prey to the tensing virtual reality adventure.

In the following study a 3D-technique is empirically tested for its usability in market research. It will be shown that the interactive 3D-simulation is not biased by the immersion it generates and provides considerably better test results than 2D-stimuli do.

Keywords: Market Research, 3D user Interfaces, virtual reality, product simulation, package simulation, test environments

Digital Peer Publishing Licence
Any party may pass on this Work by electronic means and make it available for download under the terms and conditions of the current version of the Digital Peer Publishing Licence (DPPL). The text of the licence may be accessed and retrieved via Internet at http://www.dipp.nrw.de/.

1 Introduction

The rapidly increasing development of interactive 3D-technologies is inexorably progressing and has evolved enough to expand from mere scientific visualization sectors into very new and more interdisciplinary areas. Furthermore, these 3D-techniques are an important step towards the development of enhanced virtual reality (VR)-environments: sophisticated and improved three-dimensional impressions and simulations can greatly enhance the realism of a VR-environment. 3D and VR therefore cannot be considered independent, but have to be regarded as interacting, with three-dimensional impressions being a constitutive aspect of virtual reality environments.

A lot of research in recent years has concentrated on the development of VR-techniques and -environments and while this multifaceted research *in* virtual reality is a rapidly developing field, research *with* virtual reality is only just taking off. Especially in market research processes, the adoption of VR-techniques could be expected to contain many advantages: Artificial lab environments could be designed in a more realistic manner, therewith the validity and generalisability of test results would be enhanced, the consideration of "time to the market"-factors could be improved and test results would be achieved quicker and in a cost reduced manner. The inclusion of expensive dummies and real products in a survey could be substituted by highly flexible virtual products and point of sale simulations. Especially the last point is attractive to marketing practitioners who deal with the research in packaging matters and product innovations. First studies showed that by simulating new packages in a packaging test through VR-techniques the flexibility and in

urn:nbn:de:0009-6-12901, ISSN 1860-2037

particular the cost-efficiency of the marketing research process could be substantially enhanced [HB06].

By now VR-simulations have developed sufficiently to generate persuasive test environments and interaction techniques have improved enough to provide natural and intuitive modes of interaction between a test person and its surrounding elements (e.g. technical additives like 3D-glasses or -helmets for presentation purpose or stylus or joystick for data input are no longer necessary; for details see technical appendix). Hence, many distorting influences, which emerge from an often artificial lab environment, seem to vanish with an increasing degree of reality in a market survey. Overall, VR-techniques in market research seem to offer versatile operational areas and deliver a great array of benefits. On the other hand, new techniques potentially introduce new sources of problems which in this case could pose a threat to the quality of survey results. With an increasing degree of reality, provoked by the VR-technology, a rising level of submersion into a lab survey should be expected [ITSK04]. Therewith, the quality of a test person's answers might improve or might even deteriorate instead. With a test person dipping extensively into a VR-environment, the market research task that is to be answered could fall prey to the tensing adventure of a virtual reality experience. Objects of investigation like the measurement of buying decisions or consumer preferences could suffer from an exaggerated concentration on the task that differs substantially from the often habitualised decisions consumers make in their day to day life. Therefore and in order to banish the hazard of either declining a promising new survey tool or, even worse, adopting a warping new instrument in market research, a comparative analyses seems necessary.

To analyse a 3D-technology for its usability in market research, the following test is set up:

- A Choice Based Conjoint Analysis (CBC) dealing with the measurement of consumer preferences is performed in three samples, each one dealing with one of three different stimulus presentation formats to illustrate the test objects:
  - 2D via computer-based 2D-pictures,
  - 3D via a 3D-simulation and
  - real via physical stimuli (dummies).

  One sample had to express their preferences due to 2D-pictures, one due to a 3D-simulation and the last sample judged the test objects due to real physical material on the basis of dummy test objects.

  For the 3D-simulation an interactive 3D-screen was used to build a VR-test environment [HHI07]: The displayed test objects are conveyed as spatially floating in front of the screen. This effect is generated by identifying the respondents' eyes with the help of a head tracker and projecting a separate perspective of the test object in each eye, respectively. The 3D-effect is created without the application of further technical additives (e.g. 3D-glasses) that could enforce the artificiality of the survey. The test person just sits in front of the screen and sees the test objects three-dimensionally [RS04]. Additionally, the test person has the possibility to actively control and navigate through the survey by simply pointing at the displayed tasks and objects (e.g. picking products from the shelf and putting them back with just a fingertip). This abdication of further technical additives, like a stylus etc., is achieved by using a hand tracker scanning the test persons' fingertips [dlB04, HdlB04]. The result is a virtual 3D touch screen. (For more details see the technical appendix). With this technique, virtual objects seem to be more real to a test person than with alternative artificial stimulus presentations like computer-based 2D-stimuli or other 3D-techniques which include technical additives, e.g. helmet, data gloves, etc. (Many so called 3D-online research tools, in fact, consist of two-dimensional visual effects that allow for an interaction with the environment (moving and turning the objects), but a real three-dimensionality with depth and perspective is not achieved. Those techniques will be omitted in the following.)

- In a preliminary interview questionnaire, before the CBC-tasks, every test person is asked questions dealing with his immersive tendencies to assure homogenous predispositions in the three different test groups.

- In a post-interview questionnaire the test persons are asked questions about the intensity of their just made experience to observe the varying degrees of submersion subject to the three different stimulus presentations.

With this setup, the author tries to determine the level of presence generated by the different stimulus

presentations and the potentially negative biases of the test results in a 3D-test environment. If comparable immersive tendencies can be postulated in three homogenous samples, differing intensities of the experienced submersion into the survey environment can be linked to the variant stimuli that have been employed. In a next step, the 3D-test results have to be compared to the results of alternative types of surveys (in terms of the 2D-survey) and against the results of the benchmark of reality in a lab environment (in terms of the real physical stimuli) to answer the question of negative biases to the quality of the test results in the 3D-sample.

The paper is structured as follows: After a brief literature review of the main concepts of virtual reality and the development of three fundamental hypotheses, a section introduces the methods and the test design of the study. In the next step the results and the used statistical techniques are being presented and the findings are discussed. Finally, the implications of these results for the leading hypotheses are being displayed and a summary section will conclude the paper.

2 Basic Concepts of Virtual Reality

Virtual reality can be defined in many very different ways and even the names of the numerous different approaches differ. Virtual environments, virtual worlds, artificial realities, simulated realities or synthetic environments, all of which try to describe the same phenomenon [Bio92, BKL95, Car95, Ebe97]. In the following, the variety of definitions will be subsumed under the most commonly used term "virtual reality". Corresponding to this patchwork of definitions, the techniques used to create a VR are also manifold. A useful description of the technical development was provided by Biocca ([Bio92]: p. 25):

> *"There will be no single type of VR system and no paradigmatic virtual environment. We are more likely to see tailored combinations of components and applications, each capable of producing various types of experience".*

Up to now, this statement has not lost any of its correctness [Kos03]. There continues to be a great array of different so called VR-techniques which is intensified by the fact that virtual reality is still evolving. In the following, the authors agree on the definition of Lanier, who often is seen as the father of the term. According to him a new level of experience is generated with the help of technological solutions that synthesize a new reality (Lanier in [HS91]). In his original definition, Lanier used the idea of data suits as the gate to a new, virtual reality. Of course, nowadays one has to loosen this tight definition and look at the quintessence of this idea: a technology-based gate to another reality. This definition implies two main components of VR: the help of computer-based technological solutions and a new level of experience that is based on an illusion but generates a real experience. To apply these thoughts to the current problem, we will look at the two main components of VR and analyse their occurrence in the new 3D-test environment. Lanier's technology requirement is obviously fulfilled and details can be seen in the technical appendix of this paper. On the other hand, the second component of Lanier's definition is not as self-evident but nevertheless existent. A new level of experience is generated because the test person is abducted by a virtual buying situation while physically still sitting in front of an experimental stimulus in a lab environment. Thus, in accordance with Lanier's definition, the 3D-screen seems to generate a virtual reality.

The existing virtual reality, which is created with the used 3D-technology, now implies some side-effects. One important phenomenon in this context is the degree of "presence" that is generated with virtual reality. Presence in this context can be traced back to Sheridan [She92], who shaped the term "virtual presence". This concept describes the notional attendance at a simulated, synthetically generated place while physically being in a totally different situation. When engaging in the concept of presence, one can experience the same diversity as in the definition of virtual reality. The author will follow Steuer, who defined presence as *"the sense of being in an environment"* ([Ste92]: p. 75), and Biocca, who concretised this statement with his assignment

> *"The shorter and more common term, presence, has been generalized to the illusion of 'being there' whether or not 'there' exists in physical space or not"* [Bio97].

Supporting these statements, three constitutive aspects of presence can be consulted, that have been substantiated by Slater and Wilbur [SW97]:

1. The sense of being in a special situation that is presented by the virtual environment.

2. The degree to which the virtual environment dominates the real environment.

3. The degree to which the test person remembers the virtual environment as "real".

As these three points are difficult to grasp, Sheridan's attempt to influence presence shall assist the above mentioned statements ([She96]: p. 243). According to him, the degree of presence is considered to depend on the

- *"information content of the stimulus independent of the observer*
- *ability of the observer to freely modify the 'viewpoint'*
- *ability of the observer to modify the configuration of the environment."*

To apply these thoughts to the aforementioned 3D-technology, the following can be observed: The sense and illusion of being at the point of sale and making a buying decision is presented by the virtual environment created by the used 3D-technique. For the time of the survey, the test person's reality is dominated by the virtual environment – a fact intensified by the test person's ability to freely move and observe the virtual environment from different angles. Furthermore the possibility to interact with the test environment (the test person is able to "grab" products from the simulated shelves, to put them back and grab new ones, or to put the products into the shopping basket) is provided. All in all it can be assumed that the 3D-screen creates presence. The level of presence though will still have to be determined.

In the context of presence "immersion" is a closely associated topic. Again, the relevant literature provides multiple views and definitions to this term and over the years, two main opposing understandings evolved. Witmer and Singer see immersion as

> *"... a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences"* ([WS98]: p. 227).

This definition reminds strongly of the general definitions of presence and does – in the author's opinion – not discriminate enough between the technological components that evoke presence and presence itself. Therefore in this paper the definition of Slater and Wilbur is preferred. According to them

> *"... Immersion is a description of a technology, and describes the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of a human participant"* ([SW97]: p. 604f.).

Immersion, hence, is directly connected with the technological solution creating the VR. A highly immersive technology will thus create a virtual reality that provides a strong presence. The focal point here lies in the development and analyses of technological components and applications that carry the test person to a virtual environment with the aim that – for the time being – he accepts this virtual environment as reality. To once again apply these perceptions to the current problem, it is evident, that the 3D-screen is a technology-based solution that is capable of delivering an inclusive, surrounding, and vivid illusion of a new reality as required by Slater and Wilbur.

Recapitulating, the innovative and interactive 3D-technique that is expected to enhance the degree of reality in market research processes seems to provide an immersive technology that generates a virtual reality, which builds the basis for an enhanced degree of presence in the survey. This assumed advantage over more artificial alternatives (e.g. 2D-graphics), though, has to be tested for possible biases in the survey results that could occur due to the tensing new adventure of a virtual reality experience. Objects of investigation could suffer from an exaggerated concentration on the task and therefore a comparative analysis seems necessary.

3 Hypotheses

The primary focus of this study is

1. to verify a higher degree of presence in the 3D-environment in comparison to 2D-environments and

2. to compare the test results of the 3D-test with those of a 2D-survey and additionally to those of a survey using real physical stimuli.

For validation purposes the results of the 3D-test environment will be compared to the ones of the 2D-

presentation as a lower limit and against real dummy-stimuli as an upper limit. It can be assumed that no artificial stimuli can ever beat the realistic impressions received from a physical test, but if the alternative 3D-stimuli deliver comparable results, the generalisability and therewith the validity of the more flexible and cost reduced 3D-test results can – with a good conscience – be taken as sufficient

The first hypothesis proposes a relationship between the dimensionality of the artificial stimuli involved in the survey and the degree of presence that is generated [ITSK04]. The author hypothesises that the 3D-test environment creates a higher degree of presence for the test persons than the 2D-stimulus presentation does:

H1: The dimensionality of the stimulus presentation influences the degree of presence that is created for a test person. A 3D-test environment creates a higher degree of presence than a 2D-test environment.

The next hypotheses deal with the quality of test results. The author hypothesises that the results of a market research study using 3D-stimuli and -test environments are comparable to the ones using physical stimuli and better than the ones using 2D-stimuli:

H2: The test results of the 3D-technique are roughly as good as the test results reached via a classical test involving physical stimuli.

H3: The test results gained by using the 3D-technique are better than the test results reached via a test involving 2D-stimuli.

4 Method

To test the above described interactive 3D-technology on its usability in market research processes, the following study was set up:

4.1 Sample

In November 2005 an overall sample of 181 test persons was drawn. This group emanated from a homogeneous survey population consisting of students of a medium-sized German university. This convenience sample did not distort the surveys' results, as the main survey dealing with consumer preferences was not constructed to measure consumers' attitudes towards the test product as outwardly projectable results in general, but to compare the results from the 3D-VR-technology to those from alternative 2D- and dummy-stimulus presentations.

In convenience samples, the selection of units from the basic population is guided by the principle of easy accessibility. The major disadvantage of convenience sampling is anchored in the trade-off between availability and representativeness. There is a lack of knowledge of how well the results of the sample represent the basic population as a whole. Since the main purpose of this study is not the actual measurement of market shares, but a comparison of the test results gained from different stimuli in a lab experiment, any conceivable sample, as long as it is homogenous across tests, should be feasible. In other words, convenience sampling does not pose a threat to these early test results [OK98].

Furthermore, a between-subject-design was engaged to minimize distorting learning effects and to prevent an overtaxing of the test persons' readiness and patience [HWFM93, AG91]. The test persons were randomly assigned to one of the three samples and results of 54 persons in the classical dummy-study were compared to those of 48 test persons in the 3D- and 79 test persons in the 2D-study, respectively. The smaller sample sizes in the 3D- and the dummy-case resulted from the fact that the setup was somewhat more time-consuming than in the case of the 2D-technique. Therefore, the 2D-survey started slightly earlier and more test persons could be recruited to this sample.

4.2 Measurement

The actual measurement was performed in neutral testing facilities at the university the students were recruited from and can be subdivided into three consecutive steps:

1. Preliminary Interview – Measuring respondents' immersive tendencies with a computer-based self-administered offline interview prior to the CBC-study.

2. Choice Based Conjoint-Study – Measuring respondents' consumer preferences in computer-based self-administered offline interviews. Respondents were split up into three different samples, each dealing with a different presentation format of the stimulus:

(a) artificial test objects on a 2D-screen

(b) artificial test objects on a 3D-screen

(c) physical dummy test objects

The dummy-survey was performed with assistance of an interviewer who presented the randomized product choices to the respondent. The interviewer acted according to randomized compositions of the choice tasks that were given by an assisting computer.

3. Post-Interview – Measuring the achieved degree of presence with a computer-based self-administered offline interview subsequent to the CBC-study.

Preliminary Interview

Parallel to the immersive tendencies questionnaire of Witmer and Singer [WS98], the actual measurement of the consumer preferences has been amended by a preliminary interview to assure homogenous tendencies of the test persons in all of the three groups. Albeit preferring Slater and Wilbur's definition of immersion [SW97], the author presumes the questionnaire of Witmer and Singer to be helpful, when interpreting the so called "immersive tendencies" as tendencies to plunge into a virtual environment and to experience some degree of presence in different environments where the authors themselves generally agree with:

> *"The ITQ [immersive tendencies questionnaire] was developed to measure the capability or tendency of individuals to be involved or immersed . . . "* ([WS98]: p. 230).

The original immersive tendencies questionnaire from Witmer and Singer has been shortened to a reasonable length as the actual CBC-study measuring the consumers' preferences and the post-interview dealing with the experienced submersion are still to come (see Table 1).

All three samples had to answer these questions prior to the CBC measurement of their consumer preferences.

Choice Based Conjoint-Study

The Choice Based Conjoint Analysis goes back on Louviere and Woodworth [LW83] and is nowadays the most common applied version of the traditional Conjoint Analysis [HS02]. The intention of the CBC is to determine consumers' product preferences and to express these preferences with part worth utilities. In this point it is comparable to the traditional Conjoint Analysis.

Choice Based Conjoint Analysis however adds some major advantages to the traditional Conjoint Analysis. It enhances the degree of reality of the survey and therewith the external validity of the results. CBC-surveys consist in consumers expressing their preferences by simply choosing their preferred single product concept from a variety of concepts rather than rating or ranking them. Therefore the task is much closer to a real buying decision at the point of sale in the consumers' everyday life: choosing a preferred concept is similar to what consumers actually do in the market day by day.

As the study at hand also tries to enhance the degree of reality in an experimental lab environment, the usage of a CBC-design seems consequential.

In order to validate the 3D-test environment three comparative CBC-studies (2005 Sawtooth Software, Inc.) using identical test designs were performed as follows:

- One empirical 2D-study was set up performing as a lower benchmark.
- Another empirical study was set up using real physical stimuli – this time performing as an upper benchmark.
- A third empirical analysis represented the products three-dimensionally via the 3D-screen.

In every sample 10 randomized CBC-choice tasks were specified for every respondent and in addition to the random choice tasks a holdout task was included to provide a proximal indication of validity. The test persons were confronted with 3 alternative test objects per choice task. The specific choice criterion used posed as follows: "Which of these products would you consider buying?". A None-Option was not included because respondents should not have easy access to avoidance strategies but explain their preferences, even when only of minor influence.

The product at hand is a shower gel with a fixed brand and package size. The varying attributes are "packaging" and "price". The small amount of attributes involved in the study can be traced back to the need of a constant test design in all of the three comparative surveys to gain comparable test results.

While it would not have been a great problem to simulate products with varying brand, size, packaging, and price in the 3D- and the 2D-simulations, implementing that many attributes in a dummy-test would have been. In a Choice Based Conjoint study respondents express their preferences by choosing one product concept from a number of products described by varying attributes and their levels. This task is very natural for the test person, because it can be compared very easily to their daily behaviour at the point of sale [Orm06]. But to actually adopt a CBC-study in a survey using real stimuli, one has to physically build each potential attribute combination that could occur – a task that is nearly impossible to realise against the background of practical limitations in market research. The limitations that occur from the usage of a CBC-test design so resulted in a relatively small number of used attributes and attribute levels.

The prices varied in a small range because the small and sensitive product differences generated only via the packaging should not be dominated by massive price differences to not reduce the consumers buying decision only on the price-attribute.

Post-Interview

In an additional interview after the main CBC-tasks, the test persons of each of the three samples then had to provide information about their just made virtual experience to measure the degree of achieved presence. Again, extracts of a questionnaire of Witmer and Singer were consulted to measure

> *"...the degree to which individuals experience presence in a VE"* ([WS98]: p. 230).

While the actual phrasing of the questions in some cases had to be rearranged to better suite the particular stimulus of the survey and some questions only were useful in one or two of the tests depending on the stimulus, the inner meaning of the questions was unchanged.

Additionally, the post interview included questions concerning the degree of enter-tainment of the survey in order to evaluate the indirect validity of the tests [HS04] (see Table 4).

When supposing that a survey is of a special interest to a test person and in general is enjoyable, the quality of the given answers and therefore the quality of the whole test and its results can assumed to be better than in a test in which the test person feels impelled to take part. The indirect validity therefore can be measured according to criteria as simplicity, length of an interview, entertainment value, diversification or interestingness which determine a test persons' motivation and therefore indirect the validity of the method [Ern01].

Val1	The survey was easy to handle.
Val2	The survey was too long.
Val3	The survey was interesting.
Val4	The survey was enjoyable.
Val5	The survey was diversified.

Table 4: Statements form the Post-Interview for Measuring Validity and Presence.

Besides measuring the validity of the different surveys, the five above mentioned questions will also allow a statement about the level of presence the test persons experienced. This holds because a survey that is of much interest to a test person can be assumed to create a deeper submersion into the test environment.

5 Findings

In detail the results of the above described comparison are as follows:

Preliminary Test

To assure the same homogenous immersive tendencies in all of the three samples, a comparison of the respective means has been performed based on the answers to the immersive tendencies questionnaire on a scale from "1 = very strong/very much" to "7 = very weak/not at all". In the following table the respective means in the three samples are pictured (see Figure 2).

In the following, the null hypothesis that all groups of data really are sampled from distributions that have the same mean has been tested by a One-Way ANOVA with a 95% level of confidence (see Table 5).

According to the preliminary interview the test persons in the three samples in general show the same immersive tendencies and no significant differences in the means can be identified in most of the cases. Only items it22, it24 and it27 seem to reject the null hypothesis with a p-value lower than 5%, albeit in the case of it22 only marginally.

Rejecting the null hypothesis with the One-Way ANOVA does not mean that the means of every subgroup differ from each other. ANOVA can only tell whether there is a difference between two or more of the groups but not exactly where the differences result from. A multiple comparison test therefore is

Figure 1: Levels of the Attribute "Packaging".

used post hoc to tell exactly which samples are different. With similar variances in the three different samples, the Student-Newman-Keuls-test (SNK) is used to compare all pairs of means. This test compares the differences among means to the critical points of the studentized range, trying to keep the chance of a Type I error in any comparison to be 5%.

In the existing case, SNK is used to fuse the aforementioned ANOVA-decisions and to identify the sources of the significant differences in the means. The SNK-results for the three crucial items (it22, it24 and it27) are shown in the following; for all other items SNK validates the outcomes of the ANOVA.

As can be seen, the null hypothesis in the case of item it22 can be rejected according to ANOVA as well as retained according to SNK. It thereby is technically possible to get "significant" results from a post test even when the overall ANOVA is not significant, because the ANOVA tests the null hypothesis of identical means in all of the groups while the post test tests the null hypothesis of two particular means being identical. Since the post test is more focused it has the power to find differences be-tween groups even when ANOVA did not and vice versa. Hence, in this case one follows SNK and the null hypothesis of identical means in the three groups will not be rejected.

In the case of items it24 and it27 SNK indeed seems to support the fact of significant differences in the means.

In it24 two subgroups can be identified showing that in pairwise comparison the sub-samples in the 2D- and the 3D-test show no significant differences in their means (3.73 vs. 3.92) at a p-value of .604 as well as the sub-samples in the 3D- and the dummy-test marginally show no significant differences in the means (3.92 vs. 4.59) at a p-value of .056. When comparing the mean of the 2D-sample (3.73) against the mean of the dummy-sample (4.59), however, one finds significant differences. The null hypothesis in this case has to be rejected (see Table 7).

In it27 the SNK-results imply no significant differences in the means of the 3D- and the dummy-test (1.67 vs. 1.74) at a p-value of .407, while both differ significantly from the 2D-group with a mean of 1.47. The null hypothesis of similar means in the three subgroups also has to be rejected in the case of it27 (see Table 8).

When considering the relevance of these findings for the question of homogeneous immersive tendencies in the three samples, one has to take into account that these crucial items' importances in determining the level of immersive tendencies is quite low, though, since they were mainly included in the survey as control questions to identify potential distorting external influences. Since their absolute values indicate no excessive external influence on the immersive tendencies, the differences in their means are of little consequence.

This in combination with the fact that 83% of the items show no significant differ-ences in the means at all leads the author to the conclusion of homogeneous immer-sive tendencies in the three sub-samples.

Choice Based Conjoint-Studies

After assuring homogenous immersive tendencies in all of the three samples via the preliminary interview, the test persons' product preferences gained with the aforementioned choice tasks are analysed at an aggregate level. The utility estimations achieved with the three different stimulus presentations now pose as follows (see Table 9).

In the case of the price-attribute, the comparison of the utility estimations over the three samples paints a uniform picture. Each of the three groups shows a declining preference as the price increases. This con-

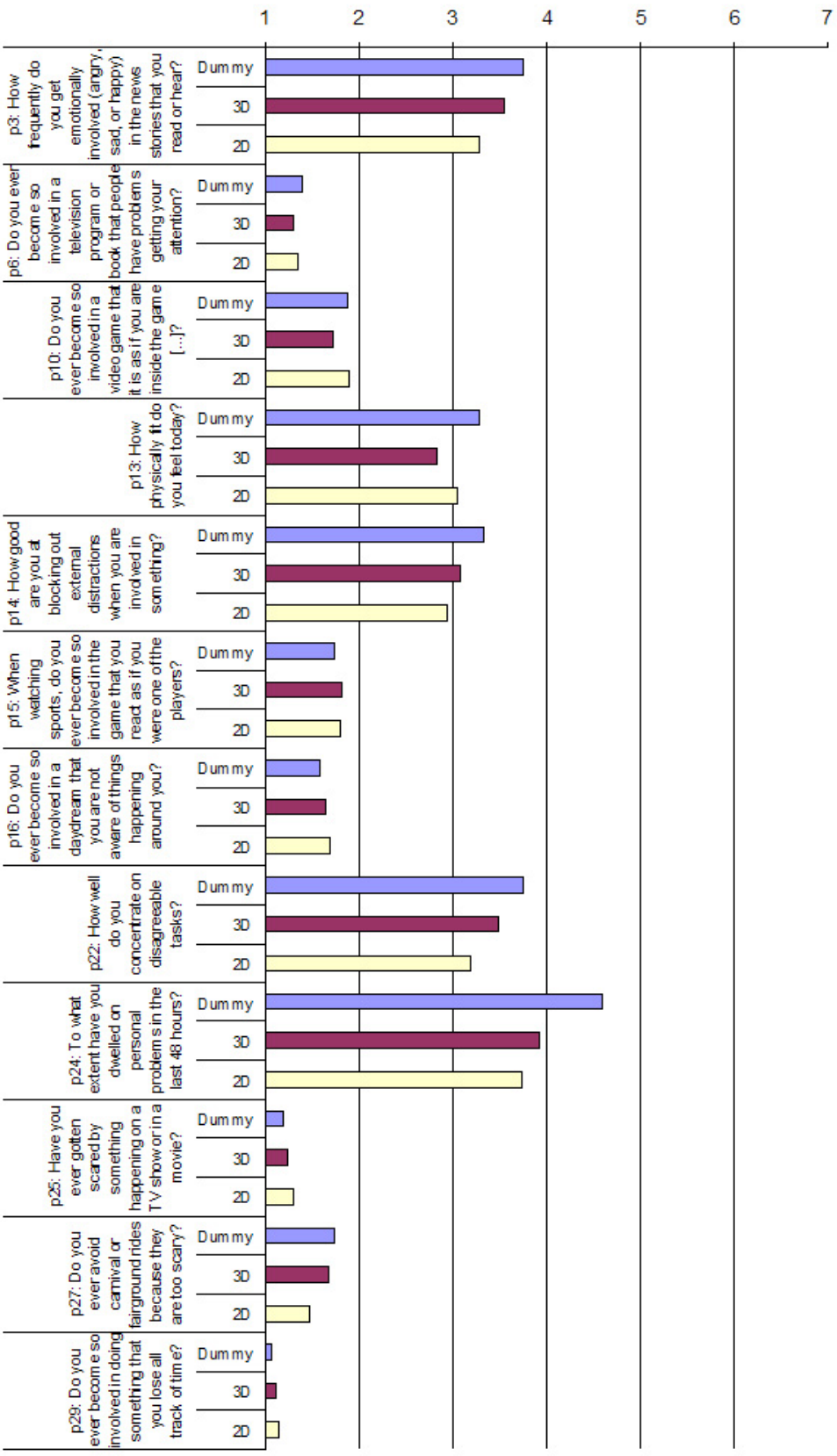


Figure 2: Means of the Answers to the Preliminary Interview.

firms prior expectations, because "price" is an ordered attribute where normally low is preferred to high. This accordance of part-worth utilities with the a-priori expectations can be seen as a confirmation of reliability [OK98].

The comparison of consumers' preferences with regard to the packaging, however, shows a different picture: While the 3D- and the dummy-utilities result in the same rank order, the utilities in the 2D-case suggest a different ranking. In other words, while the stimulus presentation in the 2D-case seems to deliver biased results when testing an attribute that depends strongly on the visual impressions, the 3D-stimulus presentation does not seem to be biased when supposing that the real physical stimuli deliver the benchmark (the test results nearest to the real consumer preferences).

To further analyse the differences in the answers of the three samples, conjoint importances have been calculated by taking percentages of the differences between the best and the worst utility for each attribute on an aggregate level. This allowed to obtain a set of attribute importance values that add up to 100%. They describe the impact – at the given range of levels – of each attribute on the consumers' decision. Here one has to keep in mind that conjoint importances depend on the respective attribute levels in the particular study. With an even narrower range of price for example, this attribute would have forfeited its importance [Orm06]. The comparability of the importances between the samples, however, remains unchanged by this fact, as the same attribute levels were included in every one of the three samples (see Figure 3).

The relative importance of the attribute "price" is bigger than the relative importance of "packaging" when again regarding the dummy-sample as the benchmark. The results in the table show that the 3D-sample displays a similar distribution of the importances. In both cases the price seems to have a bigger impact (60%) on the consumers' decision than the packaging has with only roughly 40%. In the 2D-sample, though, the results show a different picture. Here the two observed attributes both nearly express the same level of importance (50/50) and seem to have a comparable impact on the consumers' decision.

Generally, our data tell a comforting story. It suggests that the 3D-survey performs just as well as the dummy-survey with both being equally reliable and valid and resulting in very similar utilities and importances. The 2D-test results, on the other hand, are not equally good and seem to be seriously biased.

Post Interview

In a final step, the role of the virtual reality will have to be analysed specifically, in order to determine the actual degree of presence generated by the 3D-technique. Due to the fact that in the majority of cases it would not deliver further insights when looking at the questions that show no significant differences, in the following not all of the numerous post interview-questions in the three different samples will be presented. At this point, therefore, only those items will be displayed that either show significant differences between the three samples or where the lack of differences is of a special interest to the topic of this study (further details can be received by the author).

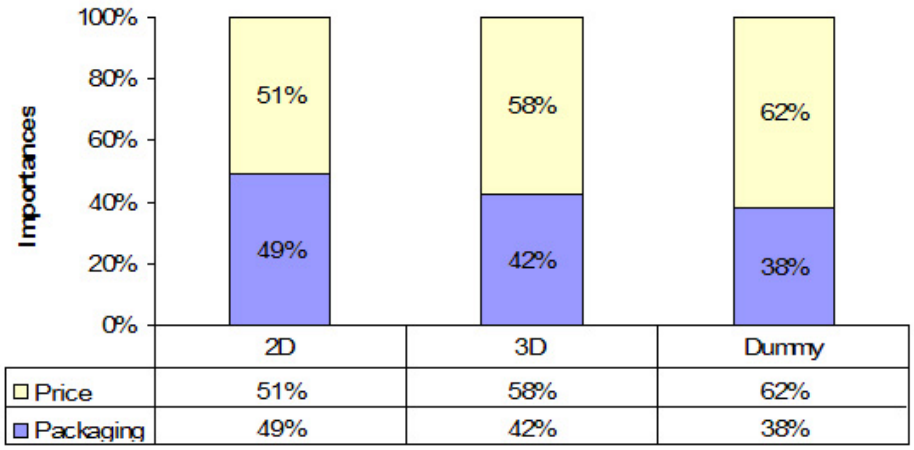
At first a look will be taken at the answers to the presence questionnaire by Witmer and Singer. Hereby the focus lies on a comparison of the 2D- and the 3D-sample, because the thought of presence in a virtual environment can not be adopted in the same way for the dummy-survey. Since the special interest of this study is the analysis of the effects of an enhanced degree of reality in a VR-market research anyway, the dummy-survey is of minor interest at this point.

In the additional presence-interview after the main CBC-tasks the test persons were asked to rate their experiences in the virtual test environment on a scale from 1 = "very much" to 7 = "not at all". While most of the cases showed no significant differences in the means, the following item was of a special interest (Table 10).

According to a t-test with a 95%-level of confidence the respective means are significantly different from each other and the null hypothesis that both groups of data are sampled from distributions that have the same mean has to be rejected. This means – with the given scale – that the test persons in the 3D-study felt significantly more involved in the virtual experience than the test persons in the 2D-study and the supposition of a higher degree of presence in the 3D-VR-environment can therefore not be rejected.

When looking at the given statements the test persons had to evaluate as a supporting measure of validity and presence (now dealing with all of the three samples again), one gets a similar impression (Table 11).

All of the three surveys were equally easy to deal with at a p-value of .334 and no significant differences could be isolated leading to the conclusion that the 3D-technique would be more complicated to handle and therefore could disturb the submersion in the test en-



	2D	3D	Dummy
Price	51%	58%	62%
Packaging	49%	42%	38%

Figure 3: Conjoint Importances in the Three Treatments.

vironment (SNK supported this result). This finding is especially interesting regarding the fact that elder 3D-techniques using a data helmet to create the three-dimensional effect resulted in considerable distortions to the test results due to problems in the handling of the technique.

In the case of the question, whether the survey was too long, a post-hoc SNK-test confirms that there are no significant differences between the three surveys (Table 12).

Since a p-value of .064 is not very comforting, one should take a closer look at the means of the three samples. With the differences in the means just not being significant, the test persons in the 2D- and in the dummy-sample could be assumed to agree stronger to the statement than the test persons in the 3D-sample on a scale from "1 = I agree very much" to "5 = I do not agree at all".

All in all, the subjective feeling of time does not seem to be influenced negatively through the stimulus presentation – this does not imply, though, that each test is equally entertaining. The only conclusion that is possible is that in this special case all of the test persons seemed to be willing to participate in the survey and did not feel bothered by it. Whether the 3D-test overall was more entertaining than the two alternatives cannot be answered by this item. Since this is a very interesting topic, though, it should be included in further research; possibly the 3D-test is thrilling enough to improve high breaking off rates and fatiguing surveys.

In the case of the following two questions the null hypothesis should be rejected according to the One-Way ANOVA and again a multiple comparison test was used to tell exactly which samples show a difference in their means. With comparable variances in the three different samples, again a Student-Newman-Keuls-test was conducted to identify the sources of the significant differences in the means.

Obviously on a scale from "1 = I agree very much" to "5 = I do not agree at all" the 3D- as well as the dummy-test were significantly more interesting to the test persons than the 2D-test, suggesting that the more realistic 3D- and dummy-stimuli lead to an improved quality of the test and its results as the test persons are more focused upon the survey. Moreover, an increased degree of interest in the survey by the test person should also lead to an improved validity of the results.

The following item supports this fact (Table 14).

On the scale from "1 = I agree very much" to "5 = I do not agree at all" the 3D-sample seems to make significantly more fun during the survey than both of the other samples with the 2D-sample being the least enjoyable. This fact not only leads to the conclusion that the quality of the answers can be enhanced when using the new technique and therefore the validity of the survey gets enhanced, but also can give a hint according to the presence on which the main focus in this study lies. When experiencing significantly more fun in a survey than one would have in alternative methods, the submerging into the tasks can also be assumed

to be deeper; the concentration on the survey will be simplified and intensified at the same time.

6 Discussion of the Hypotheses

The primary focus of this study was

1. to verify whether the 3D-environment generates a higher degree of presence than a simple standard 2D-environment and
2. to compare the test results of the 3D-test to those of the 2D-survey and additionally to those of a survey using real physical stimuli (dummy).

The first hypothesis the author examined proposed a relationship between the dimensionality of the artificial stimuli involved in the survey and the degree of presence that is generates. When looking at the results of the aforementioned analyses, this hypothesis is supported by the study. According to the post interview, the 3D-survey creates a higher degree of presence and the test persons feel more involved than in the 2D-test. When assuming a higher degree of presence in a 3D-test environment the question of possible biases to the results of a market research task comes up.

The final two hypotheses dealt with the quality of the test results, as possible biases can be detrimental to eventually drawn conclusions. The author hypothesised that the results of a survey using 3D-stimuli and -test environments are comparable to the ones using physical stimuli and better than the ones only using 2D-stimuli. The aforementioned analyses showed that the CBC-test results of the 3D-test were comparable to the benchmark of the dummy-survey and better than the ones gained with the 2D-survey and so again the hypotheses are supported.

7 Summary and Outlook

The purpose of the existing study was to analyse the effects of a VR-3D-technique regarding the degree of presence it generates and the influence on the test results it has in a market research environment.

The rising level of submersion into a lab survey that was assumed in the beginning of the study could be confirmed when using a 3D-technique in contrast to simpler artificial 2D-stimuli. In other words, as the level of realism of the artificial stimulus presentation increases, the test persons experience a more intense submersion into the survey.

Furthermore, the quality of a test person's answers was analysed, because as a test person extensively submerges into a VR-environment, the market research task could fall prey to the tensing new adventure of a virtual reality experience. The comparisons of the 3D-CBC-test results to the results of more simple artificial 2D-stimuli and to the results of the benchmark of reality in terms of the real physical stimuli show that despite the deeper submerging of the 3D-sample there are no negative biases to the quality of the choice task results in contrast to the strongly biased re-sults from the 2D-sample. The 3D-technique seems to deliver unbiased test results that outperform the results of the alternative 2D-technique in terms of quality and comparability to the dummy-benchmark.

Overall, the test results that are gained with the 3D-technique seem not to suffer from the increase in presence that is generated. In the case of utility estimations in comparison to the two control groups (2D and dummy), the results of the 3D-survey can be considered realistic, valid and generalisable, while still being less costly and quicker to obtain than dummy-results and, above all, much more flexible and useful when planning to integrate a larger number of attributes in a conjoint study. Negative biases to the test results in a quasi-real 3D-test environment do not materialize.

Of course more comparative studies will have to come, dealing with a further investigation of the kind and degree of presence 3D-techniques generate and further analysing the quality of survey results. Also the influence of presence in a market research environment certainly has to be analysed in a broader fashion before a trustful commercial use of this kind of technique is imaginable in a lab environment.

A Technical Appendix

To give a detailed explanation of the technique used in this study, the following im-portant components will be introduced [HHI07]:

The Fraunhofer HHI 3D Kiosk system offers simple intuitive handling as it presents any kind of object in a virtual manner on a large screen in photorealistic 3D quality. 3D-objects seem to float in front of the display for which no further technical additive like 3D-glasses or -helmet is necessary. The user does not need special aids (e.g. stereo glasses). He interacts by simple gestures pointing at objects in a virtual space. That kind of user interaction in combination with depth representation ensures a totally new dimension of fascination.

Head tracking is essential for showing 3D content with autostereoscopic displays. The used head tracking module tracks the user's eye position and enables the permanent detection of the test persons' position so that the three-dimensional projection can be perpetuated even with the test person moving in front of the screen. No glasses or helmets are needed to achieve the three-dimensional impression as the monitor projects a single picture in each eye separately to produce a stereovision effect.

Figure 4: Fraunhofer HHI 3D Kiosk.

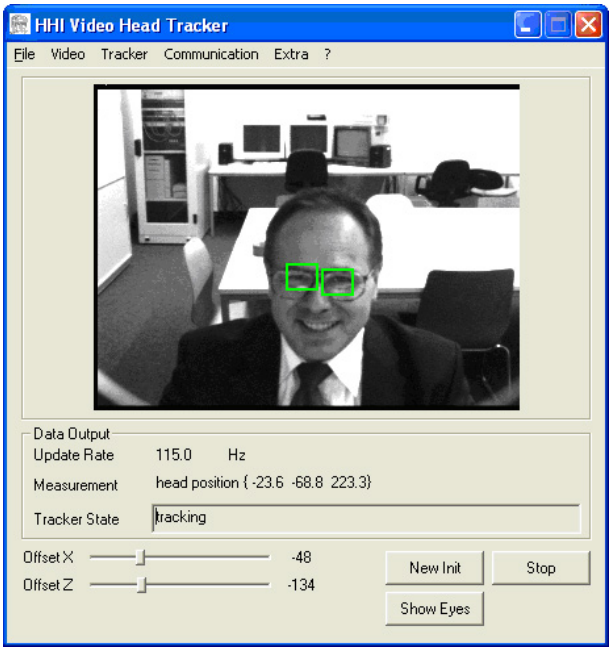


Figure 5: HHI Video Head Tracker.

A very natural pointing tool is created by combining the head (pupil) tracker with a hand (finger tip) tracker: A camera based hand tracker detects hand gestures. It recognizes the position of the finger tip, which can be used for pointing at or for moving virtual objects represented with the stereoscopic display. The test person in front of the screen is able to touch, pick and turn the three-dimensional objects as if they were real without any further support of a stylus, glove or the like (virtual 3D touch screen). With such a tool the user can easily control the position of a marker on the screen by pointing in the desired direction.

Figure 6: Virtual 3D Touch Screen.

References

[AG91] Manoj K. Agarwal and Paul E. Green, *Adaptive Conjoint Analysis versus Self Explicated Models: Some Empirical Results*, International Journal of Research in Marketing **8** (1991), 141–146, ISSN 0167-8116.

[Bio92] Frank Biocca, *Virtual Reality Technology: A Tutorial*, Journal of Communicatio **42** (1992), no. 4, 23–72, ISSN 0021-9916.

[Bio97] Frank Biocca, *The Cyborg's Dilemma - Progressive Embodiment in Virtual Environments*, Journal of Computer-Mediated Communication **3** (1997), no. 2, ISSN 1083-6101.

[BKL95] Frank Biocca, Taejong Kim, and Mark L. Levy, *Communication in the age of virtual reality*, ch. The Vision of Virtual Reality, pp. 3–14, Lawrence Erlbaum, Hillsdale, 1995, ISBN 0-8058-1550-3.

[Car95] Karen Carr, *Simulated and virtual realities – elements of perception*, ch. Introduction, pp. 1–10, Taylor & Francis, London, Bristol, 1995, ISBN 0-7484-0128-8.

[dlB04] René de la Barré, *Der Griff zum virtuellen Objekt – Interaktive 3D-Kiosksysteme mit Gesteninterpretation [Grabbing the Virtual Object – Interactive 3D-Kiosksystems with Gesture Interpretation]*, Proc. Electronic Displays 2004, 19. Konferenz fr Bildschirme und Anzeigesysteme, ihre Bauelement- und Baugruppen (Hagenburg (Germany)), 2004, in German.

[Ebe97] Samuel Ebersole, *A brief history of virtual reality and its social applications*, Colorado: University of Southern Colorad, 1997, http://faculty.colostate-pueblo.edu/-samuel.ebersole/336/eim/papers/-vrhist.html, last visited February 25th, 2008.

[Ern01] Olaf Ernst, *Multimediale versus abstrakte Produktpräsentationsformen bei der Adaptiven Conjoint-Analyse: Ein empirischer Validitätsvergleich, [Multimedial versus abstract Forms of Product Presentation in Adaptive Conjoint-Analysis: an empirical Comparison of Validity]*, Peter Lang, Frankfurt am Main, 2001, ISBN 3-631-37708-8, in German.

[HB06] Bruno Horst and Alma Berneburg, *Recht - personal - Ökologie - unternehmung: Festschrift fr prof. dr. manfred kohler zum 65. geburtstag*, ch. Virtuelle Welten virtuos genutzt: 3D-Verpackungstests zur optimalen Nutzung knapper Ressourcen [Virtual worlds used in a virtousic manner: 3D-packaging tests for an optimal use of limited resources], Shaker, Aachen, 2006, ISBN 3-8322-5104-9, in German.

[HdlB04] Klaus Hopf and René de la Barré, *3D-Kiosk mit berührungsloser Interaktion [3D-Kiosk with contact-free Interaction]*, Proc. EVA Conference 2004 (Berlin (D)), 2004, pp. 161–162.

[HHI07] Fraunhofer Heinrich-Hertz-Institut, *3d media center – description*, http://www.hhi.fraunhofer.de/en/-departments/im/products-services/-complete-interaction-systems/3d-media-center.html, 2007, last visited February 25th, 2008.

[HS91] Adam Heilbrun and Barbara Stacks, *Cyberspace - Ausflüge in virtuelle Wirklichkeiten*, ch. Was heißt "virtuelle Realität? Ein Interview mit Jaron Lanier [An Interview with Jaron Lanier: Virtual Reality], pp. 67–87, Rowohlt, Hamburg, 1991, ISBN 3-499-18185-1, in German.

[HS02] Adriane Hartmann and Henrik Sattler, *Commercial use of conjoint analysis in germany, austria and switzerland?*, 2002, http://www.uni-hamburg.de/-fachbereiche-einrichtungen/fb03/-ihm/rp6.pdf.

[HS04] Adriane Hartmann and Henrik Sattler, *Wie robust sind Methoden zur Präferenzmessung? [How robust are Methods for measuring Consumer Preferences?]*, zfbf: Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung **56** (2004), 3–22, ISSN 0341-2687, in German.

[HWFM93] Joel C. Huber, Dick R. Wittink, John A. Fiedler, and Richard Miller, *The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice*, Journal of Marketing Research **30** (1993), no. 1, 105–114, ISSN 0022-2437.

[ITSK04] Francesco Isgrò, Emanuele Trucco, Oliver Schreer, and Peter Kauff, *Three-Dimensional Image Processing in the Future of Immersive Media*, IEEE Trans. on Circuits, Systems and Video Technology **14** (2004), no. 3, 288–303, ISSN 1051-8215.

[Kos03] Christian Kosfeld, *Eintauchen in mediale Welten – Immersionsstrategien im World Wide Web [Immersing into Medial Worlds - Strategies for Immersing in the World Wide Web]*, Deutscher Universitäts-Verlag, Wiesbaden, 2003, ISBN 3-8244-4510-7, in German.

[LW83] Jordan J. Louviere and George Woodworth, *Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data*, Journal of Marketing Research **20** (1983), no. 4, 350–367, ISSN 0022-2437.

[OK98] Bryan Orme and W. Christopher King, *Conducting full-profile conjoint analysis over the internet*, http://www.sawtoothsoftware.com/-download/techpap/internet.pdf, 1998, last visited February 26th, 2008.

[Orm06] Bryan Orme, *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, Research Publishers LLC, Madison, WI, 2006, ISBN 0-9727297-4-7.

[RS04] Sylvain Renault and Oliver Stachel, *Unterstützung multimodaler und natürlicher Interaktionen durch die WORKBENCH3D bei der Modellierung und Visualisierung von VR-Umgebungen auf autostereoskopischen Endgeräten [supporting multimodal and natural interactions with the help of WORKBENCH3D for modelling and visualising vr-environments on autostereoscopic terminals]*, Internetpräsenz des 3D-Display-Innovationsforums, 2004, in German.

[She92] Thomas B. Sheridan, *Musings on Telepresence and Virtual Presence*, Presence: Teleoperators and Virtual Environments **1** (1992), no. 1, 120–125, ISSN 1054-7460.

[She96] Thomas B. Sheridan, *Further Musings on the Psychophysics of Presence*, Presence: Teleoperators and Virtual Environments **5** (1996), no. 2, 241–246, ISSN 1054-7460.

[Ste92] Jonathan Steuer, *Defining Virtual Reality – Dimensions Determining Telepresence*, Journal of Communication **42** (1992), no. 4, 73–93, ISSN 0021-9916.

[SW97] Mel Slater and Sylvia Wilbur, *A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments*, Presence: Teleoperators and Virtual Environments **6** (1997), no. 6, 603–616, ISSN 1054-7460.

[WS98] Bob G. Witmer and Michael J. Singer, *Measuring Presence in Virtual Environments: A Presence Questionnaire*, Presence: Teleoperators and Virtual Environments **7** (1998), no. 3, 225–240, ISSN 1054-7460.

Citation

Alma Berneburg, *Presence in a Three-Dimensional Test Environment: Benefit or Threat to Market Research?* Journal of Virtual Reality and Broadcasting, 5(2008), no. 1, March 2008, urn:nbn:de:0009-6-12901, ISSN 1860-2037.

it3	How frequently do you get emotionally involved (angry, sad, or happy) in the news stories that you read or hear?
it6	Do you ever become so involved in a television program or book that people have problems getting your attention?
it10	Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?
it13	How physically fit do you feel today?
it14	How good are you at blocking out external distractions when you are involved in something?
it15	When watching sports, do you ever become so involved in the game that you react as if you were one of the players?
it16	Do you ever become so involved in a daydream that you are not aware of things happening around you?
it22	How well do you concentrate on disagreeable tasks?
it24	To what extent have you dwelled on personal problems in the last 48 hours?
it25	Have you ever gotten scared by something happening on a TV show or in a movie?
it27	Do you ever avoid carnival or fairground rides because they are too scary?
it29	Do you ever become so involved in doing something that you lose all track of time?

Table 1: Immersive Tendency Questionnaire Items in the Preliminary Interview.

Price A	**Price B**	**Price c**	**Price D**	**Price E**
2.29 EUR	2.39 EUR	2.49 EUR	2.59 EUR	2.79 EUR

Table 2: Levels of the Attribute “Price”.

p1	How much were you able to control events?
p3	How natural did your interactions with the environment seem?
p8	How aware were you of events occurring in the real world around you?
p9	How aware were you of your display and control devices?
p10	How compelling was your sense of objects moving through space?
p12	How much did your experiences in the virtual environment seem consistent with your real-world experiences?
p17	How well could you actively survey or search the virtual environment using touch?
p23	How involved were you in the virtual environment experience?
p25	How much delay did you experience between your actions and expected out-comes?
p26	How quickly did you adjust to the virtual environment experience?
p28	How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?

Table 3: Presence Questionnaire Items in the Post-Interview.

		sum of squares	df	mean of squares	F	significance
it3	between groups	7,577	2	3,789	1,795	,169
	within groups	375,660	178	2,110		
	total	383,238	180			
it6	between groups	,246	2	,123	,536	,586
	within groups	40,826	178	,229		
	total	41,072	180			
it10	between groups	,376	2	,188	1,433	,246
	within groups	8,920	68	,131		
	total	9,296	70			
it13	between groups	5,061	2	2,531	1,490	,228
	within groups	302,386	178	1,699		
	total	307,448	180			
it14	between groups	5,053	2	2,527	1,361	,259
	within groups	330,350	178	1,856		
	total	335,403	180			
it15	between groups	,255	2	,127	,734	,481
	within groups	30,905	178	,174		
	total	31,160	180			
it16	between groups	,495	2	,247	1,070	,345
	within groups	41,163	178	,231		
	total	41,657	180			
it22	between groups	10,518	2	5,259	3,059	,049
	within groups	306,001	178	1,719		
	total	316,519	180			
it24	between groups	24,608	2	12,304	3,465	,033
	within groups	632,121	178	3,551		
	total	656,729	180			
it25	between groups	,373	2	,186	1,007	,367
	within groups	32,931	178	,185		
	total	33,304	180			
it27	between groups	2,651	2	1,326	5,796	,004
	within groups	40,708	178	,229		
	total	43,359	180			
it29	between groups	,300	2	,150	1,528	,220
	within groups	17,490	178	,098		
	total	17,790	180			

Table 5: Results of the One-Way ANOVA.

			subgroup for Alpha = .05.
	Stimuli	N	1
Student-Newman-Keuls-Procedure	2D	79	3.19
	3D	48	3.48
	Dummy	54	3.76
	significance		.054

Table 6: Results of the SNK-Test for Item it22.

			subgroup for Alpha = .05.	
	Stimuli	N	1	2
Student-Newman-Keuls-Procedure	2D	79	3.73	
	3D	48	3.92	3.92
	Dummy	54		4.59
	significance		.604	.056

Table 7: Results of the SNK-Test for Item it24.

			subgroup for Alpha = .05.	
	Stimuli	N	1	2
Student-Newman-Keuls-Procedure	2D	79	1.47	
	3D	48		1.67
	Dummy	54		1.74
	significance		1.000	.407

Table 8: Results of the SNK-Test for Item it27.

Attributes	**Levels**	**2D Part-Worth Utilities**	**3D Part-Worth Utilities**	**Dummy Part-Worth Utilities**
Packaging	A	-0.1165	-0.71088	-0.81228
	B	0.42309	-0.05590	0.19884
	C	-0.39962	-0.05964	-0.19857
	D	0.08818	0.82642	0.81200
Price	2.29	0.39675	0.93310	1.26370
	2.39	0.21411	0.71166	0.57831
	2.49	-0.00305	0.13540	-0.08031
	2.59	-0.16129	-0.55842	-0.41865
	2.79	-0.44653	-1.22173	-1.34305

Table 9: Estimated Part-Worth Utilities of the Three Samples.

	Stimulus	N	Mean	Standard Deviation	Mean Standard Error
p23	2D	79	3.43	1.346	.151
	3D	47	2.96	.833	.121

Table 10: Comparison of Means for Item p23 (“How involved were you in the virtual environment experience?”).

		sum of squares	df	mean of squares	F	significance
the survey was easy to deal with	between groups	.944	2	.472	1.103	.334
	within groups	75.701	177	.428		
	total	76.644	179			
the survey was too long	between groups	6.422	2	3.211	2.912	.057
	within groups	195.156	177	1.103		
	total	201.578	179			
the survey was interesting	between groups	20.886	2	10.443	13.197	.000
	within groups	140.064	177	.791		
	total	160.950	179			
the survey was enjoyable	between groups	24.999	2	12.499	15.986	.000
	within groups	138.396	177	.782		
	total	163.394	179			

Table 11: Results of the One-Way ANOVA.

	Stimuli	N	subgroup for Alpha = .05. 1
Student-Newman-Keuls-Procedure	2D	79	3.68
	Dummy	54	3.72
	3D	48	4.13
	significance		.064

Table 12: Results of the SNK-Test for Item Val2 (“The survey was too long”).

	Stimuli	N	subgroup for Alpha = .05. 1	2
Student-Newman-Keuls-Procedure	3D	48	2.40	
	Dummy	54	2.63	
	2D	79		3.19
	significance		.177	1.000

Table 13: Results of the SNK-Test for Item Val3 (“The survey was interesting”).

	Stimuli	N	subgroup for Alpha = .05. 1	2	3
Student-Newman-Keuls-Procedure	3D	48	2.21		
	Dummy	54		2.56	
	2D	79			3.10
	significance		1.000	1.000	1.000

Table 14: Results of the SNK-Test for Item Val4 (“The survey was enjoyable”).