Detecting Breaks in Presence: A Single-Session Feasibility Study for Detrended Fluctuation Analysis in Virtual Reality

Jeffrey Harris*
Department of Experimental Psychology
Utrecht University
Utrecht, Netherlands
j.j.harris@uu.nl

Ralf F.A. Cox
Department of Developmental Psychology
University of Groningen
Groningen, Netherlands
r.f.a.cox@rug.nl

Stefan Van der Stigchel
Department of Experimental Psychology
Utrecht University
Utrecht, Netherlands
s.vanderstigchel@uu.nl

Tanja Nijboer
Department of Experimental Psychology
Utrecht University
Utrecht, Netherlands
t.c.w.nijboer@uu.nl

***Abstract*—Presence is a central quality of virtual reality (VR) experiences, yet its assessment often relies on retrospective self-reports that lack objectivity and temporal resolution. To address this, we tested the feasibility of applying detrended fluctuation analysis (DFA), a method for quantifying long-range correlations in time-series data, to continuous movement data as a way to detect presence dynamics in single VR sessions. Participants engaged in a mini-game requiring visuomotor coordination while two breaks in presence (BIPs) were introduced through unexpected changes in the behavior of virtual objects. Positional data from head and controllers were transformed into velocity time-series to describe motor control and subsequently analyzed with DFA to quantify long-range correlations in movement coordination. DFA outcome variables of phases that included BIPs were then compared against phases without BIPs, Using various thresholds, motor responses to coherence-violating manipulations were classified as falling above, within, or below the variability of non-BIP phases, using various thresholds. Detection dropped at stricter thresholds, indicating that classifications were highly sensitive to threshold settings. Moreover, BIP and no-BIP phases produced predictably different outcomes, supporting the validity of the approach. As such, DFA offers a feasible way to assess presence in single VR sessions when thresholds are chosen carefully, with future work aiming to further map how BIPs impact this measure. These findings highlight both the potential and limitations of DFA for single-session presence assessment. In practice, this approach may provide a foundation for future adaptive VR systems that continuously assess presence and respond by adjusting content based on real-time user experiences.**

Keywords—presence, break in presence, coherence, detrended fluctuation analysis, long-range correlation

I. Introduction

Scientific endeavors to operationalize and assess human-computer interactions (HCI) and user experiences (UX) provide the technical foundation for building and evaluating interactive systems, user interfaces, and software tools. Virtual reality (VR) introduces a distinct challenge, namely the assessment of presence, a fundamental aspect of UX and HCI in VR systems, commonly described as the sense of being located in a virtual world. Presence is a subjective quality of an interactive experience, unique to the perspective of the user and thereby difficult to capture. Still it is crucial to assess the strength of this experience as it links to outcomes that define success across several of VR's use cases, including skill transfer in occupational training [1] and treatment response for virtual reality exposure therapy (VRET) [2], [3].

Many approaches have been proposed, yet most use retrospective self-reports [4], introducing recall and interpretation bias [5] and lacking objectivity and continuity. While novel continuous approaches such as those using physiological methods are frequently being developed and validated [6], significance testing, group comparisons, and experimental designs remain the standard for assessing measurement effectiveness. These tests, while indispensable, frequently do not reflect the reality of real-world use cases, where experiences such as presence may be tested in few, if not single, sessions and outcomes must be conclusive enough to support decisions. Particularly for adaptive VR systems, the ability to detect breaks in presence in real time for a single user is crucial. In contrast to group-level results, systems like adaptive environments or AI-driven avatar agents depend on session-level signals to adjust content in real time. In this paper, we focus on the single-session feasibility of a recently reported novel algorithmic approach for presence assessment [7] and test whether single sessions provide sufficient information to detect shifts in presence.

In addressing the limitations of self-report measures, Harris et al. [7] recently reported on a detrended fluctuation analysis (DFA) based approach that uses only the user's positional data to infer changes in presence under experimental manipulations and potentially provides an objective and continuous alternative to more common approaches mentioned above. The approach is based on the premise that DFA quantifies long-range temporal dependencies in time-series data, while changes in these dependencies can be interpreted as reflecting alterations in sensorimotor coordination that co-vary with presence [7], [8]. The approach derives from movement science, where long-range correlations during coordinated interaction vary in

979-8-3195-0467-8/26/$31.00 ©2026 IEEE

predictable ways due to different types of constraints. For example, when assessing gait in persons with Huntington's disease or Parkinson's disease compared to healthy controls [9], [10], or when differentiating athletic skill between professionals and controls [11], long-range correlations have reliably predicted group differences. In recent years, the effect of cognitive constraints on patterns of long-range temporal dependence and the robustness of the DFA approach has resulted in continuous adoption in movement science and cognitive neuroscience [12], [13], [14].

In the context of VR, system properties that shape sensory input [15], the responsiveness of environments or virtual agents [16], [17], [18], and the qualities of the user's avatar [19], [20], can be viewed as different types of external and internal constraints [21], [22]. The consideration of utilizing the DFA algorithm for presence assessments is based on the idea that these constraints influence how coordination develops over time [7], [8], and that this unfolding of coordination can be captured in the hidden structure and scaling properties of the complex motion signals by way of DFA.

In recent work by Harris et al. [7], the approach showed a promising ability to detect separation between high- and low-presence groups and detection of temporary moments with little or no presence (breaks in presence) within sessions. In addition, the prospect of assessing presence in VR through motor coordination is particularly promising as VR systems track users in multiple degrees of freedom (DOF) by default, producing a vast amount of readily available, high frequency positional data to which DFA can be applied continuously and in real time. Nonetheless, these findings were based on group comparisons with adequate statistical power.

In the current study, we are solely interested in the contrast between the phases without BIPs (i.e. non-manipulated) and the phases that include BIPs (i.e. manipulated phases), and whether these can be described as meaningfully different. BIPs are a central aspect to presence measurement as detecting them is the first step in understanding the causes of presence changes throughout continuous and highly complex VR experiences. For this reason, we aim to explore a simple yet effective approach to detecting motor responses to coherence-violating manipulations that increase BIP within continuous DFA data derived from movement. The approach investigates whether the DFA approach supports useful inference based on single sessions. We therefore evaluate single-session feasibility and sensitivity by assessing responsiveness to within-session manipulations. Based on the established robustness of DFA and its sensitivity to changes in presence when comparing experimental groups, we expect the single-session feasibility of the DFA based approach to predict shifts in presence in the context of manipulations in the majority of the effects.

II. Related Work

The study on which the present research is based [7] was originally designed to test whether DFA could detect group-level differences in motor responses to coherence-violating manipulations (a subcomponent of presence), as well as within-subject differences between phases that included coherence-violating manipulations and those that did not. The previous aim was therefore to examine whether traditional significance testing would reveal reliable between-subject and within-subject effects. The study consisted of two experiments, each employing a 2 × 5 factorial design.

In the first experiment, the experimental groups contrasted high versus low coherence, while the five within-subject levels alternated between phases during which there was no manipulation and manipulation, resulting in 5 levels (3 without manipulation, 2 with manipulation). In the second experiment, the groups combined high versus low coherence with high versus low immersion (a subcomponent of presence), while the same five within-subject levels were maintained.

Participants engaged in a ball-tossing task requiring visual scanning, locomotion, and hand-eye coordination. This task produced continuous, naturalistic movement data suitable for DFA analysis while embedding manipulations that targeted presence.

In addition, the original study also includes moment-to-moment subjective BIP measures that were correlated with DFA outcome measures to validate the DFA approach in the context of detecting coherence-violating manipulations that increase BIPs during specific moments. In the study, the subjective measure was a method called continuous response measurement (CRM) [7]. To capture the subjective outcome measure participants were presented with a video capture of their first-person experience of the VR environment after finishing the task. Using a joystick they moved a slider up and down while the video played to indicate their level of coherence felt during that moment. This recall-based strategy was the only approach that allowed for continuous subjective rating that didn't involve questionnaires which only provide a single outcome or other approaches such as buttons used to indicate presence levels live which, coincidentally might itself result in a BIP. It demonstrated limited predictive value by correlating with the DFA results on one BIP phase and one non-BIP phase. This motivated the authors to continue investigating the DFA approach further.

In the present paper, the focus is shifted from group comparisons and comparing objective with subjective measures to the feasibility of applying DFA to single-session data from individual participants. Our interest was in in testing whether DFA could detect latent patterns of motor responses to coherence-violating manipulations within a session as distinct from phases without such manipulations. To simplify the analytic framework, both experiments were combined, as the within-subject manipulations of coherence were the main focus and remained identical across experiments and conditions. Although this approach may potentially introduce interaction effects between coherence and immersion deriving from group conditions, these effects were considered negligible in light of the robust within-subject findings based on previous work by Harris et al. [7]. As we choose to focus on the within-subject effects we will not be investigating between-subject and interaction effects with alternative subcomponents of presence. For clarity, we will refer to all within-subject manipulations as breaks instead of breaks in presence (BIPs) from hereon.

III. Methods

A. Participants

Participants were between 18 and 65 years of age, fluent in English, with normal or corrected vision, no cognitive or motor impairments, and no history of mental illness. The dataset consisted of n = 61 participants with a mean (*M*) age of *M* = 25 years and a standard deviation (*SD*) of *SD* = 5. Of these, n = 18 identified as female, n = 38 as male, and n = 1 chose not to indicate gender. For n = 4 participants, gender data were lost due to technical errors in the survey application. Participants were students recruited through a participant recruitment system.

Data collection took place between October 31st, 2021, and September 23rd, 2022. This study was approved by the local ethics committee (Approval No. anonymized for review) and conducted in accordance with the Declaration of Helsinki.

B. Materials

The virtual environment was created in Unity 3D using assets from the Unity Asset Store and presented on the Meta Quest 2 headset in both the lab and the home settings. The headset recorded three-dimensional positional data from the head and both controllers at 72 Hz. Questionnaire data were collected using Qualtrics.

The virtual environment consisted of a practice room and a task room connected by an elevator (see Fig. 1). In both rooms, participants interacted with randomly colored balls and baskets while being able to see themselves in a virtual mirror. The practice room allowed participants to get accustomed to the task by tossing a total of three balls into a basket. Which ball color was requested was indicated by a display. The task room contained three baskets that activated randomly and sequentially until all nine balls were cleared. Correct and incorrect tosses triggered different forms of feedback depending on whether the current phase included a break or not. Incorrectly tossed balls would re-spawn allowing the participant to try again.

Fig. 1. Virtual environment stimuli showing all three rooms that participants experience including the practice room with one active basket (left), the elevator with mirror and buttons (middle) and the task room with one active and two inactive baskets (right).

C. Procedure

Participants signed consent forms, completed the demographics information including age, gender, VR experience, weekly play hours, average session duration, and play area size, and received instructions on the VR task. The researcher prepared the headset and started the recording before the participant mounted it. All participants began in the practice room, where they explored the environment and familiarized themselves with the task without encountering any breaks (see Fig. 2). After completing the practice room participants took a virtual elevator into the task room where the experimental task began. Finally, all participants were debriefed.

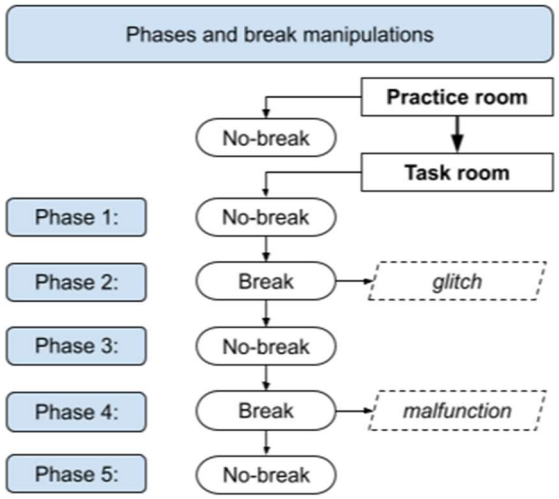

Fig. 2. The progression of phases and breaks across the different rooms (practice room and task room) in the VR environment illustrates how the different types of breaks affected the participants in-between the no-break phases, during which there were no manipulations.

IV. Manipulations

In the within-subject manipulations (breaks), which this present paper focuses on, were only manipulated via the presence subcomponent of coherence and were introduced only in the task room, where balls either suddenly shifted positions or baskets produced incorrect feedback. Coherence describes the subjective boundaries of reasonable circumstances in a VR environment, defined as those that align with prior knowledge and appear self-evident within the narrative context [23]. The first break occurred after one-third of all balls had been tossed into the designated baskets. For participants the first break appeared as a "glitch", and looked like the balls suddenly shifted or reappeared in different positions without any external event causing their movement.

After two-thirds of the balls had been tossed, a second break occurred that resembled a malfunction or feedback error. In the case that a correctly colored ball was tossed into the active basket, the basket failed to deactivate and instead respawned the same ball. This gave the participant the impression that an error was made even though the ball color matched the basket color, forcing the participant to repeat the action. The error persisted until an incorrectly colored ball was placed in the basket, at which point the toss was registered as correct. Both breaks were aimed at immediately and temporarily introducing an unreasonable inconsistency in the environment and/or task. The prediction was that this would provide a strong enough contrast in the participant's movement data compared to the non-manipulated sections, which would be detectable by DFA.

A. Data Analysis

Data were derived from the positional coordinates of the VR headset and both controllers, each with x, y, and z values, producing nine continuous streams of variables in total. The

analysis pipeline began by a row-wise removal of outliers above three median absolute deviations. The coordinates were recorded continuously to capture all types of movements and segmented into five phases. Phases 2 and 4 contained the breaks, while phases 1, 3, and 5 did not (see Fig. 2). Phases 1, 3 and 5 will also be referred to as no-break phases. Each phase began immediately after a ball entered the basket. This approach ensured that the break effects that were triggered right after the balls entered the baskets, were always captured.

To capture active control of movement, velocity time-series were calculated from the raw coordinates time-series. Velocity time-series were used to quantify active motor control, as fluctuations in velocity reflect the application of force during movement [14]. Although velocity measures are well established in motor control research [24], their use in presence studies has been limited, with only a few applications to head- or eye-tracking data [8], [25]. In this study, velocity time-series were calculated by using the change in Euclidean distance between the VR headset (head) (serving as the dynamic reference point) and the controller (hand) between all successive time points. This resulted in two velocity time-series per participant, one for the left hand and one for the right hand. After preprocessing, the length of each phase (per participant) was cut to an equal length with a minimum of 512 data points to ensure comparability and validity [26]. If a phase was too short, the outcome measure was left empty and later replaced through multiple imputation after DFA was applied.

Detrended fluctuation analysis (DFA) was applied to these velocity time-series to assess the presence of long-range correlations (pink noise), producing DFA exponents (α) as outcome measures. DFA exponents near $\alpha = 1$ indicate interaction-dominant dynamics and long-range correlations, while values near $\alpha = 0.5$ reflect randomness (Gaussian white noise) and values near $\alpha = 1.5$ reflect rigidity and short-range correlations. As a result, each participant had two DFA exponents for each phase (deriving from the right- and left-hand controller) which were subsequently averaged into a single DFA result per phase.

The aim was to understand whether the breaks represented meaningful deviations from the no-break phases by assessing how far the break phases deviated from those without breaks, expressed in terms of standard deviations from the mean of the no-break phases. To test this aim, we compared the outcome measures from break phases to the averages of the no-break phases.

To evaluate whether break phases represented meaningful deviations from no-break phases, outcome measures from each break phase were compared against the participant-specific mean and standard deviation of the three no-break phases. DFA exponents from break phases were classified as being above, within, or below the no-break range, with the range defined as the mean of phase 1, 3 and 5 plus or minus a chosen multiple of the standard deviation (0.5, 1 or 2 *SDs*). These standard deviations were not explicitly chosen with the intent of capturing outliers or based on previous approaches found in similar literature but instead served as points of reference. This ensured that deviations were expressed relative to each participant's own variability. In addition to meaningful differences, we also assessed the direction of the deviations. Based on the results of the predecessor study [7], we could make informed prediction that the breaks would result in higher DFA values, higher meaning deviating from the 0.5 range (white noise) towards the 1.0 range (pink noise). Therefore, values classified as above were judged as correctly predicted.

The classifications were then summarized both per break and across breaks, providing counts and percentages of values that fell above, within, or below the no-break range. This allowed us to quantify how often break phases produced outcomes that exceeded the range of no-break phases and to interpret these deviations as meaningful changes in presence dynamics. An interim analysis using only no-break phases was additionally performed to validate the feasibility approach and test whether break phases deviated more frequently than no-break phases.

Missing values were handled by multiple imputation in SPSS with 10 imputations and predictive mean matching. Missing movement data occurred in phases involving breaks due to participants moving too quickly through the environment.

V. Results

Fig. 3 displays the percentage of break phases classified as above, within, or below the no-break range, shown separately for both breaks combined. Across all break phases, the proportion of values classified as within the no-break range predictably increased systematically as the standard deviation threshold widened (see Fig. 3), while the proportion of values classified as above or below decreased accordingly. At the narrowest threshold (0.5 *SD*), a substantial proportion of phases were identified as deviating from the no-break range, with 46.72% falling above and a smaller 25.41% falling below. By contrast, at the widest threshold (2 *SD*), 81.97% of steps were classified as within no-break range, with only a small proportion remaining above (6.56%) or below (11.48%) no-break range (see Table 1). From this we infer the 0.5 *SD* threshold to be the most practical starting point for calibrating the breaks detection in the context of the current data set.

To validate the feasibility approach, we conducted an interim analysis using only no-break phases. For each participant, two no-break phases were used to calculate a baseline mean and standard deviation, and the remaining no-break phase was classified as above, within, or below this range, exactly as in the main analysis. This allowed us to verify whether the classification routine behaved as expected in the absence of manipulations. Specifically, we expected fewer "above" classifications when comparing no-break phases against each other.

The results supported this expectation. Across phases 1, 3, and 5 combined, 38.80% of no-break phases were classified as above threshold at 0.5 *SD*, compared to 46.72% for the combined break phases. Fig. 4 also shows that the percentages of phases above and below the 0.5 *SD* threshold are comparable and visibly closer together, which differs from Fig. 3 where the percentage of above-threshold phases was clearly higher than the within- and below-threshold percentages. At 1 *SD* this trend was no longer visible, with 26.77% of no-break phases classified as above versus 27.05% of break phases. At 2 *SD* the trend even

reversed, with 13.11% of no-break phases classified as above compared to 6.56% of break phases. What stands out here is that the test appears to de-validate the feasibility approach at wider thresholds. At 2 *SD*, no-break phases produced more above classifications than break phases. This reversal suggests that when thresholds are too broad, genuine deviations caused by breaks are absorbed into the "within" category, while normal variability in no-break phases is more likely to overshoot the threshold. In practical terms, higher *SD* settings may over-smooth the data, risk missing meaningful deviations, and create above classifications in otherwise stable phases.

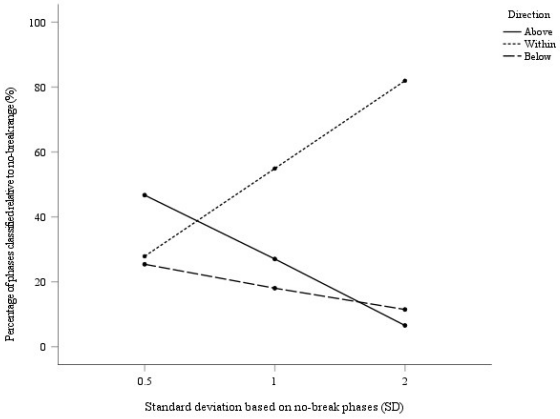

Fig. 3. To understand how the percentages of combined break phases classified relative to no-break phases across standard deviation thresholds behave, we compared those that are above, within, and below the threshold of the chosen standard deviation. The horizontal axis, which represents the standard deviation based on no-break phases, and the vertical axis, which represents the percentage of phases classified relative to the no-break range, are shown in relation to break phases

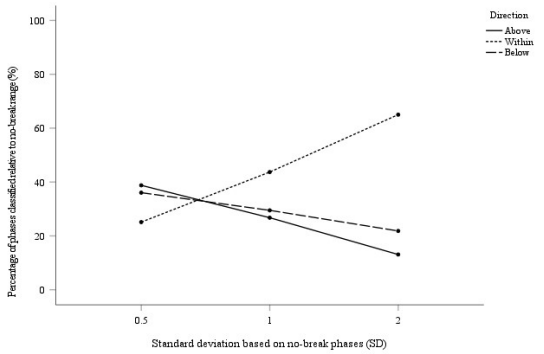

Fig. 4. The validation step compares no-break phases to other no-break phases. Therefore, the horizontal axis, which represents the standard deviation based on no-break phases, and the vertical axis, which represents the percentage of phases classified relative to the no-break range, are in relationship to other no-break phases.

In addition, we checked whether a higher percentage fell within or below the threshold, which was the case for below (36.07%) but not for within (25.14%). Notably, the individual no-break phase 3, which was positioned between the two breaks, showed an above-threshold rate of 52.46%, which was even slightly higher than the two break phases. This suggests that presence may not have fully returned to baseline after the first break, or that the short phase length led to a curvilinear progression across the entire experience

These results demonstrate that DFA classifications are sensitive to temporal dynamics and threshold settings in ways similar to hyperparameter tuning in machine learning models. At stricter thresholds, the algorithm sacrifices sensitivity for reliability, while at liberal thresholds it captures more events but risks false positives. This trade-off is similar to challenges also found in AI classification tasks and suggests that DFA could be calibrated within adaptive systems using context-specific adjustments. Finally, it should be stressed that the results, specifically the percentages at higher thresholds, are based on very few data points, which limits generalizability of these findings. Taken together, we can assume that the classification behaves in a predictable way, strengthening the interpretation of the main findings.

TABLE I. PERCENTAGES OF BREAK PHASES CLASSIFIED AS ABOVE, WITHIN, OR BELOW NO-BREAK RANGE ACROSS STANDARD DEVIATION THRESHOLDS FOR FIRST, SECOND AND COMBINED BREAKS. BREAK 1 REFERS TO THE MANIPULATION DURING PHASE 2 AND BREAK 2 REFERS TO THE MANIPULATION DURING PHASE 4.

Break	**Threshold**	**Direction of Effect**		
	Standard deviation	*Above*	*Within*	*Below*
Break 1	0.5	47.54	32.79	19.67
	1	19.67	67.21	13.11
	2	3.28	86.89	9.84
Break 2	0.5	45.90	22.95	31.15
	1	34.43	42.62	22.95
	2	9.84	77.05	13.11
Combined Breaks	0.5	46.72	27.87	25.41
	1	27.05	54.92	18.03
	2	6.56	81.97	11.48
Combined No-Break	0.5	38.80	25.14	36.07
	1	26.77	43.72	29.51
	2	13.11	65.03	21.86

VI. DISCUSSION

In the current study, we examined the feasibility of applying a DFA-based approach to assess presence in virtual environments at the level of single user sessions rather than at the group level. Feasibility was defined as the ability of DFA to distinguish phases containing breaks from non-manipulated no-break phases in a meaningful way. We operationalized a meaningful deviation as a deviation from the mean of non-manipulated phases by a multiple of its standard deviation. Thresholds of 0.5, 1, and 2 *SD* were tested, with results showing a predictable decline in detected breaks as thresholds increased. This allowed us to explore not only whether the DFA algorithm could detect breaks in single sessions, but also what threshold might balance sensitivity and reliability in a way that provides a helpful reference for tuning the detection of breaks.

At the most liberal threshold of 0.5 *SD*, less than half of the breaks (46.72%) were detected. This dropped sharply to 27.05% at the 1 *SD* threshold, which in practical terms limits the utility of the approach for decision-making if breaks are expected to strongly diverge from no-break presence.

Although group-level comparisons suggested sensitive detection rates, our results highlight that break detection in single sessions may be more nuanced. It is likely that different types of breaks vary in magnitude and may require different thresholds [27], [28], [29]. In the literature it has been frequently noted that different breaks may indeed depend on different subcomponents of presence or interacting related constructs and bodily processes [30], [31], suggesting the need for systematic mapping of break types against no-break variation. Only with such information can threshold settings be appropriately picked to reliably and reasonably detect breaks in applied settings while also avoiding detection of false positives through non-meaningful signal fluctuation. However, to achieve a more universal and dynamic approach, there will likely be a need for context-specific threshold tuning (e.g., for different BIP types, user populations, or VR applications). This would alleviate the need for fixed and rigid thresholds but require large and representative data sets. At this point, the authors would also like to stress that the *SD* thresholds were chosen to illustrate the sensitivity of the DFA approach and to propose a method to assess the feasibility of using single sessions to detect breaks. These thresholds should only be used as points of reference.

VII. Conclusion

A. Limitations

The directional prediction of increased DFA values during break phases is a limitation that warrants dedicated clarification, as it diverges from expectations derived from prior movement science literature [21], [14]. In earlier work, a clear directional hypothesis was deliberately avoided, as observed increases in DFA exponents following BIPs ran counter to these established assumptions. One possible explanation is that certain types of BIPs, like the ones discussed in this study, do not solely disrupt presence, but consequently also induce a shift toward heightened focus or engagement. In such cases, participants may temporarily rely less on automatic sensorimotor coupling and exhibit comparatively more controlled, rigid coordination patterns, reflected in higher α values. This interpretation is consistent with the idea that BIPs targeting within-system coherence may increase engagement rather than reduce it, potentially producing a comparable rise in DFA values. In contrast, other forms of disruption, such as those related to immersion, may reflect differently in the data, with DFA exponents approaching $\alpha = 0.5$, indicating more random and less structured movement. This suggests that BIPs may not uniformly reduce presence but have varied effects on coordination depending on their nature.

An overall limitation in this study and similar studies is that the breaks were predetermined and known to the researchers, which makes the breaks easier to track than in more ecologically valid settings. A critical next step is therefore to assess whether DFA can identify breaks that are not predefined, and whether thresholds can be calibrated to capture breaks of varying magnitude in non-controlled dataset while minimizing false positives.

Further, each participant contributed only five measurement points, three of which were used to estimate the mean and standard deviation against which the two break values were compared. This small number of no-break data points reduces stability and likely contributes to variability in the results. However, this reflects a realistic challenge when applying DFA to single-session analyses, where the number of outcome measures is possibly constrained. Currently, this DFA-based processing pipeline yields one outcome approximately every 6-10 seconds depending on the refresh rates of current VR headsets and the requirement of at least 512 data points for valid DFA outcome measures. VR headsets with faster refresh rates or external motion tracking could increase the number of usable data points and thereby improve threshold estimates.

A potential extension would be the use of sliding window approaches to enable more continuous estimation. However, applying DFA in this manner introduces several limitations. First, the repeated detrending within each window can suppress meaningful low-frequency dynamics, as gradual changes in presence may be interpreted as trends and removed, leaving primarily higher-frequency fluctuations. Second, the use of overlapping windows leads to substantial redundancy, as consecutive estimates are based on largely identical data. This results in highly autocorrelated α values and an apparently smooth trajectory, not because variability is preserved, but because temporal differences are blurred. Third, DFA estimates become unstable in shorter segments, as the reduced data length limits the range of scales and increases noise in the slope estimation. This often necessitates additional smoothing, which further reduces sensitivity to transient effects. Taken together, sliding window DFA may fail to capture both the slower recovery dynamics and the precise temporal onset of breaks in presence, limiting its suitability for real-time or fine-grained analyses in unstructured VR experiences. A more practical alternative may therefore be to relax the current requirement of 512 data points and systematically evaluate how robust DFA estimates remain with fewer observations, as this would directly increase temporal resolution without introducing the methodological constraints associated with sliding windows.

Another challenge is that a considerable share of the effects was not simply missed but misclassified. When both breaks were combined, 25.41% were categorized by the DFA approach as increases in presence, falling below the expected threshold. This raises concerns about the robustness of DFA for this purpose, the validity of the breaks themselves, or the possibility that breaks are subjectively experienced in different ways across individuals, a point well established in prior presence research [32]. This limitation underscores why continuous and objective alternatives to self-report have been pursued. Still, even behavioral data may reflect individual differences in how breaks are perceived, as factors such as personality, prior experience, or other traits are known to shape presence and related constructs [33].

Thus, beyond refining thresholds, it may be necessary to account for individual differences when calibrating outcome measures to ensure results that are both reliable and sensitive to user specific variability. A high misclassification rate and a low true positive rate may be addressed by applying adaptive thresholding, as outlined above, for example by integrating personality characteristics or demographic parameters. In the original work, for example, participants who played more hours per week had higher DFA effects for the first phase ($R^2 = .329$, $F(1, 32) = 15.692$, $p < .001$) [7]. This suggests that these users

may have a different starting point of baseline presence compared to less experienced users. And for real-time adaptations, eye tracking or physiological data can provide additional context about what the user is currently focusing on and how this relates to cognitive and affective states.

Our future work will focus on integrating such context aware measures into an automated approach for assessing user experience. This should also support further validation of the approach. A key objective is to extend applicability beyond constrained laboratory tasks. Current use cases remain limited and closely resemble controlled experimental settings. Expanding stepwise toward more open-ended scenarios such as training, serious games, or even entertainment will significantly increase the practical value of the approach. The main challenge is to test and validate the approach in real world VR settings, such as home use, where capturing subtle BIPs becomes both more difficult and more informative. This can possibly lead to standardized use of presence measurement within the entire industry.

B. Future Work

The next step in this research will be to improve temporal resolution through faster VR headset refresh rates and further map different types of breaks through validation against subjective presence reports to improve the sensitivity of the DFA approach. Findings from such research are likely to advance the development of adaptive and responsive VR environments as well as avatars and agents that likely require shorter feedback and response times to create vivid and dynamic game mechanics, thereby minimizing delay and improving user experience.

To address the potential sensitivity to motor movement changes in response to coherence violating manipulations, we plan to conduct research in which we induce events that change motor behavior without causing breaks in presence (e.g., increased ball toss difficulty). This allows us to test whether our approach can distinguish BIPs from simple task-induced movement changes that could otherwise confound the results.

At the same time, we will examine this from a limitation perspective. BIPs should not respond in the same way as the sanity check when motor movement changes occur without a BIP. And BIPs that are not related to task induced motor demands should still be detectable in the data. One example is audio only induced BIPs that do not have a spatial dimension, meaning the sound source of the BIP does not differ from other audio effects in the environment. This makes it possible to separate BIPs that affect motor movement from those that do not, and to test whether both types are captured by the DFA approach.

Future research may also focus on how improved methods can be supplemented with machine learning approaches that integrate additional signals such as video streams or other sensor data from modern VR headsets. These and further advancements will likely be required to achieve standardized and continuous presence measures and enable their application in practical use cases.

REFERENCES

[1] S. Grassini, K. Laumann, and M. Rasmussen Skogstad, "The Use of Virtual Reality Alone Does Not Promote Training Performance (but Sense of Presence Does)," *Front. Psychol.*, vol. 11, p. 1743, July 2020, doi: 10.3389/fpsyg.2020.01743.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Y. Ling, H. T. Nefs, N. Morina, I. Heynderickx, and W.-P. Brinkman, "A Meta-Analysis on the Relationship between Self-Reported Presence and Anxiety in Virtual Reality Exposure Therapy for Anxiety Disorders," *PLoS ONE*, vol. 9, no. 5, p. e96144, May 2014, doi: 10.1371/journal.pone.0096144.

[3] M. Price, N. Mehta, E. B. Tone, and P. L. Anderson, "Does engagement with exposure yield better outcomes? Components of presence as a predictor of treatment response for virtual reality exposure therapy for social phobia," *J. Anxiety Disord.*, vol. 25, no. 6, pp. 763–770, Aug. 2011, doi: 10.1016/j.janxdis.2011.03.004.

[4] V. Souza, A. Maciel, L. Nedel, and R. Kopper, "Measuring Presence in Virtual Environments: A Survey," *ACM Comput. Surv.*, vol. 54, no. 8, pp. 1–37, Nov. 2022, doi: 10.1145/3466817.

[5] M. Slater, "How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments," *Presence Teleoperators Virtual Environ.*, vol. 13, no. 4, pp. 484–493, Aug. 2004, doi: 10.1162/1054746041944849.

[6] D. Wang, Y. Peng, L. Haddouk, N. Vayatis, and P.-P. Vidal, "Assessing virtual reality presence through physiological measures: a comprehensive review," *Front. Virtual Real.*, vol. 6, p. 1530770, May 2025, doi: 10.3389/frvir.2025.1530770.

[7] J. Harris, J. Van Den Berg, R. F. A. Cox, S. Van der Stigchel, and T. Nijboer, "Complex Dynamical Systems Analysis as a Potential Measure for Presence," *PRESENCE Virtual Augment. Real.*, vol. 34, pp. 1–33, Aug. 2025, doi: 10.1162/pres.a.10.

[8] P. Renaud, S. Chartier, G. Albert, J. Décarie, L.-G. Cournoyer, and S. Bouchard, "Presence as Determined by Fractal Perceptual-Motor Dynamics," *Cyberpsychol. Behav.*, vol. 10, no. 1, pp. 122–130, Feb. 2007, doi: 10.1089/cpb.2006.9983.

[9] J. M. Hausdorff, "Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking," *Hum. Mov. Sci.*, vol. 26, no. 4, pp. 555–589, Aug. 2007, doi: 10.1016/j.humov.2007.05.003.

[10] J. M. Hausdorff, "Gait dynamics in Parkinson's disease: Common and distinct behavior among stride length, gait variability, and fractal-like scaling," *Chaos Interdiscip. J. Nonlinear Sci.*, vol. 19, no. 2, p. 026113, June 2009, doi: 10.1063/1.3147408.

[11] R. J. R. Den Hartigh, R. F. A. Cox, C. Gernigon, N. W. Van Yperen, and P. L. C. Van Geert, "Pink Noise in Rowing Ergometer Performance and the Role of Skill Level," *Motor Control*, vol. 19, no. 4, pp. 355–369, Oct. 2015, doi: 10.1123/mc.2014-0071.

[12] K. Linkenkaer-Hansen, S. Monto, H. Rytsälä, K. Suominen, E. Isometsä, and S. Kähkönen, "Breakdown of Long-Range Temporal Correlations in Theta Oscillations in Patients with Major Depressive Disorder," *J. Neurosci.*, vol. 25, no. 44, pp. 10131–10137, Nov. 2005, doi: 10.1523/JNEUROSCI.3244-05.2005.

[13] R. J. R. Den Hartigh, V. Marmelat, and R. F. A. Cox, "Multiscale coordination between athletes: Complexity matching in ergometer rowing," *Hum. Mov. Sci.*, vol. 57, pp. 434–441, Feb. 2018, doi: 10.1016/j.humov.2017.10.006.

[14] D. G. Dotov, L. Nie, and A. Chemero, "A Demonstration of the Transition from Ready-to-Hand to Unready-to-Hand," *PLoS ONE*, vol. 5, no. 3, p. e9433, Mar. 2010, doi: 10.1371/journal.pone.0009433.

[15] J. Hvass, O. Larsen, K. Vendelbo, N. Nilsson, R. Nordahl, and S. Serafin, "Visual realism and presence in a virtual reality game," in *2017 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Copenhagen: IEEE, June 2017, pp. 1–4. doi: 10.1109/3DTV.2017.8280421.

[16] L. Brübach, F. Westermeier, C. Wienrich, and M. E. Latoschik, "A Systematic Evaluation of Incongruencies and Their Influence on Plausibility in Virtual Reality," in *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Sydney, Australia: IEEE, Oct. 2023, pp. 894–901. doi: 10.1109/ISMAR59233.2023.00105.

[17] R. Skarbez, F. P. Brooks, Jr., and M. C. Whitton, "A Survey of Presence and Related Concepts," *ACM Comput. Surv.*, vol. 50, no. 6, pp. 1–39, Nov. 2018, doi: 10.1145/3134301.

[18] M. Slater, D. Banakou, A. Beacco, J. Gallego, F. Macia-Varela, and R. Oliva, "A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality," *Front. Virtual Real.*, vol. 3, p. 914392, June 2022, doi: 10.3389/frvir.2022.914392.

[19] K. Kilteni, R. Groten, and M. Slater, "The Sense of Embodiment in Virtual Reality," *Presence Teleoperators Virtual Environ.*, vol. 21, no. 4, pp. 373–387, Nov. 2012, doi: 10.1162/PRES_a_00124.

[20] N. Nostadt, D. A. Abbink, O. Christ, and P. Beckerle, "Embodiment, Presence, and Their Intersections: Teleoperation and Beyond," *ACM Trans. Hum.-Robot Interact.*, vol. 9, no. 4, pp. 1–19, Dec. 2020, doi: 10.1145/3389210.

[21] A. Diniz *et al.*, "Contemporary theories of 1/f noise in motor control," *Hum. Mov. Sci.*, vol. 30, no. 5, pp. 889–905, Oct. 2011, doi: 10.1016/j.humov.2010.07.006.

[22] R. F. A. Cox, R. J. R. Den Hartigh, M. J. Richardson, C. Yu, and T. D. Frank, "Complex Dynamical Systems in Human Development," *Complexity*, vol. 2019, no. 1, p. 5010413, Jan. 2019, doi: 10.1155/2019/5010413.

[23] R. Skarbez, "Plausibility Illusion in Virtual Environments," The University of North Carolina at Chapel Hill University Libraries, 2016. doi: 10.17615/2MG3-GH93.

[24] K. D. Simperingham, J. B. Cronin, and A. Ross, "Advances in Sprint Acceleration Profiling for Field-Based Team-Sport Athletes: Utility, Reliability, Validity and Limitations," *Sports Med.*, vol. 46, no. 11, pp. 1619–1645, Nov. 2016, doi: 10.1007/s40279-016-0508-y.

[25] A. Maneuvrier, N.-D.-T. Nguyen, and P. Renaud, "Predicting VR cybersickness and its impact on visuomotor performance using head rotations and field (in)dependence," *Front. Virtual Real.*, vol. 4, p. 1307925, Nov. 2023, doi: 10.3389/frvir.2023.1307925.

[26] M. Sekine, M. Akay, T. Tamura, Y. Higashi, and T. Fujimoto, "Fractal dynamics of body motion in patients with Parkinson's disease," *J. Neural Eng.*, vol. 1, no. 1, pp. 8–15, Mar. 2004, doi: 10.1088/1741-2560/1/1/002.

[27] J. Chung, "Analysis of Break in Presence During Game Play Using a Linear Mixed Model," *ETRI J.*, vol. 32, no. 5, pp. 687–694, Oct. 2010, doi: 10.4218/etrij.10.1510.0054.

[28] B. Rey, V. Parkhutik, and M. Alcañiz, "Breaks in Presence in Virtual Environments: An Analysis of Blood Flow Velocity Responses," *Presence Teleoperators Virtual Environ.*, vol. 20, no. 3, pp. 273–286, June 2011, doi: 10.1162/PRES_a_00049.

[29] B. Liebold, M. Brill, D. Pietschmann, F. Schwab, and P. Ohler, "Continuous Measurement of Breaks in Presence: Psychophysiology and Orienting Responses," *Media Psychol.*, vol. 20, no. 3, pp. 477–501, July 2017, doi: 10.1080/15213269.2016.1206829.

[30] O. I. Caldas, M. Mauledoux, O. F. Aviles, and C. Rodriguez-Guerrero, "Breaking presence in Immersive Virtual Reality toward behavioral and emotional engagement," *Comput. Methods Programs Biomed.*, vol. 248, p. 108124, May 2024, doi: 10.1016/j.cmpb.2024.108124.

[31] B. Rey, V. Parkhutik, and M. Alcañiz, "Breaks in Presence in Virtual Environments: An Analysis of Blood Flow Velocity Responses," *Presence Teleoperators Virtual Environ.*, vol. 20, no. 3, pp. 273–286, June 2011, doi: 10.1162/PRES_a_00049.

[32] D. Weibel, B. Wissmath, and F. W. Mast, "Immersion in Mediated Environments: The Role of Personality Traits," *Cyberpsychology Behav. Soc. Netw.*, vol. 13, no. 3, pp. 251–256, June 2010, doi: 10.1089/cyber.2009.0171.

[33] I. Alsina-Jurnet and J. Gutiérrez-Maldonado, "Influence of personality and individual abilities on the sense of presence experienced in anxiety triggering virtual environments," *Int. J. Hum.-Comput. Stud.*, vol. 68, no. 10, pp. 788–801, Oct. 2010, doi: 10.1016/j.ijhcs.2010.07.001.