DrawingPresence: A Method for Assessing Temporal Fluctuations of Presence Status in a VR Experience

CHRISTIAN MAI, LMU Munich, Germany
NIKLAS THIEM, LMU Munich, Germany
HEINRICH HUSSMANN, LMU Munich, Germany

One of the main measures to evaluate a head-mounted display (HMD) based experience is the state of feeling present in virtual reality (VR). The detection of disturbances of such an experience that occur over time, namely breaks in presence (BIP), enables the evaluation and improvement of these. Existing methods do not detect BIPs, e.g., questionnaires, or are complex in their application and evaluation, e.g., physiological and behavioral measures. We propose a handy post-experience method in which users reflect on their experienced state of presence by drawing a line in a paper-based drawing template. The amplitude of the drawn line represents the state of presence of the temporal progress of the experience. We propose a descriptive model that describes temporal variations in the drawings by the definition of relevant points over time, e.g., putting on the HMD, phases of the experience, transition into VR, and parameters, e.g., the transition time. The descriptive model enables us to objectively evaluate user drawings and represent the course of the drawings by a defined set of parameters. Our exploratory user study ($N = 30$) showed that the drawings are very consistent between participants and the method is able to securely detect a variety of BIPs. Moreover, the results indicate that the method might be used in the future to evaluate the strength of BIPs and to reflect the temporal course of a presence experience in detail. Additional application examples and a detailed discussion pave the way for others to use our method. Further, they serve as a motivation to continue working on the method and the general understanding of temporal fluctuations of the presence experience.

CCS Concepts: • **Human-centered computing → Usability testing**.

Additional Key Words and Phrases: virtual reality, presence, head-mounted displays, usability, user experience, method

ACM Reference Format:
Christian Mai, Niklas Thiem, and Heinrich Hussmann. 2019. DrawingPresence: A Method for Assessing Temporal Fluctuations of Presence Status in a VR Experience. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 146 (December 2019), 21 pages. https://doi.org/10.1145/3369827

1 INTRODUCTION

The availability of consumer-grade head-mounted displays (HMDs) leads to a growing number of users in the research community and gives industrial users new commercial perspectives. To improve virtual reality (VR) experiences for these domains, methods and tools for evaluation are needed. A unique measure used for VR systems is the quantification of the user's state to feel present in VR. Feeling present in VR is a cognitive state that is created by feeling spatially located in the virtual world, feeling involved in the actions and being able to generate a sense of the things happening around oneself [27].

Authors' addresses: Christian Mai, LMU Munich, Munich, Germany, christian.mai@ifi.lmu.de; Niklas Thiem, LMU Munich, Munich, Germany, niklas.thiem@outlook.de; Heinrich Hussmann, LMU Munich, Munich, Germany, heinrich.hussmann@ifi.lmu.de.

2474-9567/2019/12-ART146 $15.00
https://doi.org/10.1145/3369827

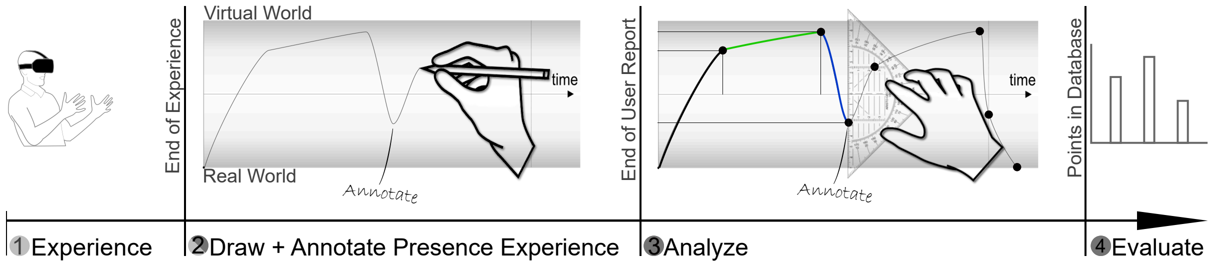

Fig. 1. We propose a post-experience method to collect data about the temporal course of presence during an HMD based experience. After the experience (1), the user *draws* the course of the presence state (2) into a template we provide. The horizontal axis represents the temporal progress, the vertical axis the experienced state of presence. The user *annotates* the drawing (2), in particular breaks in presence. An expert then analyses the drawings by identifying distinctive points and phases (3). Our *descriptive model* describes the distinctive points of a presence experience in the drawings. The model enables to store the drawings as compact numerical data in a database, for storage, comparison and *evaluation* (4) purposes.

Previous research showed that the state of feeling present in VR varies over time [6, 14, 31, 33, 34]. Particularly negative effects on presence in the course of an experience are sometimes noticed by a user, so-called breaks in presence (BIPs) [31]. Mostly known are abrupt BIPs, e.g., people acting in the real world touching an HMD user. However, cognitive processes like getting bored in a VR experience make a user's attention shift between VR and the real world, over a longer period of time [33].

The detection and removal of individual BIPs contributes to a better overall experience [1, 4, 17, 31]. Practicable and applied methods for the detection of BIPs are currently limited to interviews.

Due to ease of use, today more than 90% of studies rely on questionnaires [5, 9, 25], to quantify the presence experience. However, questionnaires are not able to detect single BIPs [4, 6, 17]. Interviews give more detailed insights into the experiences of the users, but are only capable to a limited extent of supporting a continuous and structured report. Hence, they bear the risk to miss the report of BIPs. Further, they do not quantify the presence state, and the analysis of interviews is cumbersome. Physiological measurements enable a continuous and objective evaluation of the users' state of being present. On the other hand, physiological measures are complicated to use, intrusive to a user and the association of the measured signal and the state of presence is often not clear [6, 17]. Behavioral observations analyze users' state of being present based on their actions and are unobtrusive to a user. They do not enable a continuous evaluation and do not fit to a wide range of experiences, as they rely on a certain behavior pattern in a defined situation, e.g., the threat of being hit by a car [17]. In summary, a method is missing that is able to detect BIPs with the practicability of questionnaires and the detailed insight of interviews.

A promising approach to a feasible method are user self-reports of BIPs, as they correlate with the presence value used in presence questionnaires [31]. In existing concepts VR users' are asked to orally report attention shifts towards the real world [31], draw a line [6, 16] or manipulate a slider [4, 10] to express their presence experience. Current approaches are either just in a conceptual state [6], are time expensive in their application [4, 10], do not offer a detailed evaluation of the temporal course of the presence state [31] and might miss certain types of BIPs which develop over a longer period of time [31], e.g., due to boredom.

Hence, we aim at providing a BIP detection method suitable for everyday practice. In addition, we found throughout our work that our method shows potential to assess the strength of BIPs and represent the course of temporal fluctuations. Although, this ability was not the initial goal of our work, the results show great potential for further research and applications. We propose a post-experience method that builds upon on the idea of drawing the temporal course of a presence experience [6] into a template. Users can express the presence state by changing the amplitude of the drawn line either towards the real world or the virtual world (Figure 1(2)). The method is a an improvement of existing methods, as it is not intrusive to the user, can reliably detect BIPs, is fast and easy to use and reflects temporal fluctuations of the presence experience. To enable storage of the data in a database for quantitative evaluation and automation purposes, we propose a descriptive model (Figure 1(3), Figure 3) that we derive from related work. It replaces the users' drawings by describing the important points that occur during a VR experience, e.g., the point describing the end of the transition into VR (e.g., [34]).

In an application example with 30 participants not familiar with the drawing method, we demonstrate its feasibility and potential. In the progress of a 15 minutes long experience, participants were exposed to five BIPs with varying intensity [17] and type [27]. Participants were not asked for or being informed about the appearance. We found that the method allows participants to report temporal progress according to the expected cognitive model derived from related literature. 100% of introduced BIP could be detected and additionally we identified BIPs that we did not find in advance by using established methods. Further, there is evidence that the drawings reflect the strength of a BIP.

With our work we contribute:

- A drawing template and method to collect user reports about their presence experience over time.
- A model describing temporal variations of presence using parameters to replace the drawings by compact numerical data for storage, comparison and evaluation purposes.
- A user study ($N = 30$) that demonstrates that the drawings reflect the expected cognitive model of a VR experience, that the method is superior to previous approaches in the detection of BIPs and that it has the potential to quantify BIPs and the temporal course.

2 BACKGROUND

This section discusses the term presence, followed by the introduction of works that have inspired our method.

2.1 Presence

There are multiple definitions of the term *presence* (e.g. [21, 28, 32, 38]). Hence, we refer to the most common ones in research, according to their application in the form of questionnaires [5, 9, 23], by Witmer et al. [38] and Schubert et al. [28]. We distinguish presence from the term immersion [29]. Immersion describes the measurable attributes of a system enabling to evoke the feeling of being present. The sense of presence is a cognitive construct that integrates incoming stimuli with user attributes, leading to a variable state of presence. The term presence is often referred to a feeling of "being there", the spatial presence [28] or place illusion [30]. Other forms of presence can be experienced like involvement [28] in a story, the general plausibility [30]/ realness [28] of an experience and social presence [8]. In our work, we follow a similar definition by Schubert and colleagues [28], who conducted a meta-analysis on the presence construct including Witmer's results. The orientation on Schubert's three-component scale gives us the confidence to design our method in such a way that it corresponds to most users' cognitive model of presence.

2.2 Assessing Temporal Variations of Presence

In related work we found it particular interesting that self reports depend on the point in time they are conducted in, the time span of a measure and the scale of the measurement.

User report on the presence state during the course of an experience reduces the risk of inaccurate recall or completely missing BIPs. However, continuous reporting interrupts the experience [31]. An example is the manipulation of a mechanical slider which redirects a user's attention towards the real world and occupies the hand for other tasks [4, 10]. BIPs tend to be overemphasized here as users are made aware of them [31]. Further, users being new to a evaluation method, are challenged in reporting an absolute value, before they had a reference experience. In contrast, during a post-experience method, participants can revise their ratings and relate them to each other. A post-experience method might suffer from inaccurate recall and bear the risk of missing the report BIPs [4, 6, 17]. Watching the recording of an experience can support recall [4], which is a time-consuming process. We aim for a post-experience method to avoid distracting users from the experience being evaluated.

Slater and colleagues' presence counter [31] asks participants to report the experience of BIPs when they appear, at a specific point in time. Users do not consciously perceive the state of presence in which they find themselves, but notice a disturbance of the state [31]. Although, this gives a discrete number to work with, information around the measures are lost. Examples are the transition into and out of VR and slow attention shifts, e.g. due to boredom, as they might be superimposed by a short-term disturbance. On the other hand, there is work showing the possibility of participants to express differences in the attention shifts between experiences [6, 14]. Hence, we speculate that users are able to express a continuous presence experience if they are provided with the right tools. Garau and colleagues [6] proposed drawing a continuous line representing the presence state. They found that the drawing fostered the participants' self-reflection and discussion in interviews. The drawing template, method and evaluation was not grounded in related work and was left in a conceptual state. Hence, the collected drawings were very diverse and did not show a course of the experience one would expect in regards to related work. However, a benefit of a meaningful visualization of a continuous measure enables a quick overview on an experience [4, 10]. The overview improves the recall of what you have experienced and thus prevents you from forgetting something in the report. Illogical reports and BIPs become visible very quickly. Hence, we follow Garau's [6] proposal of a line drawing method.

The presence state can be expressed binary [31] or on a scale [4, 10, 17]. Slater's presence counter [31] is a simple to use binary measure that can estimate the presence state. However, it is not able to determine the strength of a BIP [17]. Brogni could confirm that there is a correlation that a growing number of BIPs leads to a lower presence state in VR [1]. As a result, finding, our main goal in this paper, and removing BIPs will increase the overall presence experience. Chung shows that participants are able to report the strength of BIPs by manipulating the position of a slider along a continuous scale [4]. Garau reports that participants indicate BIPs by drawing peaks in a hand-drawn line that indicates the state of being present [6]. As Chung's method is very time consuming, we follow the promising approach of drawing a line, in which participants mainly report the important shifts in their presence experience and leave out parts without any changes in between. The line drawing approach, however, is not yet understood in its possibility to detect BIPs and estimate their strength.

To inform our design, we were inspired by Kujala and colleagues [16]. They propose a post experience graph drawing method similar to Garau [6] that successfully assesses user experience over several months (e.g., [26, 36]). Participants left comments on the curve describing relevant points and primary reported about the hedonic quality of a system [16, 26]. Kujala's method does not fit the context of a VR experience due to the long time-span and the different topic. Further, we expect a better remembrance of pragmatic qualities of a VR experience, as the time between experience and evaluation is shorter.

Depending on the data, there are different approaches to analysis and formal description. As analysis strongly depends on the type of data collected, we focus on the formal description of temporal effects. Slater's presence counter generates an overall presence value, but does not enable the identification of single BIPs [31]. Liebold and colleagues discuss the rationale of BIPs and by this make it possible to estimate the strength of a BIP [17]. Chung and colleagues describe the temporal course of a BIP by the relative change of the presence value and the recovery time from a BIP [3, 4]. Hence, Chung's approach misses a number of other values, e.g., the absolute value of the lowest point of the BIP or the time it takes for a BIP to develop. Ultimately slow evolving BIPs, e.g., caused by boredom, can lead to them not being detected. Hence, to the best of our knowledge, a full conceptualization of the temporal progress of a presence experience is missing.

In summary, we found that user self-reports post experience are a feasible approach to assess a presence experience, in particular to identify BIPs and understand temporal variations. We aim at elaborating Garau's method to enable reliable drawings and provides a comprehensive method for analyzing the data.

3 DRAWING THE PRESENCE EXPERIENCE

We combine the concept of drawing a line into a diagram indicating the presence state [6] with previous concepts that use line drawings as a user experience evaluation method for long term interactions with products [16].

3.1 Design Process of the Drawing Template

We designed the drawing template in an iterative design process based on 60 collected drawings. The drawings were collected in four different user studies applying commercial (Job Simulator VR [19]) and customized experiences. The drawing template was improved between sessions based on participants' feedback. All variations were designed with the goal to enable users' reflection on their VR experience without any moderation. Additionally, we conducted interviews about the course of drawing and the participants' understanding of the drawing template. The template for the drawings included event ticks along the time axis for about one third of the participants (similar to Figure 6). The other two thirds did not have event ticks because the experiences were non-linear (Figure 2). In the final template (Figure 2), we added a dot that indicates the starting point for the drawing. Further, we introduce a linear, gradual transition from pure white to 25% grey from the time axis towards the extremes (See Figure 2). By this, we give the user orientation while drawing along the time axis.

3.2 The Drawing Template

Previous work uses a drawing area with a horizontal time axis and the presence state on top of it [3, 6, 11].

We propose a scale as shown in Figure 2, with the time axis indicating the middle between being present in the virtual and the real world. By introducing a middle line, we provide a balanced diagram that represents a common description of going over to another world when transitioning. The y-axis we chose to be $+/-40mm$. The time-axis we chose to be 200mm between the point $P_{\text{transition}}$ and the $P_{\text{physical exit}}$ (Figure 3). The end of the experience is indicated by a dashed line with a symbol below it, indicating the moment of HMD removal (Section 4). The recall of a presence experience can be supported by adding event ticks to the timeline [6]. Events in an experience are defined by significant predefined tasks or actions, e.g., achieving a goal such as building a furniture or an upgrade of a virtual character to a new level. Ideally, these event ticks have the same order and relative distribution on the timeline as they appeared during the VR experience. Event ticks might influence drawing the line as participants might feel the need to "fit" the lines between the ticks. If there is no possibility to provide event ticks, the supervising person can manually note down significant events during the experience and mark them on the time-axis afterwards. Alternatively, screenshots can be collected for later viewing during the experience. The simplest form of presentation is then on a digital drawing surface.

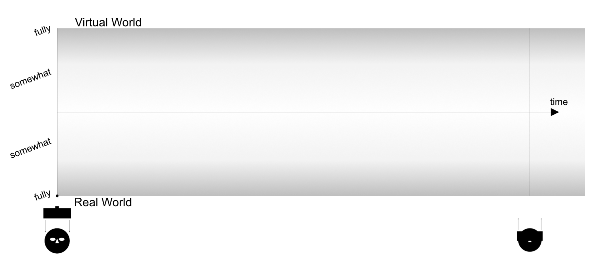

Fig. 2. The basic drawing template (Download: 3) can be used to collect user reports of temporal variations in a HMD based presence experience. A definition of the term presence, in our case the IPQ [28], has to be provided to the participant.

3.3 Application of the Drawing Method

We imagine three ways of using the drawing method in the following. Due to the complexity and page limitations, we only explore the first suggestion. Hence, the other two examples might be incomplete in the current form and need exploration and further refinement in future work.

3.3.1 First: BIP detection. To improve a VR experience, the method enables a quick detection of BIPs in an expert review or user study. We gave the blank drawing template (Figure 2) to the participant immediately after the experience. In the exploration example we present in this paper, the participant was informed that the drawing starts at the position being fully present in the real world at the moment they put on the HMD – $P_{transition}$ –. Based on Schubert's [28] work we asked: *Draw the feeling of spatially being in the virtual world, being involved in the experience and interaction and how plausible the experience was to you.* In cases where the terms were unclear to the participants, the examiner gave an example. We described spatial presence as how much one felt to be surrounded by the virtual objects, or the virtual world in general compared to the real world. Involvement was clarified by explaining that it asks for the attention one devoted to the virtual world, how much one felt involved into the actions and progress of the story in the virtual world. Plausibility, as a substitution for the original *realness* in the IPQ, asked if whether the experience made sense in general or whether things have happened that are not logical. An example was a ball that falls upwards, which can make sense in the virtual situation, but if it did not, then it was not plausible. After drawing, we recommend to discuss the drawing with the participant to gain further insights and prepare for the labeling process. The aim is to assign the same labels between the participants to corresponding BIPs. For long or very complex experiences where the user's actions are unpredictable, we recommend collecting additional data, such as a recording of the VR experience or surroundings, to retrieve the moments when a BIP arose.

3.3.2 Second: Detailed Analysis. A *detailed analysis* aims to better understand the temporal course of the presence experience of a particular event. To achieve this, the supervising person may give additional information about a particular phase of interest during the experience to the participant. For instance, the participant may get introduced to the concepts of mental and physical exit from Knibbe [14] (Figure 3). Then the participant would draw the line. In discussion with the supervising person, details of the drawing line are elaborated, and comments about certain key events, e.g., what happened in VR when the phase began, are made by the participant. With the comments for the events, the supervising person could, for example, review the recorded experience and get a better impression and even measures of the temporal course.

3.3.3 Third: Comparison. A *comparison* makes it possible to contrast the temporal course of different variants of the same experience in a drawing. In particular, it addresses the detection of BIPs which affect only a short part of an experience and, hence, are not reflected in an overall presence value. An example would be the design of a smoother transition into the VR. To compare these phases, the participants will use the same drawing template several times. After experiencing the first condition they are asked to draw their line according to our *BIP detection* method. After experiencing the second condition, they will be given the same template again and add a line for the second experience. Discussing the temporal progress in reference to the first experience helps to identify possible differences. Previous work argues not to conduct within subject design when doing research on presence [13]. Chung recommends to be aware of possible effects and counterbalance the order of conditions [4] .

The *detailed analysis* and the *comparison* can be combined to improve the outcome of the drawings.

3.4 The Analysis of Drawings

In order to derive data from the drawings that enable numerical analysis, descriptive points in the drawing must be identified. First, the examiner marks descriptive points P_i (Figure 3, black points). A definition of the descriptive points is given in the following chapter (Section 4). Then all points are measured and the values

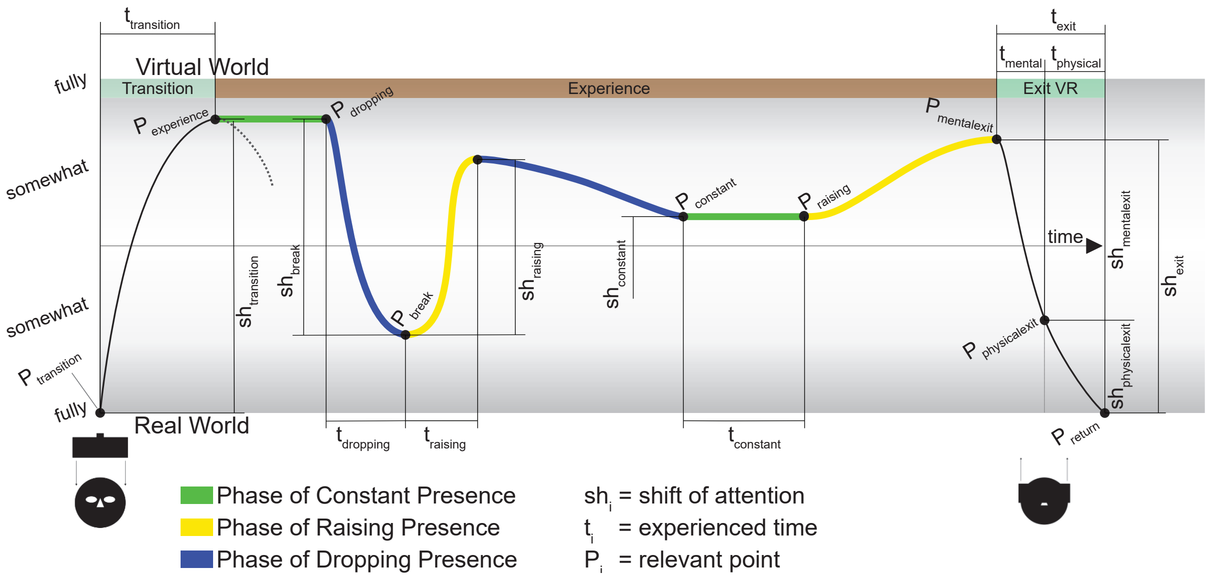

Fig. 3. The drawing shows the phases experienced during a VR experience. For reasons of readability, the description shows only one example for each point and parameter. Capital letters indicate Points that define the course of the experience. Lower case letters encode the resulting parameters.

are stored in a database. The position along the time axis starts with 0mm at the line indicating putting on the HMD. The presence state is measured from the middle line, in our case 40mm upwards and -40mm downwards. To balance templates of different sizes, we normalize the values collected in the previous step, by dividing the measured position for time by the distance between the lines indicating *taking on the HMD* and *taking of the HMD*, in our case 200mm. The presence measure is divided by the distance between the middle line and the maximum positive distance, in our case 40mm. From these values, the parameters t_i and sh_i (Figure 3) are calculated by subtracting the presence values of the according P_{break} from $P_{dropping}$, which leads to a negative value.

In order to label the BIPs, the examiner should use annotations and, if existing, the event ticks. The combination of both gives the label. For instance, the exemplary label *NoScreenBallThrow*, is a screen black out when the participant threw a ball in the VR experience. Without event ticks, examiners need to use the comments and notes they made as recommended above.

For statistical analysis, the data should be treated as being ordinal although the taken measures are in millimeters. Similar to questionnaires using continuous scales, the presence value will not be equally distributed. For instance, a change from 0.1 to 0.5 might express less impact on the users experience then the change from 0.5 to 0.6. The application example in this paper shades some light on these effects.

4 THE DESCRIPTIVE MODEL

Theoretical and practical findings influence our definition of the descriptive model for temporal variations in a presence experience. From literature we derive the **three main phases** of (1) transitioning into VR [6, 24, 31, 33–35], the (2) experience itself [6, 30, 31] and (3) exiting VR [14] (Figure 2). In the following, we define the phases, their start and end points P_i and important stages in a VR experience. In the following, we will call an attention shift towards the real world a dropping presence and attention shift towards the virtual world raising presence, as it is the common language in the community.

4.1 Phases of a VR Experience

Transition into VR – The transition into the virtual world can be divided into physical and virtual [12, 33]. During the physical transition, the users habit themselves to the hardware by putting on the HMD [12, 18, 33]. The virtual transition follows the habituation, as soon as the user can see a picture of the VR [12, 33]. The transition is not accompanied by a distinct perception of the user [30], hence is is difficult to define a starting point based on a cognitive model, such as by Wirth and colleagues [37]. We argue that an HMD based experience is visually dominated [15] and one of the critical elements in a VR experience is to maintain a stable sensorimotor loop [30], mainly based on the user's visual perception. Habituating with the hardware will influence the transition phase [12, 18, 33], but does not provoke a strong reaction in the user's presence experience. We argue that taking on the HMD and perceiving a picture is the only conscious action that marks a distinct point in time, clearly identifiable for a supervising person and a participant. Since we are looking for a reliable BIP detection method, this moment has proven to be the least misleading for the participants and is therefore used in our template. If the transition to VR is of interest, the template we propose must be extended to the left. Point $P_{transition}$ marks the start of transitioning. The end of a transitional phase is defined by any point P_i, indicating a change in the course of presence (Figure 3, green).

The VR Experience - The experience phase is the summary of all stages of constant presence and temporal variations while acting in VR. The first P_i defines the start of the VR experience, $P_{experience}$, after the transition phase. The end of the VR experience is the beginning of the last phase of dropping presence, indicated by the point $P_{mentalexit}$. We could not define a presence parameter for the experience phase that would resemble an overall presence rating. The experienced time is described by $t_{experience}$.

Exiting VR – The transition out of VR is the final drop in presence [14]. A VR experience ends after the VR exit phase. We assume that the *VR exit* phase follows a similar process as the transition phase but in reverse order, as shown by Liebold [17]. Liebold [17] and Slater [31] describe the transition out of VR as a conscious act in which the users suddenly realize that they have focused on the VR. Knibbe [14] collected user reports about the experience of exiting VR and found a mental transition – getting aware that the VR experience ends – followed by a physical transition – taking off the HMD. But they also found reports that did not distinguish between these two parts. The moment at which the user removes the HMD and can visually perceive the real world is chosen as an anchor point for the model. A *VR exit* phase starts at the beginning of the last *phase of dropping presence* which is called $P_{mentalexit}$. The end of the *mentalexit* is defined by the end of the user's visual immersion. The start of the $P_{physicalexit}$ is always situated on the crossing of the drawn line and the auxiliary line. At $P_{physicalexit}$ the user might not be fully present in the real world yet. By this, we take into account reports of possible aftereffects like the need for reorientation [14]. The *VR exit* phase ends at the point P_{return}, describing the moment when participants report being fully present in the real world, again.

4.2 Stages in the Progress of a VR Experience

In the following, we provide our definition for important stages in a VR experience.

Constant Stage – Constant stages of presence do not change in the reported presence level and therefore have a slope close to zero. The starting point of a constant presence phase is the first on the curve that does not show a change in presence towards the one before it. The end of a phase of constant presence is the last point on the curve before the slope changes from 0.

Raising Stage – The start of a phase of raising presence is the first point on the curve that does not show a slope. The end of a raising stage can be any P_i.

Dropping Stage – The start of a stage of dropping presence is the first point on the curve that shows a negative slope. The end of a dropping stage can be any P_i.

4.3 Parameters

Parameters are relative measures between two relevant points P_i (Figure 3, lower case letters). The proposed concept includes known parameters, reveals new parameters and offers the opportunity to derive measures of a presence experience, e.g., ratios.

People need different amounts of time to feel present in VR, which is represented by $t_{transition}$. Breaks in Presence (BIPs) [31] are represented by the Point P_{break} and the parameter sh_{break}. Chung [3, 4] defined a parameter similar to sh_{break} as the intensity of an attention shift. The parameter $t_{dropping}$ helps to differentiate between sudden BIPs and longer lasting effects, that let the user's attention drift slowly back to the real world [6, 17]. Further, the relation between sh_{break} and a following $t_{raising}$ gives insights into the intensity of the break as recovery time was shown to be slightly correlated with the intensity of the break [4]. We expect, according to Slater [31], that $t_{dropping}$ in general will be shorter then $t_{raising}$. $t_{transition}$ is expected to be longer than t_{exit}. Exiting VR is defined by three temporal parameters, total time t_{exit}, mental transition t_{mental} and $t_{physical}$ (see Section 4.1).

5 EXPLORATION OF THE METHOD

In the previous sections we introduced a method to collect and store user reports about temporal fluctuations during a VR experience. Although we argue that our method can be used in different ways (Section 3.3), in this work, we focus on exploring it as a BIP detection method. In addition, this chapter serves as an application example for our method and its possibilities of storing and evaluating data. The goal of the exploration is to gain insights into the reliability of the drawing method and the quality of detecting and evaluating BIPs.

5.1 Measures

To gather insights in the reliability of drawings, we argue that the drawings need to follow a certain progress. The drawings will have strong interpersonal variations and will be dependent on the experience implemented in our study. To make the results more general, we will look for the occurrence of key characteristics, such as transition into VR and BIPs, as defined in Section 4. The characteristics we expect a user to draw are:

a: Line drawing starts at the point $P_{transition}$.
b: Point P_{return} exists.
c: A Break in presence is a dropping line (sh_{break}).
d: The experience phase covers most of the time axis.
e: An attention shift towards the virtual world takes a longer time span than towards the real world (e.g., [4, 6]).

If the users draw these characteristics, our method is reliable as it enables them to represent their experience according to the cognitive model we expect from the related work. To gain insights on the possibilities of the method as a BIP detector we will count the detected breaks. Furthermore, we analyze possible order effects, e.g., how an early occurrence of a BIP affects the probability for it to get reported. Finally, we gain insights in the possibility to reflect the strength of a BIP. Additionally, we collect qualitative feedback to gain better insights into the understanding and usage of the drawing method by the participants.

5.2 Task

We designed a linear experience consisting of six subtasks to be solved by the user that did not include any BIPs. To achieve that, we went through three iterative design steps with a total of 18 participants including four experts. In the process we removed all unintended BIPs. We used the virtual experience test [2] and the igroup presence questionnaire (IPQ) [28] during the design process. We conducted semi-structured interviews asking the participants for the positive and negative aspects of the experience, reminders of the real world and flaws in the design of the virtual world. In a final evaluation with eight novice participants the experience was rated to have a high presence on the IPQ ($G1 : 4.6$; $SP : 5.1$; $INV : 4.8$; $REAL : 3.1$) and no events influencing the experience were reported. The dimensions of the virtual room in the experience were $3.5x2.5meters$ and fitted into the lab space with $4.0x3.0meters$ (Figure 4).

During the experience, the participants fulfilled six tasks. The first was a paper-tossing task, in which they threw paper balls in a garbage bin. The second task was to find two pictures of cats in the room and hang it on the corresponding spot on the wall. In the third task participants had to find and tidy up pens. In the fourth task the goal was to find books and put them in the right order into a bookshelf to complete the word *reality*. In the fifth task they had to solve a riddle which solution was to use a switch on the wall. The sixth task was also a riddle which solution was to turn on the music and make it play Beethoven.

5.3 Stimuli – Breaks in Presence

As stimuli we used BIPs and varied their intensity, which is a combination of violations of the user's cognitive model about the VR and the strength of the distracting stimulus [17]. The duration was chosen to be between 2 and 4 seconds, according to related work [3, 6]. The breaks targeted spatial presence (SP), realness (R), and involvement (I) as defined by Schubert [28] and were designed to meet ecological validity. The five BIPs used are ordered from the strongest to the weakest in the following:

(1) BIP Cable Malfunction (SP,R,I): A four second long *blue screen* error is shown in the HMD and the examiner asks the participants verbally not to move, followed by touching and moving the cable attached to the HMD to "solve" the issue.
(2) BIP Whitescreen (SP, R, I): A four-second long white screen with a beeping audio track, similar to Garau [6].
(3) BIP Teleport (SP, R): The users are teleported for four seconds to an empty virtual room, having the same dimensions as the main room, but showing pictures of the physical room surrounding them on the wall.
(4) BIP Failed Interaction (R, I): Users had to hang two framed pictures on the wall by holding them to a mark and releasing the grab handle. The second picture was made not to stick before the seventh attempt.
(5) BIP Unrelated Controller Vibrations (R): The right controller vibrated for four seconds when it was not used for any purpose.

The breaks were triggered between 5 and 10 seconds after a task was solved and the appearance of the BIPs was counterbalanced in three groups. The concept was to have a weak and strong break at the beginning and the end in the sequence of BIPs, in one condition each. Further, the strong breaks were clearly separated once and one time close together. The resulting orders of BIPs are: A (2-4-5-3-1), B (5-4-2-1-3) and C (1-4-3-5-2).

5.4 Apparatus

We used an HTC Vive HMD with according controllers[1] , the TP-Cast wireless system [2] and Bose QC25 noise cancelling Headsets. The virtual experience was run at 90 frames per second using Unity 5.6 on a PC with a NVIDIA GTX 1080 graphics card and an Intel I7 processor. The presented virtual scene was optimized to evoke presence according to the five parameters introduced by Schubert [28]. *Dramatic Involvement* was realized by giving the user the overall task to clean up the office, as the boss will arrive soon. *Quality of Immersion* was addressed by optimizing the visual appearance of the VR scene regarding 3D models, texturing and lighting. In addition, spatialized sound was used with three ambient sound sources and twenty audio sources. The controllers vibrated when an object was touched. To support *Exploration of VE*, all objects, but heavy furniture

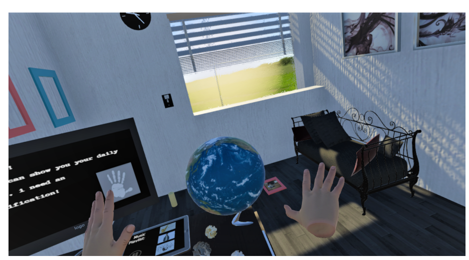

Fig. 4. Start position in the main room.

[1]https://de.wikipedia.org/wiki/HTC_Vive, accessed: October 18th, 2019
[2]https://www.tpcastvr.com, accessed: October 18th, 2019

was movable or reactive to actions (e.g., spinning a globe on the desk, Figure 4) and behaved physically to support *Predictability*. *Interface Awareness* was optimized by enabling basic interactions with just one button all the time – the users had virtual hands that supported a grasping animation –, and a wireless HMD.

5.5 Participants

30 participants (11 females, 19 males, age: 19 to 40) that did not take part in the iterative design process or had knowledge about the method, were invited by mailing lists, paper displays, social media, and personal social network. They had backgrounds from diverse fields. 12 participants had never used an HMD before, the rest used HMDs less than one time a year. Participants were rewarded study credits or 10 euro according to their choice.

5.6 Procedure

Participants were welcomed, and the purpose of the study was explained as being a user experience study about interaction in VR. The whole study took about 60 minutes (15 minutes in VR). The participants were welcomed, informed about anonymous storage and evaluation of tracked movement data, recording of screen captures and sound and were asked to fill out a consent form (10 minutes). The usage of the HMD, noise canceling headset and controllers was explained to and tested by the participants in the standard grey Steam VR environment. After the successful introduction, they removed the HMD again and were led to a starting position marked on the ground. The participants were instructed that the examiner will not help them during the experience. They were informed that the session will end after the last task, which will be indicated by a prompt that shows up. No information about possible BIPs was given. They started the experience independently by taking on the HMD. They started in a transitional room that had the same size and layout as the final room, but only a few objects were presented on a desk in front of the user. The objects were used to again train the interaction in VR. The participants were guided by instructions, which were presented textually on a monitor on the desk, during the whole experience. After the tutorial, users were teleported into the main room by pressing a virtual button themselves.

After the experience, the drawing task according to the procedure described in Section 3 was conducted. The participants conducted the drawing without external help (2-5 minutes). When finished, the examiner conducted an interview about the drawings to record details and to assign the appropriate drawn BIPs to the actual ones (10 minutes, 1-3 minutes were used to label the BIPs). Finally, the participants were asked to fill out the IPQ [28], demographic questionnaire, followed by a debriefing (15 minutes).

5.7 Results

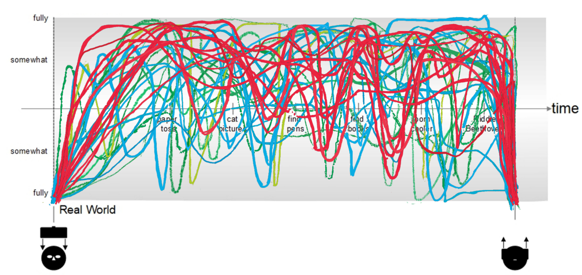

Fig. 5. All drawings collected in the study. It illustrates the need for our descriptive model based on numerical values that enable an evaluation independent of the drawings. Colours represent the counterbalanced groups ($green = A, red = B, blue = C$). Annotations removed for better readability.

To showcase our method as a BIP detector, we present the processing of the collected drawings step-by-step. By doing so, we want to communicate the possibilities of using the data. We encourage future users of the method to suggest alternative ways on using and analyzing the data.

Figure 5 shows all collected drawings. Since all the counterbalanced groups, indicated by colour, are shown in the figure, it represents the possible result of a real application where the occurrence of BIPs is unpredictable. It is also an example for the diversity in user drawings. In this form it is very difficult or even impossible to work with the data. To obtain data that can be evaluated, we use our proposed method and the descriptive numerical model to analyze it. Figure 6 shows exemplarily the result of identifying descriptive points in the drawings. In particular, in Figure 6 the points indicating the end of transitioning into and the begin of exiting VR are shown. The values of these points are stored in the database together with the annotations describing them. One point P_{return} could not be described by our model. It did not show the end of the line in a position that is defined as being fully present in the real world and therefore was excluded. For the transition into VR, we found a Gaussian distribution for the points $P_{experience}$. The points $P_{mentalexit}$ do not show a distinct distribution, but the points P_{return} are all at the same position.

Next, the parameters based on the stored data are calculated. As we focus on BIPs, we will use the parameter sh_{break} as an example. Table 1 presents the outcome of a statistical evaluation of the drawings based on the database for the single counterbalanced groups. It shows the evaluation of points (P_{break}), derived parameters (sh_{break}) and their detection rate according to the order of appearance during the experience. Since now we only have numerical data, the next step is to derive a visualization of the results. Figure 7 is en example for such a visualization using our drawing template. It shows the aggregated data of all groups for the introduced BIPs as Box-Plots.

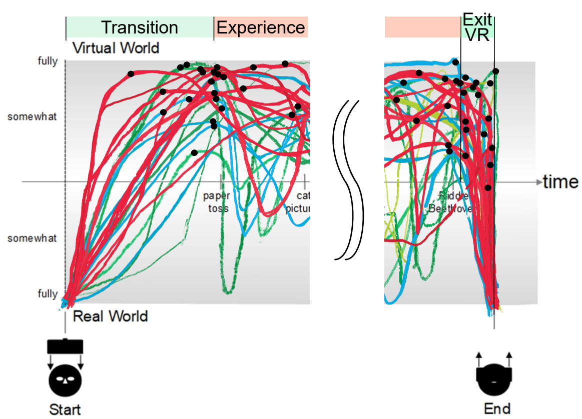

Fig. 6. In line with related work, transition into VR is reported to be longer than the transition out of VR. Black dots show the transition points $P_{experience}$ (left) and $P_{mentalexit}$ (right). Counterbalanced groups: *green* = *A*, *red* = *B*, *blue* = *C*.

5.7.1 The Transitional Phases. Figure 6 shows the transition and exit phases, without the experience in between. Six participants experienced a disturbance during transition, not caused by the BIPs we introduced intentionally. We excluded these drawings in Figure 6 (left). The unintended disturbances are caused by the HMD not fitting and readjustments by the participants. One very shallow transition drawing was caused by boredom during the instructional part, leading to a low feeling of involvement. The last one did not give a justification. Exiting VR was influenced for one participant (group C) as the BIP showing a white screen happened as last BIP, before the end of the experience. It was removed in Figure 6 (right picture) as no exit phase was drawn according to our definition. The excluded cases are still visible in the overview of drawings (Figure 3).

Further, Figure 6 is an example for the application of the analysis (Figure 1). Because we are able to store the data in a database, we can calculate parameters, for instance on the experience of time. In average participants used 21% (SD = 10%) of the timeline to report the transition into VR ($t_{transition}$). Further, they used 8% (SD = 5%) to express the transition out of VR (t_{exit}). As a result the *experience* phase took 71% of the available space. $P_{transition}$ was always situated at the defined position, as it was an instruction. Participants finished their drawings in average at P_{return} 99.6% (SD = 1%) of the timeline, and at −93% (SD = 3%) of the negative presence axis.

In no cases, participants reported separating their exit experience in a *mental* and *physical* exit. 25 of the drawings show a drop which accelerates continuously over time and shows a uniformly accelerated slope of the curve between the points $P_{mentalexit}$ and P_{return}. Five participants drew their experience as a sudden breakpoint at which $P_{mentalexit}$ is at the same value on the timeline as P_{return}.

5.7.2 Breaks in Presence. Detection Frequency – The detection frequency is the percentage of participants reporting a BIP. We report the detection frequencies in Table 1. Strong BIPs – *Cable Malfunction*, *White Screen* –

Table 1. Detection rate and reported intensity of the BIP for the groups. The strongest – Cable Malfunction – is in the top row, the weakest – controller vibration – in the bottom row. POS describes the appearance in the counterbalanced groups, whereas 1 is the first place and 5 is the last place. The values sh_{break}, P_{break} (Figure 3) report the averages.

	Randomization											
	A				B				C			
Breaks in Presence	POS	Detect. (%)	sh_{break}	P_{break}	POS	Detect. (%)	sh_{break}	P_{Break}	POS	Detect. (%)	sh_{break}	P_{Break}
Cable Malfunction	5	100	-0.33	-0.72	4	90	-0.28	-0.33	1	70	-0.28	-0.34
White Screen	1	60	-0.38	-0.4	3	70	-0.25	-0.03	5	70	-0.28	-0.29
Teleport	4	30	-0.45	-0.15	5	50	-0.18	0.08	3	40	-0.3	-0.11
Failed Interaction	2	70	-0.2	-0.23	2	90	-0.15	0.23	2	50	-0.34	-0.2
Vibration	3	20	-0.1	0.19	1	0	-	-	4	20	-0.22	0.09

are more likely to be reported than weak BIPs, no matter if they appear early or late during the experience. *BIP Failed Interaction* stands out from this with a high detection rate in two of the counterbalanced groups. Although, we defined it as a weak BIP and the disturbance resulting from it was perceived by the participants as such during the experiment (Figure 7), the detection rate was similar to the stronger BIPs. On average across all groups, a single participant was able to identify 2.8 BIPs (SD = 0.9).

Order effects – The order might show an effect in our study, but the results are not clear as the number of detected BIPs is very low in some cases, which prevents further statistics. The two strongest breaks are 10% and 40% more likely to be reported when in the last position of the BIPs compared to being the first. The weaker breaks do not clearly show an effect. The weakest break was not reported, when appearing in the first position – Position 1, group B. However, the *BIP Failed Interaction*, which was always at the same position during the experience, shows a variation in the detection rate between 50% to 90%.

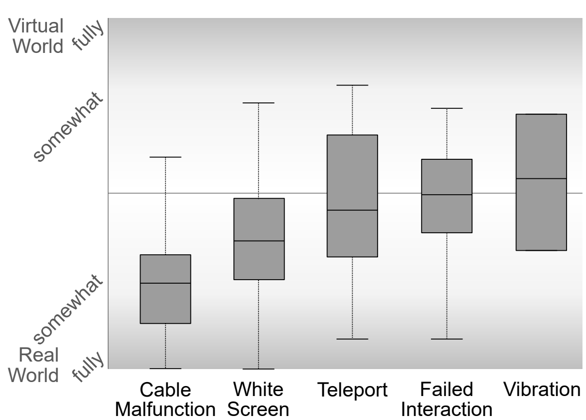

Fig. 7. Average of P_{break}s for the induced BIPs across all groups. The impact of BIPs is reported according to our prediction, indicating the ability to reflect the strength of BIPs.

Intensity – The intensity for P_{break} in the single groups can be found in Table 1. We do not use it for further evaluation of trends, as the number of participants within the single balanced groups showed to be to small. A trend in the order of the BIPs' strength is visible, when averaging over all participants (Figure 7). It fits the prediction on strength we derived from related work in Section 5.3. The weaker breaks, e.g., the unrelated controller vibration, are drawn with a tendency of their median towards the *virtual world*. In particular, the strongest of the cable malfunction is reported to be closer to the real world. We did not find a tendency for the relative parameter sh_{break}.

Time Parameters – We emphasize that the temporal arrangement for the detection of BIPs in our study is not of particular interest. However, it gives an estimate on how precisely users are able to report the appearance of a BIP relative to the events in the experience, if the method is used in a real world application. We counted the correct positioning of $P_{dropping}$ or the following P_i. BIPs were counted as positioned correctly if one of the points was situated within 5mm (2.5%) before and 25mm (12.5%) after the according task tick. We found that participants had 97 of the 118 detected breaks positioned correctly. As a result, in 18% of the cases the relation between the event in VR and the drawing would not be clear.

We further compared the time it takes that a BIP comes into full effect, $t_{dropping}$, to the time participants need to recover from it, $t_{raising}$. We calculated the average of each and then divided the result of $t_{dropping}$ by $t_{raising}$. The resulting ratio is 0.8, which means that a dropping phase is drawn 20% shorter than a raising phase.

5.8 Other Anecdotal Findings

Stimulated by drawing the line, many users started to reflect on their experience. In the course of the drawing, they went back to earlier parts and adjusted them. By adjusting they expressed, for example, a stronger influence of a preceding BIP, the sequence of BIPs or the relative position to the event ticks.

5.9 Limitations

The report of a longer transition time into VR then to exit VR, is in line with the related work. However, there might be a bias introduced by the drawing template. In the template the ticks for the tasks – e.g., putting a picture on the wall – are distributed along the time axis relative to their appearance in the experience. The first task in the experience – paper toss – was much later in time, than it took between the last task – Riddle Beethoven – and the end of the experience. Therefore, there was much more space to draw the transition into VR than the exit of VR, which might lead to the current result (Figure 6). We do not cover drawings that show a step raising at the beginning and then change to a more shallow curve when transitioning into VR. We decided to stick to our decision and used the first point of constant presence or dropping presence for the parameters $t_{transition}$ and $sh_{transition}$. However, people might express that they felt present in the virtual world very fast, but became more and more confident over time, as known from other studies [19]. The experience in this study took about 15 minutes. A longer experience might amplify possible inaccurate recall of the experience. An analysis of longer experiences should include more than 10 participants.

5.10 Remark: Faster Analysis Process

One of the main goals for the method was to provide a tool for everyday practice. However, manual transfer of the data points into a standardized data format took about 5 minutes per drawing in our study. Therefore, we provide an Android application on Git that enables the user to draw on a tablet[3], either using a finger or a pen. The sketch is analyzed immediately, for the descriptive points, main phases and parameters according to our descriptive model. The result of this analysis is provided to the users after they have finished the drawing, providing the opportunity to leave comments to each point and phase. The raw data, the processed data and the comments are stored in a comma separated value file (.csv) for further analysis. Hence, our method puts minimal time effort on the participants and the examining person.

6 DISCUSSION

First, we will reflect on the results of our application example, followed by a general discussion that puts our method in relation to existing approaches.

6.1 Discussion of the Application Example

Users are enabled to reflect the course of a presence experience reliably – All collected drawings fulfill our five prerequisites (Section 5.1) on reliable reports of a presence experience. Hence, our method enables participants to

[3]https://github.com/petarzoric/VRGraphAnalyzer.git

report according to the cognitive model we derived from related work (Section 4). We have, therefore, achieved our goal to develop Garau's [6] idea into a functioning method.

In particular, we could show that all participants started at $P_{transition}$ (prerequisite a) and end with P_{return} (prerequisite b). Only one case did not fit our model, as raising presence was reported again after the last dropping phase without a particular reason (Figure 3, counterbalance Group C). All participants understood a BIP as an attention shift towards the real world and drew it as a dropping line (prerequisite c). Most of the template (71%) was used to express the temporal variation during the experience phase (prerequisite d). Participants reported an attention shift towards the real world to happen faster than towards the virtual world (prerequisite e), which is in line with Chung's findings [4]. In particular, $t_{transition}$ was 38% longer than t_{exit} and to recover from a break ($t_{raising}$) took in average 20% longer than the break ($t_{dropping}$).

We want to stress at this point that the participants had minor experience in HMD usage. The drawings were not moderated by the examiner and no information about expectations on the drawings or the BIPs in the study were given. Hence, the result confirms that the combination of instructions and the template we provide offers reliable results independent from the examining person. However, it is a novel method, hence, we recommend to have an experienced person present during conduction to discuss the drawings.

Drawing Method is a tool to detect BIPs – The detection rate of nearly 100% in each counterbalanced group confirms our second goal for the application example of showcasing the ability to detect BIPs of different type [28] and strength [17]. Only the weakest BIP was missed, when it occurred at the beginning of the experience (Table 1). Similar to other post-experience methods, the drawing method suffers from inaccurate recall by the participants [6]. In average, a participant reported 55% of the BIPs. Besides the outlier, at least 20% of the participants could identify the weakest BIP and 60% to 100% reported the strongest. Stronger BIPs might have the tendency to be reported more often. Further, we found BIPs that we did not find using established methods during the design phase. In particular, in the transition phase, 7 participants reported a disturbance with effects on the further course of the experience. This disruption even led to the first BIP not being reported.

Each counterbalance group included 10 participants, which discovered all BIPs in a 15 minutes long VR experience. Fewer participants lead to the risk of not finding weak disturbances appearing early in an experience. Longer experiences might demand more participants to securely detect all BIPs. Although, we could detect BIPs of a variety of types and strength, misinterpretations of our definition of the term presence, given to the participants, might be possible. In summary, our method is able to detect a wide range of BIPs. Hence, it can be used in an iterative design processes to improve the quality of VR experiences, e.g., by expert reviews based on our method within a company. The drawings, detected BIPs and evaluation results can be used to document the work's progress and communicate it to others.

Drawings indicate the intensity of a BIP – The expected strength of the BIPs gets visible when the drawings of all participants are averaged (Figure 7). In contrast, the single counterbalanced groups do not show a clear trend for the intensity of a BIP (Table 1). The points P_{breaks} represent BIPs the best, which is in line with Chung's suggestion [4]. We could not find a clear trend for the parameter sh_{break}, similar to Chung, as there were interactions with preceding or following events. Besides the low number of participants or ordering effects, future work should look into the definition of outliers that we did not remove in our study.

From this, we conclude that our drawing template has potential to evaluate the intensity of a BIP. However, the sample size needs to be larger than for the detection of BIPs. Therefore, we argue to use BIP intensity evaluation in the current state as a warning system. For example, in a public deployment users could be asked to draw their experiences on a tablet computer next to the deployment, as a quality management system. The application could analyse reported BIPs and instantly inform a responsible person.

Drawing the temporal progress has potential – People used the drawings to differentiate between slow and fast attention shifts during the presence experience. Currently we do not know how effects caused by the post-experience design of our method will affect reports, e.g., the compression of the temporal course. An example is the time a BIP takes to develop to its full extent, which might be drawn using a smaller amount of the timescale if many BIPs are reported. In addition, reports may vary based on the individual perception of each participant and their personal preference for the use of the template. Therefore, we argue not to overemphasize the results of individual drawings. Instead, analogous to theoretical considerations such as by Garau [6], we argue to compare variants of experiences against each other in one drawing. Additionally, we recommend to counterbalance when testing design variants.

In contrast to research on exiting VR [14], our method did not reveal a two-stage structure of the exit experience. Our template was thought to support the report of the cognitive and physical exit from VR, by providing a line that indicates the separation. We assume that the more detailed reports about two stages in Knibbe's work were caused by emphasizing the exit phase in the discussion with the participants [14]. In our study, 83% of the participants drew the exit as a continuous curve that might include the aforementioned separation (Figure 6). However, the continuous curves might be a result of a more natural or comfortable drawing behavior and are not related to the experienced changes. In general, as our study was not designed to evaluate the transitional phases, certain parameters might need clarification in future studies. For example, the parameter $t_{transition}$ explains the shift of attention to the first point of maximum presence. However, crossing the middle line separating real world from the VR in our template, could already represent a sufficient shift in attention towards the virtual world, e.g., to ensure behavioral realism.

In summary, our method has the potential to provide additional information about the subjectively experienced temporal course of presence. We have indications that users express differences in the temporal progress between experiences, which is a qualitative rating. Whether quantitative measures, like the slope of a line-graph, correlate with the users' experience or not, cannot be derived from our study. An exemplary challenge is that people need time to transition and show behavioral realism in VR. Currently studies either start with the task from the beginning or use an undefined time to let participant transition. Hence, the question is when the task that is central to a study should ideally start in order not to interfere with the transition phase. The drawing method helps to define the moment more precisely, which prevents noise in the study results and safes time for a participant.

6.2 General Discussion on the Drawing Method

DrawingPresence has the ease of use for everyday practice – Our method requires the same effort and time for participants and examiners to conduct as questionnaires and interviews, since both are pen and paper based. Further, as our method helps participants to focus on the phenomenon of presence, they can communicate more in less time than they would in interviews. Similar to other self-reporting (e.g., [4, 11, 17]) and physiological measures (e.g., [17]), our method demands extra time to label the reported BIPs. Labeling also requires experience with the method to ensure are correct assignment and meaningful names for analysis. Still the process will be faster than analysing an interview that needs to be transcribed and coded first. Further, our method adds security to the labeling process, as it can be done in conjunction with the participant straight after drawing. The descriptive points are transcribed into a database which takes about 20 to 30 seconds to define a point and capture its coordinates and annotations. Depending on the number of reported breaks it is slightly faster to transcribe Likert-scale questionnaires, as only individual numbers have to be transcribed. However, it is difficult to compare the results, as questionnaires do not contain information about the temporal course. As a remark, transcription of drawings is automated and instantly if our provided Android App[3] is used, which makes it as fast to analyze as a questionnaire. Further, our method outperforms other self-reporting methods in terms of ease of use. For example, Chung uses a post-experiment method that requires all participants to review their entire experience [4]. Ijsselstein states that a detailed analysis of the self reported data collected over the whole experience is too much effort [10]. In summary, our method places little additional demands on the participant and examiner compared to established methods and therefore can be an alternative.

Drawing presence is a user experience method to detect BIPs – Our method is the first user experience method that provides a process to detect BIPs in everyday practice. Since less than 10 participants are required, it corresponds to other UX methods in terms of the number of participants needed [22]. Compared to the established interviews, we found that drawing introduces additional security in finding BIPs. In particular, in line with Garau's [6] findings, we found that drawing the line supports users' reflection on the experience. Additionally, the drawings reveal gaps or illogical descriptions in the users' reports. In the discussion between an examiner and a participant it serves as a common frame of reference. Interviews do not offer such guidance. Hence, interviews bear the risk of missing the report of BIPs, as participants often digress in their reports. As a result, our method outperforms established methods such as the virtual experience test [2], semi-structured interviews and the IPQuestionnaire [28], which is similar to reports about shortcomings of these methods in previous work [19]. In particular, our method simplified the report of complex interactions between BIPs and revealed very early appearing BIPs .

The method comes with additional benefits – Our model of the temporal progress in a presence experience can be used in other methods, e.g., physiological measures, to store meaningful data in a compact way. For instance Chung missed in his work the analysis of the value P_{Break}, as his conceptualization did not include it [4]. Further, we found good indications that we can estimate the impact of a single BIP on the user. The estimate can be used to prioritise development goals based on the impact rating, which supports traditional expert ratings. If there is the need for a quantification of an overall presence value, our method can be combined with Slater's presence counter [31] to get a rough estimate.

6.3 Application Examples

To better understand the new possibilities our method offers in practice, we present application examples.

Designing BIPs – The usage of HMD based VR systems in everyday contexts bears novel challenges like an increasing number of disruptions by surrounding people in the real world. An exemplary disrupting moment is the need of a real world user to talk to the HMD user [7, 20]. With the ongoing research on interruptions, more detailed design solutions have to be expressed and explored. In the particular example the temporal progress of the disruption could either be smooth and slow or hard and sudden [14]. Our method can support development teams in achieving these design goals.

Designing Transitions – Knibbe relied on interviews [14] to better understand the transitioning out of VR and the difference between design alternatives. The differences between the concepts could be explored in more depth by letting participants draw several lines of each condition in the same template (Section 3.3).

Combining with physiological measures – Physiological measures, e.g., as used by Liebold [17], do not accurately detect BIPs, and include false positives. Our method could have helped to better interpret the measured data. The drawings summarize the BIPs experienced by the user and thus implicitly exclude the false positives. The moment of occurrence can be depicted relatively in the drawings and precisely assigned by means of the corresponding comments. With the help of this information it becomes easier to find actual disturbances in the measured data, in particular if the experience was very long. Further, a strong reaction in physiological data does not necessarily mean a BIP is experienced as strong by the user. A comparison of the physiological data and the drawn intensity of a BIP might generate additional insights supporting the interpretation of physiological data.

Lightweight feedback tool – There is a current trend to deploy HMDs in the public, e.g., for marketing purposes. However, due to its high psychological impact on the user, a faulty system can cause a bad reputation of the provider and might even harm the user. Hence, we suggest to put a tablet computer next to a deployment to collect feedback in form of drawings. By this we provide a lightweight feedback tool that can be run without supervision and may lead to more answers, as this type of survey is new and interesting for the user. Even more, it might be possible to analyse the data in real time and inform a technician if unexpected BIPs show up.

Detecting relevant details – Mai and colleagues researched effects from HMD usage in public [19]. They compared the three conditions of having a VR experience in a separate room, a physical separation by a barrier tape in public and being in public without any separation to others. They could not find a difference in the measures they applied. These were either objective measures, e.g., movement speed, and subjective, e.g, questionnaires on emotions and presence. Conducted interviews were accompanied by an early version of our drawing method. When discussing their drawings, participants articulated that they struggled in transitioning to VR as they perceived several BIPs caused by actions in the real world. Further, they reported that over time, the highly immersive experience enabled them to fully focus on the VR and they forgot about the surrounding. These findings were missing in the usual measured data. Hence, in line with our experiment, without the use of the drawing method, BIPs negatively affecting the experience in the beginning would not have been found.

7 FUTURE WORK

Our work motivates research on the presented method and the general temporal course of feeling present in VR.

The main question for future use as a BIP detection method is how many BIPs can be reliably detected. The focus should be on issues arising from the post-hoc design of the method. An important question is how long a VR experience can be for the method to remain effective. The challenge is that in longer experiences the interplay of length, type and strength of BIPs can have stronger impact on the reporting rate. In particular, we assume that a BIP of the same type and strength will be reported less in longer experiences. However, if the same type of BIP is stronger, we can only speculate whether the reporting rate raises. Additionally, when strong BIPs are removed we cannot estimate the effect on the report of weaker BIPs. The effect could be that weaker BIPs become more prominent for the participants and therefore have a higher reporting rate. Finally, in our study people were not asked to look out for BIPs. If the method is used in a design process, it is possible to inform participants in advance to search for BIPs. Hence, it must be better understood what effect it has on the detection rate and reported strength of BIPs when there is a defined order to seek BIPs.

To support recall of BIPs our template includes ticks indicating the subtasks in the experience. While our scenario followed a linear story and the BIPs appeared close to the tasks ticks, other experiences might follow a different structure. We cannot estimate the impact on the recall rate of BIPs when leaving out ticks, but without the ticks it gets difficult to find the moment in an experience when the BIP appeared. Today, we rely on the examiner to provide enough expertise during the labeling of BIPs and making additional comments to correlate the drawings with the actual experience. If event ticks cannot be prepared in advance, screenshots or short animations of the experience could be taken by the examiner or automatically presented in the android application to support the participants. We do not know about their ability to support recall when created in a non-linear story line in which the same event tick might occur several times and time intervals between ticks might vary.

There are open challenges regarding the interaction design to be optimized for everyday practice. One key challenge is the labelling by the participant, as it was for other methods before (e.g., [3, 4, 11]). We argue to work on a framework that clusters BIPs based on their type or by the user cognitive model.

Further, at present our method is not optimised for barrier-free usage, e.g., by people with visual impairments or other constraints. With the current form of our method there is a risk that these people can not make use of it to give feedback. Hence, we would like to encourage researchers to adapt the method to the specific needs in their field and make it available to others. Finally, we need additional insights in the application of our model, e.g., the application examples we provide (Section 6.3).

Research applying our method could be used in the future to better understand the design of experiences, allowing the HMD user to quickly recover from a BIP. Future work should explore the report of BIPs with more variations of type and strength. A prominent and timely question would be to understand the impact of a by-standing person addressing an HMD user in different ways.

8 CONCLUSION

More than 90% of studies using HMDs apply questionnaires as a main measure to assess the feeling of being present in virtual reality. Physiological measures show the potential to assess temporal fluctuations of the presence experience, in particular to detect breaks in presence (BIPs), but are hardly not used. One reason not to use them is the complexity of their application. However, the detection and resulting ability to remove BIPs is crucial to improve the overall presence experience. Post-experience user self-reports are a promising approach, however, there is no method available that fits the needs of everyday practice. Therefore, we propose a novel post-experience method that collects user self-reports on the temporal course of their presence experience by drawing a line. The amplitude of the line indicates the experienced state of being present in VR.

In order to enable the evaluation of the drawings, we propose a descriptive numerical model that represents a VR presence experience from beginning to end. In particular, the model describes based on related work the relevant points and phases of the temporal course in a presence experience. As a result the descriptive model replaces the drawings by compact numerical data for storage, comparison and evaluation purposes.

In an exploratory user study, we demonstrate the feasibility and reliability of the method. In particular, we were able to show that our method is capable of securely detecting a variety of BIPs. The type, strength and timing of the occurrence of BIP affected the reporting rate, but a group of 10 participants could identify them all. In addition, we found that participants' self reflection on their experience was stimulated by the drawing. As a result, they were able to give a very detailed expression, even about interactions of BIPs. Additionally, we have good indications that the method can be used to estimate the strength of a BIP.

The immersive quality of today's HMD based systems is very high. Established questionnaires and interviews may soon no longer be sufficient to identify and understand BIPs covering only a short period of an experience. The focus of this work was on the use of the method as a BIP detector. However, our findings motivate further work on the method to develop it to a tool that is able to assess the strength and temporal course of BIPs. With our method we help to push the boundaries of the current understanding of the users experience in VR. We want to encourage other researcher to build upon our method and apply it to make VR even better.

ACKNOWLEDGMENTS

We would like to thank Petar Zoric for his great efforts in developing the Andriod application for our method.

REFERENCES

[1] Andrea Brogni, Mel Slater, and Anthony Steed. 2003. More breaks less presence. In *Presence 2003: The 6th Annual International Workshop on Presence (Presence '03)*. 1–4.

[2] Dustin Chertoff, Brian Goldiez, and Joseph LaViola. 2010. Virtual Experience Test: A Virtual Environment Evaluation Questionnaire. In *Proceedings of the 2010 IEEE Virtual Reality Conference (VR '10)*. IEEE Computer Society, Washington, DC, USA, 103–110. https://doi.org/10.1109/VR.2010.5444804

[3] Jaeyong Chung and Henry Gardner. 2009. Measuring Temporal Variation in Presence During Game Playing. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and Its Applications in Industry (VRCAI '09)*. ACM, New York, NY, USA, 163–168. https://doi.org/10.1145/1670252.1670287

[4] Jaeyong Chung and Henry Gardner. 2012. Temporal Presence Variation in Immersive Computer Games. *International Journal of Human-Computer Interaction* 28 (2012), 511–529. https://doi.org/10.1080/10447318.2011.627298

[5] James Cummings and Jeremy Bailenson. 2015. How Immersive Is Enough? A Meta–Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology* 19, 2 (2015), 272–309. https://doi.org/10.1080/15213269.2015.1015740

[6] Maia Garau, Doron Friedman, Hila Ritter Widenfeld, Angus Antley, Andrea Brogni, and Mel Slater. 2008. Temporal and Spatial Variations in Presence: Qualitative Analysis of Interviews from an Experiment on Breaks in Presence. *Presence: Teleoperators and Virtual Environments* 17, 3 (2008), 293–309. https://doi.org/10.1162/pres.17.3.293

[7] Ceenu George, Philipp Janssen, David Heuss, and Florian Alt. 2019. Should I Interrupt or Not?: Understanding Interruptions in Head-Mounted Display Settings. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. ACM, New York, NY, USA, 497–510. https://doi.org/10.1145/3322276.3322363

[8] Carrie Heeter. 1992. Being There: The Subjective Experience of Presence. *Presence: Teleoperators and Virtual Environments* 1, 2 (1992), 262–271. https://doi.org/10.1162/pres.1992.1.2.262

[9] Dimitri Hein and Christian Mai. 2018. The Usage of Presence Measurements in Research: A Review. In *Proceedings of the International Society for Presence Research Annual Conference (Presence '18)*. The International Society for Presence Research.

[10] Wijnand Ijsselsteijn and Huib. de Ridder. 1998. Measuring temporal variations in presence. In *Papers from the workshop on presence in shared virtual environments*. Ipswich, UK,.

[11] Wijnand IJsselsteijn, Huib de Ridder, Jonathan Freeman, and Steve Avons. 2000. Presence : concept, determinants and measurement. In *Human Vision and Electronic Imaging V (Proceedings of SPIE - The International Society for Optical Engineering)*, B.E. Rogowitz and T.N. Pappas (Eds.). SPIE, San Jose, USA, 520–529.

[12] Sungchul Jung, Pamela Wisniewski, and Charles Hughes. 2018. In Limbo: The Effect of Gradual Visual Transition Between Real and Virtual on Virtual Body Ownership Illusion and Presence. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR'18)*. Tuebingen/Reutlingen, Germany, 267–272. https://doi.org/10.1109/VR.2018.8447562

[13] Pankaj Khanna, Insu Yu, Jesper Mortensen, and Mel Slater. 2006. Presence in Response to Dynamic Visual Realism: A Preliminary Report of an Experiment Study. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '06)*. ACM, New York, NY, USA, 364–367. https://doi.org/10.1145/1180495.1180569

[14] Jarrod Knibbe, Jonas Schjerlund, Mathias Petraeus, and Kasper Hornbæk. 2018. The Dream is Collapsing: The Experience of Exiting VR. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 483, 13 pages. https://doi.org/10.1145/3173574.3174057

[15] Luv Kohli. 2013. *Redirected Touching*. Ph.D. Dissertation. Chapel Hill, NC, USA. Advisor(s) Brooks, Frederick.

[16] Sari Kujala, Virpi Roto, Kaisa Väänänen-Vainio-Mattila, Evangelos Karapanos, and Arto Sinnelä. 2011. UX Curve: A Method for Evaluating Long-term User Experience. *Interacting with Computers* 23, 5 (Sept. 2011), 473–483. https://doi.org/10.1016/j.intcom.2011.06.005

[17] Benny Liebold, Michael Brill, Daniel Pietschmann, Frank Schwab, and Peter Ohler. 2017. Continuous Measurement of Breaks in Presence: Psychophysiology and Orienting Responses. *Media Psychology* 20, 3 (2017), 477–501. https://doi.org/10.1080/15213269.2016.1206829

[18] Christian Mai and Mohamed Khamis. 2018. Public HMDs: Modeling and Understanding User Behavior around Public Head-Mounted Displays. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays (PerDis '18)*. ACM, New York, NY, USA. https://doi.org/10.1145/3205873.3205879

[19] Christian Mai, Tim Wiltzius, Florian Alt, and Heinrich Hußmann. 2018. Feeling Alone in Public. Investigating the Influence of Spatial Layout on Users' VR Experience. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction (NordiCHI '18)*. ACM, New York, NY, USA. https://doi.org/10.1145/3240167.3240200

[20] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2143–2152. https://doi.org/10.1145/2702123.2702382

[21] Marvin Minsky. 1980. Telepresence. *Omni* 2, 9 (1980), 45–52.

[22] Jakob Nielsen and Thomas Landauer. 1993. A Mathematical Model of the Finding of Usability Problems. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. ACM, New York, NY, USA, 206–213. https://doi.org/10.1145/169059.169166

[23] Niels Nilsson, Rolf Nordahl, and Stefania Serafin. 2016. Immersion Revisited: A Review of Existing Definitions of Immersion and Their Relation to Different Theories of Presence. *Human Technology* 12 (2016), 108–134. https://doi.org/10.17011/ht/urn.201611174652

[24] Sebastian Oberdörfer, Martin Fischbach, and Marc Erich Latoschik. 2018. Effects of VE Transition Techniques on Presence, Illusion of Virtual Body Ownership, Efficiency, and Naturalness. In *Proceedings of the Symposium on Spatial User Interaction, SUI 2018, Berlin, Germany, October 13-14, 2018*. 89–99. https://doi.org/10.1145/3267782.3267787

[25] Christine Rosakranse and Soo Youn Oh. 2014. Measuring Presence: The Use Trends of Five Canonical Presence Questionaires from 1998–2012. In *In Proceedings of the 15th International Workshop on Presence (ISPR'14)*.

[26] Farrukh Sahar, Jari Varsaluoma, and Sari Kujala. 2014. Comparing the Effectiveness of Electronic Diary and UX Curve Methods in Multi-component Product Study. In *Proceedings of the 18th International Academic MindTrek Conference: Media Business, Management, Content & Services (AcademicMindTrek '14)*. ACM, New York, NY, USA, 93–100. https://doi.org/10.1145/2676467.2676500

[27] Thomas Schubert. 2003. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie* 15, 2 (2003), 69–71. https://doi.org/10.1026//1617-6383.15.2.69

[28] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments* 10, 3 (June 2001), 266–281. https://doi.org/10.1162/105474601300343603

[29] Martijn Schuemie, Peter van der Straaten, Merel Krijn, and Charles van der Mast. 2001. Research on Presence in Virtual Reality: A Survey. *CyberPsychology & Behavior* 4, 2 (April 2001), 183–201. https://doi.org/10.1089/109493101300117884

[30] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364, 1535 (2009), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

[31] Mel Slater and Anthony Steed. 2000. A Virtual Presence Counter. *Presence: Teleoperators and Virtual Environments* 9, 5 (Oct. 2000), 413–434. https://doi.org/10.1162/105474600566925

[32] Mel Slater and Sylvia Wilbur. 1997. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 6, 6 (1997), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

[33] Daniel Sproll, Jacob Freiberg, Timofey Grechkin, and Bernhard Riecke. 2013. Poster: Paving the way into virtual reality - A transition in five stages. In *IEEE Symposium on 3D User Interface 2013 - Proceedings (SUI'13)*. 175–176.

[34] Frank Steinicke, Gerd Bruder, Klaus Hinrichs, Anthony Steed, and Alexander Gerlach. 2009. Does a Gradual Transition to the Virtual World increase Presence?. In *2009 IEEE Virtual Reality Conference (VR'09)*. 203–210. https://doi.org/10.1109/VR.2009.4811024

[35] Dimitar Valkov and Steffen Flagge. 2017. Smooth Immersion: The Benefits of Making the Transition to Virtual Environments a Continuous Process. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. ACM, New York, NY, USA, 12–19. https://doi.org/10.1145/3131277.3132183

[36] Jari Varsaluoma and Farrukh Sahar. 2014. Measuring Retrospective User Experience of Non-powered Hand Tools: An Exploratory Remote Study with UX Curve. In *Proceedings of the 18th International Academic MindTrek Conference: Media Business, Management, Content & Services (AcademicMindTrek '14)*. ACM, New York, NY, USA, 40–47. https://doi.org/10.1145/2676467.2676485

[37] Werner Wirth, Tilo Hartmann, Saskia BÃűcking, Peter Vorderer, Christoph Klimmt, Holger Schramm, Timo Saari, Jari Laarni, Niklas Ravaja, Feliz Gouveia, Frank Biocca, Ana Sacau, Lutz JÃďncke, Thomas Baumgartner, and Petra JÃďncke. 2007. A Process Model of the Formation of Spatial Presence Experiences. *Media Psychology* 9 (05 2007), 493–525. https://doi.org/10.1080/15213260701283079

[38] Bob Witmer and Michael Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (June 1998), 225–240. https://doi.org/10.1162/105474698565686