



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Impact of Alignment Edits on the User Experience of 360-degree Videos

Lucas dos Santos Althoff

Tese apresentada como requisito parcial para
conclusão do Doutorado em Informática

Orientadora
Prof.a Dr.a Mylène C. Q. Farias

Brasília
2023

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

dA467i dos Santos Althoff, Lucas
Impact of Alignment Edits on the User Experience of
360-degree Videos / Lucas dos Santos Althoff; orientador
Mylène C. Q. Farias. -- Brasília, 2023.
95 p.

Tese(Doutorado em Informática) -- Universidade de
Brasília, 2023.

1. Quality of Experience. 2. Vitual Reality. 3.
360-degree Videos. 4. User Experience. 5. Video Edits. I. C.
Q. Farias, Mylène, orient. II. Título.



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Impact of Alignment Edits on the User Experience of 360-degree Videos

Lucas dos Santos Althoff

Tese apresentada como requisito parcial para
conclusão do Doutorado em Informática

Prof.a Dr.a Mylène C. Q. Farias (Orientadora)
Universidade de Brasília

Prof.a Dr.a Célia Ghedini Ralha
Universidade de Brasília

Prof. Dr. Rudinei Goularte
Universidade de São Paulo

Prof.a Dr.a Débora Christina Muchaluat Saade Prof. Dr. Bruno Macchiavello
Universidade Federal Fluminense Universidade de Brasília

Prof. Dr. Ricardo Pezzuol Jacobi
Coordenador do Programa de Pós-graduação em Informática

Brasília, 20 de Dezembro de 2023

Dedication

I dedicate this work to Dom and Daniela.

Thank you for being there every step of the way.

To Vera and Mario who gave me the ruler and the compass.

Acknowledgement

First and foremost, I am thankful to my supervisor Dra. Mylène C. Q. Farias for the confidence and knowledge transmitted over these five years. Her mentorship has left an indelible mark on my academic and personal development.

I would like to especially thank my parents, Vera L. dos Santos and Mario C. Althoff, who not only supported me emotionally but also took care of my son while I pursued research at the XLIM Lab in France, your generous contributions made this endeavor possible. My deepest thanks also go to my brother, sister, and close friends for their unwavering support and patience.

I also would like to thank colleagues who contributed intellectually to my work, providing inestimable insights and thoughts about my research. In special, José S. Cerqueira, Flávio A. Daltro, Safaa Azzakhnini, Andre Costa, Henrique D. Garcia, Dario D. R. Morais, Myllena Prado, Gabriel Araújo.

Special thanks to my senior research collaborators, Prof. Dr. Marcelo M. Carvalho, Prof. Dr. Li Weigang, Prof. Dr. Chaker Larabi, and Alessandro Rodrigues, for their shared expertise and collaborative spirit.

I am grateful for the financial support from institutions, including the Coordination for the Improvement of Higher Education Personnel (CAPES), Fundação de Apoio a Pesquisa do Distrito Federal (FAP-DF), and the DPI/DPG/UnB - Decanates of Research and Innovation and Postgraduate Studies of the University of Brasilia.

To all those who played a role in shaping my journey, thank you for helping me flourish both academically and personally.

Finally, I want to express my heartfelt thanks to my late graduation supervisor, Ivan Soares Ferreira, who initially offered me the shoulders to see further.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por meio do Acesso ao Portal de Periódicos.

Resumo

Esta pesquisa aborda desafios fundamentais na melhoria da experiência dos usuários em vídeos 360°, especificamente a predição imprecisa do campo de visão do usuário e a compreensão da narrativa. Em vídeos 360°, o conteúdo possui elementos dentro e fora do campo de visão dos usuários. Como consequência, acompanhar a narrativa se torna uma tarefa complexa e dependente da navegação do usuário no conteúdo. Para superar esses desafios, edições de alinhamento ajustam o campo de visão do usuário alinhando-o com uma região de interesse pre-determinada. Essa tese examina como as edições de alinhamento em vídeos 360° impactam a Qualidade de Experiência (QoE). Para investigar os efeitos das edições de alinhamento na QoE dos usuários, conduzimos uma série de experimentos aplicando as recomendações mais atuais da União Internacional de Telecomunicação (ITU). A pesquisa em edições de alinhamento ainda é restrita, a única edição investigada detalhadamente funciona com cortes e alinhamento instantâneo. Neste trabalho, nós propomos uma nova edição de alinhamento gradual, chamada Fade-rotation, que replica o comportamento natural de piscar os olhos para reduzir o desconforto causado com a rotação do conteúdo. Testamos essa abordagem sob uma variedade de condições, e avaliamos seu impacto a partir dos dados de movimento de cabeça e das notas de cada vídeo aferidas pelos participantes. Aplicamos as duas principais metodologias para experimentos subjetivos de QoE, coletamos dados de 108 participantes, cobrindo 5 tipos de edições de alinhamento em 12 conteúdos diferentes. Os resultados foram encorajadores, eles confirmaram que o mecanismo proposto (Fade-rotation), com velocidade de rotação abaixo de 20°/s, atinge um nível de conforto e presença semelhante à edição de alinhamento mais consolidada na literatura (Snap-change). Além disso, todas as edições de alinhamento testadas reduziram a velocidade do movimento de cabeça após a edição, confirmando a utilidade dessas edições para o uso na transmissão de vídeo sob demanda. Finalmente, observamos que o Fade-rotation pode atingir uma redução na velocidade do movimento de cabeça até 8% maior do que a técnica do Snap-change, e uma notável tendência de o Fade-rotation implicar em maiores notas de sensação de presença do que o Snap-change.

Palavras-chave: Qualidade de Experiência, Video 360-graus, Edição de alinhamento, Realidade Virtual, Experiência do Usuário.

Resumo Expandido

Impactos das Edições de Alinhamento na Experiência de Usuário de Vídeos 360 graus

O consumo de vídeos de 360° tem crescido rapidamente, impulsionado pela acessibilidade de dispositivos de realidade virtual (HMD, do inglês Head Mounted Displays), e a produção crescente de conteúdo imersivo de alta qualidade [1]. No entanto, a área da Realidade Virtual Cinemática (CVR, do inglês Cinematic Virtual Reality) enfrenta dois grandes desafios para alcançar maiores públicos. Primeiro, a questão da imprevisibilidade na orientação dos usuários, que leva à narrativas inefetivas quando o criador do conteúdo não prevê corretamente a direção de visão dos espectadores [2, 3, 4]. Em segundo lugar, as técnicas de melhoria de transmissão de vídeos 360° pela internet dependem fundamentalmente da previsibilidade do movimento do espectador. Portanto, ambos desafios se originam da imprevisibilidade do comportamento dos espectadores.

Contrastando com os filmes tradicionais, a estrutura dos filmes imersivos baseados em vídeos de 360° permitem que os espectadores ajam como a câmera, ampliando a liberdade para explorar a cena, ao mesmo tempo em que introduzem dificuldades na criação de uma narrativa coesa [5, 6]. Muitas técnicas tradicionais de edição (*e.g.*, ângulos de câmera, zoom, fade, corte) podem se tornar ineficazes no cenário de 360°, levantando questões sobre como criar narrativas para esse tipo de mídia imersiva. Visando prevenir a perda de informações importantes para a compreensão da história, variadas estratégias de construção de cena foram estabelecidas para esse novo formato usando atratores visuais, porém ajustes na construção da cena não conseguem garantir a direção de visualização dos espectadores [7, 8, 9].

Nesta tese nós investigamos as edições de alinhamento como mecanismo de combate ao problema da previsibilidade da visualização dos espectadores. As edições de alinhamento consistem no redirecionamento do Campo de Visão (FoV, do inglês Field of View) do espectador durante a reprodução do vídeo 360°. Esse mecanismo apresenta uma vantagem com relação ao uso de atratores visuais pois garantem a direção de visualização dos espectadores, são compatíveis com outros mecanismos de guias de visualização, e interferem

apenas pontualmente na experiência dos usuários (UX, do inglês User Experience). Finalmente, as edições de alinhamento são promissoras na efetividade da narrativa e também na melhoria da transmissão de vídeos 360° [10].

Até o presente, apenas um tipo de edição de alinhamento para CVR foi investigado sistematicamente na literatura [10]. Os autores desse estudo inaugural se restringiram a propor um alinhamento instantâneo aplicado na transição de cenas, seguindo as diretrizes da diretora de CVR J. Brillhart [?]. Brown *et al.* destacaram a importância de os criadores de CVR possuírem uma gama de opções de ferramentas de melhoria de UX. Visando estender o número e opções de edições de alinhamento, essa tese propõe um novo mecanismo de edições de alinhamento baseado no alinhamento gradual, ao invés de instantâneo.

O movimento relativo dos elementos visuais provocado por um alinhamento gradual pode ativar o mal estar cibernético e gerar desconforto nos espectadores. No entanto, até onde sabemos, nenhum teste empírico investigou essa suposição diretamente para edições de alinhamento em CVR. Ademais, a rotação gradual possui uma importante vantagem em termos de imersão; ao incorporar a transição de cena com o movimento da cena, espera-se conservar a sensação de presença, diferente dos alinhamentos instantâneos que podem quebrar a imersão ao realizar cortes bruscos. Com base em estudos direcionados à redução do mal estar cibernético, assumimos que rotações graduais podem ser ajustadas para serem tão confortáveis quanto o redirecionamento instantâneo do FOV dos espectadores [11, 12, 13]. Garantindo os mesmos níveis de Qualidade da Experiência (QoE, do inglês Quality of Experience).

Neste estudo, concentramos nossos esforços em dois objetivos principais: desenvolver e avaliar um novo mecanismo gradual offline de edição de alinhamento para vídeos de 360° e avaliar o impacto dessas edições na QoE e no comportamento dos usuários. Os objetivos específicos incluem avaliar a aceitabilidade da edição de alinhamento proposta em relação ao senso de presença e o conforto dos usuários, tendo como base comparativa a edição de alinhamento instantânea [10]; comparar as métricas de movimento da cabeça entre a edição proposta e a edição de alinhamento instantânea; determinar um intervalo seguro de velocidade de rotação para a edição gradual proposta. Para alcançar esses objetivos específicos, coletamos um conjunto de dados de QoE por meio de experimentos subjetivos. A base de dados resultante desta tese é aderente às últimas recomendações da União Internacional de Telecomunicações (ITU, do inglês International Telecommunication Union), é a maior base de dados sobre edições de alinhamento offline em termos de quantidade de participantes, e diversidade de conteúdo. Além disso, a nossa base de dados é composta por dois conjuntos de dados, cada um referente a um dos dois experimentos conduzidos. Foram conduzidos dois experimentos com os métodos de avaliação

mais tradicionais Estímulo Único (SS, do inglês single stimulus) e Estímulo Duplo (DS, do inglês double stimulus), visando construir uma base completa.

Metodologia

A metodologia desta pesquisa é estruturada em três componentes: a edição de alinhamento proposta, a plataforma desenvolvida e os procedimentos experimentais conduzidos. O desenho do novo mecanismo proposto nesta tese é orientado para promover um alinhamento suave e ininterrupto que previna ou reduza o mal estar cibernético, inspirado na ação natural humana de piscar [11]. A nova edição de alinhamento apresentada é chamada “Fade-rotation” (FR), ela combina uma rotação horizontal do quadro de 360° com um efeito de fade-in fade-out. O “Fade-rotation” representa um tipo de edição de alinhamento gradual, com a rotação ocorrendo ao longo de um determinado intervalo de tempo. Essas edições podem ser implementadas enquanto você assiste ao vídeo (online) ou antes de assisti-lo (offline). Em nosso estudo, implementamos apenas a versão offline, ou seja, aplicamos essas edições aos vídeos antes que as pessoas os assistissem. A versão offline é adequada para pesquisas piloto, de aceitabilidade da solução e definir parâmetros gerais. Com essa decisão, evitamos implementar a versão online sem confirmar a viabilidade da edição de alinhamento em termos de QoE.

Para realizar nossos experimentos, precisamos de uma plataforma que atenda aos seguintes requisitos:

- Reproduzir vídeos em 360° em uma ampla variedade de HMDs.
- Conter implementados os métodos experimentais SS e DS.
- Possuir questionários 3D incorporados no reprodutor de vídeo.
- Coletar tanto dados de avaliação subjetiva quanto dados de movimento da cabeça.

Após uma busca na literatura e nos repositórios abertos, não encontramos uma solução aberta disponível para a avaliação subjetiva de vídeos em 360° que satisfaça os requisitos de nosso caso de uso [14, 15, 16, 17]. Portanto, como parte das contribuições desta tese, desenvolvemos uma aplicação web para coletar dados e conduzir os experimentos com usuários.

A plataforma de avaliação subjetiva de vídeos monoscópicos em 360° desenvolvida é denominada Mono360. Nossa plataforma integra um reprodutor de vídeo 360° com um módulo de questionários. Utilizando uma arquitetura cliente-servidor, emprega tecnologias de código aberto, com o back-end executando o framework Yii2 do PHP, uma interface front-end usando o Javascript e Bootstrap e um banco de dados relacional PostgreSQL. O reprodutor de vídeo, operando no navegador do HMD, utiliza a API WebXR para a transmissão e aquisição de dados, e sua implantação é gerenciada com o Docker Compose para possibilitar a portabilidade.

Foram executados dois experimentos com dois métodos diferentes. Em ambos, utilizamos a metodologia experimental descrita na Recomendação ITU-T P.919 [18]. A execução completa do experimento SS levou aproximadamente 37-40 minutos, o de DS durou 57 minutos. Ambos experimentos foram possuíam duas sessões de avaliação de vídeo, com um intervalo de descanso entre as duas sessões. Durante o experimento, os participantes estavam sentados em uma cadeira giratória. Os participantes que usavam óculos ou lentes os mantiveram durante toda a sessão. Os experimentos foram desenhados com oito fases: (1) instruções, (2) treinamento, (3) primeira sessão, (4) primeiro questionário de mal estar cibernético (SSQ, do inglês cybersickness questionnaire), (5) intervalo, (6) segunda sessão, (7) segundo SSQ e (8) finalização. Nas duas sessões, os participantes assistiram aos em ordem aleatória, atribuindo notas a cada vídeo assistido. Mais especificamente, os participantes assistiram a metade das condições na primeira sessão, completaram o primeiro SSQ, retiraram o HMD, fizeram uma pausa de 5 minutos para evitar carga cognitiva excessiva [18], e assistiram outra metade de condições na segunda sessão. No final, os participantes completaram o questionário pós-experimental com perguntas adicionais sobre o experimento, como insights pessoais e comentários. A implementação dos questionários foi totalmente automatizada, sem intervenção do experimentador.

Após assistir a cada vídeo, os participantes foram solicitados a avaliar atributos do conteúdo usando o controle do dispositivo, apontando um raio virtual nos botões da interface. No experimento SS os participantes avaliaram três atributos de cada vídeo: experiência geral, desconforto e presença. No experimento DS o atributo experiência geral foi excluído. Ambos experimentos são do tipo within-subjects, o que significa que todos os participantes avaliaram todas as condições de teste. No experimento SS utilizamos a Avaliação Categórica Absoluta com Referência Oculta (ACR-HR), que requer que os participantes pontuem todos os vídeo, sem saber se eram os originais ou os processados, usando uma escala discreta de degradação de 1 a 5 [19, 20].

Um resumo dos parâmetros da edição de alinhamento utilizados no experimentos é apresentado a seguir:

- SC $t_1 = 15\text{ s}$, $\Delta T_{edit} = 0\text{ s}$, $\Delta T_{fade} = 0\text{ s}$, e alinhamento instantâneo.
- FR $t_1 = 14\text{ s}$, $\Delta T_{edit} = 2\text{ s}$, $\Delta T_{fade} = 1\text{ s}$, e $\omega = 10^\circ/\text{s}$, $20^\circ/\text{s}$, $40^\circ/\text{s}$, $60^\circ/\text{s}$.

No experimento de SS escolhemos o conteúdo com três tipos de movimento de câmera distintos (acelerados, movimento uniforme, fixos). No experimento de estímulo duplo optamos por conteúdos com personagens que interagissem com a câmera em proximidade próxima ao ponto de edição, com o objetivo de intensificar a atenção dos participantes para o que se assume ser a visualização inicial pouco antes do início da edição. A interação

escolhida entre personagem e câmera foi considerada, para buscar incentivar os usuários a identificar uma narrativa dentro da limitada duração do clipe. Dentre os 12 vídeos brutos utilizados nos experimentos, dez foram extraídos de dois conjunto de dados Directors Cut [21], UTD [22] disponíveis na literatura e dois vídeos foram fornecidos estúdio Caixote XR ¹, sendo esses conteúdos exclusivos sem publicações relacionado a eles. As edições de alinhamento foram implementadas manualmente e adicionadas aos vídeos originais usando os parâmetros de rotação dentro do efeito “VR projection” ² do Adobe Premiere Pro.

Resultados

Para realizar as análises estatísticas, formulamos hipóteses visando cobrir todos os objetivos específicos:

H1 : O nível de conforto de FR é equivalente ao de SC;

H2 : SC tem um efeito negativo maior na presença do que FR;

H3 : O alinhamento da ROI impacta nas pontuações de presença, conforto e experiência;

H4 : Edições de alinhamento reduzem a velocidade de movimento da cabeça do espectador após a edição.

No experimento SS, **H1** foi aceita para FR10 em vídeos de movimento de cena fixa e para FR10 e FR20 em vídeos de movimento constante. No entanto, rejeitamos **H1** para qualquer conteúdo de vídeo com movimento dinâmico de cena e para FR com velocidade angular superior a 40°/s. Em termos práticos, para reprodutores de vídeo que não têm a capacidade de considerar o movimento da cena durante a reprodução, recomendamos evitar FR20, FR40 e FR60, pois eles possuem uma probabilidade maior de causar desconforto ao espectador. Em vez disso, optar por FR10 ou a abordagem SC é preferível, pois apresentam uma probabilidade menor de efeitos desconfortáveis. Para vídeos caracterizados por movimento constante da câmera, sugerimos o uso de edições de FR com uma velocidade angular inferior a 20°/s, pois isso pode aprimorar a experiência do espectador, minimizando o risco de desconforto. Em essência, essas descobertas destacam a importância de selecionar uma estratégia FR apropriada, levando em consideração o movimento da câmera, para otimizar a experiência e o conforto do espectador.

No experimento SS testamos a hipótese **H2** agrupando as pontuações de presença por tipo de edição e aplicamos o teste t de Welch para todos os pares. Não encontramos diferenças estatisticamente significativas ($p < 0.05$). Portanto, rejeitamos a hipótese **H2**,

¹<https://caixotexr.com/>

²<https://creativecloud.adobe.com/cc/learn/premiere-pro/web/vr-projection>

confirmando que o SC e o FR não tiveram um efeito distinguível na presença. Considerando as hipóteses **H1** e **H2** no experimento DS, confirmamos que o tipo de edição teve um efeito estatisticamente significativo nas pontuações de diferença de presença e conforto. Em termos de conforto, FR e SC tem as médias indistintas apenas quando FR tem velocidade de rotação de $10^\circ/\text{s}$ (FR10), confirmando assim **H1** para essa condição. Em termos de presença, **H2** não foi estatisticamente confirmada para nenhum caso; no entanto, FR10 e FR20 tiveram DMOS de presença mais baixos do que SC.

Para o experimento SS, analisamos o desempenho de alinhamento a partir do classificador A , que classifica cada observação de um vídeo no experimento como $A = 1$ (alinhado) ou por $A = 0$ (não alinhado). A única condição em que o par de conjuntos alinhado e não alinhado ($p < 0.05$) teve uma diferença significativa entre eles foi para FR 10° no atributo de experiência. Portanto, exceto pela pontuação de experiência FR 10° , o desempenho de alinhamento A não teve impacto nas pontuações subjetivas, satisfazendo parcialmente **H3**. Para completar a análise de **H3**, realizamos um teste post hoc Tukey HSD em todas as combinações possíveis de A com conteúdo (21 comparações), e de A com tipo de edição (15 comparações), bem como de A com tipo de movimento de cena (6 comparações). No total, realizamos 42 comparações, todas elas não significativas. Portanto, não foram encontradas diferenças estatisticamente distinguíveis entre o grupo não alinhado e o grupo alinhado. Com isso, cumprimos a **H3**. No experimento DS, nenhum efeito significativo do desempenho de alinhamento (A) relacionado ao conforto e à presença foi detectado, resultando em $\chi_r^2 = 0.0286$ ($df = 1$, valor de $p = 0.866$) e $\chi_r^2 = 0.61$ ($df = 1$, valor de $p = 0.435$), respectivamente. Portanto, rejeitamos o impacto do desempenho de alinhamento sobre as pontuações de conforto ou presença, reforçando a conclusão do experimento SS.

No experimento SS, todos os tipos de edição que mostram redução na velocidade média de movimento da cabeça são: FR $10^\circ = 14.9^\circ$, FR $20^\circ = 9.5^\circ$, FR $40^\circ = 26.7^\circ$, FR $60^\circ = 33.1^\circ$, Snap-change = 21.5° . Para todos os tipos de edição, há uma redução na velocidade de movimento da cabeça que pode estar relacionada a uma fixação em uma ROI, reduzindo o comportamento exploratório em concordância com a literatura [10]. Com esses resultados, provamos **H4**, que afirma que edições de alinhamento reduzem a velocidade de movimento da cabeça. Por outro lado, para o experimento DS, considerando a hipótese **H4**, confirmamos 18 casos em que a velocidade da cabeça diminuiu. Detectamos duas condições em que a velocidade da cabeça aumentou, ambas para FR20 nos vídeos Vaude e Amizade2. Nenhum tipo de edição teve redução significativa para todos os vídeos testados. No entanto, FR10 teve redução significativa, exceto para o vídeo BSB. O vídeo Paris teve a maior redução geral na velocidade da cabeça.

Em termos mal estar cibernético os níveis identificados em ambos os experimentos

foram incipientes. Mais de 90% dos participantes relataram nenhum ou leve desconforto. Apenas um participante, do experimento SS, relatou sintomas graves causados pelo vídeo Jet. Este participante mencionou ter fobia de altura, após a conclusão do experimento. Essas condições individuais são conhecidas por causar diferenças no conforto e tendência a desencadear ciberdesconforto em RV [23]. No DS, os sintomas foram mais comuns após a primeira sessão do que após a segunda sessão, o que pode ser devido ao fato de que a maioria dos participantes era composta por usuários novatos e com experiência moderada em VR, sendo a primeira reação à tecnologia imersiva potencialmente mais desconfortável. Assim, após a primeira sessão, os participantes estariam mais propensos a sintomas leves e moderados.

Conclusões

Esta tese apresentou a técnica de edição de alinhamento FR desenvolvida para aprimorar a experiência de visualização de vídeos em 360°. O FR utiliza um ponto de gatilho predeterminado, permitindo que cineastas definam os tempos de edição com antecedência. Sua eficácia foi avaliada por meio de experimentos com usuários e uma análise comparativa com a edição de alinhamento instantânea SC [10]. Foram considerados em nossa análise os impactos na UX a partir o julgamento de atributos de QoE (presença, conforto, experiência, ciberdesconforto) e do comportamento de movimento da cabeça. As principais conclusões do estudo são:

1. Com base no feedback subjetivo, as edições de alinhamento testadas não degradaram significativamente o conforto ou a presença dos usuários, com muitos participantes sem perceber as edições.
2. O conteúdo do vídeo e o movimento da cena influenciaram significativamente as avaliações dos usuários, destacando o impacto do movimento do conteúdo no conforto, presença e experiência geral.
3. Uma edição FR com velocidade de rotação maior ou igual a 20°/s deve ser evitada para conteúdos com movimento de cena dinâmico. Uma velocidade de rotação de 10°/s ou um SC é preferível para diminuir a probabilidade de desconforto.
4. O alinhamento entre a RoI e o FoV reduziu a velocidade do movimento da cabeça após a edição, sendo que alinhamentos graduais alcançaram uma velocidade 8% menor do que edições instantâneas.
5. O alinhamento entre a RoI e o FoV não impactou significativamente a presença, o conforto e a experiência.

6. Embora seja necessária uma validação estatística adicional, os resultados de ambos os experimentos sugerem que o FR implica um maior senso de presença do que o SC.

Finalmente, este trabalho lança as bases para outras investigações potenciais. Entre outras possibilidades, futuras linhas de pesquisa incluem:

1. Investigar a versão online do FR.
2. Implementar métodos de automação para edições de alinhamento.
3. Ampliar o conhecimento sobre o impacto dos parâmetros de FR, como duração da edição, velocidade de rotação não uniforme e direção da rotação.
4. Expandir o número de participantes no conjunto de dados para melhor distinguir as pontuações.
5. Realizar estudos adicionais com usuários para avaliar os efeitos de ΔT_{fade} na QoE.
6. Integrar atributos adicionais de QoE, como atenção ou emoção, na análise de edições de alinhamento.
7. Analisar fatores culturais e demográficos usando nosso conjunto de dados.

Abstract

This research addresses fundamental challenges in improving the user experience in 360° videos, specifically the imprecise prediction of the user’s field of view and narrative comprehension. In 360° videos, content includes elements both within and outside users’ fields of view. As a result, tracking the narrative becomes a complex task, dependent on user navigation within the content. To overcome these challenges, alignment edits adjust the user’s field of view by aligning it with a predetermined region of interest. This thesis examines how alignment edits in 360° videos impact the Quality of Experience (QoE). To investigate the effects of alignment edits on users’ QoE, we conducted a series of experiments applying the latest recommendations from the International Telecommunication Union (ITU). The research on alignment edits is still limited; the only extensively investigated edit operates with cuts and instant alignment. In this work, we propose a new gradual alignment edit, called Fade-rotation, which mimics the natural blinking behavior to reduce discomfort caused by content rotation. We tested this approach under various conditions and evaluated its impact based on head movement data and participant-rated scores for each video. We employed the two main methodologies for subjective QoE experiments, collecting data from 108 participants, covering 5 types of alignment edits across 12 different content pieces. The results were encouraging, confirming that the proposed mechanism (Fade-rotation), with rotation speed below $20^\circ/\text{s}$, achieves a level of comfort and presence comparable to the more established alignment edit in the literature (Snap-change). Additionally, all tested alignment edits reduced head movement speed after the edit, confirming the utility of these edits for on-demand video streaming. Finally, we observed that Fade-rotation can achieve up to an 8% greater reduction in head movement speed compared to the Snap-change technique, and a notable tendency to Fade-rotation imply higher sense of presence than Snap-change.

Keywords: Quality of Experience, 360° Video, Alignment Edit, Quality Assessment, Virtual Reality.

Contents

1	Introduction	1
1.1	Contextualization	1
1.2	Goals and Contributions	4
1.3	Publications	5
1.4	Dissertation Outline	7
2	Background	8
2.1	Multimedia Systems and Applications	8
2.1.1	Human-centered Multimedia Applications	10
2.1.2	Streaming 360° videos Use-Case	11
2.2	Subjective QoE Assessment	13
2.3	Storytelling and Editing for 360° Videos	18
3	Related Works	20
3.1	Alignment in Immersive Cinematography	20
3.2	Viewing Guidance	21
3.3	Alignment Edits	23
4	Proposed Solution	26
4.1	Fade-rotation Parameters for User Studies	26
4.2	Mono360 Web-application	30
5	QoE Assessment Experiments	34
5.1	Single Stimulus User Study	34
5.1.1	Tested conditions	34
5.1.2	Experimental procedures	37
5.2	Results	40
5.2.1	Opinion score analysis	45
5.2.2	Head motion analysis	50
5.3	Double Stimulus User Study	56

5.3.1	Content preparation	57
5.3.2	Procedures	59
5.3.3	Data preparation	60
5.4	Results	61
5.4.1	Difference opinion score analysis	63
5.4.2	Head motion analysis	69
6	Conclusions	77
6.1	Final remarks	77
6.2	Limitations and Future work	79
	References	80
	Appendix	94
A	Free and Informed Consent Term	95
A.1	Free and Informed Consent Term	95
A.2	Laboratory setup of the experiments	96
B	Mono360 Details	98
B.1	Survey interface	98
B.2	Recruitment Page	98

List of Figures

1.1	The typical rendering procedure of a 360° video by wrapping the video frame as a texture of a 3D visual sphere and projecting the user’s FoV.	2
1.2	Illustration of the two types of alignment edits investigated. Left: video frames prior to the alignment edits and right after it lining up the user FoV with a specific RoI. Right: top-down perspective of the RoI motion across an alignment edit.	3
2.1	Example of immersive experiences, across various media formats and devices. On the vertical axis, we put the milgram’s virtuality continuum. On the horizontal axis, the interactive continuum.	9
2.2	Two outputs from the video player prototype with alignment edits implemented, which is the target applications of our work.	10
2.3	Example of a QoE-aware network application, within a user-centered design.	11
2.4	Diagram of the main components of a subjective experiment.	14
2.5	Most common scales used in subjective experiments. Examples of continuous scales are shown.	16
2.6	The mode of presentation of four subjective experiment methodologies.	16
2.7	ACR ordinal scale with the subjective bias and unevenly distributed scores.	17
2.8	Sketch storyboards of camera displacement from two sequences in Alfred Hitchcock’s Rope.	18
2.9	Diagrams of RoI positioning around the viewer. Left: Staging zones in the full 360 view. Right: Proxemics.	19
3.1	Two basic types of alignment edit investigated in this dissertation.	24
3.2	Tool for RoI annotation of 360° content	24
4.1	Fade-rotation alignment edit aligns the RoI with viewer FoV. For simplicity, we illustrate a fixed viewer FoV.	27
4.2	Reference coordinate system, defined in terms of the render sphere, the red dot represents the origin of the equirectangular frame.	27

4.3	Two Fade-rotations included in a video timeline, representing the temporal edit structure of a video with multiple alignment edits.	29
4.4	Applied parameters to the video stimuli of the user study: a) instant alignment Snap-change settings; b) gradual alignment Fade-rotation settings. . .	29
4.5	Mono360 architecture.	31
4.6	Tools for capturing and saving experimental data. Subjective rating scores are captured from an embedded user interface without removing the HMD.	32
5.1	Video-stimuli of the subjective experiment, organized by camera motion type. Top: the user FOV at the center point (initial head position). Bottom: the pre-defined target ROI.	35
5.2	Spatial and temporal activity indexes of videos from the user study.	35
5.3	Editing setup for preparing the videos, showing the editing controls for applying the parameters for FR60 and FR10.	37
5.4	Procedure of the experiment, and the subject rating time structure.	38
5.5	Instruction phase views.	38
5.6	Scores for the QoE attributes (presence, comfort and experience) measured in the user study. The scores are grouped by video content. In our user study, each participant rated each video six times. Best viewed in color. . .	41
5.7	Mean opinion scores for presence, comfort, and experience for each video sequence.	43
5.8	Presence and comfort MOS barplots, grouped by edit type and video-content.	45
5.9	Presence and comfort MOS barplots, grouped by edit type and scene motion.	46
5.10	The Mean Opinion Score (MOS) of presence and comfort for each Fade-rotation (FR) rotation speed tested in the study. Two baseline conditions are depicted: snap-change (dashed line) and no edit (solid line) for each video.	49
5.11	Distribution of all cybersickness symptoms.	50
5.12	CDF of the head speed measured 1s after the edit for each video-content.	52
5.13	Boxplot of head speeds measured 1s after the edit, for each video-content grouped by edit type.	53
5.14	Data transformation pipeline for the alignment state (A) computation.	54
5.15	Possible states of alignment: $A = 1$ (first row) when alignment is successful, and $A = 0$ (second row) otherwise. The mean distance between user FOV and ROI just after edit is used to compute the A . We applied a distance threshold of $\tau < 60^\circ$ to classify each trial in terms of A	55

5.16	Boxplots of the participants head speed 1 s after edit for those in the “aligned” ($A = 1$) and “non-aligned” ($A = 0$) groups. The circles shows the mean values.	56
5.17	Illustration of the video content utilized in the experiment, featuring the identification of the assumed viewport and the designated target RoI for each video.	57
5.18	SI and TI indexes of the original videos from DS experiment.	58
5.19	Procedure of the experiment, and the assessment methodology of the applied DS method for comfort and presence QoE attributes.	59
5.20	Assessment questions for comfort and sense of presence QoE attributes. Measurement of the difference between original (A) and processed (B) video.	60
5.21	Difference count histogram grouped by video. Left: Comfort difference count. Right: Presence difference count.	62
5.22	Histogram of difference scores grouped by edit type. Left: Comfort difference count. Right: Presence difference count.	62
5.23	Difference Mean Opinion Scores (DMOS) for each experiment condition	64
5.24	Pairwise comparison of the difference between edit types. The left plot refers to comfort differences, and the right plot to the presence differences.	66
5.25	Pairwise comparison between videos aggregated by resolution level. The left plot refers to comfort differences, and the right plot to the presence differences.	67
5.26	Barplot with the count of cybersickness symptoms intensity	67
5.27	SOS hypothesis for SS, DS, for both comfort and presence data. For the SS experiments we use MOS, whereas DMOS is used for DS experiments.	69
5.28	CDF of the head speed for each video content. Left: CDF measured 1s before the edit. Right: CDF measured 1s after the edit.	70
5.29	Boxplot of head speeds measured 1s after the edit, for each experiment condition.	71
5.30	Boxplot of the head speed difference, computed from the subtraction of the head speeds 1s before the edit and after the edit, for each participant.	72
5.31	Count bars for aligned trials ($A = 1$), and non-aligned trials ($A = 0$) for each video.	73
5.32	Boxplot of head speeds measured 1s after the edit separated in align states facets.	76
A.1	Participant wearing the HMD, and watching a experiment’s video.	97
A.2	Participant wearing the HMD, and watching a experiment’s video.	97

B.1	Welcome page.	99
B.2	Pre-questionnaire.	99
B.3	Free and Informed Consent Term.	100
B.4	Introduction of the training.	100
B.5	Instructions	101
B.6	Session starting page.	101
B.7	Loading session.	102
B.8	Recruitment page	102

List of Tables

2.1	Specification of the system to deliver 360° videos in three use-case scenarios.	12
4.1	Setup table for the QoE assessment experiments, showing the fixed parameters.	33
5.1	Subjective assessment measures	39
5.2	Experiment population summary for both devices.	40
5.3	Correlation between QoE attributes, with data aggregated by Edit type. In bold we highlight the moderate or strong correlations.	44
5.4	Paired Kruskal-Wallis test with FDR adjusted p-values for presence and comfort scores.	47
5.5	Participant’s population. The VR familiarity is categorized into “Novice” (1st experience), “Moderate” (1 or 2 experiences), and “Extensive” (more than 3 experiences).	61
5.6	Difference scores proportion for both QoE attributes, where “Worst” refers to the difference scores of -1, -2, -3. “Better” refers to scores 1, 2, 3.	61
5.7	The coefficients of the correlation between the difference scores of presence and comfort. Top: grouped by video. Bottom: grouped by edit type.	63
5.8	Friedman rank sum test for the experiment factors and variables.	64
5.9	Pairwise comparison with edit type as factor, for Comfort (top) and Presence (bottom) attributes. Applying Wilcoxon Rank Sum test with adjusted p-values using FDR correction. In bold, the comparisons statistical significant or close to significant $p\text{-value} \leq 0.1$	65
5.11	Summary of Pairwise comparisons using Wilcoxon rank sum test with continuity correction considering the effect of video over the alignment state	73
5.10	Difference in Head Speed (HS) between 1s before and 1s after the alignment edit with Standard Error (SE). In bold, the conditions where happened a significant reduction or increase in mean head speed.	75

Acronyms

ABR Adaptive Bit Rate.

ACR Absolute Category Rating.

ACR-HR ACR with Hidden Reference.

CCR Comparison Category Rating.

CDF Cumulative Distribution Function.

CGI Computer Generated Imagery.

CVR Cinematic Virtual Reality.

DASH Dynamic Adaptive Streaming over HTTP.

DCR Degradation Category Rating.

DMOS Degradation Mean Opinion Scores.

DoF Degree of Freedom.

DRL Deep Reinforcement Learning.

DS Double Stimulus.

DSCQS Double Stimulus Continuous Quality Scale.

DSIS Double Stimulus Impairment Scale.

FDR False Discovery Rate.

FoV Field of View.

FR Fade-Rotation.

HMD Head-Mounted Displays.

ITU International Telecommunication Union.

KPI Key Performance Indicators.

M-ACR Modified Absolute Category Rating.

MOS Mean Opinion Score.

MPD Media Presentation Description.

POI Point of Interest.

PVS Processed Video Sequence.

QoE Quality of Experience.

QoS Quality of Service.

RoI Regions of Interest.

SAMVIQ Subjective Assessment Methodology for Video Quality.

SC Snap-Change.

SI Spatial Information.

SOS Standard deviation of Opinion Score.

SRC Source Reference Sequence.

SS Single Stimulus.

SSCQE Single Stimulus Continuous Quality Evaluation.

SSQ Simulation Sickness Questionnaire.

TI Temporal Information.

UX User Experience.

VQA Video Quality Assessment.

VR Virtual Reality.

Chapter 1

Introduction

1.1 Contextualization

Virtual Reality (VR) has become increasingly popular, offering immersive experiences. The VR market is projected to surge from US 28.42 billion in 2022 to an estimated US 87 billion by 2030 [1]. This rise is fueled by factors such as the affordability of Head-Mounted Displays (HMD), the growth of metaverse solutions, and the rising production of high-quality VR content, notably 360° videos [24].

The 360° videos offer exciting possibilities for immersive storytelling, providing a platform to create realistic environments and engaging experiences. The process of rendering these videos involves wrapping the video content onto a virtual sphere and projecting it within an HMD, as illustrated in Figure 1.1. This setup empowers viewers with control over the camera direction, allowing them to explore the scene freely [25]. The result is a seamless interaction between the viewer and the content, enhancing the overall immersive experience [8, 26]. For producers aiming to create enjoyable immersive experiences, it is crucial to comprehend viewer behavior and perception in 360° videos. In the realm of Cinematic Virtual Reality (CVR) storytelling, understanding how viewers follow storylines, perceive scene transitions, and respond to content manipulations is essential for creating engaging narratives [7].

Despite the maturity of the VR industry, streaming 360° videos are in the early development stages. This is primarily due to challenges associated with streaming such content over typical residential broadband Internet connections. However, there is substantial potential for growth in this domain [27]. Addressing two pivotal questions becomes crucial for ensuring a high-quality viewing experience of streaming 360° videos. First, to what extent can the design of 360° video content be improved to enhance user Quality of Experience (QoE)? Second, how can the delivery of resource-demanding 360° videos be improved over the Internet?

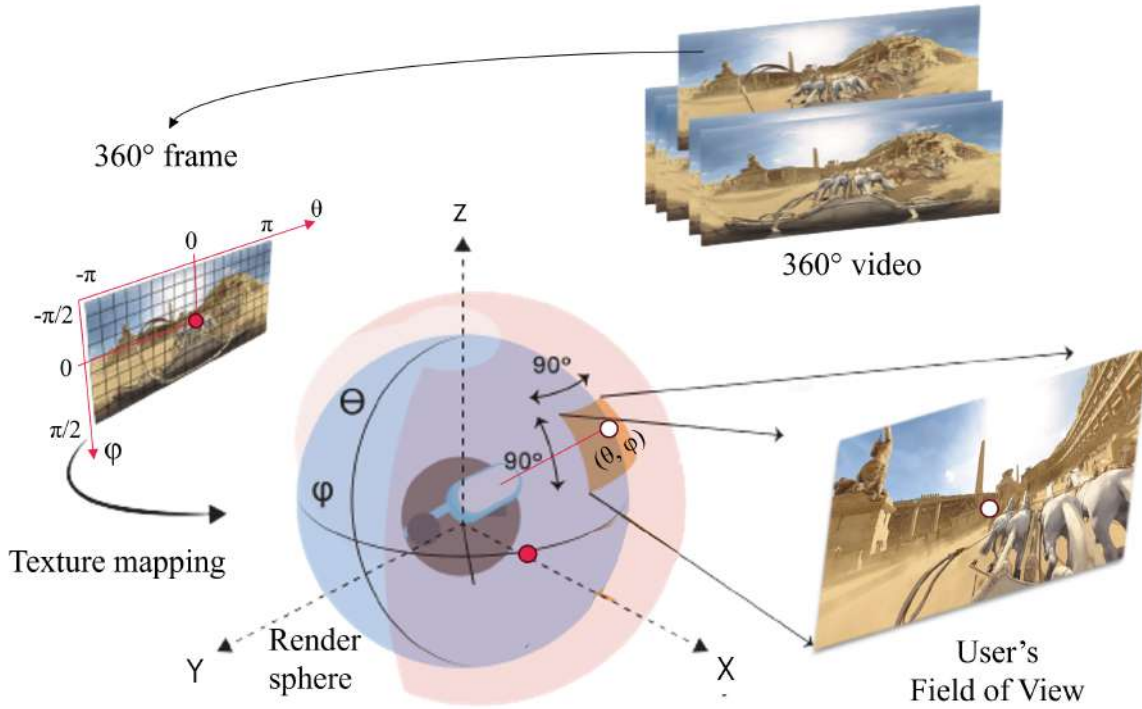


Figure 1.1: The typical rendering procedure of a 360° video by wrapping the video frame as a texture of a 3D visual sphere and projecting the user’s FoV.

Alignment edits have the potential to address both aforementioned questions. Alignment edits involve redirecting the user’s Field of View (FoV) during video playback, as depicted in Figure 1.2. This study delves into understanding user behavior and the subjective evaluation of *alignment edits* on the QoE in 360° videos. Among various techniques proposed for visual guidance in the literature [2, 28, 29, 30, 31], our focus is on alignment edits. These edits stand out due to their substantiated evidence in improving video transmission, and storytelling [32, 33], yet being compatible with other visual guidance techniques, and providing planable Regions of Interest (RoI) visualization for content creators.

Figure 1.2 illustrates the two fundamental types of alignment edits explored in this study: instant and gradual edits. Both approaches aim to align the user’s FoV with a predefined RoI at a specific timestamp. The use of content alignment holds the promise of improving gaze prediction, facilitating more efficient utilization of network resources, and potentially enhancing user QoE [10]. These alignment edits can be activated either in real time by the video player system or seamlessly integrated directly into the original video content. To our knowledge, Dambra *et al.* [10] were the first to investigate real-time alignment editions for streaming Cinematic Virtual Reality (CVR), their technique (called Snap-change) aimed at instantly directing viewers to RoI, this alignment edit allows the

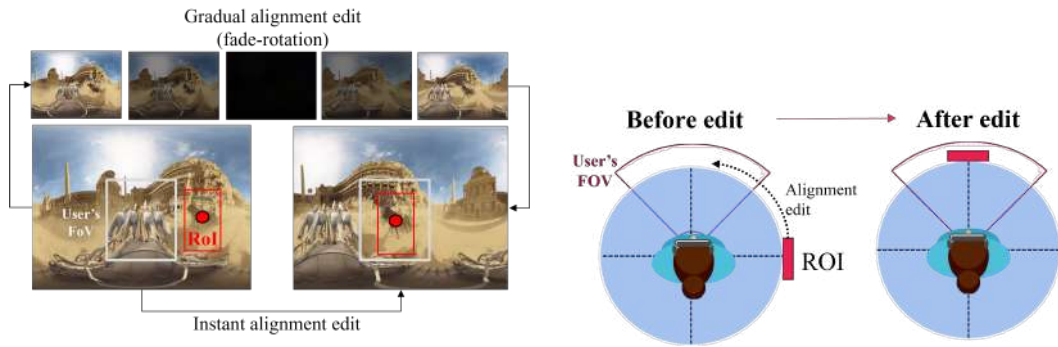


Figure 1.2: Illustration of the two types of alignment edits investigated. Left: video frames prior to the alignment edits and right after it lining up the user FoV with a specific RoI. Right: top-down perspective of the RoI motion across an alignment edit.

VR content creator to drive the user’s attention actively, avoiding observers missing out plots of the storyline.

An outstanding reason for the huge lack of research on techniques that gradually orient viewers watching immersive experiences, is the fact that induced external motion (or motion parallax) can imply cybersickness and discomfort on viewers [34, 12]. Thus, for the best of our knowledge, there is no research on gradual align edits for 360° videos. Following evidence from similar studies, an important assumption of this work is that gradual rotations can be tuned to be as comfortable as instantaneously redirecting the FoV of viewers [11, 12, 13]. Although often the CVR has multiple RoIs in a scene, in this study we considered the simplest case of one RoI alignment with the viewport; *i.e.*, we do not consider the multiple RoI alignment problem. In addition, to avoid intrusive rotations, we highlight that short-duration rotations should be prioritized [35].

Evaluating user perceptions plays a crucial role in optimizing multimedia systems and applications, spanning across telecommunication and signal processing fields [36, 37]. This evaluation involves subjective experiments, systematically assessing multimedia content and delivery systems in controlled environments. Our study aligns with this methodology, seeking to delve into user responses and derive QoE scores from the presented stimuli. Subjective experiments employ two main method types: those with explicit references Double Stimulus (DS) and those with implicit references Single Stimulus (SS). While SS methods are simpler to design and provide a viewing experience close to the real watching experience [36], DS methods become essential when very subtle differences are expected, or when fidelity checks, and discrimination with the reference are important. Moreover, when the experiment will not cover the complete QoE scale [38, 39]. Recent studies also suggest that quality assessment of 360° videos tends to be more reliable in DS than SS experiments [40, 41], thus a complete investigation should account with both methods.

Immersive 360° videos introduce new aspects influencing QoE, such as gaze navigation, HMD, and interactivity [42, 43, 44]. In the traditional multimedia context, quality mainly focuses on visual quality, evaluating sensitivity to impairments in images or videos. However, 360° videos broaden this definition to include user interaction. The most recognized QoE standard definition of QoE is “the user’s degree of delight or annoyance based on the fulfillment of expectations regarding utility and enjoyment in light of the user’s personality and current state,” defined in European network on quality of experience in multimedia systems and services (COST Action IC 1003)[45]. In terms of immersive media experiences, the notion of QoE is closely related with the feeling you are really there [46]. This feeling of "being there" includes knowing where you are in the virtual world, often called the sense of presence or immersion. It also involves taking control of a virtual body, if there is one, and feeling like you can move around and interact with things in the virtual world.

Algorithms for on-demand 360° video streaming are very dependent on head motion; therefore, the impacts of content on viewer behavior concern the research community [47, 48]. This is why, streaming application are a proficient use-case for applying alignment edits [33]. In that context, enhancing real-time User Experience (UX) in multimedia systems requires system adaptability. The adaptability is usually defined by policies to optimize resource usage and avoid UX degradation [49, 50, 24, 51, 37]. This adaptability is particularly critical for achieving user-centric design goals in 360° video streaming applications. These applications rely on UX data, such as the user’s FoV and head motion, to guide graphical computations and traffic management operations. However, accurately estimating QoE is a complex task. While some QoE attributes like visual quality are reliably predictable, others, such as the sense of presence and emotions, pose challenges for precise estimation [52].

1.2 Goals and Contributions

In this study, our focus revolves around two goals:

- Develop and evaluate a new gradual offline alignment edit method for 360° videos.
- Assesses the impact of such edits on the users’ QoE and behavior.

Our specific goals are as follows:

- Evaluate the acceptability of the proposed alignment edit concerning user’s sense of presence and comfort.

- Compare the proposed alignment edit with a baseline in terms of three fundamental QoE attributes (sense of presence, comfort, and cybersickness).
- Compare the proposed alignment edit with a baseline using head motion metrics.
- Determine a secure interval of rotation speed for the proposed gradual offline alignment edit.

To accomplish the aforementioned specific goals, we collected a QoE dataset from a set of subjective experiments with the following key features: multi-factorial, containing measures for more than one attribute of QoE; updated, based on the last International Telecommunication Union (ITU) experiment recommendations; large (in terms of participants), the larger dataset for alignment edits technique; diverse (in terms of content), allowing the evaluation of alignment edits over several conditions; complete (in terms of methodology), providing a QoE assessment with the two fundamental assessment methodologies- *i.e.*, SS, DS.

In this work, we conducted a set of QoE studies designed to collect substantial data for investigating the alignment edits. This dataset contains different QoE attributes. It is consistently updated to align with the latest recommendations from the ITU experiments. Notably, it stands out as a large-scale dataset in terms of participant numbers, providing a robust resource for the thorough evaluation of alignment edits. The dataset covers various content conditions, facilitating a comprehensive assessment of alignment edits. Furthermore, it is methodologically comprehensive, offering QoE assessments through two fundamental methodologies: SS and DS.

The contributions of this work can be summarized as follows:

1. The gradual alignment edit Fade-rotation method.
2. The dataset to evaluate user's QoE and behavior.
3. The web-platform to collect experimental data.
4. The set of metrics to evaluate the QoE behavior.
5. The evaluation using both SS and DS methods.

1.3 Publications

During my doctoral studies, I published one international journal paper and six conference proceedings papers, covering topics related to this dissertation as well as other fields resulting from academic collaborations across disciplines and research projects.

Publications that are part of this dissertation

- **Lucas S. Althoff**, Mylène C. Q. Farias, Alessandro R. Silva and Marcelo M. Carvalho, “Impact of Alignment Edits on the Quality of Experience of 360° Videos,” in IEEE Access, vol. 11, pp. 108475-108492, 2023, doi=10.1109/ACCESS.2023. (A3 Journal - 0.926 SJR - Open Access).

Contribution: Proposed a new alignment edit mechanism, performed two subjective experiments, and made a comparative analysis between the proposed and the literature edit firmly establishing the viability of the new technique [53].

- **Lucas S. Althoff**, Henrique D. Garcia, Dario D. R. Morais, Sana Alamgeer, Myllena A. Prado, Gabriel C. Araujo, Ravi Prakash, Marcelo M. Carvalho, Mylène C. Q. Farias, “Designing an user-centric framework for perceptually-efficient streaming of 360° edited videos,” in Electronic Imaging, 2022, pp 394-1 - 394-7, doi=10.2352/IQSP-394. (International Conference - Open Access)

Contribution: Designed a framework for user-centered 360-degree video adaptive transmission, combining modules for head motion prediction, automatic edit based in saliency map prediction [27].

- Myllena A., **Lucas S. Althoff**, Sana Alamgeer, Alessandro R. e Silva, Ravi Prakash, Marcelo M. Carvalho, Mylène C. Q. “360RAT: A Tool for Annotating Regions of Interest in 360-degree Videos,” in Brazilian Symposium on Multimedia and the Web WebMedia 2022, doi=10.1145/3539637 (A4 conference - **Awarded as best paper**).

Contribution: Provide a software tool to accelerate data annotation in 360-degree videos, a user study resulted in an annotated dataset with a strong correlation between RoI maps and saliency models, indicating a link between the annotated RoI and the saliency properties of the content [54].

- Morais, D. D., **Althoff, L. S.**, Prakash, R., Carvalho, M. M., & Farias, M. C. “A Content-Based Viewport Prediction Model,” in Symposium on Image Quality and System Performance, Electronic Imaging, 2021, doi=10.2352/ISSN.2470-1173.2021.9.IQSP-255. (International Conference - Open Access)

Contribution: The Most Viewed Cluster algorithm (MVC) is proposed. Performed analysis on the head motion data [55].

Publications that are not part of this dissertation

- **Lucas S. Althoff**, Mylène C. Q. Farias, Li Weigang. “Once Learning for Looking and Identifying Based on YOLO-v5 Object Detection,” in WebMedia 2022,

doi=10.1145/3539637.3557929 (A4 conference).

Contribution: Developed a once learning procedure with YOLO deep-learning model [56].

- Li Weigang, Luiz Martins, Nikson Ferreira, Christian Miranda, **Lucas Althoff**, Walner Pessoa, Mylene Farias, Ricardo Jacobi, Mauricio Rincon. “Heuristic Once Learning for Image & Text Duality Information Processing,” in IEEE UIC: International Conference on Ubiquitous Intelligence and Computing, 2022, doi=1109/2022 (B1 conference).

Contribution: Performed analysis showing the predictive capacity of a YOLO model into a once-learning task [57].

- José A. S. de Cerqueira, **Lucas S. Althoff**, Paulo Santos de Almeida and Edna Dias Canedo. “Ethical Perspectives in AI: A Two-folded Exploratory Study From Literature and Active Development Projects,” in HICSS: Hawaii International Conference on System Sciences, 2021, doi=10.24251/HICSS.2021.639 (A1 conference)

Contribution: Designed the study approach and performed the bibliometric analysis [58].

1.4 Dissertation Outline

In Chapter 2, we explore the fundamentals of multimedia systems, underscoring the importance of subjective experiments for optimizing multimedia systems. Chapter 3 provides a detailed overview of the viewing guidance techniques and researches related to alignment edits. Moving to Chapter 4, we elaborate on the proposed alignment edit investigated in this dissertation, showcasing the parametrization of the Fade-rotation. Chapter 5 presents both user experiments conducted, covering the procedures, test conditions, and the results yielded from the data analysis. Section 5.1 unveils the first user experiment based on SS methodology. Section 5.3 shows the procedures, conditions, and results from the user experiment conducted with DS methodology. Additionally, a comparison between both experiments is offered. Finally, Chapter 6 encapsulates the conclusions drawn from the results obtained in the experimental chapters, providing a cohesive wrap-up to the dissertation.

Chapter 2

Background

In Section 2.1, we outline the context within which this work operates, focusing on multimedia systems and applications and the imperative need for their optimization. Moving to Section 2.2, we provide an overview of the foundational principles underlying subjective experiments. These principles are instrumental in shaping our experiments, particularly those centered on alignment edits. Finally, in Section 2.3, we present fundamental concepts related to alignment edits and 360° cinematography. This section serves to provide a contextual understanding of the core elements pertinent to our exploration.

2.1 Multimedia Systems and Applications

Currently, the most prevalent immersive media formats are omnidirectional videos and images [59], categorized by their visual (monoscopic, stereoscopic, 360 degrees, 180 degrees) and audio profiles (directed audio, spatial audio) [60]. These formats, coupled with the Degree of Freedom (DoF) of motion they offer (3DoF, 6DoF), define the UX. As previously mentioned, the omnidirectional viewing experience, unlike regular 2D screen media, encourages a heightened level of immersion. When users wear a HMD, they can alter their point of view by directing their heads within an enclosed virtual space.

Immersive media, supported by various devices, spans across different mediums. Figure 2.1 illustrates immersive experiences categorized by the level of virtuality and interactivity they offer to consumers [61, 62]. In 360° videos viewed through HMD, the level of virtuality surpasses that of conventional videos, offering heightened interactivity by allowing users to choose the direction of their gaze. This medium facilitates the distribution of CVR experiences, primarily based on Computer Generated Imagery (CGI) or live-action content. The virtuality continuum increases with the integration of more virtual elements, while interactivity is associated with the number of reactive actions users can undertake.

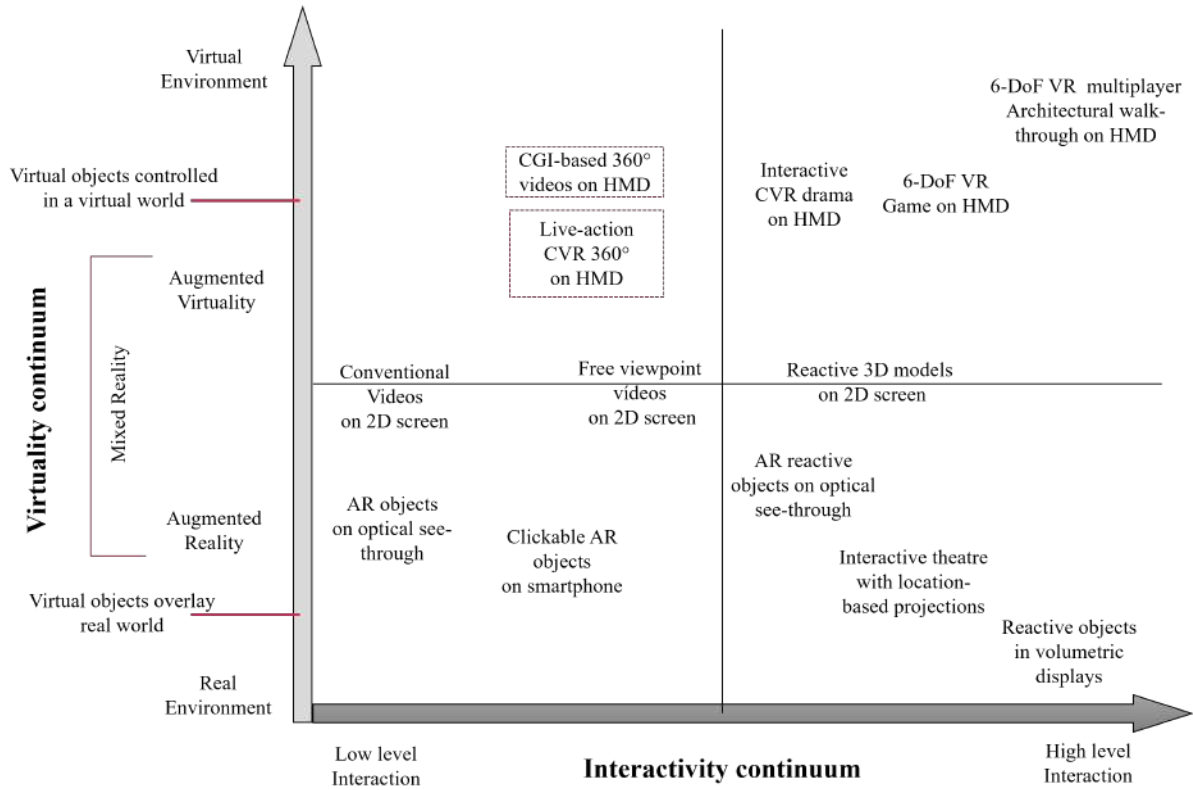


Figure 2.1: Example of immersive experiences, across various media formats and devices. On the vertical axis, we put the milgram’s virtuality continuum. On the horizontal axis, the interactive continuum. Adapted from [61] [62].

The ongoing development of new devices and displays continues to shape the landscape of these experiences.

Although our research on alignment edits had few constraints, with potential applicability across various use cases, our research was guided by a specific target application. This prototype took the form of a 360° video player equipped with the capability to initiate alignment edits at controlled timestamps. Section 2.1.2 delves into the requirements of a significant extension to this initial use case — a video player incorporating Adaptive Bit Rate (ABR) functionality for adaptive streaming combined with alignment edits, some examples of such video players were recently investigated showing promising results [33, 10]. Figure 2.2 represents two cases of a video player implementing alignment edits. Incorrectly applied edits have a detrimental impact on the UX. Conversely, precise alignment that meets user expectations enhances the overall QoE, directly influencing the UX improvement.

In the 360° video viewing task, viewers have two main behaviors: exploratory, where the gaze navigates the content freely searching for some interesting spot, and fixation, in which the act of focusing on some RoI. This selective process is the underlying process

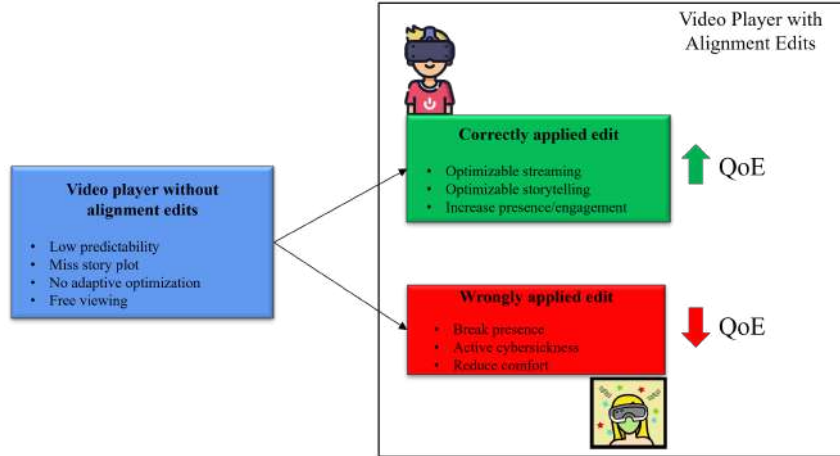


Figure 2.2: Two outputs from the video player prototype with alignment edits implemented, which is the target applications of our work.

of the sequence of gaze directions [4, 63]. Deciding how to align RoI in a scene is a semantic choice; in consequence, if something captures user attention in the wrong way that filmmakers expect, viewers can miss notable events that help understand the story and enjoy the best from the content, thereby likely degrading the UX. In summary, the higher the DoF of 360° videos, the higher the chances of allowing the sensation of immersion, although it also amplifies the probability of missing notable events [3].

Relative motion between content and user gaze is a research interest in VR, and it can be the most significant factor impacting users QoE [64]. The sensation of self-movement is crucial to understanding the effects of relative motion. It is known that the interaction between visual and vestibular systems triggers self-motion, for example, in flying scenes or when a virtual object crosses the viewport [65]. Serrano *et al.* (2020) [66], measured just noticeable differences in the lateral shifts of the head, showing that it is possible to implement real-time optimizations based on content parameters, specifically object distances, that would lead to an imperceptible translation gain.

2.1.1 Human-centered Multimedia Applications

Multimedia systems and applications typically incorporate functionalities to optimize a set of Key Performance Indicators (KPI). The choice of the KPI guiding optimization varies depending on the specific goals of the multimedia application. These applications monitor a wide range of parameters, including content-based properties (*e.g.*, media format, genre, semantic information), network properties (*e.g.*, bandwidth, latency, stall events), and display settings (*e.g.*, resolution, pixel density).

In general, the optimization of multimedia services involves two facets: service provider optimization linked with Quality of Service (QoS), and UX improvement associated with

QoE [67, 68]. The correlation between QoE and QoS properties has been extensively explored in regular multimedia contexts.¹ Shaikh *et al.* (2010) [69] examined the reciprocal relationship between QoS and QoE in multimedia systems, describing it as an exponential decay.

The evaluation of media quality from the end-user perspective holds significant importance for multimedia applications. Figure 2.3 illustrates a QoE-centered application, emphasizing the enhancement of UX through QoE-aware traffic management. In this human-centric scenario, it becomes imperative to assess the UX under diverse conditions [24]. Moreover, it is worth noting that research on QoE-centered applications is currently active, given the substantial volume of content being actively produced [70, 71]. In their 2023 work, Hoßfeld *et al.* [72] delve into the challenges associated with quantifying user-perceived QoE for network operators. The study addresses the limitations of existing QoE models, underscoring issues such as time-consuming development and a lack of universal applicability.

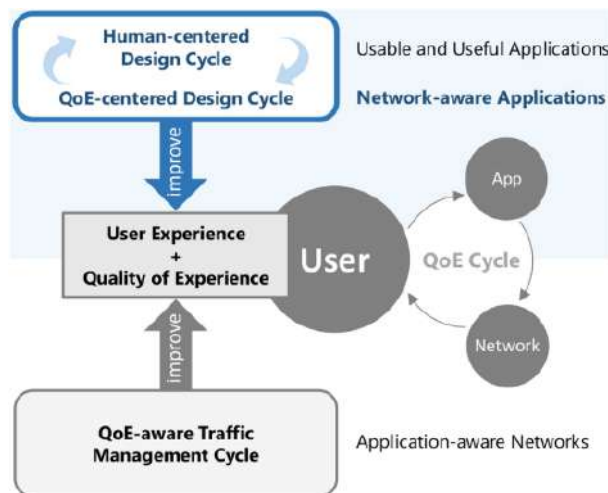


Figure 2.3: Example of a QoE-aware network application, within a user-centered design. Figure extracted from [73]

2.1.2 Streaming 360° videos Use-Case

Since streaming 360° videos is a pertinent use-case for alignment edit, in this section we detail the QoS requirements within this use case. When low latency is important, 360° videos require huge amounts of bandwidth. Taking into account the service specification shown in Table 2.1, watching a 2K video in the HMD requires streaming a 8K video for the full visual sphere. This requirement limits the user reach and, according to the global

¹By regular multimedia, we refer to images, videos, audio, and text formats experienced without special immersive devices, e.g., embedded 2D displays, plain web browsing, etc.

Internet index [74], only 50 countries in the world have minimum entry-level bandwidth connections that allow 2K resolutions. This means that the deployment of immersive multimedia applications is still inaccessible to most Internet users, especially in countries where the communication infrastructure is precarious [75].

Table 2.1: Specification of the system to deliver 360° videos in three use-case scenarios [76].

Aspects	Entry-level	Advanced	Ultimate
Video Resolution	8K mono (7680x3840)	12K mono (11520x5760)	24K stereo (23040x11520)
HMD FoV Resolution	90x90 (1920x1920)	120x120 (3840x3840)	120x120 (7680x7680)
Pixel density (/°)	21	32	64
Color representation	8 bit, 4:2:0	10 bit, 4:2:0	12 bit, 4:2:0
Frame rate	30	60	120
Compression ratio (Estimated)	165:1 (H264)	215:1 (HEVC/VP9)	350:1 (H266)
Compressed Bitrate (full 360)	64 Mbps	279 Mbps	3.29 Gbps

Presently, most Internet video streaming applications rely on protocols like DASH [77]. With DASH, the video file can be partitioned into segments corresponding to short fixed time intervals of playback, termed “chunks” [78]. Each chunk of the video is encoded under different bit rates to accommodate diverse network conditions and device requirements. Additionally, in the DASH-SRD extension [79], each video frame can be divided into tiles of fixed dimensions, forming sets of tiled video segments. Each tile has independent elements of the DASH standard adaptation set, with its position in the frame described by a property element supplementary to Media Presentation Description (MPD). Subsequently, multiple versions of the segmented and tiled video, at different bit rates, are stored on a given server. During video playback, the application’s client side initially requests an MPD XML manifest, followed by sequential requests for each video segment (complete frames or subsets of tiles) based on the appropriate quality defined by the ABR algorithm and the MPD. Consequently, the visual quality perceived by the end user depends significantly on the implemented ABR algorithm. Due to this, the scientific community has actively contributed to the design of ABR algorithms for 360° video transmission [80, 81].

In terms of service requirements, 360° videos require significantly more data than a regular 2D video [82], with this additional information being used to render the entire scene, allowing the user to “look around,” as shown in Figure 1.1. Typically, to have

a viewport resolution of 4K (3840×2160 pixels), we must have at least a total spatial resolution of 12K (11520×6480 pixels), from which most of the information will be ignored by the user [83]. Such a high resolution imposes significant challenges for streaming use-cases. According to Netflix [84], 4K video streaming requires a connection of at least 25 Mb/s, but the average broadband connection in the US is about 18.7 Mb/s, while in Brazil it is just 5.2 Mb/s [74].

2.2 Subjective QoE Assessment

Subjective experiments play a crucial role in enhancing multimedia systems by evaluating the perceptual impact of algorithms or systems on UX (UX). These studies involve human judgments, inherently relying on subjective assessments. The entire domain of objectively assessing image and video quality draws heavily from data collected through subjective experiments, which are continuously evolving [85, 86]. For instance, this data is fundamental for compression, reconstruction, enhancement, and tone-mapping algorithms [87, 88].

Figure 2.4 illustrates the key stages of a typical Video Quality Assessment (VQA) experiment, where participants evaluate various video sequences. The video sequences consist of the Source Reference Sequence (SRC) (reference videos) and the Processed Video Sequence (PVS) (processed videos). The algorithm or system applied to transform SRC into PVS is precisely the one under investigation. The output of the VQA experiment involves averaging scores from individual participants per video sequence, referred to as Mean Opinion Score (MOS) in SS methodology or Degradation Mean Opinion Scores (DMOS) in DS methodology.

Beyond evaluating only visual quality, the subjective QoE experiments consider multiple attributes contributing to the overall UX. Tatsuya *et al.* (2021) emphasized that historically, attention has focused on service quality rather than a user-centric evaluation index. This is partly due to the subjective nature of QoE, closely tied to user perception and expectation, making quantitative and comparative analysis challenging [68]. In Perez, P. *et al.* (2022), authors describe the variety of QoE attributes in the immersive communication systems [37]. Recently, comprehensive reviews about the attributes of QoE showed that most of the evaluating indicators of QoE are distributed in the network layer and application layer, and are less studied from the perspective of users or services [89, 90, 91].

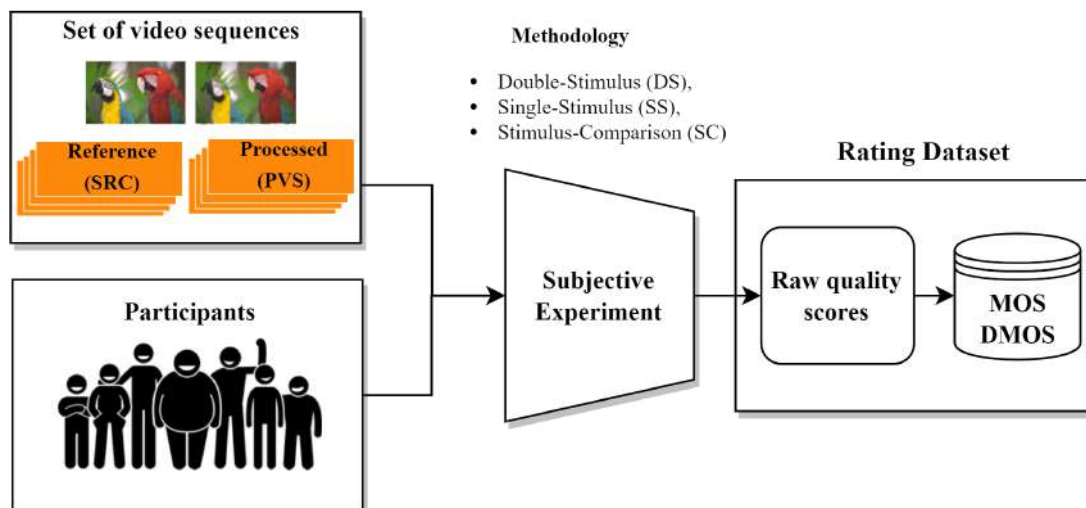


Figure 2.4: Diagram of the main components of a subjective experiment.

The main subjective experiments guidelines are the recommendations from the International Telecommunication Union (ITU)². Those guidelines compile the methods and best practices validated in a vast number of studies. Since Images and videos are visual contents to be transmitted by visual communication systems [92], traditionally ITU has the role to regulate the standards, and recommendations for audiovisual quality in multimedia services. The ITU-T recommendations that will be used in this dissertation are:

- ITU-T BT.500 [36] - Methodologies for the subjective assessment of the quality of television images.
- ITU-T P.800 [38] - Methods for subjective determination of transmission quality
- ITU-T P.910 [93] - Subjective video quality assessment methods for multimedia applications.
- ITU-T P.913 [94] - Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment
- ITU-T R.919 [95] - Subjective test methodologies for 360° video on head-mounted displays
- ITU-T [96] G.1011 - Reference guide to quality of experience assessment methodologies

²ITU is the United Nations specialized agency for information and communication technologies. More information on: www.itu.int

- ITU-T [97] G.1035 - Influencing factors on quality of experience for virtual reality services

Recently, ITU-T R.919 (2021) [95] brought an inaugural recommendation for subjective experiments on 360° videos, formulated through an inter-lab experiment and standardizing experiment protocols for short videos (less than 30s) [18]. ITU-T R.919 comprehensively covers all phases of a subjective experiment, encompassing preparation, conduction, and statistical data treatment. Related to QoE concept, ITU-T G.1011 (2016) [96] brings an overall description of QoE assessment. Further, ITU-T G.1035 (2021) [97] describe the design and measurement of attributes QoE in the context of VR applications. Also useful for conducting the experiments and analysis in Chapter 5, Clause A1-2.3 of ITU-R BT.500-14 [36] shows the experimental methods, post-screening, and traditional outlier removal procedures, that are complemented in Annex A of ITU-T P.913 [94]. For MOS computation, the recommendation is found in Clause 12 of ITU-T P.800.2 [38].

Methodologies for subjective quality assessment are rating and ranking methods [36, 38]. Three aspects characterize an experiment methodology: 1) the mode of stimuli presentation 2) the scale 3) the duration of the stimuli. There are two basic experiment methodologies SS and DS, defined by the mode of stimuli presentation to the subjects, whether the stimuli are presented isolated (SS), or in pairs (DS). The most common SS methodologies are: the Absolute Category Rating (ACR), where subjects measure the quality of a given PVS, not including the SRC; or the ACR-HR in which the reference video is included as a freestanding stimulus for rating like any other. In DS methodologies, participants always watch a pair of SRC and PVS, the most common methodologies are: the Double Stimulus Impairment Scale (DSIS), with the reference preceding the processed video; the Double Stimulus Continuous Quality Scale (DSCQS) where the order of the reference video is randomized, and the Comparison Category Rating (CCR) where the processed video is randomized. Figure 2.5 presents the three basic types of scales (absolute category, impairment, and comparison scale).

When designing a subjective experiment, it is important to decide about the sequence of the three tasks subjects will perform: 1) watching the stimuli 2) waiting for the next stimuli 3) rating the stimuli. Figure 2.6 illustrates the mode of presentation of four important methodologies - ACR, DSIS, M-ACR, SAMVIQ. In Subjective Assessment Methodology for Video Quality (SAMVIQ) subjects can choose the order of tests and correct their votes, as appropriate [98]. Whereas the modified ACR (M-ACR) was designed for 360°, in this methodology participants watch the content twice before judging its quality, the rationale behind this design is to enable users to navigate through all the content before judging it [40].

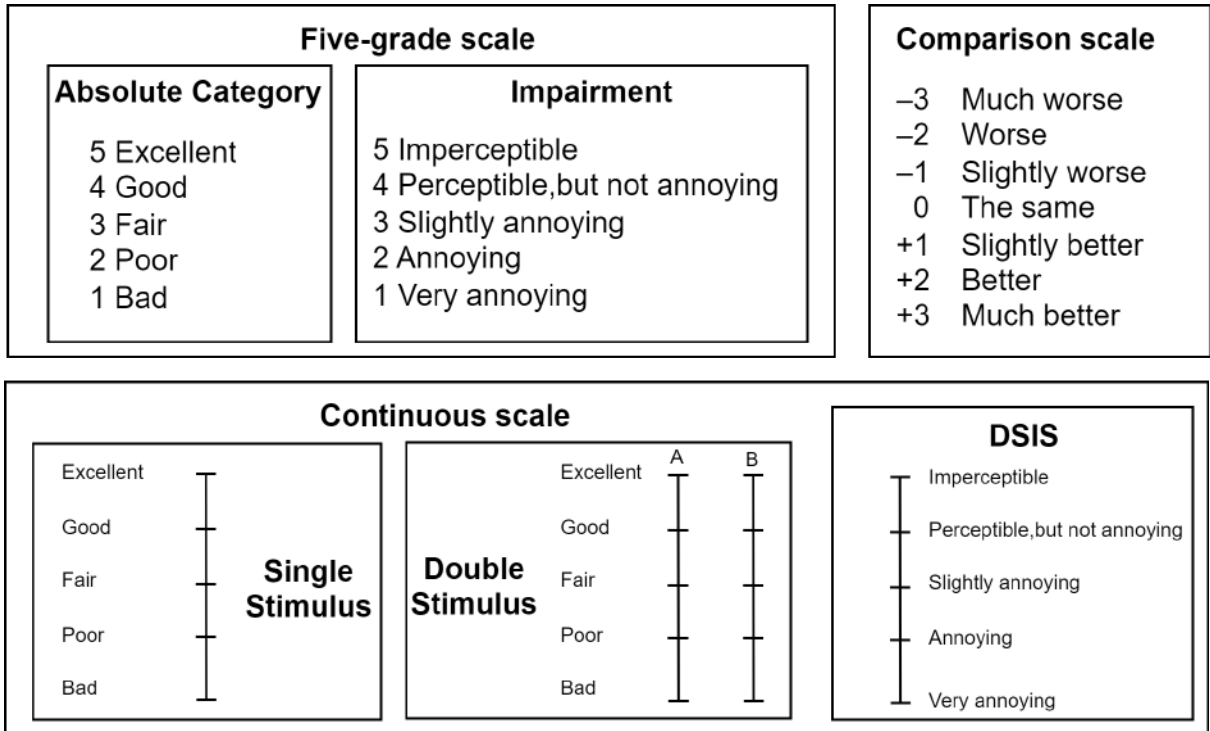


Figure 2.5: Most common scales used in subjective experiments. Examples of continuous scales are shown.

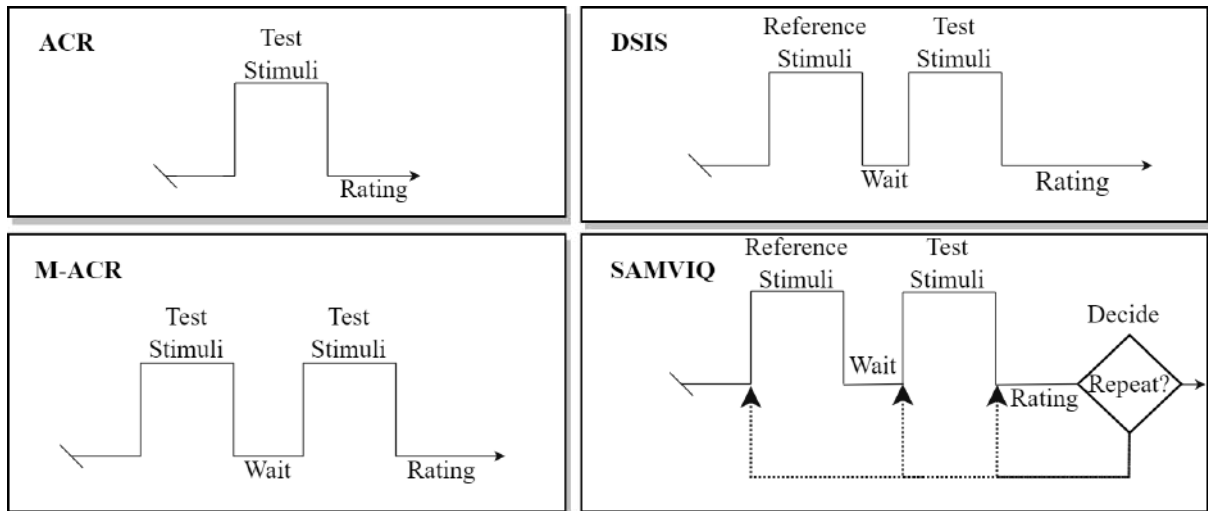


Figure 2.6: The mode of presentation of four subjective experiment methodologies.

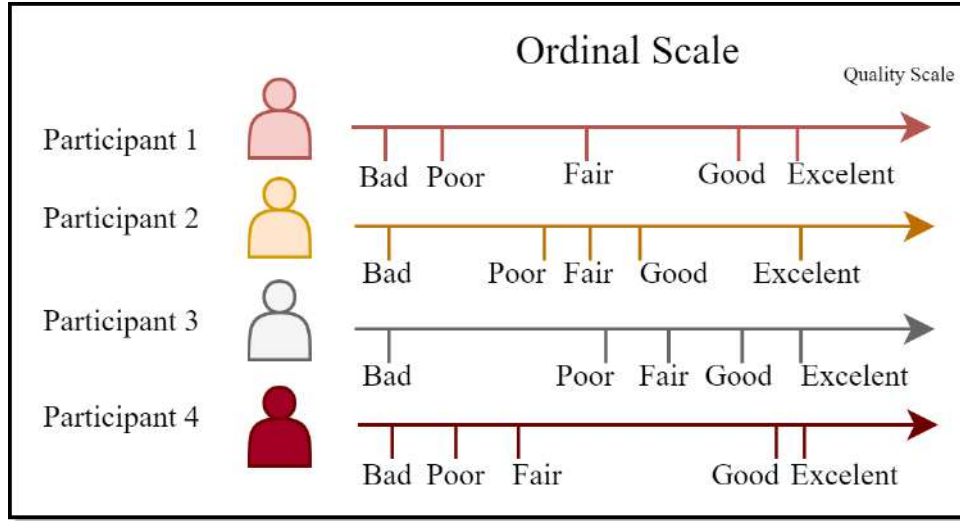


Figure 2.7: ACR ordinal scale with the subjective bias and unevenly distributed scores.

There are valid variations of the main methodologies, for example, the ACR-HR that extends from ACR method, where in ACR-HR the original version of each video is included as a freestanding stimulus for rating like any other. In the study by Freitas *et al.* [99], authors tested modifying the scale labels by using single frames as a quality ruler for VQA, showing promising results for this semantic labeling compared to the traditional SS and DS methodologies. Another specific variation of SS, is the Single Stimulus Continuous Quality Evaluation (SSCQE), where subjects measure the video across time [100]. Tominaga *et al.* [101] compared eight subjective assessment methods, confirming that ACR is the best choice in terms of total assessment time, the difficulty of the evaluation, and statistical reliability. Several studies have explored the pros and cons of each methodology [102, 101, 103, 104].

Figure 2.7 illustrates the ACR ordinal scale. Data acquired from participants of subjective experiments measure order, however, it does not measure the difference between values like an interval scale does. Each subject perceives the quality scale differently, because of that the distance between two points on the scale is not consistently informative [105]. This fact has implications in data analysis, distributions of ordinal data do not satisfy normality assumptions, suggesting that MOS scores cannot be analyzed with parametric tests. Brunnström *et al.* (2018) [106] investigated the impacts of approaching quality scores using parametric and non-parametric tests, demonstrating that parametric test results are similar to non-parametric approaches. However, the authors emphasize the need for caution in analyzing parametric tests and recommend confirmation with non-parametric tests. Ongoing efforts are underway to model distributions on an ordinal scale [107].



Figure 2.8: Sketch storyboards of camera displacement from two sequences in Alfred Hitchcock’s *Rope*. From [62].

2.3 Storytelling and Editing for 360° Videos

A key goal of cinematography guidelines is to establish rules to achieve the feeling of continuity of the scene and the coherence of the aesthetics [108]. Figure 2.8 shows the sketch storyboard in Alfred Hitchcock’s *Rope*, to achieve scene coherency, a director should respect the “180° rule,” which restricts camera positioning across the action axis. Moreover, to achieve continuity of action, typically directors start action in one shot and immediately continue it after a cut. Furthermore, the 180° rule creates a virtual stage where the action unfolds [109], and the action cut simulates the biological motion tracking processes [110]. However, in immersive storytelling, the role of the directors had a drastic change, since the viewer frame is not fully controlled anymore [111, 112].

Immersive media allow viewers to act as the camera, enhancing the freedom to explore the scene, and also introducing difficulties in creating a coherent narrative [5, 6]. Many traditional editing techniques (*e.g.*, camera angles, zooms, fade, cut) may become ineffective in the 360° scenario, raising questions on how to create narratives for this type of immersive media. The development of new guidelines for immersive media supports directors to improve UX [114], [7]. Guidelines suggest 360° video filmmakers planning the RoI positioning in terms of the narrative importance [8], as illustrated in Figure 2.9. The sequence of action should take into account how close to the camera it will happen, and how far from other potential RoIs it happens, since the chance of distraction of the viewer’s attention is higher when RoI is closer [4, 48].

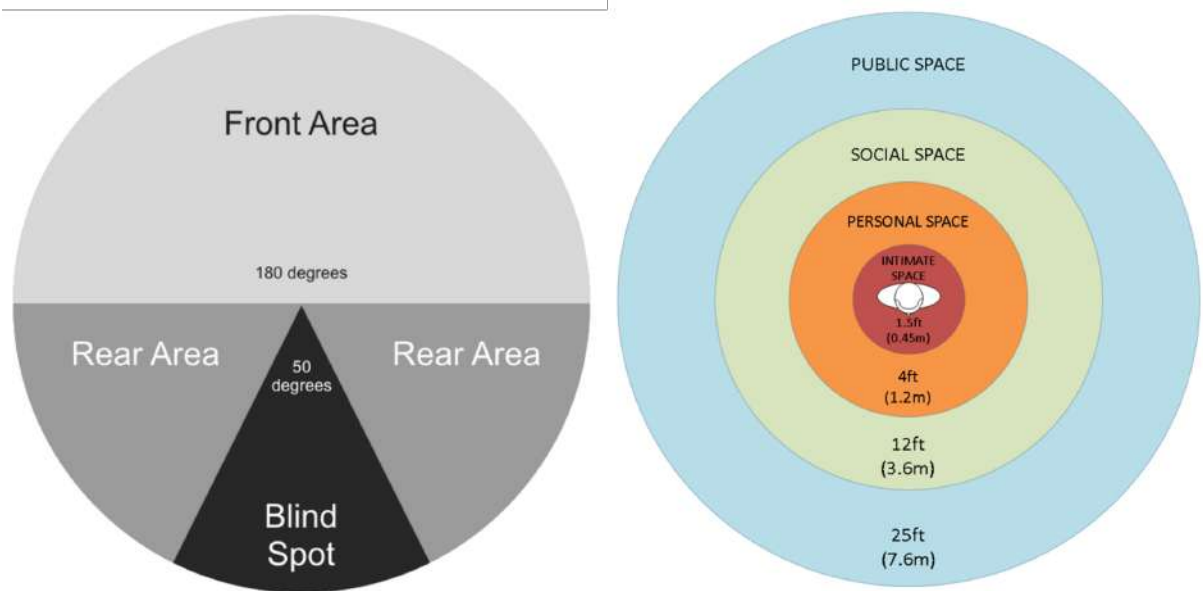


Figure 2.9: Left: Diagrams of RoI positioning around the viewer. Staging zones in the full 360 view. Right: Proxemics. Based on [9, 113].

Chapter 3

Related Works

The current chapter covers the works related with our investigation. In Section 3.1, we describe the cinematography studies in immersive media, focusing on the evaluation methods and the role of the editing strategies. Moving to Section 3.2, we introduce the classification of viewing guidance techniques, delineating the context where our proposed technique is inserted. The principles underlying alignment edits, and the distinction between its two versions conclude the chapter, in Section 3.3.

3.1 Alignment in Immersive Cinematography

One of the precursors of the idea of alignment between scenes in immersive cinematography was J. Brillhart [?] whose created an editing principle for CVR called Probabilistic Experiential Editing, a procedure that generates scene edits by estimating which areas of the content are more salient or perceptually important to the storyline. Another significant concept of CVR cinematography is the temporal and spatial density of the story [9] that corresponds to the quantity, positioning, and frequency of Point of Interest (POI) for a given story timeline. This study also examined viewers' tolerance for spatio-temporal story density. Aitamurto *et al.* [3] examined variations in spatiotemporal viewing conditions in CVR, testing how it triggers the psychological condition of Fear Of Missing Out (FOMO), resulting in anxiety and degrading viewer enjoyment. Fearghail *et al.* [114, 21] investigated how predicted visual attention could help directors perform automatic content analysis, forecasting where the users should direct their gaze.

Cinematography studies argue that scene montage and editing could prevent viewers from missing the plot and promoting engagement in CVR [?, 7]. The authors from [115] propose Adaptive Playback Control (APC) to guide content creators in designing CVR storytelling, in the context of cultural heritage guided tours. The recommendations emphasize the importance of considering viewer tendencies, suggesting more viewer

control for higher engagement and enjoyment, particularly in educational contexts. In 2018, Michael Gødde et al. [9] found, from a user study with 50 participants, that for scenes with high spatial-temporal semantic density, a significant part of the audience can miss the plot. For example, they found a case where 80% of the participants could not correctly answer story-based questions such as “What happened to the main character?” “Why did the character become aggressive?”

In this work, we are interested in studying the effect of using editing strategies to attract the viewer’s attention in 360° videos, we review techniques with this same goal but with different approaches. The work of Kjaer *et al.* [116] explored the effects of editing by adhering to both the principles of attention and match-in-action, they considered the effect of cut frequency on viewer disorientation and found evidence that editing does not pose an statistical significant issue. Speicher *et al.*[117] suggested viewing guidance techniques as a post-production resource that can be implemented in video players to expand accessibility [118]. Another approach, in the montage or post-processing of the content, is to manipulate/edit footage aligning the potential POI across shots. In the work of Pavel *et al.* [119], the authors analyze an additional shot orientation technique that helps viewers visualize all critical information in 360° video stories. Sitzmann *et al.* [48] studied head and eye tracking data to examine the effects of content on user behavior. They conducted an analysis based on time to find RoI and the gaze stabilization metrics, identifying that those metrics were impacted by the number of RoI in the scene and the RoI displacement. Going further in the head motion analysis in CVR contents, Marañes *et al.* [4] analyzed a huge head-tracking dataset (more than 1000 head scanpaths) to evaluate *movie cut* edits impacts on user behavior, resulting on proposing several metrics to support data-based decision-making for film creators.

3.2 Viewing Guidance

Assuming that the content is organized and/or manipulated to engage the audience, two types of techniques are used for viewing guidance: active and passive techniques [120, 7]. The passive guidance can use either diegetic or non-diegetic attractors. Diegetic attractors are those elements included in the scene that are part of the fictitious story and are inserted to capture the viewer’s attention [7], *i.e.*, elements in the content environment like a character crossing the scene. Furthermore, non-diegetic attractors are elements inserted in viewers’ display and outside the fictitious story that acts as visual cues to guide viewers to important parts of the content, *e.g.*, visual effects like arrows, radar, blinking dots [121]. The efficacy of passive viewing guidance strategies is often measured in terms of the viewer behavior, for example, [122] proposed a systematical way of searching for

the optimal class of trajectories using head trajectory data, a complete evaluation of these techniques [123] and a compilation of the available techniques are described in [120].

Active viewing guidance techniques support viewers in run-time, for example, by managing the camera to follow specific targets [124, 125] or by manipulating the luminosity or saturation of the scene [126]. Compared to passive techniques, the active techniques act at conducting viewers' gaze, meaning that they can provide a full predictability of gaze direction often required for streaming purposes. However, if intrusive, it can be inconvenient and annoying for viewers. In terms of UX, avoiding discomfort and cybersickness is fundamental for enabling the acceptability of CVR content. Moreover, the immersion or sense of presence is often determinant to emulate a successful scene [127, 128]. In particular, the activation of cybersickness through motion scenes imposes one of the main challenges to QoE enhancement, since it can degrade the sense of presence and lead to discomfort. Therefore, it is essential to consider the trade-off between cybersickness and presence when designing any mechanism that acts on the sense of motion of viewers [129], [130]. Matching the content and the motion of the viewer in real-time is challenging, although critical, as it can encourage spatial presence and empower the psychological experience of spatial fusion [131].

Given the importance of cybersickness in the UX of immersive content, techniques for reduction cybersickness reduction has been very active, especially for VR games [132, 133, 134, 135]. Closer to our study interests, the authors from [11] proposed a technique that acts on gaze direction. To avoid activating cybersickness, when rapid head movement occurs, the system triggers this technique. When activated, the user's screen illumination decreases, while discrete angular offsets (of 25°) occur. This bio-inspired solution simulates the blinking of the eyes, and a UX experiment proved the reduction of cybersickness by up to 40% for a first-person VR shooting game.

Other type of active techniques are the "Autopilot mechanism", which are based on the generation of camera paths from a 360° video, these systems have a wide set of applications, for example managing the camera in runtime to support viewers to follow specific targets [124], or simplifying the task of watching 360° videos on monitors [136]. Su *et al.* [137] proposed a precursor data-driven algorithm that increases the correlation between videos shot by people and those generated automatically. In [124] a method based on Deep Reinforcement Learning (DRL) was applied, in which an agent identifies and tracks specific objects of interest. In [138] and [136] saliency and optical flow maps are used to calculate optimized camera paths for multiple RoI. Another important application of autopilot techniques is video summarization [139] which can be useful for alleviating film-making work and improving viewer QoE and engagement [140], [7], [141]. A downside of these mechanisms is that they significantly reduce viewers' control which implies loss of

immersion. Yet, the multi-RoI transition system shown in [136] illustrates how alignment edits could function in the context of multi-RoI decision-making.

3.3 Alignment Edits

Alignment edits are active techniques that fall in between the passive and autopilot approaches. They are more intervening than a traditional guiding mechanism but impose less disturbance into UX than “Autopilot mechanisms”. Dambra *et al.* (2018) [10] were the first to propose them to streaming CVR contents, running user studies examining how alignment edits impact viewing behavior for CVR content. In 2019 Cao *et al.* [142] explored three types of transition effects (portal, fade, cut), but did not relate them to aligning or guiding purposes, and did not observe a conclusive reduction in story recall. Dambra *et al.* (2018) [10] argued that the potentiality of cybersickness activation should exclude the usage of mechanisms based on gradual rotation. However, to the best of our knowledge, no empirical test satisfies this assumption directly. Besides the potentiality of comfort impairment, the gradual rotation has advantages in terms of immersion; by embedding scene transition with scene motion, it is expected to conserve the sense of presence.

Figure 3.1 shows a variety of alignment edits we will investigate in this study. The one we suggest, called “Fade-rotation”, is inspired by a method to reduce feeling sick in VR gaming [11]. In “Fade-rotation”, we slowly turn the video while quickly making it briefly disappear and reappear, mimicking a blink of the eye. This way, the change is subtle. It is a gentler approach compared to the second, which is the “Snap-change”, where the video suddenly changes direction and was proposed by Dambra *et al.* (2018) [10]. Both methods let us adjust the video without limiting how users explore it.

Alignment edits, mainly the online version, introduce opportunities for customized improvements. For instance, adaptive accommodation of people susceptible to cybersickness [143, 144], or prone to diverge from a pre-designed storyline (individuals with low reaction time [145]). Furthermore, they also enable service and system improvements, e.g., the efficacy of streaming applications. The authors of [10] and [146] examined the online version of the instantaneous alignment edit, they observed that such automatic edits reduce the exploratory behavior, improving streaming metrics. Furthermore, an important insight from [10] is that instant edits can be imperceptible to viewers, showing that, if inserted in specific content conditions, edits can even not be noticed while incorporating technical improvements.

The online version of alignment edits require annotated videos, defining timestamp where edits should be triggered. Thus a complete automatic implementation of align-

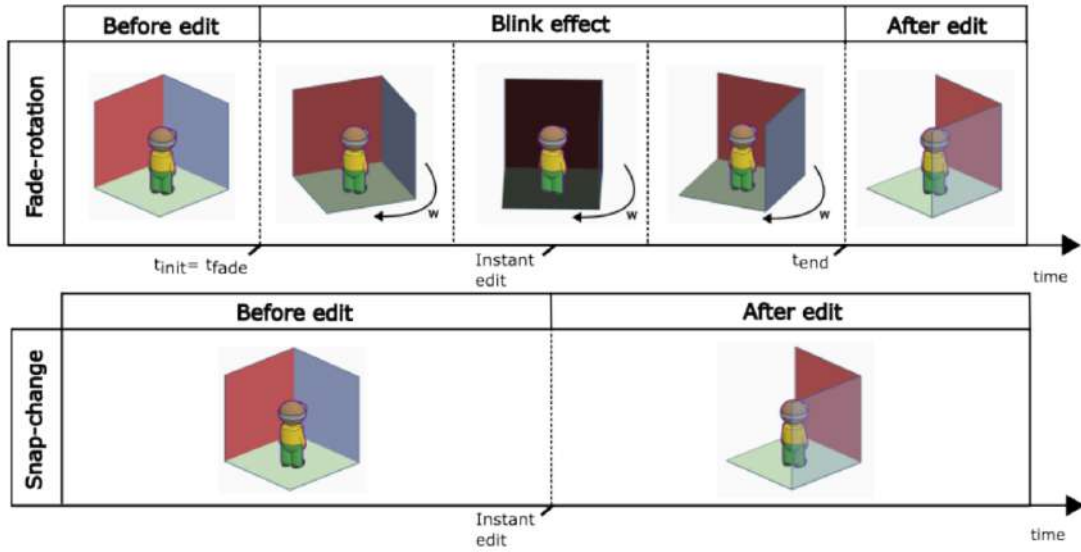


Figure 3.1: Two basic types of alignment edit investigated in this dissertation. Figure extracted from [33].



Figure 3.2: Tool for ROI annotation of 360° content. Figure extracted from [54].

ment edits would require data-driven video description (annotation). A big roadblock for this field is the absence of high-scale datasets with ROI annotation and alignment edits, recently open-source software was published to that end [54]. Figure ?? shows an example of annotating tool specially crafted for 360° content, which is fundamental preparing datasets for supervised learning models. Furthermore, advancements in story summarization of 360° videos can benefit the task of automatic annotation respecting an underlying story, consequently benefiting automatic editing based on machine learning models [141, 139].

On the quest for producing big datasets to enable automatic enhancements of 360° videos, several recent studies have focused on conducting subjective experiments to measure the quality of 360° videos. Elwardy et al. (2022) investigated the minimal number

of participants necessary for providing reliable and precise quality scores [147]. The authors conducted an analysis based on the Standard deviation of Opinion Score (SOS) and demonstrated that the minimal number of subjects may range between 7 and 23, depending on the selected SOS threshold of the study. Singla et al. (2019) [40] proposed and examined the M-ACR, a modified version of the ACR method, specifically designed for 360° videos. In this method, each video is presented twice, aiming to alleviate the effects of scanpath in the measurement of quality. This study also compared ACR, M-ACR, and DSIS methods, concluding that the DSIS method provides more accurate quality scores.

No official recommendation describes experiments for long-duration 360° videos yet. In that direction, Orduna et al. (2023) [148] investigated three methodologies (ACR, SSCQS, and SSDQS) for quality experiments on videos with long duration (approximately 5 minutes-long), they also evaluated UX attributes such as presence, attitude, and attention. In another work [149], the authors compared M-ACR, M-ACR-HR methods concluding that for experts and novice VR users, a larger number of video pairs can be differentiated by the M-ACR-HR method compared to the ACR-HR method which translates to higher reliability of the M-ACR-HR method, while for moderate users ACR-HR still have higher discriminability. Moreover, studies on subjective experiments for other immersive media have also indicated that DSIS is more accurate, underscoring the importance of double stimuli for a more precise quality score in 3D graphics [150].

Considering the potential benefits of alignment edits in enhancing immersive cinematography and the lack of exploration into gradual variations, this thesis aims to introduce a novel gradual alignment edit. To assess its effectiveness, we conduct a comparative analysis with the existing "Snap-change" alignment edit, which is the only type of alignment edit studied thus far [10]. Our focus is on the offline application of these techniques, leaving the exploration of their online counterparts for future research. Chapter 4 provides an in-depth description of our proposed solution.

Chapter 4

Proposed Solution

This chapter introduces the proposed Fade-rotation alignment edit. We present the parameters used for comparing Fade-rotation with the competitor Snap-change in the user studies (see Chapter 5 for results).

4.1 Fade-rotation Parameters for User Studies

Alignment edits involve video transitions that incorporate a frame rotation to align the user’s view with a specific RoI. Our goal is to create a smooth and uninterrupted edit that reduces cybersickness, inspired by the natural human action of blinking when viewing visual material. To accomplish this and inspired by [11], we introduce a novel alignment edit called Fade-rotation, which combines a horizontal rotation of the 360° frame with a fade-in fade-out effect. The Fade-rotation transition represents a type of gradual alignment edit, with the rotation occurring over a certain time interval. These edits can be implemented either while you are watching the video (online) or before you watch it (offline). In our study, we only implemented the offline version, meaning we applied these edits to the videos before people watched them. The offline version is adequate for investigating the acceptability of the solution and defining overall parameters. With this decision, we avoid implementing the online version without confirming the viability of the alignment edit. Therefore, the Online FR, which is applied in video playback time, is an object of future work.

Figure 4.1 shows an illustration of Fade-rotation performing an alignment between RoI and viewers FoV, in a sequence of frames when the viewer is looking statically at the center of the 360° video frame, the horizontal distance between the user’s view and the RoI is reduced. Note that the RoI gradually moves into the user’s FoV, while simultaneously a “blink-of-the-eye” effect is applied.



Figure 4.1: Fade-rotation alignment edit aligns the RoI with viewer FoV. For simplicity, we illustrate a fixed viewer FoV.

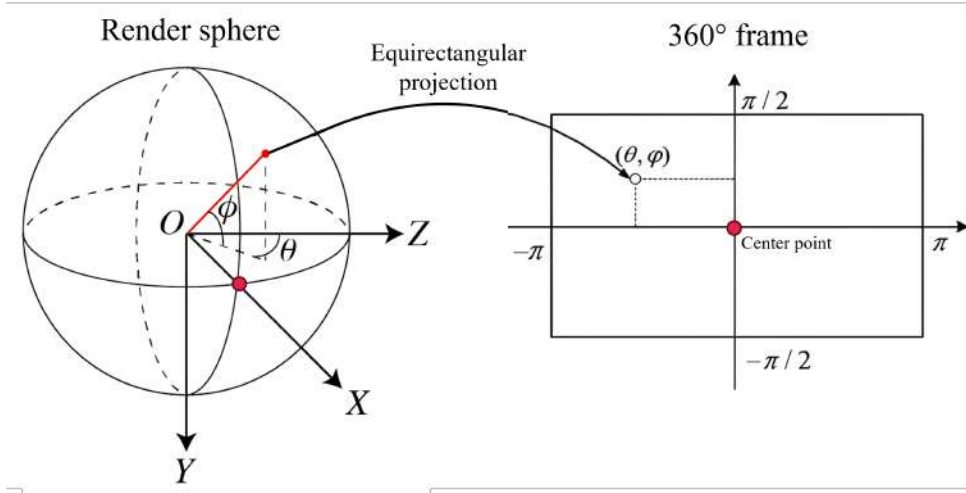


Figure 4.2: Reference coordinate system, defined in terms of the render sphere, the red dot represents the origin of the equirectangular frame.

Figure 4.2 shows the reference coordinate system used in this work, which has an origin in the center of the render sphere. In this system, the HMD is positioned at the origin, and the 360° frame is projected in the shell of the render sphere with a fixed radius (R). We anchor the center point of the 360° frame at $x = R, y = 0, z = 0$. With this anchoring, we have a single coordinate system to describe the content positions, the head directions, and the camera FoV consistently. To map back-and-forth the positions from the render sphere and the 360° frame, we use the equirectangular projection. In this projection, the azimuthal angle (θ) varies within the interval $[-\pi, \pi]$ (in radians) and corresponds to horizontal (side-to-side) head movements around the y -axis. The polar angle (ϕ) varies within the interval $[-\pi/2, \pi/2]$ (in radians) corresponding to vertical (up-down) head movements. The center of the 360° frame is fixed at $\theta = 0, \phi = 0$, and corresponds directly to a fixed position at the reference coordinate system $x = R, y = 0, z = 0$; establishing a fixed reference.

The alignment edit method has three parameters: the total duration of the rotation

edit (ΔT_{edit} in seconds), the duration of the fade-in/fade-out effect (ΔT_{fade} in seconds), and the angular speed of the 360°-frame rotation (ω in degrees/s). These parameters can be adjusted and combined to obtain the desired transition behavior. Some examples of alignment edits include:

1. instant alignment (Snap-change): $\Delta T_{edit} = 0$, $\Delta T_{fade} = 0$, and high values of $|\omega|$;
2. gradual-rotation alignment: $\Delta T_{edit} > 0$, $\Delta T_{fade} = 0$, and $|\omega| > 0$;
3. Fade-rotation alignment: $\Delta T_{edit} \geq \Delta T_{fade} > 0$ and $|\omega| > 0$.

In this study, we focus on investigating Snap-change and Fade-rotation. We deliberately excluded the general “gradual-rotation alignment” because studies have shown that it implies a negative impact on user comfort [11, 12]. For studying Fade-rotation, we analyze only the rotation speed (ω) parameter, fixating the parameters ΔT_{edit} and ΔT_{fade} . By concentrating on $|\omega|$ parameter, we aim to simplify the experiment design reducing the number of controlled parameters. In addition, since this is a precursor study on Fade-rotation, we considered that rotation speed should be prioritized in detriment of the edit duration. With that, we will be able to propose Fade-rotation and to determine a secure interval of rotation speed for it. Finally, edit duration (ΔT_{fade}) and the time interval between edits should be investigated in a future work.

To fix t_1 and θ_T parameters we consider two facts. First, the accommodation time around 14 and 16 seconds before the edit [48], so we set $t_1 = 15s$ for SC and $t_1 = 14s$ for FR. Second, for our offline alignment edit we must fix an “assumed viewport” RoI. To improve chances that participants would be gazing the “assumed viewport” at t_1 , we took advantage of one fact: when viewers watch a 360° video, generally she/he looks towards the center of the frame, regardless of the content [48]. In Section 5.1 we described the rationale behind fixing the “assumed viewport”. The Fade-rotation edit should be implemented at pre-selected timestamps of the video (t_1, \dots, t_N). Figure 4.3 illustrates the video structure, which contains alignment edits between video shots. Applying the edit can be a player’s decision to enable real-time streaming optimization. Furthermore, optimization models can automatically determine whether to trigger alignment edits, while still respecting the cinematographic choices of content creators.

Figure 4.4 provides an illustration of the Snap-Change (SC) and Fade-Rotation (FR) edit techniques. When editing the source videos, we considered the following visual equivalence rule: if two viewers were looking at the same frame location at the start of an edit, they should end up at the same frame location regardless of the type of edit executed. We define θ_T as the total angular displacement between an initial RoI and a target RoI after an edit.

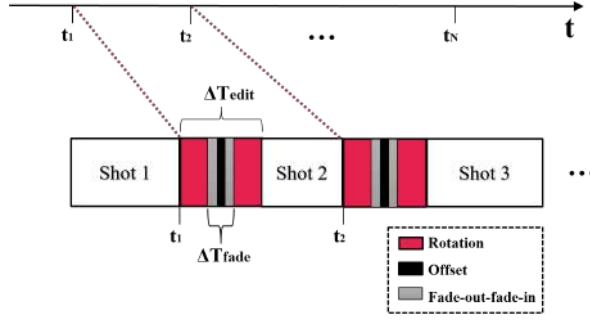


Figure 4.3: Two Fade-rotations included in a video timeline, representing the temporal edit structure of a video with multiple alignment edits.

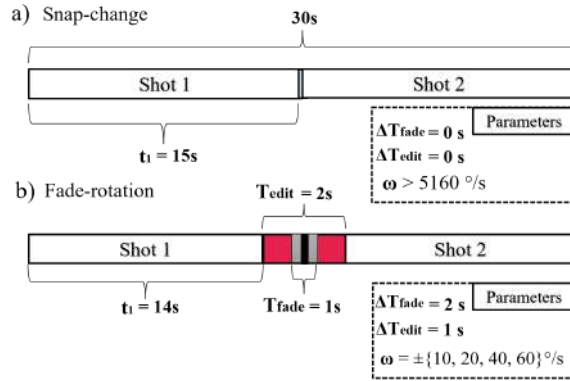


Figure 4.4: Applied parameters to the video stimuli of the user study: a) instant alignment Snap-change settings; b) gradual alignment Fade-rotation settings.

In the offline alignment edit case, we cannot control where users are looking at the video when the edit is executed. Because of this, we selected the initial RoI as an assumed viewport, where if viewers were looking at it they would watch the target RoI as expected. To increase the chance that users watched the initial RoI, we selected videos in such a way that there was one meaningful object in the center of the frame at the time of editing.

The gradual alignment edit tested in this work has several rotation speeds, which affects the angular displacement that can be achieved in a given time interval. For a video, the angular displacement achieved with the FR edit method is given by $\theta_r = \omega \cdot \Delta T_{edit}$, which may be higher or lower than the target total angular displacement θ_T . This requires that a small offset rotation be applied to the video, which is done at the exact moment the frame is completely black. The value of the offset rotation is simply the difference between θ_T and θ_r .

4.2 Mono360 Web-application

As part of our proposal, we developed a web-application for gathering data and conducting the experiments. For running our experiments we need a platform that fulfills the following requirements:

- Plays 360° videos for a big set of HMDs.
- Implements both SS and DS methods.
- Have embedded 3D questionnaires in the video player.
- Gathers both subjective rating data and head motion data.

For the best of our knowledge, there is no open solution available for subjective evaluation of 360° videos fulfilling the requirements of our use case [14, 15, 16, 17]. Specifically, the ALTRUIST [17] platform does not have an integrated 360° video player, the AV-track360 [14] does not have integrated questionnaires for gathering subjective ratings, while MIRO360 [15] and TOUCAN-VR [16] platforms do not implement the DS method, and their software architecture is hard to customize. Therefore, we developed an open platform to perform the experimental procedures.

Our platform is called Mono360, standing for “monoscopic 360° video subjective assessment tool”. Mono360 consists of a 360° video player integrated with a survey module and a relational database. The Mono360 application was designed to be web-based, providing a flexible, portable and robust solution for conducting subjective experiments with immersive multimedia.

Figure 4.5 illustrates the main components of the application, emphasizing that the deployment of the application is managed with docker compose, to ensure portability. It is based on a client-server architecture and uses only open-source technologies. The back-end of the application executes the PHP’s Yii2 framework¹ on the server side, while the front-end interface uses Bootstrap² framework. The database is a Postgresql³ relational database, while the video player runs on the HMD’s browser, and uses WebXR⁴ API to transmit data from the device to the browser.

For rendering 360° videos, we used the Three.js⁵ library, which is based on the WEBGL2 renderer. The rendering procedure consists of decoding the video texture into two spheres corresponding to both eye screens. For that, we implemented the render sphere

¹<https://www.yiiframework.com/>

²<https://getbootstrap.com/>

³<https://www.postgresql.org/>

⁴<https://immersiveweb.dev/>

⁵<https://threejs.org>

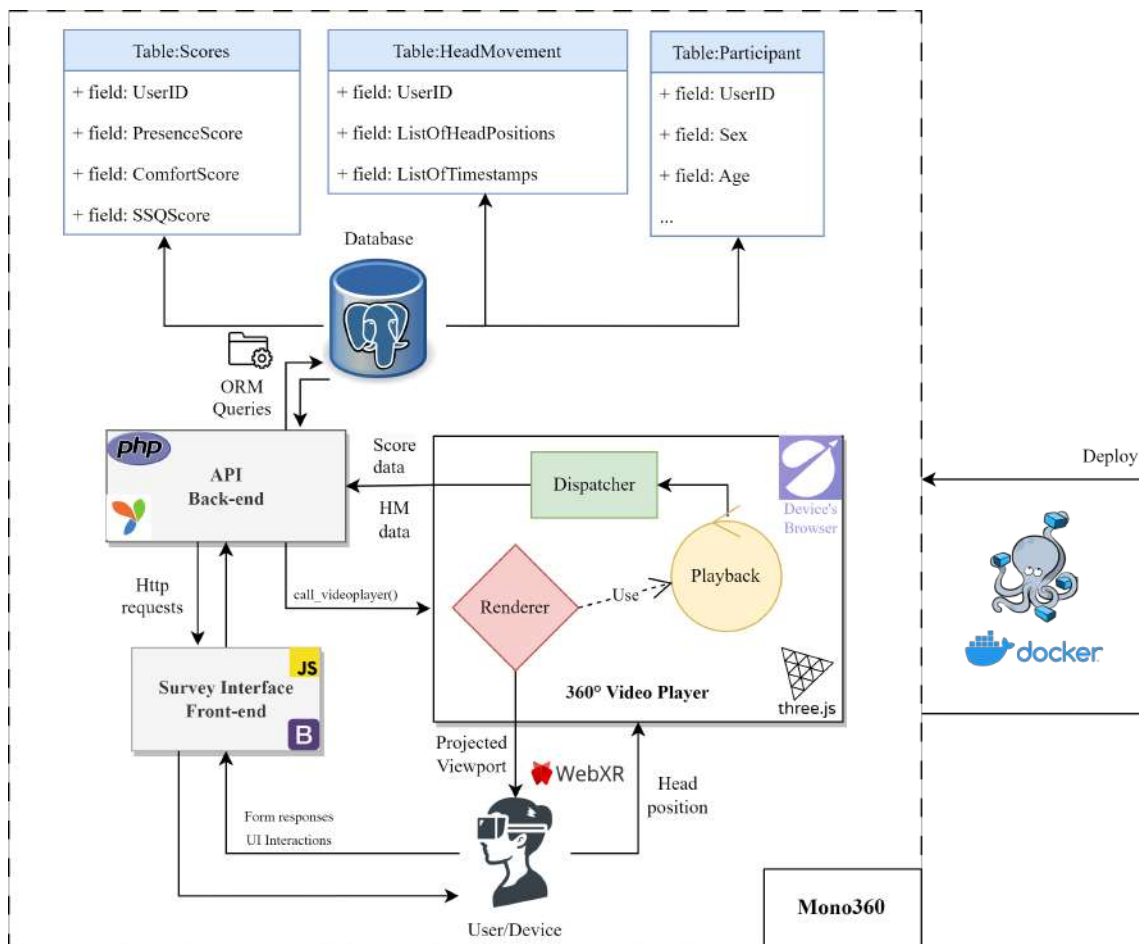


Figure 4.5: Mono360 architecture.

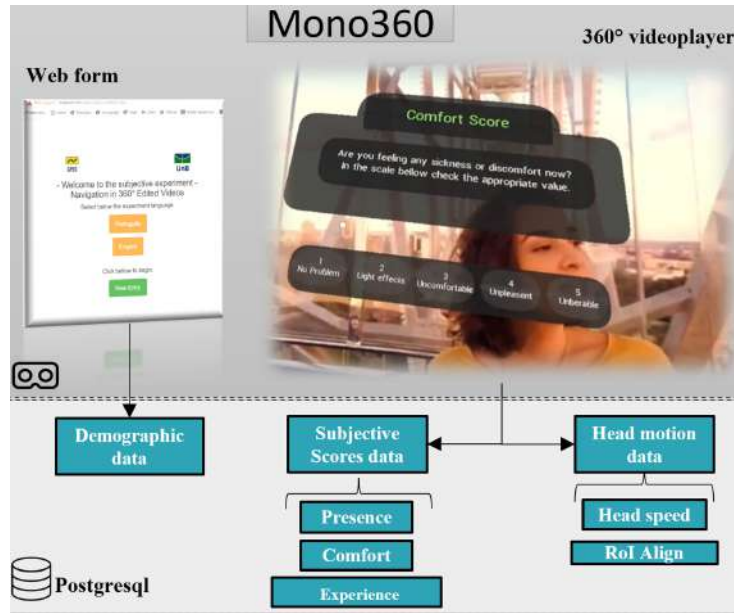


Figure 4.6: Tools for capturing and saving experimental data. Subjective rating scores are captured from an embedded user interface without removing the HMD.

with the “SphereGeometry” class from Three.js with $radius = 500$, $widthSegments = 60$, $heightSegments = 40$. The video decoding process is managed by the device’s browser, so our platform is compatible any HMD which has a browser compatible with the standard WebXR device API⁶. For the SS, we adjusted the procedures across the two different devices used (Meta Quest 2, Oculus Rift S).

The embedded 3D questionnaires are coded with Three-mehs-ui⁷ library. Figure 4.6 shows a question and scale embedded in the video player interface. Mono360 is able to gather data from three sources, the embedded 3D questionnaires where participants judge the videos, the web forms where participants fulfill personnel and the cybersickness information, and the head tracking data from the device. For more details on survey interfaces of the Mono360, please refer to the Appendix B.

We turned Mono360 available⁸ for reproducing the experiment or to reuse in other researches, it stands with Apache License, thus it is free to use with academic purposes. The parameters setup for conducting the experiments can be found in Table 4.1, this table refers to the value used for each variable as well as showing if that is a fixed variable. All fixed parameters were repeated for both subjective experiments.

⁶<https://caniuse.com/webxr>

⁷<https://github.com/felixmariotto/three-mesh-ui>

⁸<https://osf.io/kn27r/>

Parameter	Symbol	Fixed	Value
Visual sphere radius	r	✓	500
Visual sphere type		✓	“SphereGeometry”
Sphere segmentation		✓	width = 60, height = 40
Edit timestamp	t_1	✓	14 s (FR), 15 s (SC)
Angular displacement	θ_T		in Section 5.1.1, and Figure 5.17
Fade effect duration	ΔT_{fade}	✓	1 s
Edit duration	ΔT_{edit}	✓	2 s
FR rotation speed	$ \omega _{FR}$		10°/s, 20°/s, 40°/s, 60°/s
SC rotation speed	$ \omega _{SC}$	✓	> 5160° (instant)
Start direction		✓	center point ($\theta = 0, \phi = 0$)
Type of edit			“Fade-rotation,” “Snap-change”
Content			$video_1, \dots, video_{12}$
Temporal resolution		✓	60 fps
Spatial resolution			in Sections 5.1.1, and 5.3.1
Encoding codec		✓	H.264
Encoding target quality		✓	40 kbps

Table 4.1: Setup table for the QoE assessment experiments, showing the fixed parameters.

Chapter 5

QoE Assessment Experiments

In Chapter 2, we introduced the protocols for conducting subjective QoE experiments and obtaining useful information about UX in multimedia applications. In line with the goals shown in Chapter 1, the objective of the current chapter is to investigate the effects of alignment edits on user’s QoE and behavior. As aforementioned, two experiments were conducted with different methods. First, we describe the SS experiment, showing its preparation, procedure, and results. Second, we describe the DS experiment.

5.1 Single Stimulus User Study

To carry out the SS user study, we formulate four research hypotheses aiming to cover all specific goals from Chapter 1:

H1 : The degree of comfort of Fade-rotation is equivalent to that of Snap-change;

H2 : The Snap-change has a higher negative effect in presence than Fade-rotation;

H3 : The ROI alignment impacts presence, comfort, and experience scores;

H4 : Alignment edits reduce the viewer’s head movement speed after the edit.

5.1.1 Tested conditions

When selecting the experiment’s video content, we chose videos that have three types of camera motions: static, steady, and dynamic. Static refers to videos that were shot with a fixed camera, steady refers to videos where the camera is in motion for most of the scene (independent of direction), and dynamic refers to videos that contain camera acceleration and content motion [65, 12]. Figure 5.1 shows snapshots of the six videos selected for the experiment, where four videos were chosen from the datasets Directors Cut [21] and



Figure 5.1: Video-stimuli of the subjective experiment, organized by camera motion type. Top: the user FOV at the center point (initial head position). Bottom: the pre-defined target ROI.

UTD [22] (360partnership, Jet, Dance, and Cart) and the other two videos were provided by filmmakers from Caixote XR studio ¹(Amizade and Park).

The chosen set of video stimuli covers a wide range of spatial and temporal information, including outdoor and indoor content. Figure 5.2 shows the Spatial Information (SI) versus the Temporal Information (TI) for each video, computed with *siti-tools*.² These metrics indicate the amount of spatial and temporal dynamic in a video sequence.

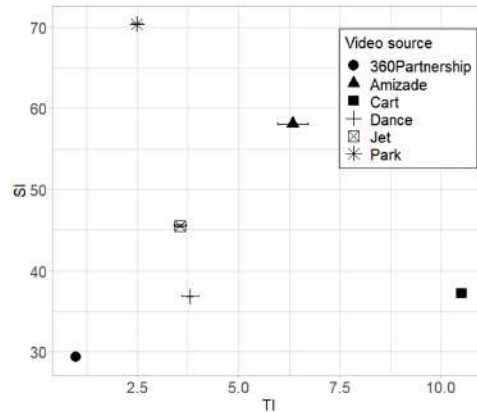


Figure 5.2: Spatial and temporal activity indexes of videos from the user study.

When selecting the content, no criteria were used for the number of ROIs in the scene, but we looked for scenes that had relevant moving objects that could capture the viewer’s attention until the intervention occurred. To select the target ROI for the alignment edit, we watched the original videos with an HMD and decided which parts of the content were

¹<https://caixotexr.com/>

²siti-tools is the official tool to compute SI and TI, conformant with ITU P.910 recommendation [93], found at: github.com/VQEG/siti-tools.

perceptually important. To prevent temporal and content bias, we avoided ROIs at the beginning of the video and made sure that transitions occurred within the same duration and started at the same video timestamp. The audio track was removed from the videos, which were encoded with the H.264 codec at 40 kbps (target quality), 60 frames per second (fps), using equirectangular projection at 3840×1920 resolution.

The implementation of the edit took into consideration the experiment parameters, described in Chapter 4. Since this is our first look into "Fade-rotation," we decided to set a secure interval for the rotation speed. Instead of conducting a separate experiment just for this purpose, we relied on the investigation of nausea scores by rotational speed from Farmani’s study [11]. Based on their findings, we set 60°/s as the maximum rotation speed, as participants withdrew from the experiment at speeds exceeding 65°/s in Farmani’s work. To keep each participant’s SS experiment under 50 minutes, we limited ourselves to four levels of rotation speed. Consequently, we chose 10°/s and 20°/s for lower speeds, and 40°/s and 60°/s for higher speeds.

A summary of the alignment edit parameters used in the user SS experiment (see Figure 4.4) is as follows:

- Snap-change (SC): $t_1 = 15s$, $\Delta T_{edit} = 0s$, $\Delta T_{fade} = 0s$, and the angular speed (performed) being equal to $\theta_T \cdot 60^\circ/s$.³
- Fade-rotation (FR): $t_1 = 14s$, $\Delta T_{edit} = 2s$, $\Delta T_{fade} = 1s$, and $\omega = 10^\circ/s, 20^\circ/s, 40^\circ/s, 60^\circ/s$.

The alignment edits were manually implemented and added to the source videos using Adobe Premiere software⁴ and the “VR projection” plugin. Each clip underwent editing operations using rotation parameters within the “VR projection” effect⁵ of Adobe Premiere Pro, as described in Chapter 4. The editing setup is illustrated in Figure 5.3.

Before editing, we first selected an “assumed viewport” and a “target ROI” for each video based on two criteria: temporal proximity to the edit and a minimum 60-degree angular displacement between them in the θ axis. This selection aimed to align with participants’ attention focus relevant to the storyline [48]. To select the assumed viewport, we watched the original videos with an HMD searching for important character’s interactions or a plot twist (scene event important to the storyline). Next, we selected the target ROI as a relevant point of view 2 seconds after the edit. We trimmed a 30s long clip from the original videos. The total alignment angle θ_T for each video is: Park = 180°, Jet = 180°, 360Partnership = 170°, Dance = 86°, Amizade = −120°, and Cart = 120°.

³To compute ω , consider that the rotation is performed in the interval between two frames. Since the video has 60 frames per second, a single frame occurs in $1/60 = 0.0167s$. For the target angular displacement θ_T , the angular speed is $\theta_T \cdot 60^\circ/s$.

⁴<https://www.adobe.com/br/products/premiere>

⁵<https://creativecloud.adobe.com/cc/learn/premiere-pro/web/vr-projection>

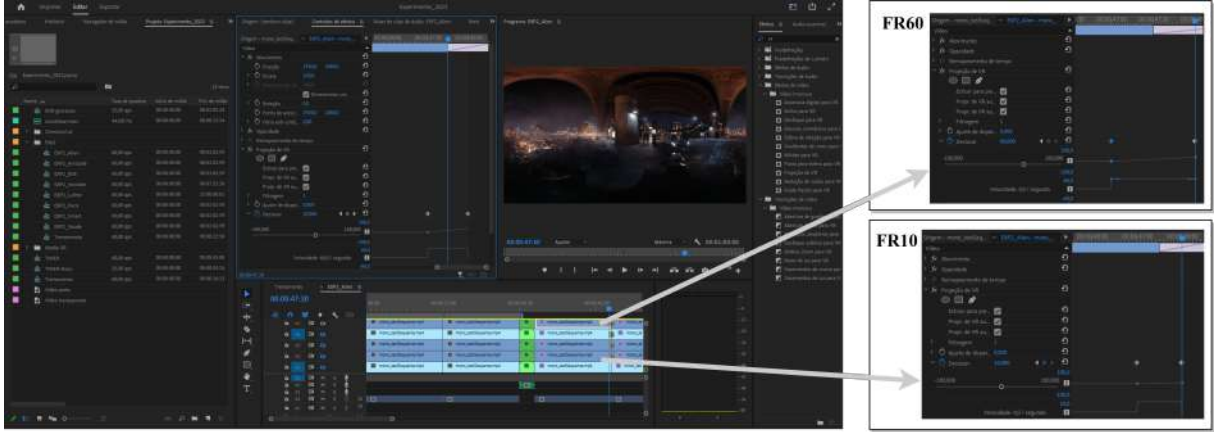


Figure 5.3: Editing setup for preparing the videos, showing the editing controls for applying the parameters for FR60 and FR10.

After trimming the clips, we applied an offset in the original video, so that the assumed viewport would be in center position and t_1 . As mentioned in Chapter 4, this was done to increase the probability of having the viewers looking at the desired frame center at t_1 . Thus, at t_1 all videos had an RoI at the center point ($\theta = 0^\circ$, $\phi = 0^\circ$). To achieve that, we adjusted the initial frame center of the two videos by applying an initial offset in the source video, namely Park (-180°) and Dance (-86°). The other four videos already had an RoI in the frame center point at t_1 . Afterwards, for FR edits, we applied the transition called “to black” centered at $t = 15s$ in the editing software. This transition reduces linearly the luminosity of the frames until it gets completely black, next it increases the luminosity until it gets the original frame luminosity.

5.1.2 Experimental procedures

We use the experimental methodology described in ITU-T Recommendation P.919 [18]. A full run of the experiment took approximately 37-40 minutes. The experiment was spread over two periods of time (sessions), with two different HMDs in each period. At first, participants used Oculus Rift S, while in the second participants used Meta Quest 2. During the test, participants were seated in a swivel chair. Participants who wore glasses or lenses kept them on throughout the session. As shown in Figure 5.4, the experiment had eight phases: (1) instructions, (2) training, (3) first session, (4) first cybersickness questionnaire (SSQ), (5) rest, (6) second session, (7) second SSQ, and (8) finalization.

In the instructions phase, participants had to select the language of the experiment (Portuguese or English), sign a consent form, read safety protocols, and complete a screening pre-questionnaire and a consent form. Figure 5.5 shows screenshots of the instructions phase, we emphasized “name” field is not obligatory. The pre-questionnaire contained de-



Figure 5.4: Procedure of the experiment, and the subject rating time structure.



Figure 5.5: Instruction phase views.

mographic and visual aptitude questions and was based on the ITU-T 919 recommendation [95]. The consent form can be found in Appendix A.1, also the sequence of the surveys interfaces can be found at Appendix B. Following the instructions, the participants participated in a training session, where they watched a training video and simulated rating the videos by interacting with the interface. The participants were given the opportunity to repeat the training until they felt confident to proceed to the experimental session.

In the first and second sessions, participants watched the 36 videos in a randomized order, giving attribute scores to each watched video. More specifically, participants watched 16 videos, completed the cybersickness questionnaire, removed the HMD, took a 5-minute break to avoid excessive cognitive load [18], and watched 20 videos. In the end, participants completed the post-questionnaire with additional questions about the experiment, such as personal insights and comments about the experiment. The implementation of the questionnaires was fully automated and no intervention from the experimenter was required. After viewing each video, participants were asked to rate attributes of the content using the device controller, by pointing a virtual raycast in the interface's buttons.

The experiment was within-subjects, which means that all participants evaluated all test conditions. We used the Absolute Category Rating with Hidden Reference (ACR-HR), which requires participants to score the processed video sequences (PVS) and the corresponding source video sequences (SRC) using a discrete degradation scale ranging from 1 to 5 [19, 20]. As the name suggests, in the ACR-HR methodology, the reference video is not identified. The participants rated three attributes of each video: overall expe-

Table 5.1: Subjective assessment measures

Questions	Scale	Attribute
Are you feeling any sickness or discomfort now? In the scale below, check the appropriate value.	(1) Unbearable (2) Unpleasant (3) Uncomfortable (4) Light effects (5) No Problem	Comfort [19]
To which extend do you feel present in the virtual environment, as if you were really there? In the scale below, check the appropriate value.	(1) Nothing (2) Little much (3) Reasonably (4) Very much (5) Entirely	Presence [20]
Evaluate the overall experience when watching the video In scale the below, check the appropriate value.	(1) Bad (2) Poor (3) Fair (4) Good (5) Excellent	Experience [18]
Evaluate the following symptoms: Nausea, Vertigo, Sweating, Stomach awareness, Increase in salivation, Difficulty in concentration.	For each symptom: (1) None (2) Slight (3) Moderate (4) Severe	Cyber-sickness [152]

rience, discomfort, and presence. Table 5.1 presents the questions and the specific scoring scales used for each of the three attributes [151, 18]. The questions for Comfort, Presence and Experience were embedded in the video player interface, as shown in Figure 4.6. For more details on survey interfaces of the Mono360, please refer to the Appendix B.

The videos were presented in random order [93] to prevent or minimize temporal bias, memory-related impacts, among other issues. However, based on Farmani *et al.* [11], who proposed a method to alleviate induced cybersickness during subjective experiments involving rotations, we refined the randomization process excluding videos with rapid “Fade-rotations” (angular speed exceeding $40^\circ/\text{s}$) from the initial set of 8 videos.

Table 5.2 presents a summary of the characteristics of the pool of participants for each experiment. We recruited 40 and 23 participants for the first and second experiments, respectively. The sampled population had a wide variety of ages and HMD experience, and the proportion of women was greater than 40% in both experiments. In total, we collected 6,804 opinion scores and 1,300-2,000 head tracking samples per video watched. We prioritized recruiting participants outside of the university to improve population

Table 5.2: Experiment population summary for both devices.

Device	Num. of Particip.	Age			Prop. of Women	1st time VR users
		Avg.	Min.	Max.		
Rift S	40	35.62	15	65	60.0%	55.0%
Quest 2	23	29.56	18	41	43.0%	60.0%
Total	63	33.40	15	65	53.8%	56.8%

sampling. The complete SS dataset containing the analysis code, the experiment videos, the QoE data, the head motion data, and the state of the mono360 at the end of the experiment is publicly available ⁶.

5.2 Results

The experiment contained six videos and six types of edits. The edits are the following:

- Fade-rotation with 4 rotation speeds: 10°/s, 20°/s, 40°/s, and 60°/s, referred as FR10, FR20, FR40, and FR60;
- Snap-change, referred as SC;
- No Edits, referred as NONE.

Therefore, each participant assessed a total of 36 videos. We gathered a total of 6,804 scores, collecting scores from 63 participants for the attributes experience, discomfort, and presence. We first examine the distribution of the subjective scores for the three attributes. Figure 5.6 shows histograms containing the distribution of the presence, experience, and comfort scores grouped by video content. Applying a Shapiro test, we confirmed that the distribution is non-normal ($P < 0.05$), this signifies the need for employing non-parametric tests in our subsequent analyses.

For each studied attribute q (comfort, presence, and experience, $1 \leq q \leq 3$) and each j -th video sequence, we compute the MOS for the pool of N participants:

$$MOS_{j,q} = \frac{1}{N} \sum_{i=1}^N x_{i,j,q}, \quad (5.1)$$

and the SOS

$$SOS_{j,q} = \sqrt{\frac{\sum_{i=1}^N (x_{i,j,q} - MOS_{j,q})^2}{(N - 1)}}, \quad (5.2)$$

⁶<https://osf.io/yftv7/>

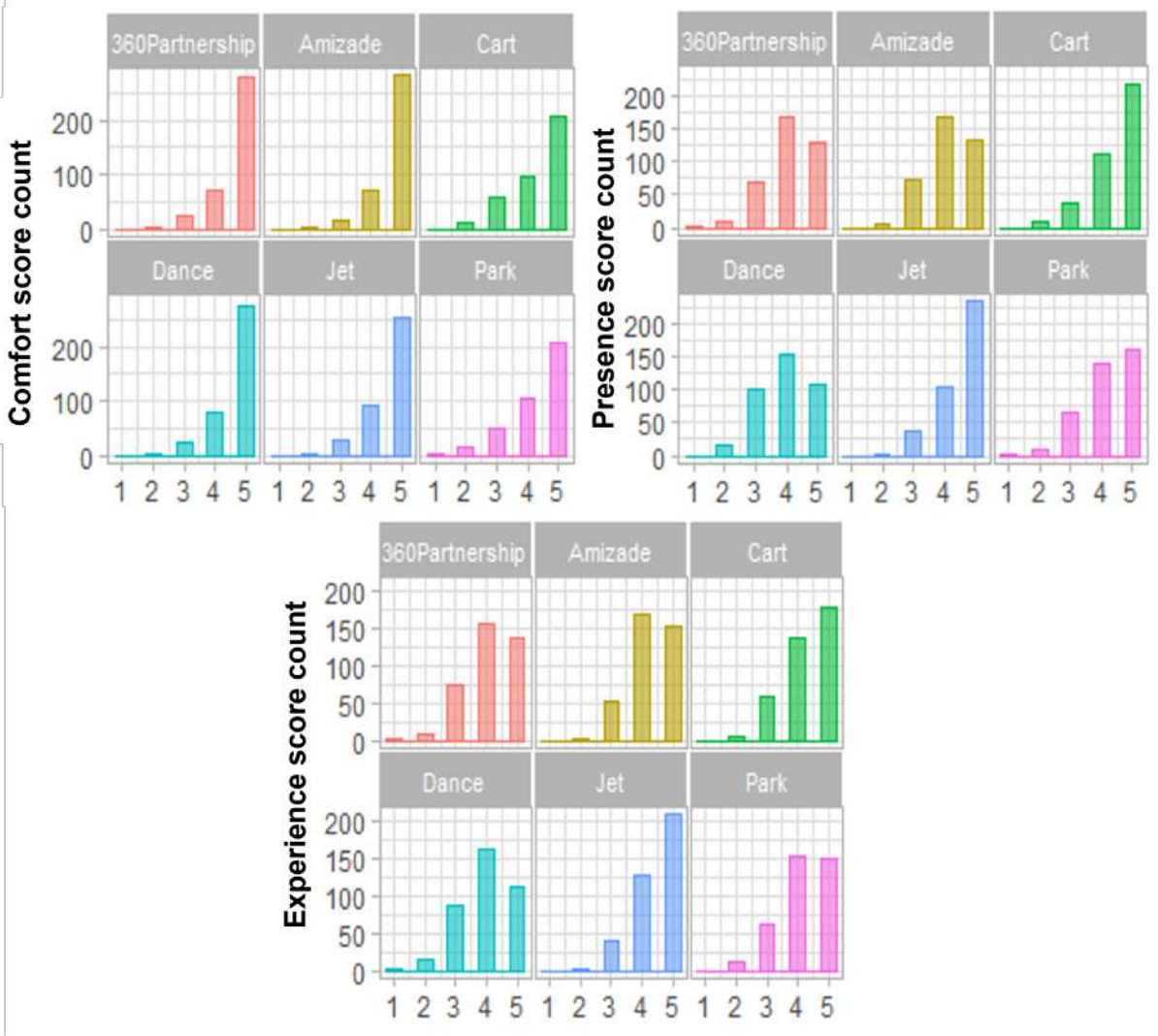


Figure 5.6: Scores for the QoE attributes (presence, comfort and experience) measured in the user study. The scores are grouped by video content. In our user study, each participant rated each video six times. Best viewed in color.

where $x_{i,j,q}$ represents the score given by the i -th participant to the q -th attribute corresponding to the j -th video sequence. For the tests, we consider a margin of error of 0.05 and a confidence interval of 95%, computed as follows:

$$CI_{j,q} = MOS_{j,q} \pm \frac{1,96 \cdot SOS_{j,q}}{\sqrt{N}}. \quad (5.3)$$

Next, we performed the Welch t-test to identify if the attribute scores of the data acquired in October and November 2021 (with different devices, the Rift and Quest 2) are statistically different. For that, we perform a pairwise comparison between the two sub-experiment groups. The Welch t-test is adopted because the samples are not balanced and the subsets are of different sizes. The test shows that for the presence scores, there are

no significant differences ($P > 0.05$) between the Rift or Quest 2 groups. For the comfort scores, when a pairwise comparison grouped by edit type was performed, a significant difference was found for the FR20 group. However, no significant differences were observed for all other cases.

Figure 5.7 shows the MOS values grouped by video for each measured attribute. Notice that comfort achieved scores greater than 4 for all content, indicating that users felt high levels of comfort for the different types of content motion, and for the several alignment edits. The highest comfort scores were for Amizade, followed by Dance, while Cart had the lowest comfort scores because it has a strong camera acceleration. In terms of the attribute presence, only Dance had scores less than 4, while the best scores were for Jet, which is the only video with presence higher than comfort. In terms of experience, the scores followed the same tendency of presence scores, where the highest score was for Jet, and the lowest score for Dance. Dance and Amizade were the only videos in which the experience scores were higher than presence scores. This similar trend in experience and presence scores is expected, since viewers expect media to make them feel immersed, and QoE itself consists of the viewers expectations [46]. This trend will be further evaluated by measuring their correlation coefficients.

We observed a relevant pattern in the data: videos characterized by minimal camera movements (Dance, 360Partnership, and Amizade) exhibited a substantial discrepancy between the “comfort” and “presence” scores (refer to Figure 5.7). In contrast, videos featuring more pronounced camera acceleration (Jet, Cart, and Park) displayed a “comfort” and “presence” difference of less than 0.12. Notably, among these, the video Cart stood out, being the sole instance where the presence value surpassed comfort. This indicates that videos with intense camera motion tend to yield lower comfort scores and higher presence scores. This observation underscores the significance of considering scene motion when incorporating alignment edits into the video.

From the feedback provided by the participants, other aspects of the content decrease the perceived presence. For example, in Dance, ten participants reported that this video lacks realism because the dancers in the video seemed out of scale, causing strangeness. This is observed in the data that show low average scores for presence. Another feedback provided by the participants is that the video content that resembled conventional 2D videos reduced their sense of presence. This was true for the videos Dance, 360Partnership, and Amizade, as mentioned by participants. For example, in “Amizade, some participants reported feeling outside of the car, while others reported that the content of 360Partnership appeared artificial because they felt smaller. These situations illustrate how content can break the feeling of ‘being there’ (presence), corroborating recent studies on realism in VR [153].

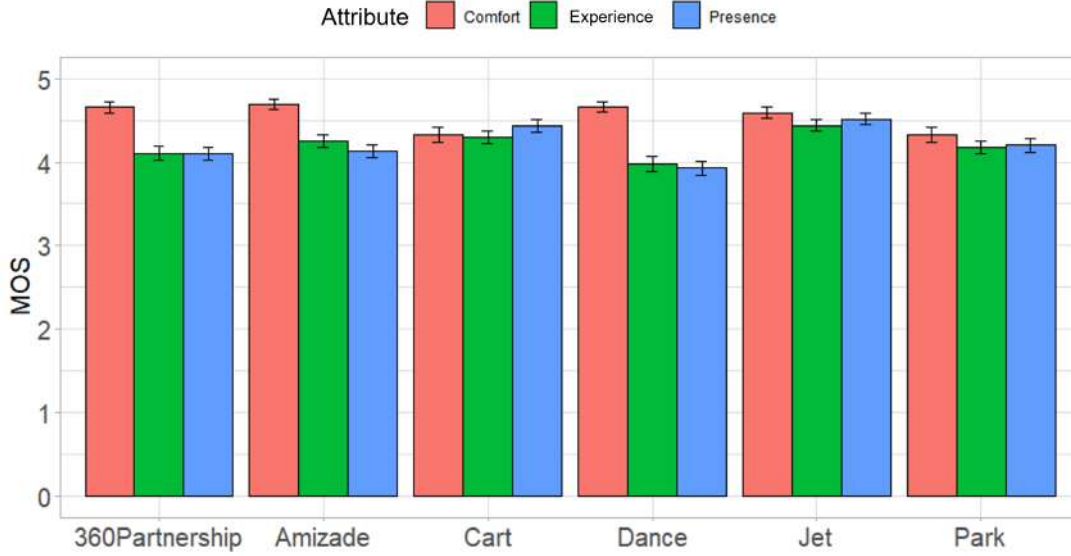


Figure 5.7: Mean opinion scores for presence, comfort, and experience for each video sequence.

Before performing the hypothesis analysis, we checked the reliability of the collected scores. For this, we computed the correlation between the various attribute scores (presence, comfort, and experience). As suggested by the ITU guidelines [18], the correlation between attribute pairs can be computed with the conventional Pearson linear correlation coefficient (PLCC):

$$PLCC(x, y) = \frac{\sum_{i=1}^N (x(i) - \bar{x})(y(i) - \bar{y})}{\sqrt{\sum_{i=1}^N (x(i) - \bar{x})^2} \sqrt{\sum_{i=1}^N (y(i) - \bar{y})^2}}, \quad (5.4)$$

and the Spearman rank correlation coefficient (SRCC):

$$SRCC(x, y) = 1 - \frac{6 \sum_{i=1}^N (R_x(i) - R_y(i))^2}{N(N^2 - 1)}, \quad (5.5)$$

where x and y are vectors of length N that represent the two variables being compared, \bar{x} and \bar{y} are the mean values of x and y , respectively, and $R_x(i)$ is the rank of the i -th value of x , and $R_y(i)$ represents the rank of the i -th value of y . To interpret the correlation values, we follow the convention of Schober *et al.* [154], where values below 0.1 are considered negligible, values between 0.1 and 0.69 are considered moderate, values between 0.7 and 0.89 are considered strong, and values over 0.9 are considered very strong.

Table 5.3 shows the pairwise correlation comparison of attribute scores under the same

Table 5.3: Correlation between QoE attributes, with data aggregated by Edit type. In bold we highlight the moderate or strong correlations.

Edit type	Comparison	PLCC	p-val	SRCC	p-val
Gradual FR10	comfort/presence	0.021	0.681	0.021	0.688
	comfort/experience	0.136	0.01	0.143	0.01
	presence/experience	0.713	0.001	0.694	0.001
Gradual FR20	comfort/presence	0.191	0.001	0.142	0.001
	comfort/experience	0.339	0.001	0.319	0.001
	presence/experience	0.720	0.001	0.709	0.001
Gradual FR40	comfort/presence	0.205	0.059	0.175	0.001
	comfort/experience	0.396	0.001	0.360	0.001
	presence/experience	0.732	0.001	0.736	0.001
Gradual FR60	comfort/presence	0.173	0.001	0.132	0.01
	comfort/experience	0.363	0.001	0.328	0.001
	presence/experience	0.716	0.001	0.710	0.001
SC	comfort/presence	0.045	0.385	0.028	0.591
	comfort/experience	0.150	0.01	0.157	0.01
	presence/experience	0.674	0.001	0.658	0.001
NONE	comfort/presence	0.001	0.988	0.007	0.887
	comfort/experience	0.236	0.01	0.225	0.01
	presence/experience	0.672	0.001	0.675	0.001

edit conditions. A negligible correlation was found between presence and comfort scores for three edit types (FR10, SC, and NONE) and a weak correlation ($CC < .2$) for FR20, FR40, FR60. This shows that the participants were able to distinguish presence from comfort. A weak correlation was found between comfort and experience, and a strong correlation between presence and experience for all edit types. This result appears to be due to ambiguities in the definition of the overall experience for immersive experiences [18].

Figure 5.8 shows MOS values for different edit types grouped by video content. We notice that the comfort MOS is higher than 4 for all cases, while the comfort MOS for dynamic motions tends to be lower than for the other scene motions. We used the Kruskal-Wallis (KW)⁷ test to determine whether there are significant differences between two or more independent groups, verifying the effect of video-content on comfort and presence.

⁷All statistical the statistical analysis was conducted using built-in packages from R - <https://www.R-project.org/>

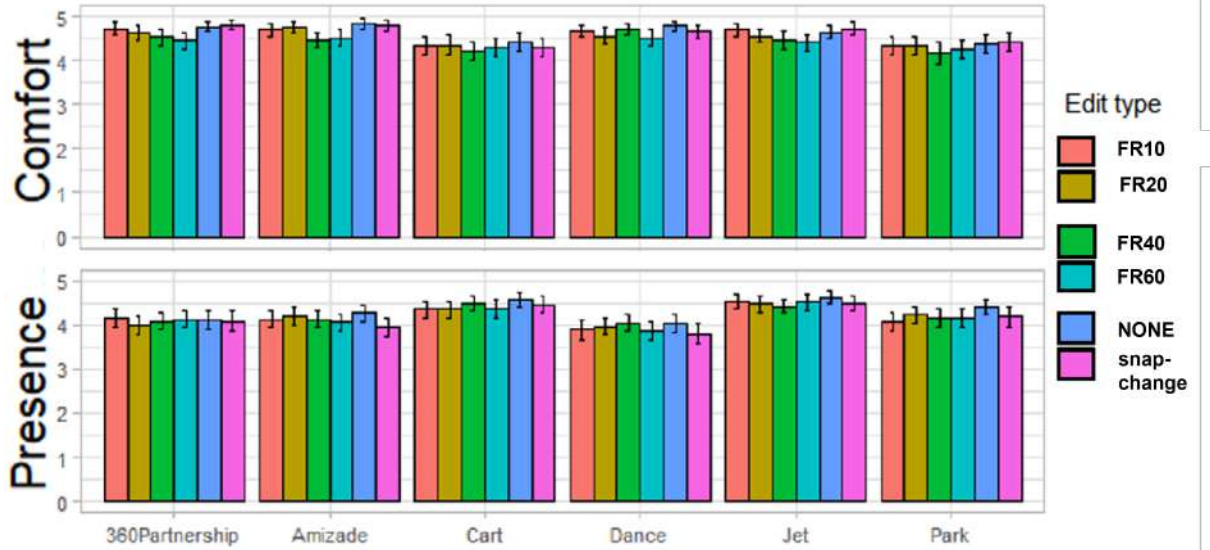


Figure 5.8: Presence and comfort MOS barplots, grouped by edit type and video-content.

We found a statistically significant effect of content on presence ($\chi^2 = 20.376, df = 4, p < 0.001$) and comfort ($\chi^2 = 31.423, df = 4, p < 0.001$).

Figure 5.9 illustrates the MOS for different edit types grouped by scene motion. Specifically, the comfort MOS exhibits a discernible decline, correlating with the rotational speed of gradual edits. To analyze this trend, we grouped the scores by edit types and performed a KW test to examine the relationship between comfort scores and rotation speed values. The results show a significant impact of the rotation speed on comfort ($\chi^2 = 12.511, df = 3, p < 0.01$). In contrast, the effect on presence was found to be non-significant ($\chi^2 = 0.236, df = 3, p > 0.05$), implying that the type of edit does not impact presence.

Finally, we performed a multiple pairwise comparison for all groups, the post-hoc test for Kruskal-Wallis, to analyze whether the attribute scores given to a pair of videos are statistically different. The results of this test are shown in Table 5.4. Note that there is no significant statistical difference of comfort for video-pairs with the same motion type, such as Dance/360Partnership, Amizade/Jet, and Cart/Park. This suggests that the camera’s dynamic categorization (static, steady, dynamic) accurately classified the content, at least in terms of its impact on comfort. In terms of presence, the videos with no significant difference are 360Partnership/Amizade/Park/Dance, and separately Jet/Cart. In terms of genre, Jet and Cart are action videos.

5.2.1 Opinion score analysis

To test if “the degree of comfort of Fade-rotation is equivalent to that of Snap-change” (hypothesis **H1**), we use the comfort scores shown in Figure 5.6, grouping them by edit

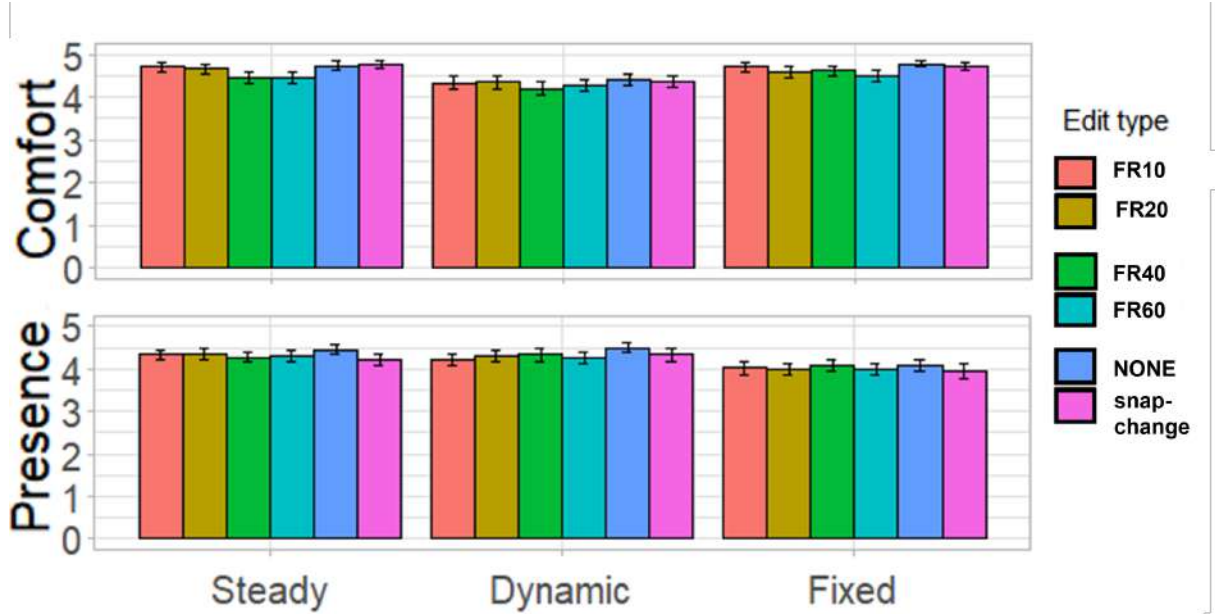


Figure 5.9: Presence and comfort MOS barplots, grouped by edit type and scene motion.

type. We test the statistical difference between two sets of comfort scores, comparing Snap-change with each Fade-rotation type. For this analysis, we used Welch’s t-test with FDR correction. We found a significant difference ($p < 0.05$) between Snap-change and FR40, and between Snap-change and FR60.

Complementing these results with a pairwise test, we also consider grouping the data in terms of the scene motion category. First, for dynamic scene motion, all comparisons had significant differences. Second, for the steady scene motion, a significant difference was found in the instant-FR40 and instant-FR60 pairs. Third, for fixed scene motion, a significant difference was found in the instant-FR20, instant-FR40, and instant-FR60 pairs.

From all the comparison results presented in the last two paragraphs, **H1** is accepted for FR10 in fixed-scene motion videos and for FR10 and FR20 in steady motion videos. However, we reject **H1** for any video content with dynamic scene movement, and for FR with angular speed greater than $40^\circ/\text{s}$. In practical terms, for video players that lack the ability to account for scene motion in playback time, we recommend steering clear of FR20, FR40, and FR60, as they carry a higher likelihood of causing viewer discomfort. Instead, opting for FR10, or the Snap-change approach, is preferable, as they exhibit a lower probability of discomfort-inducing effects. For videos characterized by steady camera motion, we suggest employing Fade-rotation edits with an angular speed of less than $20^\circ/\text{s}$, as this can enhance the viewer’s experience while minimizing the risk of discomfort. In essence, these findings underline the importance of selecting an appropriate FR strategy, taking into account camera motion, to optimize the viewer’s experience and

Table 5.4: Paired Kruskal-Wallis test with FDR adjusted p-values for presence and comfort scores.

Camera Dynamic	Video	Comparison Video	p-value	
			Comf.	Pres.
Fixed	Dance	360Partnership	1	0.748
		Amizade	1	0.711
		Jet	1	< 0.001
	360Partnership	Amizade	0.184	0.914
		Jet	0.817	< 0.001
Steady	Amizade	Jet	0.108	< 0.001
		Cart	< 0.001	< 0.001
		Park	< 0.001	0.712
	Jet	Cart	< 0.05	0.968
		Park	< 0.01	< 0.01
Dynamic	Cart	Park	0.667	< 0.01
		Dance	< 0.01	< 0.001
		360Partnership	< 0.01	< 0.001
	Park	Dance	< 0.001	0.377
		360Partnership	< 0.01	0.677

comfort.

Next, we investigate the Fade-rotation scores relative to two baselines: the original version of the videos and the Snap-change version. Figure 5.10 shows the scores for the four types of “Fade-rotation” for each video-content, with the baselines shown as straight lines. This graph provides a visual comparison of multiple conditions. In terms of comfort, we see that FR10 had no significant difference ($p > 0.01$) compared to Snap-change, for any video content. Furthermore, no significant differences were observed between Fade-rotation and Snap-change for Cart and Park.

Snap-change had the worst comfort scores for videos Cart and Park. As the Cart scene takes place, the viewer becomes a participant in a chariot race, while echoes of cheers reverberate as the race unfolds inside a coliseum. In the case of Cart, instant edit was uncomfortable because it was combined with a strong camera translation when the chariot was turning. In the case of Park, the viewer shares a Ferris wheel cabin with a young woman. As the cabin ascends, the edit takes place. Comfort tends to decrease with the rotation speed for all video-content; however, specific conditions can break this trend. For example, for the video Dance we expected a decrease in comfort. But, surprisingly, there is a peak for the FR40 edit, showing that there is a non-trivial relationship between the rotation speed of the FR and the content; other similar cases happened for Cart, Park,

and Amizade in FR60.

In terms of presence for the instant edit, we notice a relatively low average score for Dance, 360Partnership, and Amizade. These video-content had the lowest scene motion. Feedback from the participants indicated that when watching the Dance video, the instant alignment edit interrupted the change between the dancing groups, which caused the loss of the sense of presence. Dance and Amizade have fixed cameras and indoor scenes. It is not clear what attributes lead to a higher sense of presence; however, from the presence MOS values, we observe that videos Cart, Jet, and Park engaged them. It seems that interactions of the characters are not enough to promote a high sense of presence, given that for Dance, and 360Partnership there were people interacting with the camera and performing actions. However, they had a fixed camera and resulted in the lowest presence scores.

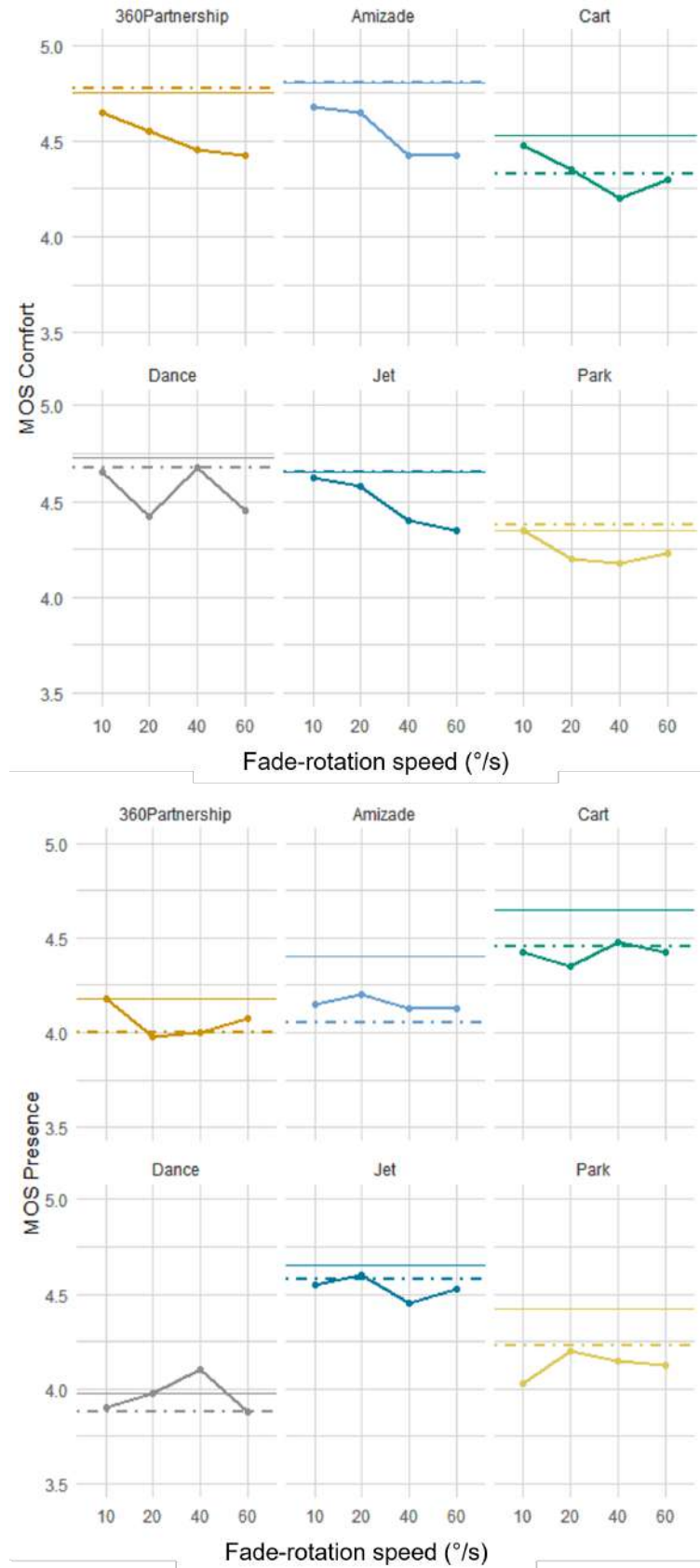


Figure 5.10: The Mean Opinion Score (MOS) of presence and comfort for each Fade-rotation (FR) rotation speed tested in the study. Two baseline conditions are depicted: snap-change (dashed line) and no edit (solid line) for each video.

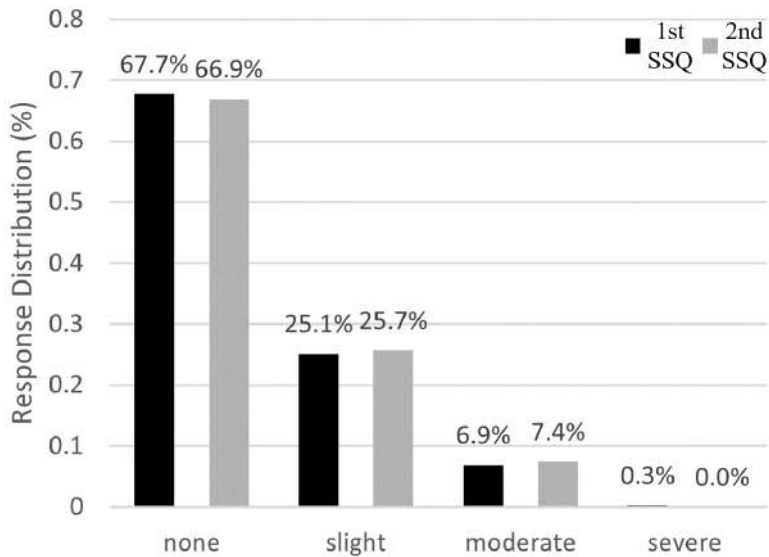


Figure 5.11: Distribution of all cybersickness symptoms.

To test hypothesis **H2**, we group the presence scores by edit type and apply the Welch t-test for all pairs. We did not find statistically significant differences ($p < 0.05$). Therefore, we rejected the hypothesis **H2**, confirming that the Snap-change and Fade-rotation did not have a distinguishable effect on presence.

As mentioned above, the cybersickness questionnaire consisted of four possible levels of symptoms. Participants filled out the questionnaire after the first and second video sessions. Figure 5.11 shows the frequency of these 4 levels of symptoms for these two instants. Note that the responses are similar results for the pre and post questionnaires, with more than 90% of the participants reporting none to a slight discomfort. Only one participant reported severe symptoms caused by the Jet video. This participant mentioned that he/she had height phobia. These individual conditions are known to cause differences in comfort and tendency to trigger cybersickness in VR [23].

5.2.2 Head motion analysis

The head motion analysis is performed using the head tracking data and two distance metrics. The two distance metrics are: *i*) the distance between the gaze position and a target in the video content, and *ii*) the distance between two head tracking samples. As discussed in Section 5.1.2, gaze positions are recorded using normalized screen coordinates (X, Y) , with the origin in the upper left corner of the 360° frame, spanning the interval $X, Y \in [0, 1]$. To convert the stored gaze position to the reference coordinate system (see Figure 4.2), we convert the normalized screen coordinates to Eulerian coordinates (ϕ, θ) by rescaling them to the appropriate intervals: $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $\theta \in [-\pi, \pi]$. With the

rescaling procedure, the center point of the 360° frame matches the reference coordinate system, as shown in Figure 4.2.

The collected head-tracking data consists of the intersection points between the HMD’s gaze direction and the spherical shell defined by the render sphere. To compute the spherical distance, we use the orthodromic distance metric, which is given by:

$$d(u, u') = 2R \cdot \arcsin \left[\frac{c(u, u')}{2R} \right], \quad (5.6)$$

where $c(u, u')$ is the Euclidean distance between two points on the spherical surface u, u' , given by:

$$c(u, u') = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}.$$

Rondon *et al.* [155] found that the orthodromic metric is the most suitable distance metric for spherical surfaces. It can handle the periodicity of the latitude, while fitting the spherical geometry distance problem more accurately. Furthermore, Rossi *et al.* [156] showed that the orthodromic distance is a reliable proxy of the viewport overlap. To appropriately compute the orthodromic distance, we convert the gaze positions to 3D Cartesian points of the spherical surface. Thus, after this transformation, each data point has the form $u = (x, y, z, t)$, where t is its time coordinate.

Figure 5.12 depicts the empirical Cumulative Distribution Function (CDF) representing the average head speed for each video content. The CDF is based on the accumulation of the counts of a head speed value from each sample. In our analysis, we pinpointed outliers characterized by exceptionally high head movement speeds. By examining the CDF, we ascertain that a suitable threshold for filtering out these outliers is 150°/s. This value effectively encompasses the majority of the typical head speeds recorded. Note that these outliers are rare and typically arise from inaccuracies in the head-tracking system. The HMD’s tracking system is equipped with Micro Electro-Mechanical System (MEMS) sensors for orientation data collection. Although head tracking reliability has seen significant improvements in the last decade, certain issues such as drift, tilt, and stationary jitter can still affect data quality [157]. We established a head-speed threshold of 150°/s and excluded data points above this threshold, which allowed us to keep more than 99.9% of the dataset. Figure 5.13 shows the mean head speed for fixed-motion videos after removal of the outliers, considering the head tracking data for the entire video, most of the mean head speed are between 60 and 80°/s, and the Amizade had the less varied speeds, indicating that viewers were more focused in this video.

To analyze the head tracking data, we calculate the distance between the gaze direction and the ROI at any given time t . For each experiment trial, defined by the i th participant

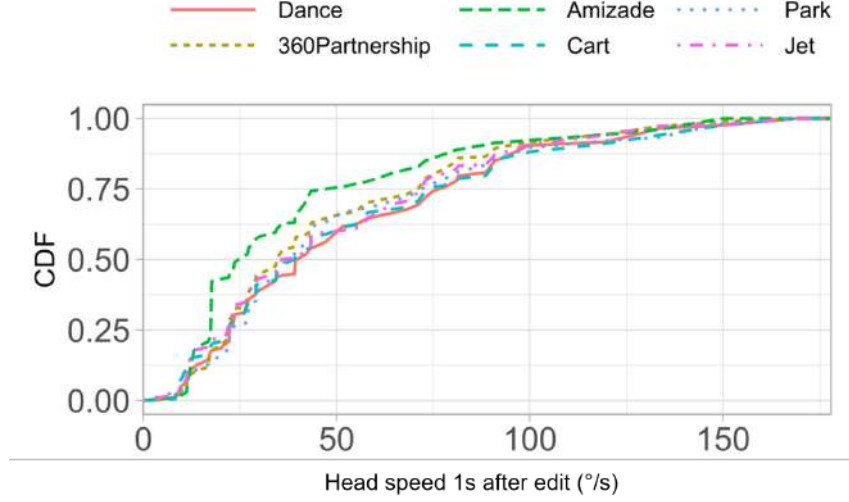


Figure 5.12: CDF of the head speed measured 1s after the edit for each video-content.

and the j th video, we collect a matrix U_{ij} of gaze positions that is expressed as follows:

$$U_{ij} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_{N_{ij}}], \quad (5.7)$$

where i corresponds to the i th participant, j to the j th video, and N_{ij} to the total number of gaze positions for a single trial of the experiment. We can define each collected gaze position \mathbf{u}_k as a 4D vector, represented as \mathbf{u}_k , encompassing the 3D k -th spatial coordinates (x_k, y_k, z_k) and the temporal component (t_k) of the sample:

$$\mathbf{u}_k = \begin{bmatrix} x_k \\ y_k \\ z_k \\ t_k \end{bmatrix}.$$

To execute the analysis, we need not only the gaze positions but also the ROI positions for each time t . Thus, similarly, we generate a matrix (V_{ij}) of ROI positions \mathbf{v}_k , containing the same number of samples as the user gaze matrix (U_{ij}) .

Now, let $\mathcal{T}_{ij} = [\tau_1, \tau_2, \dots, \tau_{N_{ij}}]$ be an array of time samples from the gaze position data collected in each trial. We define the observation window around a time τ as

$$[\tau_0, \tau_f] = \left[\tau - \frac{\Delta\tau}{2}, \tau + \frac{\Delta\tau}{2} \right],$$

where the time window starts at $k_0 = \text{closest}(\mathcal{T}_{ij}, \tau_0)$ and ends at $k_f = \text{closest}(\mathcal{T}_{ij}, \tau_f)$. $\Delta\tau$ corresponds to the time interval around τ to be used in the analysis. The function

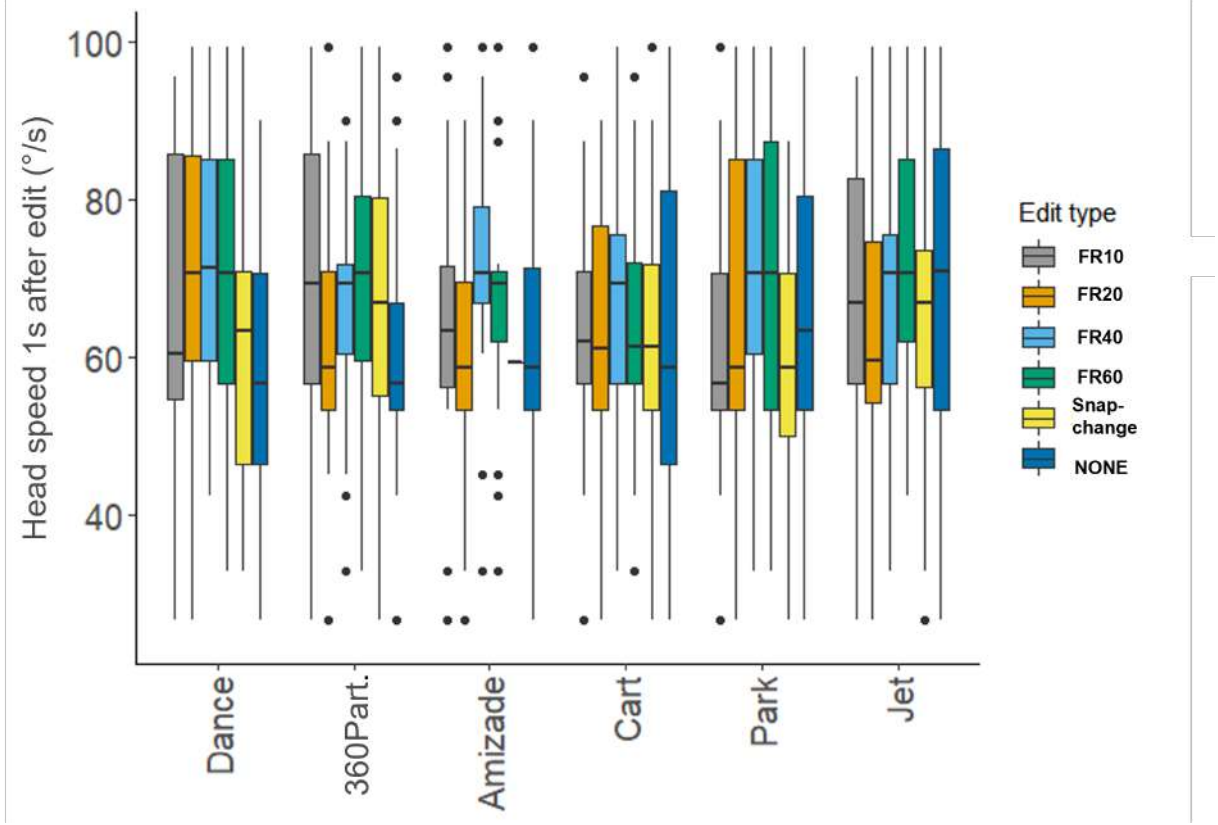


Figure 5.13: Boxplot of head speeds measured 1s after the edit, for each video-content grouped by edit type.

`closest()` returns the element x_i that minimizes $|x_i - y|$:

$$\text{closest}(x, y) = \text{argmin}_i(|x_i - y|).$$

For a single trial, defined by the i th participant and the j th video, the distance between the gaze of the user (U) and the ROI (V) at time t is given by the mean orthodromic distances around the observation window:

$$\bar{d}(U, V) = \frac{1}{N} \sum_{k=k_0}^{k_f} d(u_k, v_k). \quad (5.8)$$

From the gaze-ROI distance, we analyze the research hypothesis H3 and H4. To perform statistical t-tests, we group the experimental trials according to gaze-ROI alignment. To this end, we propose an alignment function A_{ij} that attributes one of two states, “aligned” or “non-aligned,” to each trial according to the following equation:

$$A_{ij} = \begin{cases} 1, & \text{if } \bar{d}(U_{ij}, V_{ij}) < \delta; \\ 0, & \text{otherwise;} \end{cases} \quad (5.9)$$

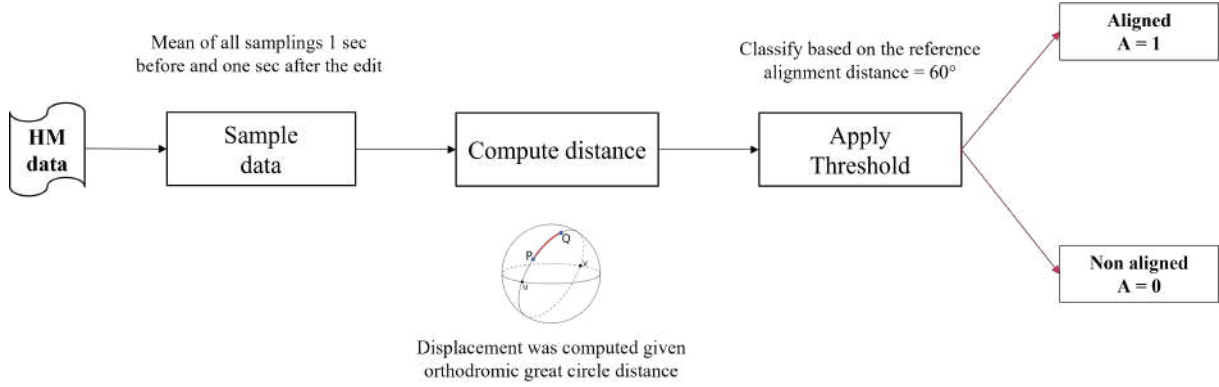


Figure 5.14: Data transformation pipeline for the alignment state (A) computation.

where δ is a threshold value for the maximum distance between gaze and ROI. This value corresponds to a radius around the point of perfect intersection into which we consider gaze and ROI aligned. Figure 5.15 illustrates these two alignment states. The data transformation pipeline for the head motion data to compute the alignment states A are illustrated in Figure 5.14.

To classify each trial in terms of alignment, we calculate the alignment just after the edit ($t = 15s$ for Snap-change, $t = 16s$ for Fade-rotation). As shown in Figure 5.15, for all videos rotations, if the participant was looking at the center point ($\theta = 0, \phi = 0$) at time t , she/he would be perfectly “aligned” with the target ROI at the end of the rotation. We classify each trial by computing $\bar{d}(U, V)$ at t . We fixed $\Delta t = 250ms$ (equivalent to approximately 30 samples for the typical data sample frequency) and the tolerance region $\tau = 60^\circ$. We chose this tolerance region because both devices used in the experiment have more than 90° FoV. Therefore, if the participant’s gaze direction is within 60° , the ROI will be within the FoV [156, 158, 159].

To analyze the effects of alignment on subjective scores, we consider the alignment state A (see Figure 5.15), which can be “aligned” or “non-aligned”, depending on whether the ROI was within an angular distance of 60° or not. Then we group the “aligned” or “non-aligned” cases per edit type, resulting in 2 unbalanced sets per edit type. For each condition, we perform Wilcoxon rank sum tests (with continuity correction) to analyze the differences between “aligned” and the “non-aligned” sets. There are 15 conditions, resulting from 5 types of edit (NONE not considered) and 3 attributes (presence, comfort, experience). Thus, we applied the t-test between two sets “aligned” and “non-aligned” for each condition, and for each attribute. The only condition where the pair of sets “aligned” and “non-aligned” ($p < 0.05$) had a significant difference between them was FR10 in the experience attribute. Therefore, except for the FR10 experience score, the gaze-ROI alignment classification had no impact on subjective scores, partially fulfilling **H3**.

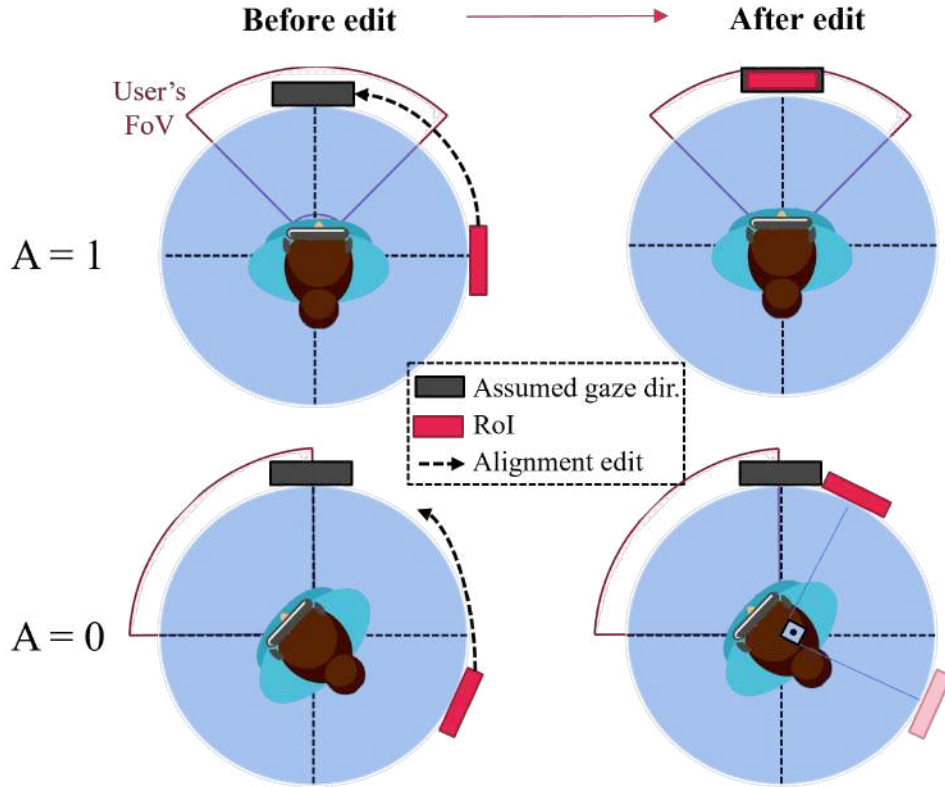


Figure 5.15: Possible states of alignment: $A = 1$ (first row) when alignment is successful, and $A = 0$ (second row) otherwise. The mean distance between user FOV and ROI just after edit is used to compute the A . We applied a distance threshold of $\tau < 60^\circ$ to classify each trial in terms of A .

To complete the **H3** analysis, we performed a Tukey HSD post hoc test on all given combinations of A (“aligned” represented by $A = 1$ and “non-aligned” represented by $A = 0$) and contents (21 comparisons), of A and edit type (15 comparisons), as well as of A and scene motion type (6 comparisons). In total, we performed 42 non-significant comparisons. No statistically distinguishable differences were found between the group “non-aligned” and the group “aligned”. With that, we fulfill **H3**.

We tested the effect of A on the reduction in head motion. For that, first we computed the head movement speed of users at 1 second before and 1 second after the edit. The head movement speed for the i -th participant watching the j -th video at time t is given by:

$$\bar{s} = \frac{1}{N} \sum_{k=k_0}^{k_f} \frac{d(u_k, u_{k+1})}{t_{k+1} - t_k}, \quad (5.10)$$

where $T_{ijk} = \{t_{ij1}, \dots, t_{ijk}, \dots, t_{ijN_{ij}}\}$ are the timestamps inside the temporal window Δt around t , $N_{ij}(t)$ is number of samples inside Δt , and d is the orthodromic distance metric (see Eq. (5.6)).

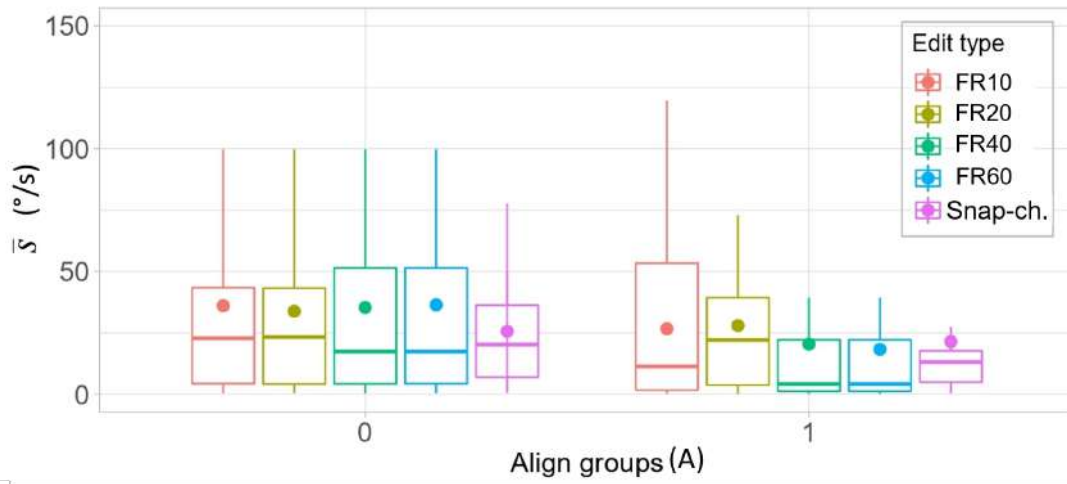


Figure 5.16: Boxplots of the participants head speed 1 s after edit for those in the “aligned” ($A = 1$) and “non-aligned” ($A = 0$) groups. The circles shows the mean values.

Since head movement speed data are continuous, we perform an Anova Omnibus test between the “aligned” ($A = 1$) and “non-aligned” ($A = 0$) groups. The Anova Omnibus test returned $F(30.95, 1, p < 0.01)$, meaning an F-test with 30.95 degree of freedom and a p-value lower than that of the significance level, which can be interpreted as a significant reduction in head speed after alignment edit. Therefore, aligning with ROI just before editing reduces the speed of head movement, allowing viewers to stabilize their view in the region. Figure 5.16 shows the boxplot of the head speed distribution at 1 second after editing, for the two A groups, grouped by edit type. For the aligned group, FR40 has the lowest values (8% lower than Snap-change), followed by FR60. The edit types that show a reduction in the average head movement speed are: FR10 = 14.9° , FR20 = 9.5° , FR40 = 26.7° , FR60 = 33.1° , Snap-change = 21.5° . For all edit types, there is a reduction in head movement speed that may be related to a fixation on an ROI, which reduces exploratory behavior in agreement with the literature [10]. With these results, we prove **H4**, which states that alignment edits reduce head movement speed.

5.3 Double Stimulus User Study

As aforementioned in Chapter 1, the DS methods are especially recommended for experiments where the scale is not completely covered. Given our focus on evaluating the sense of presence and comfort attributes, we aim to avoid conditions with excessively low comfort levels. Thus, to improve the robustness of our findings from the initial SS user studies (Section 5.1), we have chosen to conduct a DS experiment, maintaining alignment edits

	Alien	Amizade2	BSB	Paris	Smart	Vaude
Assumed Viewport						
(Position in SRC)	(theta = 80°, phi = 4°)	(theta = -34°, phi = 0°)	(theta = 0°, phi = 0°)	(theta = 80°, phi = 0°)	(theta = -43°, phi = -3°)	(theta = -10°, phi = -11°)
Target RoI (position)						
(Position in SRC)	(theta = -126°, phi = -17°)	(theta = 38°, phi = 0°)	(theta = -77°, phi = 0°)	(theta = 0°, phi = 0°)	(theta = 42°, phi = 0°)	(theta = -69°, phi = 0°)
Assumed viewport action	Lady searching	Showing a picture	Conversational lady	Tour guide walking away	Lively talking driver	Conversational lady
Target ROI action	Alien attack	Taking a selfie in a moving car	Lady approximating with a headphone	View of Tour Eiffel	Street band playing	Lady driving a bicycle

Figure 5.17: Illustration of the video content utilized in the experiment, featuring the identification of the assumed viewport and the designated target RoI for each video.

with parameters described in Chapter 4 and incorporating a new set of video content, covering additional conditions.

5.3.1 Content preparation

For the DS experiment, we selected video content prioritizing clips with engaged characters in close proximity to the edit point, aiming to enhance participants’ attention to the assumed initial viewport just before the edit initiation. The chosen character-camera interaction was intended to augment the perception of a narrative storyline within the limited clip duration. Snapshots of the selected videos, including Vaude, Smart, Alien, and Cart from the Directors Cut dataset [21], Amizade2 and BSB from Caixote XR studio ⁸, and Paris from Corbillon’s [160] dataset, are shown in Figure 5.17.

The source videos underwent the same edit operations from the SS experiment, as described in the Section 5.1.1. As like in the SS experiment, assumed viewport and “target RoI” for each video were selected based on two criteria: proximity to the edit and a minimum 60-degree difference in the θ axis. This selection aimed to align with participants’ attention focus relevant to the storyline [48], as shown in Figure 5.17.

With the exception of the Alien and Paris videos, all other clips feature character interactions with the camera precisely at the moment when the edit is triggered. For example, in the Amizade2 video, the characters engage in conversation with the camera and showcase a picture, while in Smart, the driver smiles and communicates directly

⁸<https://caixotexr.com/>

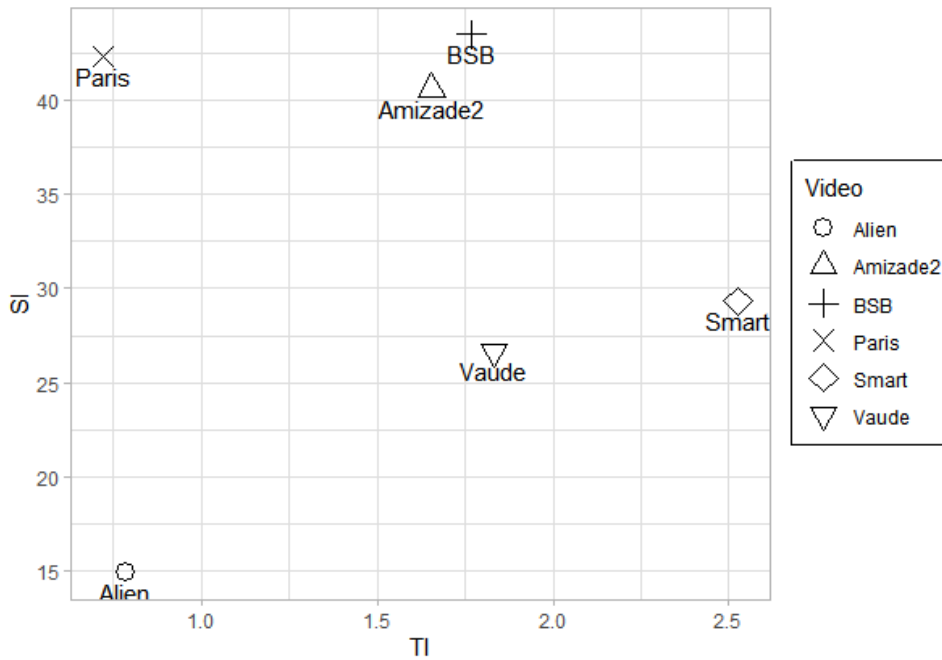


Figure 5.18: SI and TI indexes of the original videos from DS experiment.

with the camera. In Alien, the primary character (a blond lady) is positioned in the opposite direction of the alien, which unexpectedly attacks the lady. In this scenario, the assumed viewport is the lady, and the target ROI is the alien. The edit occurs just moments before the attack, suggesting that the alignment edit should prove beneficial in preventing participants from missing the pivotal plot development. In the Paris video, the narrator moves away from the camera during a scene transition in the original video. However, observers focusing on the narrator do not face the Tour Eiffel. In this case, the assumed viewport is the narrator, and the target RoI is the Eiffel Tower.

The experiment evaluates five alignment edits, including four Fade-rotation variations (FR10, FR20, FR40, FR60) and the SC, mirroring the SS experiment, for details refer to Section 5.1.1. Figure 5.18 displays the Spatial Information (SI) and Temporal Information (TI) indexes for the original videos, calculated using the recommended ITU algorithm ⁹.

In this experiment, we adopted the same video encoding approach as employed in the SS experiment, the processed videos were encoded using the H.264 codec at 40 kbps (target quality), 60 frames per second (fps), and employed equirectangular projection. In this DS experiment, we retained the audio track from the videos. This decision stems from our interest in evaluating the impacts of alignment edits on a viewing experience of higher fidelity, once the DS experiment goal is to complement the SS, with this decision, we will confirm whether the conclusions from SS will hold for higher fidelity use case. The

⁹<https://github.com/VQEG/siti-tools>

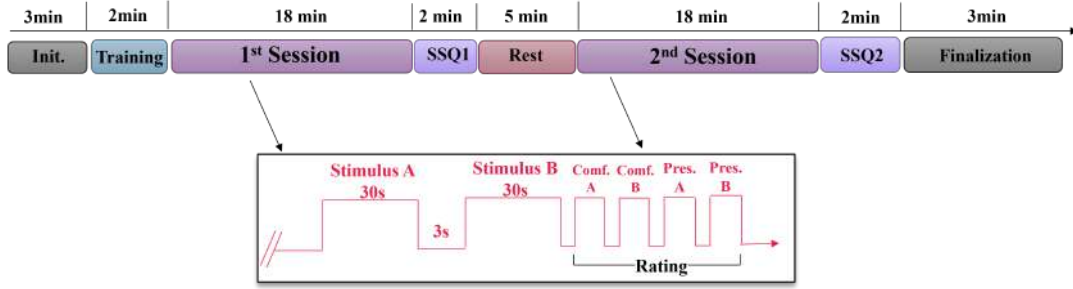


Figure 5.19: Procedure of the experiment, and the assessment methodology of the applied DS method for comfort and presence QoE attributes.

video resolutions were the same as the source video: Paris, Amizade2, and Vaude are at 3840×1920 , Alien is at 3412×1920 , BSB is at 3448×1920 , and Smart is at 2880×1440 .

5.3.2 Procedures

The main DS methods are the Degradation Category Rating (DCR) and the Double Stimulus Continuous Quality Scale (DSCQS). In DCR method, the reference is presented first to participants, this is done to anchor the ratings with the reference signal, which is a feature we need in our method. However, the DCR method were designed to measure the quality degradation. Thus, we adjusted the DCR scale to make it possible to assess alignment edit that either enhances or degrades the participant's QoE attributes. Specifically, in this experiment we aim to evaluate three QoE attributes- comfort [19], sense of presence [20], and cybersickness [152]. Finally, to enable comparisons, we apply the same scales used in the SS experiment, which is shown in experiment Table 5.1.

Figure 5.19 illustrates the procedure and the assessment method of our experiment. The DS method we applied the following steps, participants watched the SRC (A) followed by the Processed Video Sequence (PVS) (B). After watching video B, they assess both videos, first evaluating the video A in terms of comfort and presence, then evaluating video B, following the same order of attributes to avoid misinterpretation [36]. Participants were aware of that order, like in the DCR method. After watching both videos, participants judge them using the QoE. A full run of the experiment took approximately 53 minutes. The procedure was the same of the SS experiment. Participants were seated in a swivel chair. Participants who wore glasses or lenses kept them on throughout the session. As shown in Figure 5.19, the experiment had eight phases: (1) initialization, (2) training, (3) first session, (4) first SSQ, (5) rest, (6) second session, (7) second SSQ, and (8) finalization.

For detailed description of each procedure step please refer to Section 5.1.2. The experiment procedure was executed in Mono360 software. For details on Mono360 please refer to the Appendix B and Section 4.2.

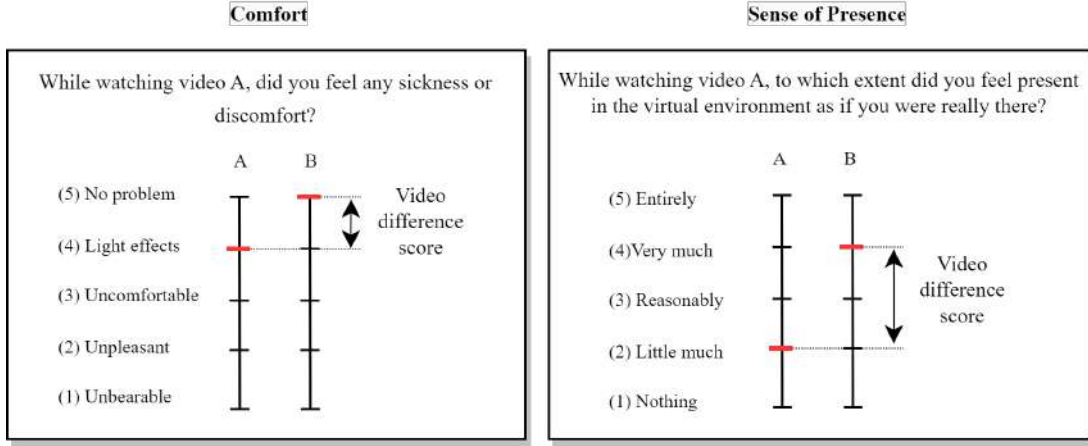


Figure 5.20: Assessment questions for comfort and sense of presence QoE attributes. Measurement of the difference between original (A) and processed (B) video.

5.3.3 Data preparation

The questions and the scales used by participants to assess the video pairs are shown in Figure 5.20. For each pair of videos (PVS and SRC) measured by participants, we compute the video difference score. Consider a PVS j with a corresponding SRC i judged by subject k with scores m_{ij} and m_{ijk} respectively, the differential opinion score for that case is:

$$d_{ijk} = m_{ik} - m_{ij}. \quad (5.11)$$

The DMOS is the average of individual difference scores across the scores from all experiment participants. Before computing DMOS, a re-scaling was performed using a linear mapping the range $[-3, 3]$ to $[1, 5]$.

We recruited 45 participants for the DS experiment. The experiment was conducted from July 7th to 29th, 2023. The demographic distribution of the participants is detailed in Table 5.5. The sampled population exhibited a diverse range of ages and varying levels of experience with HMD. Throughout the experiment, we gathered a total of 5400 opinion scores and obtained 1300 to 2000 head tracking samples for each video viewed. To run the experiment, we used the Mono360 web-application, for details about Mono360 please refer to Section 5.1.2. The complete DS dataset contains the experiment videos, the QoE data, the head motion data, and the state of the Mono360 at the end of the experiment. The dataset is publicly available ¹⁰.

¹⁰<https://osf.io/5fa7y/>

Number of participants		
45		
Proportion of Women(%)		
31		
Age		
Avg	Min.	Max.
35.53	13	76
Familiarity with VR (%)		
Novice	Moderate	Extensive
42	47	11

Table 5.5: Participant’s population. The VR familiarity is categorized into “Novice” (1st experience), “Moderate” (1 or 2 experiences), and “Extensive” (more than 3 experiences).

	Worst	Neutral	Better
Presence	381 (28%)	775 (57%)	194 (14%)
Comfort	262 (19%)	990 (73%)	98 (7%)

Table 5.6: Difference scores proportion for both QoE attributes, where “Worst” refers to the difference scores of -1, -2, -3. “Better” refers to scores 1, 2, 3.

5.4 Results

We first examine the distribution of the subjective difference scores for presence and comfort attributes. A useful way to consider difference scores is by classifying each comparison as “Worst” (meaning -1 or -2 or -3 score), “Neutral” (meaning 0 score) and “Better” (meaning 1 or 2 or 3 score). Table 5.6 shows the general proportions for both presence and comfort attributes. The majority of the comparisons fall in Neutral, mainly for comfort. In this overall analysis, “Worst” has more than double the number of “Better” comparisons, showing that it was more common to make wrong alignment edits.

Figure 5.21 depicts histograms containing the difference score count grouped by video content. We notice that, for all videos, the most common difference is 3 (“Neutral”) the case where PVS and SRC had no perceptual difference. In terms of comfort, the video “Smart” had the worst result with 59 cases where the processed video had the worst score than the original video. While the video “Vaude” had the best count for differences higher than 3. In terms of presence, again “Smart” had the worst difference count, while “BSB” had the better difference count. The video “Amizade2” had the highest number of no difference cases for both comfort and presence.

Applying a Shapiro test, we confirmed that the distribution of difference scores are non-normal ($P < 0.05$), indicating that inference should use non-parametric tests, this show up the need for employing non-parametric tests in our subsequent analyses. Figure 5.22 presents the difference count histograms by edit type.

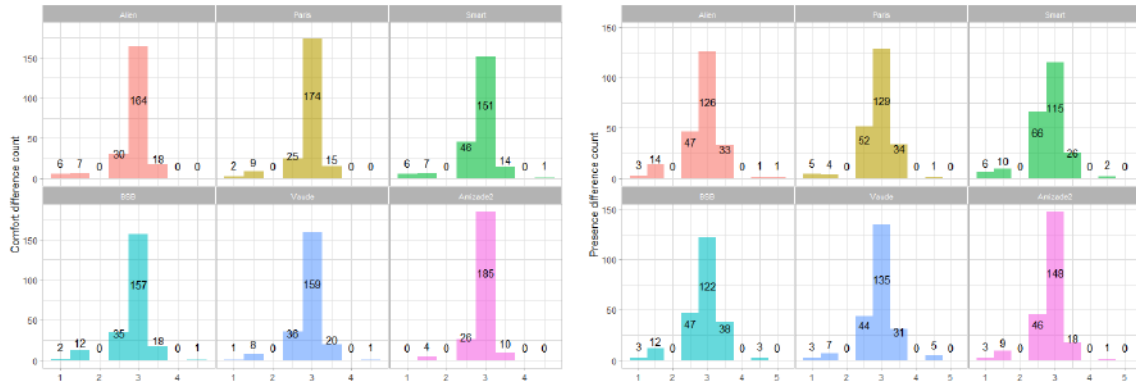


Figure 5.21: Difference count histogram grouped by video. Left: Comfort difference count. Right: Presence difference count.

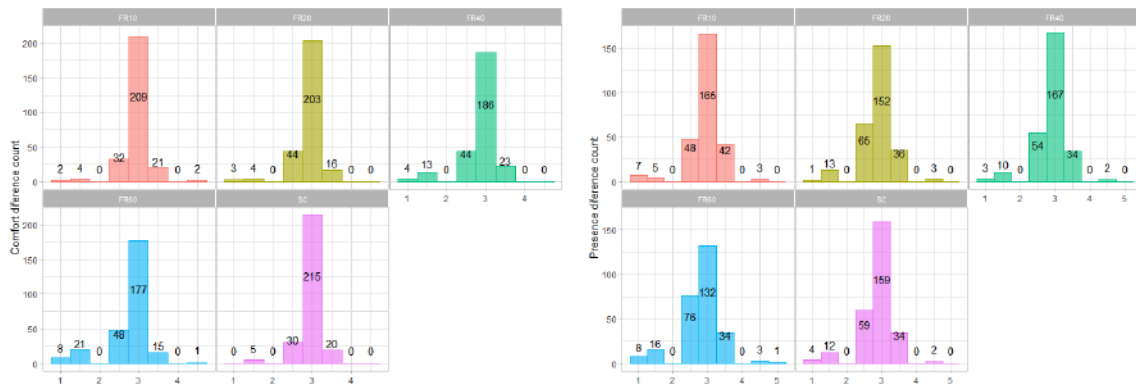


Figure 5.22: Histogram of difference scores grouped by edit type. Left: Comfort difference count. Right: Presence difference count.

5.4.1 Difference opinion score analysis

Before performing the inference analysis, we perform a sanity check with the collected difference scores. For this, we computed the correlation between the difference scores of presence and comfort. Table 5.7 shows three correlation coefficients in the data grouped by video, the correlation is positively weak for all Videos. We observe the same result when data is grouped by edit type, with a tendency that the higher the FR speed the higher the correlation. One possible interpretation is that the higher the rotation in 360°, the more representative comfort is.

Video	Pearson	Spearman	Kendall
Alien	0.44	0.37	0.34
Paris	0.31	0.32	0.30
Smart	0.29	0.33	0.29
BSB	0.39	0.42	0.38
Vaude	0.35	0.37	0.34
Amizade2	0.29	0.34	0.32
Edit type	Pearson	Spearman	Kendall
FR10	0.27	0.30	0.28
FR20	0.29	0.29	0.27
FR40	0.40	0.40	0.37
FR60	0.44	0.44	0.40
SC	0.29	0.32	0.29

Table 5.7: The coefficients of the correlation between the difference scores of presence and comfort. Top: grouped by video. Bottom: grouped by edit type.

Figure 5.23 shows the barplots containing DMOS values and confidence intervals separated by video, the edit types are separated by color. In general, as expected from the distributions, the DMOS is close to three for all conditions. In terms of comfort, we notice that tendency for DMOS to decay with the rotation speed of the FR. For FR60, this tendency holds for all cases. The SC has very similar results with FR10, based on confidence interval both alignment edits can not be discriminated. In terms of presence, the difference between DMOS across test conditions is even less distinguishable. However, we notice a tendency for FR10 to surpass SC in “Smart,” “BSB,” and “Vaude” videos; thus, reinforcing the findings from the SS experiment.

Further from the score distribution, we must infer which effects of dependent variables are statistically significant, for each factor of the study. Given that we have a non-parametric distribution of difference scores, we must conduct non-parametric test. Moreover, our experiment is withing subjects, thus, an adequate omnibus variance test is the Friedman rank sum test with six combinations of experiment factors and variables. As experiment factors, we consider Edit type, Video, and Video Resolution, whereas the



Figure 5.23: Difference Mean Opinion Scores (DMOS) for each experiment condition

Comfort and Presence difference scores are the dependent variables. Table 5.8 shows the results of the overall effect test, where we consider statistically significant those with p-value less than 0.05. The overall variance test shows that Edit type is statistically significant for both comfort and presence. Showing that a great amount of the difference score variance can be explained by the Edit type.

In terms of the video, the Friedman test does not yield a statistical significant effect, indicating that content is not a relevant factor for the difference in comfort and presence. Further, we classify the data, by grouping the videos in terms of their resolution, to perform the t-test in this aggregation. We justify that by the correlation found between overall experience and presence observed in Section 5.1, this correlation made us suspect about the influence visual quality had in comfort and presence. Conducting again the Friedman we found a non-significant difference close to the significance cut line. For that, we decided running the pairwise comparison for the Resolution classified data.

Variable	Factor	χ^2	df	p-value
Diff Comfort	Video	7.29	5	0.20000
Diff Presence	Video	5.4	5	0.36900
Diff Comfort	Edit type	19.2	4	0.00073
Diff Presence	Edit type	10	4	0.04020
Diff Comfort	Resolution	4.65	2	0.09790
Diff Presence	Resolution	5.3	2	0.07080

Table 5.8: Friedman rank sum test for the experiment factors and variables.

Given the aforementioned overall variance tests, we now perform the pairwise comparison to search for the specific test conditions that explain the effects observed. Adequate to our type of data (non-parametric ordinal data), we select Wilcoxon rank sum test (Mann-Whitney U test) to perform the pairwise comparisons. We applied the False Discovery Rate (FDR) method for p-value correction. Table 5.9 shows the pair comparison results for Edit type factor, containing 10 pair comparisons. We notice that the pairs with significant (or close to significant $p\text{-value} < 0.1$) difference involve FR60, thus showing that this edit implies a distinguishable user assessment. These results also solidify the evidence that “Fade-rotation” and “Snap-change” are equally viable for 360° videos.

Comfort				
	FR10	FR20	FR40	FR60
FR20	0.09			
FR40	0.07	0.69		
FR60	< 0.001	0.03	0.09	
SC	0.99	0.09	0.07	< 0.001
Presence				
	FR10	FR20	FR40	FR60
FR20	0.16			
FR40	0.40	0.59		
FR60	0.01	0.16	0.10	
SC	0.16	0.98	0.59	0.16

Table 5.9: Pairwise comparison with edit type as factor, for Comfort (top) and Presence (bottom) attributes. Applying Wilcoxon Rank Sum test with adjusted p-values using FDR correction. In bold, the comparisons statistical significant or close to significant $p\text{-value} \leq 0.1$.

Figure 5.24 presents the pairwise comparison by Edit type pairs, the graph shows the difference between each pair DMOS and their respective 95% CI. When the p-value is less than 0.05 or the confidence interval does not include zero, a statistical significant difference is identified. In terms of comfort, FR10 had the higher DMOS when comparing with FR20, FR40, FR60. However, only the data from FR60 and from FR10 had a statistical significant distinction. When comparing FR to SC, except for the case of FR10 which the same DMOS was found, the higher rotation speed the worst comfort difference. A statistically significant difference was found for the pair SC-FR60, and an almost significant result was found for SC-FR40. Showing a tendency that Fade-rotation edits with rotation speed higher or equal to 40°/s imply a comfort decay.

In terms of presence, a significant difference was found for the pairs FR10-FR60, FR20-FR60, and FR40-FR60, showing that the sense of presence has a significant decay for rotation speed of 60°/s. An important tendency was revealed by the pairs SC-FR10,

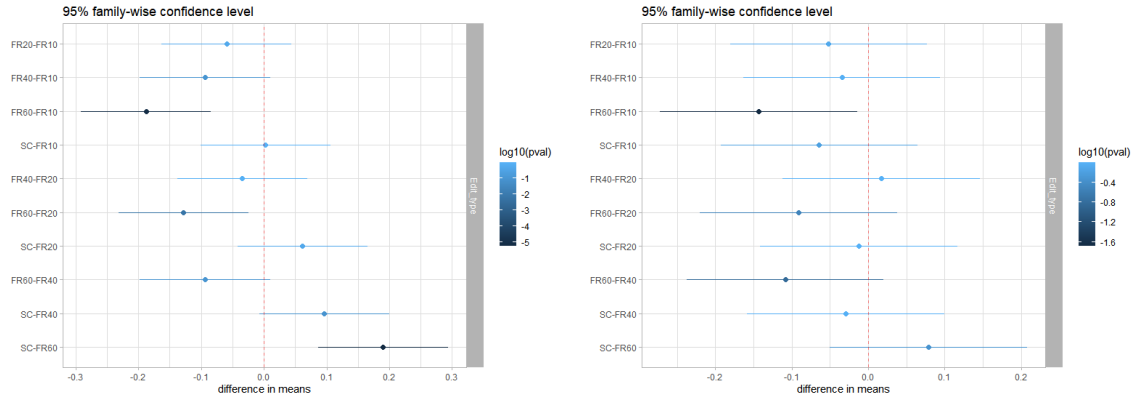


Figure 5.24: Pairwise comparison of the difference between edit types. The left plot refers to comfort differences, and the right plot to the presence differences.

SC-FR20, and SC-FR40; we notice that even though these pairs did not offer significant difference, all of them show higher sense of presence than Snap-change. This leads us to an important finding: when the rotation speed is less than $60^\circ/s$, Fade-rotation tends to imply a higher sense of presence.

When executing the pairwise comparison by video content, we observe three cases where the p-value is close to statistical significance Smart-Paris, Smart-Vaude, and Smart-Amizade2. All these cases contains the video Smart (the one with lower resolution) and the videos with higher resolution. Moreover, participants pointed out in the feedback that they observed a lesser visual quality in the Smart video. Therefore, we suppose that the visual quality played an important role in the comfort difference observed for those pairs. This fact indicates that, although video content had no significant effect on comfort and presence, we have to check whether the resolution of the videos is an important factor. To investigate this, we aggregate the scores in three levels of resolution, where *High resolution* refers to videos Vaude, Amizade2, and Paris; *Middle resolution* refers to BSB, and Alien; *Low resolution* refers to Smart. Figure 5.25 shows the pairwise comparison for the video resolution classification. In terms of comfort, the video resolution pair *High-Low* had a significant difference. For presence, both *Middle-Low* and *High-Low* pairs had significant difference.

Figure 5.26 shows the count barplot of the symptom intensities for the pre-ssq and post-ssq, surprisingly the symptoms were more common after the first session than after the second session, which could be due to the fact that the population of participants had a majority of novice and moderate users, the first reaction to the immersive technology can be a little uncomfortable. Thus, after the first session, participants would be more prone to slight and moderate symptoms.

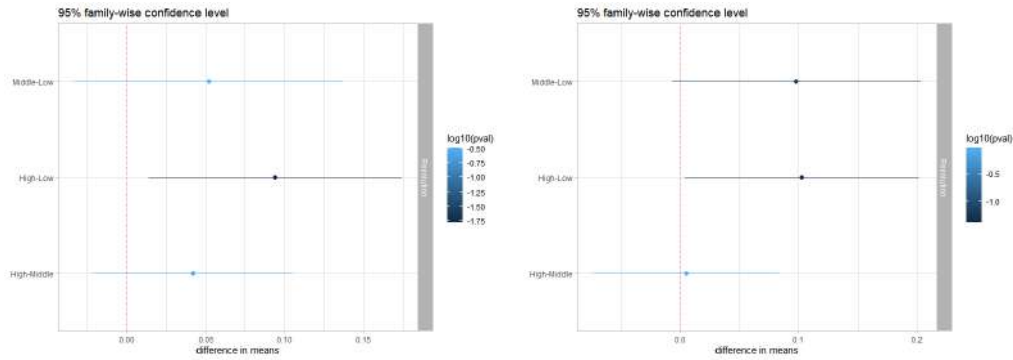


Figure 5.25: Pairwise comparison between videos aggregated by resolution level. The left plot refers to comfort differences, and the right plot to the presence differences.

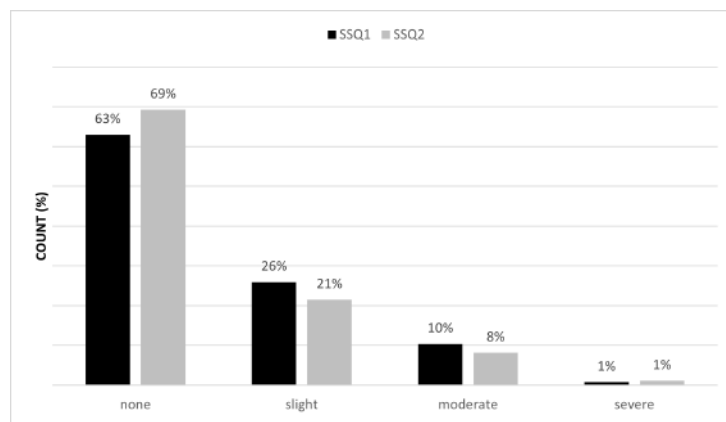


Figure 5.26: Barplot with the count of cybersickness symptoms intensity

Comparison with SS experiment

Before comparing the outputs of both experiments, it is crucial to highlight the various modifications made to the components of the DS experiment in comparison to the SS experiment. In the DS experiment, specific adjustments were implemented, such as utilizing only one device, incorporating videos with audio, introducing new source videos, and not placing the initial RoI in the center position. These modifications were justified by certain considerations. First, the device factor was deemed non-contributory to comfort or presence. Second, the absence of audio was found to disrupt and hinder participants from focusing on the content. Moreover, given the longer duration of the DS experiment, engaging content with audio was necessary to maintain participant attention, providing insights into a more realistic use case. The reduction in modifications to the source video aimed at aligning the experiment with a more authentic scenario. Third, the fact that RoI is not always at the center point is closer to a realistic use case. Therefore, the main purpose of DS experiment was to investigate alignment edits in a more realistic use case and to offer other perspectives derived from the SS experiment’s hypothesis, thereby

either reinforcing or challenging its conclusions.

Considering hypothesis **H1** and **H2**, in experiment DS we confirmed that the Edit type had a statistically significant effect on presence and comfort difference scores. In terms of comfort, Fade-rotation is statistically equivalent to Snap-change only when Fade-rotation has 10°/s rotation speed (FR10), confirming **H1** for that condition. In terms of presence, **H2** was not statistically confirmed for any case, however, FR10 and FR20 had lower presence DMOS than SC.

Formulated by Hofffeld et al. (2011) [161], the *SOS hypothesis* is a general relationship between SOS and MOS (or DMOS). The quadratic coefficient a is computed as the polynomial regression from MOS and SOS data points from the dataset, and it is very useful to compare datasets. The quadratic function mapping MOS to SOS is given by:

$$SOS(DMOS)^2 = a(-DMOS^2 + 6DMOS - 5). \quad (5.12)$$

In comparing the results of the SOS analysis for two sets of user experiments, it's crucial to consider the variation in subjective methodologies. As the original paper suggests, the "a" parameter in the SOS method can be influenced by the type of user study. Figure 5.27 illustrates the polynomial regression for each attribute in both SS and DS experiments. Notably, the calculated "a" coefficients for comfort in SS and DS were 0.41 and 0.11, respectively, and for presence, they were 0.31 and 0.14. It's noteworthy that the "a" value in SS exceeds the typical range expected in quality assessment experiments, emphasizing the influence of different measurement methodologies, such as QoE. Despite this, DS demonstrates superior performance for both comfort and presence attributes, aligning with the general trend observed in SOS parameter comparisons across diverse user studies.

In justifying the differences observed in the SOS analysis between the two sets of user experiments, it is crucial to consider the specific attributes under evaluation. The SOS method primarily aims at evaluating quality, a parameter that may have more straightforward interpretations in studies focusing on traditional quality assessments. In our study, however, we navigate a distinct landscape by assessing attributes such as comfort and sense of presence. Unlike traditional quality metrics, these attributes are inherently subjective and may present a greater degree of complexity for participants to consistently assess. Furthermore, it's imperative to note that our study intentionally avoids covering the entire range of the scale for these attributes. This strategic choice is made to prevent potential disengagement from participants, emphasizing a careful balance to ensure survey engagement while still capturing meaningful insights. As a result, the divergence in the SOS parameters can be attributed to the nuanced nature of our evaluation criteria, marking a departure from conventional quality assessments and presenting a unique set of challenges and considerations in the analysis.

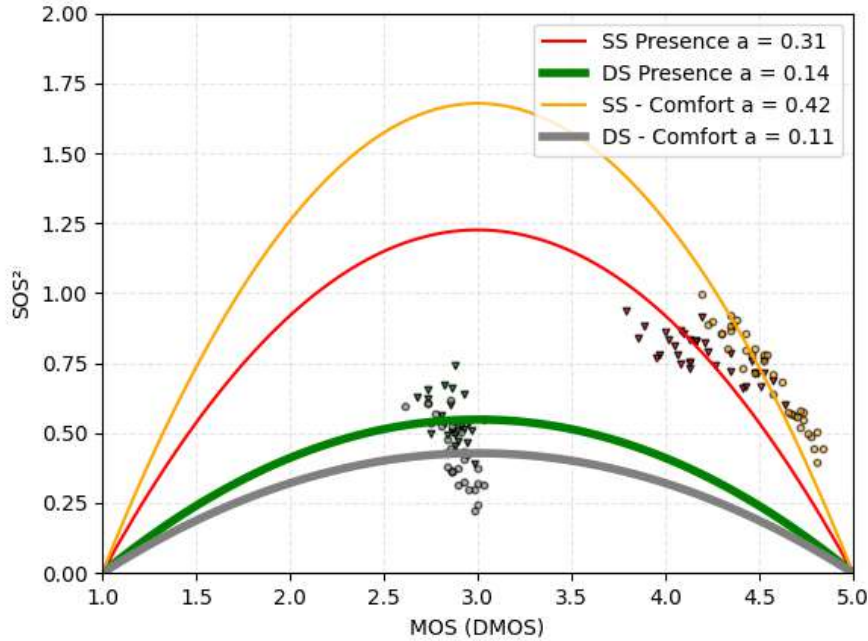


Figure 5.27: SOS hypothesis for SS, DS, for both comfort and presence data. For the SS experiments we use MOS, whereas DMOS is used for DS experiments.

5.4.2 Head motion analysis

In this study, we performed data manipulation on an experimental dataset using the *dplyr* package in R and *pandas* in Python. The raw Head Motion dataset, contains the DataFrames with the corresponding timestamp and coordinates for each trial. In average, the HMDs position sample rate was 188 samples/s. The data was organized as a DataFrame, containing information on video categories (Alien, BSB, Vaude, Amizade2, Paris, and Smart), edit types (FR10, FR20, FR40, FR60, SC), angular speeds recorded at different timestamps for relevant one second intervals close to the edit timestamps (Speed46, Speed47, Speed48, Speed49, Speed50, Speed51), and the angular distance from the participant’s viewport center point to the assumed viewport at fixed timestamps (Distance47, Distance48, Distance49, Distance50, Distance51). Please refer to Section 5.2.2 for details on the computation of the viewport distance and head speed for our sampled data.

For conducting head motion analysis, we first investigate the distributions of head speeds one second after the edits, because head motion is a relevant measure for streaming applications. Figure 5.28 depicts the CDF of the head speed for each video content, the graphs shows the speed CDF one second before and after the edit, and the x axis are limited to the portion where 75% of the speeds are included. The two sets of CDFs

exhibit similar trends, suggesting a shared underlying pattern in head speed distribution. However, a noticeable distinction emerged as the CDF corresponding to the “After Edit” condition displayed a steeper slope, indicative of a higher likelihood of observing slower head speeds. Moreover, from the data, we again observed that the threshold of $150^\circ/\text{s}$ speed would retain more than 99% of the data, because of that, for more consistent analysis we filtered out speeds higher than this threshold. We conducted a two-sample Kolmogorov-Smirnov test to assess the potential differences between the distributions of head speeds before and after the edits. The test yielded a test statistic $D = 0.015471$ with an associated p-value of 2.961×10^{-10} . The null hypothesis, which posits that the two distributions are identical, was rejected based on the very low p-value. The alternative hypothesis, suggesting a difference between the distributions, was favored. These results indicate that there is strong statistical evidence to support the assertion that the head speed distributions before and after the experimental condition differ.

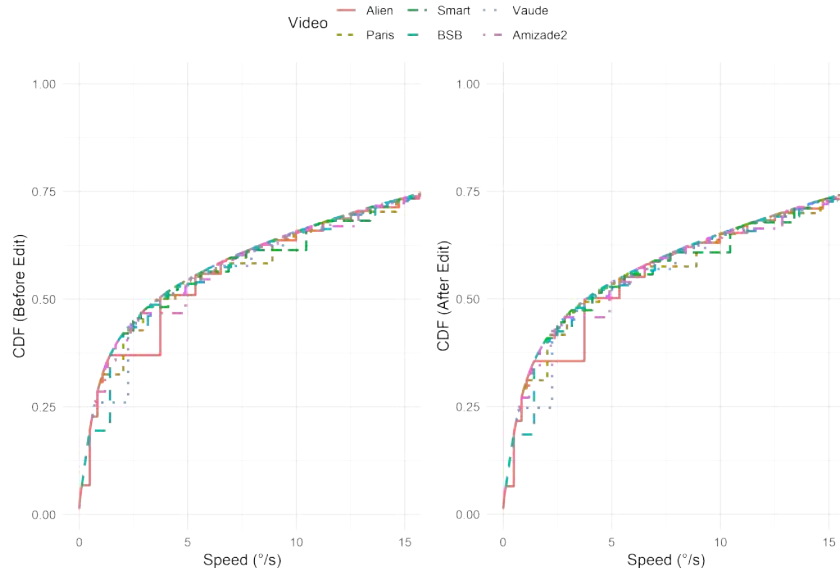


Figure 5.28: CDF of the head speed for each video content. Left: CDF measured 1s before the edit. Right: CDF measured 1s after the edit.

The boxplot in Figure 5.29 illustrates the distribution of head speeds one second after the edits. From the graph, we observe a big variety of head speeds. The videos BSB, Smart, and Amizade2 had less variance than Alien and Paris which indicates that for those videos participants navigate more actively in the content. Moreover, Figure 5.30 visually represents the difference in head speed observed 1 second after and 1 second before applied edits. Overall, no distinct reduction in head speed is evident across the distributions. However, a detailed examination reveals nuanced patterns. Notably, for videos Paris and Smart, FR10 edits exhibit a discernible average reduction in head speed. Conversely, in the case of video Vaude, an increase in head speed is observed following

FR10 edits. In terms of SC edits, it do not distinctly reduce the average head speed in any video, suggesting that this type of edit did not have a pronounced impact on head motion. Furthermore, for Alien and Paris videos, FR20 edits amplify average head speeds rather than reduce them. These findings highlight the importance of considering video-specific characteristics when evaluating the impact of different edit types on head speed.

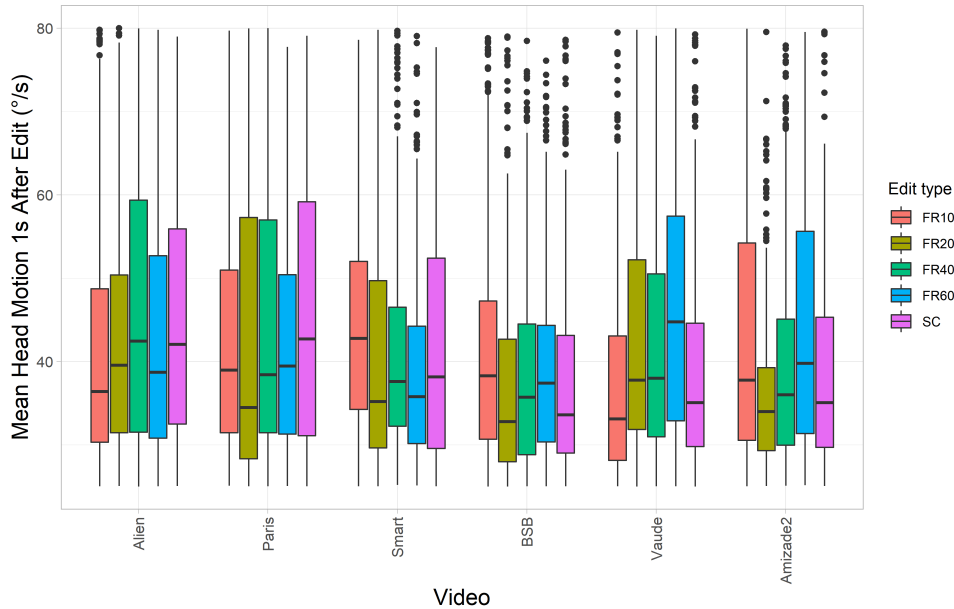


Figure 5.29: Boxplot of head speeds measured 1s after the edit, for each experiment condition.

To test whether the alignment edit implied a reduction in head speed, we compute the difference between the mean head speed 1s before and the mean head speed 1s after the alignment edit. Table 5.10 shows the result for the 30 conditions tested. A significant reduction or increase in mean head speed happened when the head speed difference minus the head speed difference standard error crossed zero. In Table 5.10, we emphasize the conditions found with significant reduction or increase. Considering hypothesis **H4**, in experiment DS we confirmed 18 cases where the head speed decreased. We detected two conditions where the head speed increased, both for FR20 in “Vaude” and “Amizade2” videos. No edit type had significant reduction for all videos tested. However, FR10 had significant reduction except for “BSB” video. The video “Paris” had the highest overall reduction in head speed.

Like in Section 5.2.2, we classify each experiment trial in terms of alignment the alignment state “A.” We compute the alignment by measuring the angular distance between the center points of the participants viewport and assumed viewport. The assumed viewport angular positions for each video are shown in Figure 5.17. Thus, if the participant’s

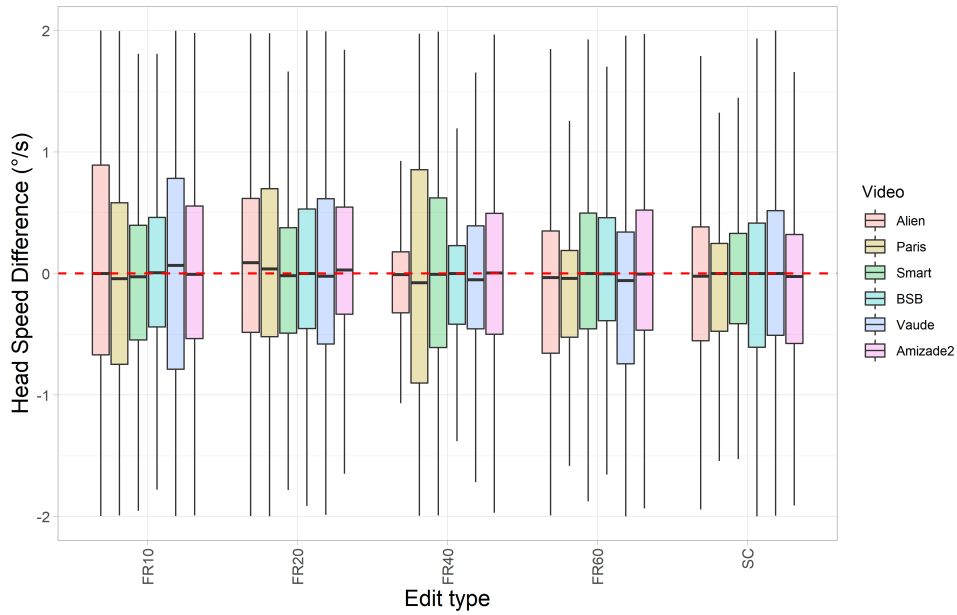


Figure 5.30: Boxplot of the head speed difference, computed from the subtraction of the head speeds 1s before the edit and after the edit, for each participant.

viewport center fall within the radius of 60° around the assumed viewport center, she/he would be “aligned” ($A = 1$) with the assumed viewport.

The Friedman rank sum test was conducted to assess the effect of alignment state (A) on head speeds before edits. The test yield no significant difference ($\chi_r^2 = 0.209$, $df = 1$, p-value = 0.647). When examining the impact of head speeds after edits, the Friedman test demonstrated a significant difference among alignment states ($\chi_r^2 = 3.93$, $df = 1$, p-value = 0.047). This indicates a substantial effect of alignment on head speeds after edits. When viewers look at RoI, aligning with the narrative plot, it is expected that viewers change behavior by fixating longer at the identified RoI. This result underscores the influence of RoI alignment on head speeds, mainly in post-edit scenarios. Considering hypothesis **H4**, in DS experiment we confirmed that the alignment state (A) had a statistically significant effect on head speed after the edit, reinforcing the conclusion from the SS experiment.

Table 5.11 presents the pairwise comparisons effect of video over the alignment state, from 15 comparisons only 5 did not show a significant difference. Figure 5.31 shows the count of aligned and non-aligned trials for each video. From those 5 significant cases, three contains the video “Smart” against other videos with a high proportion of aligned trials, while the other “Paris” vs. “Alien” relates to two videos with lower proportion of aligned trials. Moreover, except from “Paris” and “Alien”, respectively with 39% and 45% of the trials aligned, the assumed viewport successfully called participant’s attention.

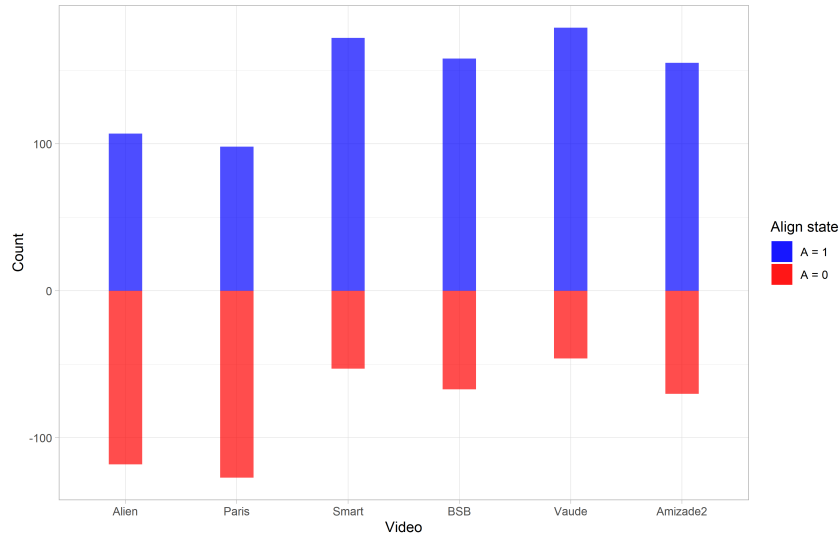


Figure 5.31: Count bars for aligned trials ($A = 1$), and non-aligned trials ($A = 0$) for each video.

Table 5.11: Summary of Pairwise comparisons using Wilcoxon rank sum test with continuity correction considering the effect of video over the alignment state

Video Comparison	p-value
Paris vs. Alien	0.423
Smart vs. Alien	3.4×10^{-8}
BSB vs. Alien	1.4×10^{-7}
Vaude vs. Alien	1.4×10^{-11}
Amizade2 vs. Alien	8.6×10^{-6}
Smart vs. Paris	4.9×10^{-10}
BSB vs. Paris	1.3×10^{-9}
Vaude vs. Paris	6.8×10^{-14}
Amizade2 vs. Paris	1.4×10^{-7}
BSB vs. Smart	0.393
Vaude vs. Smart	0.393
Amizade2 vs. Smart	0.189
Vaude vs. BSB	0.058
Amizade2 vs. BSB	0.53
Amizade2 vs. Vaude	0.016

The mean head speed after the edit varied across different edit types and alignment states, as illustrated in the corresponding boxplot (refer to Figure 5.32). For edit type FR10, the mean head speed was $13.6^\circ/\text{s}$ for non-aligned instances ($A = 0$) and $12.2^\circ/\text{s}$

for aligned instances ($A = 1$). Similarly, edit type FR40 showed a mean head speed of $13.1^\circ/\text{s}$ for $A = 0$ and $12.2^\circ/\text{s}$ for $A = 1$. However, FR20 and SC increased the mean head speed in aligned instances, indicating the influence of both edit type and alignment state on head speed after the edit.

Finally, we must check if the alignment state impacts the comfort and presence difference scores. Whereas the data is ordinal we performed again the Friedman test, including the alignment state A as the factor and the difference score as a variable. No significant effect related to comfort and presence, respectively resulting in $\chi_r^2 = 0.0286$ ($df = 1$, p-value = 0.866), and $\chi_r^2 = 0.61$ ($df = 1$, p-value = 0.435). Therefore, this evidence suggests that alignment performance did not affect the sense of presence and comfort while watching the content. Considering hypothesis **H3**, in DS experiment we rejected the impact of the align performance over comfort or presence scores, reinforcing the conclusion from the SS experiment.

Table 5.10: Difference in Head Speed (HS) between 1s before and 1s after the alignment edit with Standard Error (SE). In bold, the conditions where happened a significant reduction or increase in mean head speed.

Video	Edit type	HS before	HS after	HS diff.	Reduction (%)
Alien	FR10	18.23	16.57	-1.66	-9.08
Alien	FR20	15.86	16.62	0.76	4.77
Alien	FR40	15.49	15.43	-0.07	-0.44
Alien	FR60	16.78	14.30	-2.48	-14.79
Alien	SC	13.92	12.89	-1.03	-7.40
Paris	FR10	21.29	13.79	-7.51	-35.25
Paris	FR20	16.05	13.64	-2.41	-15.00
Paris	FR40	16.87	13.25	-3.61	-21.41
Paris	FR60	13.56	10.04	-3.52	-25.98
Paris	SC	14.31	12.10	-2.20	-15.38
Smart	FR10	11.60	9.96	-1.64	-14.14
Smart	FR20	9.12	9.67	0.56	6.12
Smart	FR40	12.66	10.48	-2.18	-17.22
Smart	FR60	11.02	7.92	-3.10	-28.11
Smart	SC	12.77	12.09	-0.68	-5.34
BSB	FR10	11.71	12.07	0.36	3.09
BSB	FR20	9.66	7.38	-2.28	-23.62
BSB	FR40	11.38	9.36	-2.02	-17.73
BSB	FR60	10.18	10.53	0.36	3.49
BSB	SC	10.51	10.93	0.41	3.95
Vaude	FR10	15.42	13.98	-1.43	-9.30
Vaude	FR20	15.35	16.57	1.22	7.92
Vaude	FR40	14.25	13.71	-0.54	-3.82
Vaude	FR60	16.29	13.45	-2.83	-17.40
Vaude	SC	14.53	12.27	-2.26	-15.56
Amizade2	FR10	10.73	9.90	-0.83	-7.77
Amizade2	FR20	8.97	10.33	1.36	15.20
Amizade2	FR40	12.12	12.66	0.55	4.51
Amizade2	FR60	10.52	9.98	-0.53	-5.07
Amizade2	SC	10.78	8.38	-2.40	-22.25

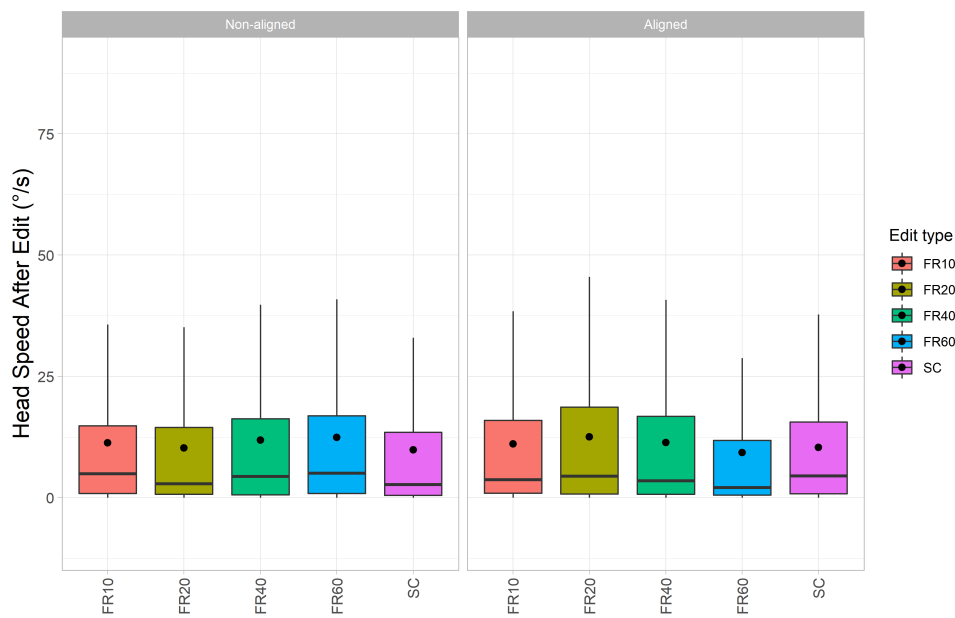


Figure 5.32: Boxplot of head speeds measured 1s after the edit separated in align states facets.

Chapter 6

Conclusions

6.1 Final remarks

This study introduces the “Fade-rotation” alignment editing technique tailored for enhancing the 360° video-watching experience. Unlike comparable methods relying on user head movements as triggers, “Fade-rotation” employs a predetermined trigger point, enabling filmmakers to predetermine edit times. Evaluating its effectiveness involved user studies and a comparative analysis with the “Snap-change” edit method. Categorized as alignment edits specific to 360° videos, our analysis compared two types: “Snap-change” and “Fade-rotation”, considering impacts on UX, QoE (presence, comfort, experience, cybersickness), and head movement behavior.

Key Conclusions:

1. Based on subjective feedback, alignment edits tested did not significantly degrade users’ comfort or presence, with many participants not noticing edits.
2. Video content and scene motion significantly influenced user ratings, emphasizing the impact of content motion on comfort presence, and overall experience.
3. A “Fade-rotation” edit with a rotation speed greater or equal to 20°/s should be avoided for dynamic scene motion contents. A 10°/s rotation speed or a “Snap-change” is preferable to lower discomfort probability.
4. The alignment between ROI and FOV reduced head movement speed post-edit, with gradual alignments achieving an 8% lower speed than instant edits.
5. The alignment between ROI and user FOV did not impact presence, comfort, and experience significantly.

6. Although further statistical validation is needed, results from both experiments suggest that “Fade-rotation” implies a higher sense of presence than “Snap-change”.

The inclusion of a DS experiment enhanced robustness of our findings, in this occasion we selected videos with active characters near the edit point. Notably, the study identifies that rotation speeds equal to or higher than $40^\circ/\text{s}$ in “Fade-rotation” edits imply a comfort decay. Moreover, a decay in the sense of presence is observed for rotation speeds of $60^\circ/\text{s}$. Although not conclusive, we observed a tendency in both SS and DS experiments, “Fade-rotation” tends to evoke a higher sense of presence compared to “Snap-change”, especially when rotation speeds is $10^\circ/\text{s}$. While video content itself does not significantly affect comfort and presence, a resolution-based analysis indicates that video resolution plays a crucial role. The barplot of symptom intensities reveals a surprising trend, with symptoms being more common after the first session, suggesting that initial discomfort in novice users might contribute to this pattern. Another important finding include the identification of significant differences in head speed distributions before and after alignment edits, supported by a Kolmogorov-Smirnov test. Moreover, the results emphasized the impact of alignment states on head speeds, both pre and post-edit, underscoring its influence on UX. Additionally, the study highlighted the significant effect of video content on alignment states, emphasizing its role in determining participant attention. Insights into mean head speeds across different edit types provided perspectives on the effects of alignment performance. Finally, an investigation into the impact of alignment on comfort and presence difference scores revealed no significant effects, reinforcing the results from the DS experiment. Overall, these findings provide valuable insights into the nuanced impact of editing types, rotation speeds, and video resolution on UX and comfort in immersive environments.

In summary, this thesis contributed with:

1. We propose a new alignment edit “Fade-rotation”, based on Farmani *et al.* (2020) [11], and recommend parameters for automating it based on evidence collected in user studies.
2. We provide the dataset containing rating data for five QoE attributes, collected from a total of 108 participants, covering 12 different contents and five alignment edit types.
3. We develop a web-platform to run the complete subjective experiments, respecting the last ITU’s recommendations for QoE assessment.
4. We propose and apply a set of metrics to examine head motion based on alignment states, which can be useful for QoE and behavior cross-analysis.

5. We apply a comprehensive comparison between results from SS and DS, increasing the evidence to sustain our findings.

6.2 Limitations and Future work

Our decisions in designing the experiment were centered on addressing our research objectives; however, these choices come with certain limitations. A significant constraint in our study is that we exclusively explored the offline version of alignment edits. Although we attempted to address this limitation by analyzing the alignment performance, future investigations should delve into online alignment edits, as only the online version can ensure RoI visualization. Furthermore, some factors remained uncontrollable with offline edits, such as rotation direction. In the offline version, we had to incorporate an offset within the rotation and were unable to vary the duration of the rotation. Another noteworthy limitation was the utilization of different video content in the experiments. It is essential to emphasize that this decision was made because we deemed dataset diversity more strategic than making direct comparisons between both experiments. In fact, during the DS experiment, we were able to indirectly gather evidence aligning with the findings from the SS experiment, albeit with different content. This strategic decision allowed us to test a more extensive range of content (12 video clips). Nevertheless, it remains imperative for future research to validate whether the conclusions hold true for SS experiments.

Finally, this work lays the foundation for other potential investigations. Among other possibilities, future research avenues include:

1. Implementing automation methods for alignment edits.
2. Extending the knowledge about the impact of “Fade-rotation” parameters, *e.g.* edit duration, non-uniform rotation speed, rotation direction.
3. Expanding the number of participants in the dataset, to better distinguish the scores.
4. Conducting additional user studies to assess the effects of ΔT_{fade} on QoE.
5. Integrating additional QoE attributes like attention or emotion into alignment edits analysis.
6. Analyzing cultural and demographic factors using our dataset.

References

- [1] *Virtual reality (vr) market size, share & trends analysis report by technology (semi & fully immersive, non-immersive), by device (hmd, gtd), by component (hardware, software), by application, by region, and segment forecasts, 2023 - 2030*. Report ID: GVR-1-68038-831-2. Number of Pages: 300, 2022. <https://www.grandviewresearch.com/industry-analysis/virtual-reality-vr-market/methodology>. viii, 1
- [2] Nielsen, Lasse T, Matias B Møller, Sune D Hartmeyer, Troels CM Ljung, Niels C Nilsson, Rolf Nordahl, and Stefania Serafin: *Missing the point: an exploration of how to guide users' attention during cinematic virtual reality*. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*, pages 229–232, 2016. viii, 2
- [3] Aitamurto, Tanja, Andrea Stevenson Won, Sukolsak Sakshuwong, Byungdoo Kim, Yasamin Sadeghi, Krysten Stein, Peter G Royal, and Catherine Kircos: *From fomo to jomo: Examining the fear and joy of missing out and presence in a 360° video viewing experience*. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021. viii, 10, 20
- [4] Marañes, Carlos, Diego Gutierrez, and Ana Serrano: *Towards assisting the decision-making process for content creators in cinematic virtual reality through the analysis of movie cuts and their influence on viewers' behavior*. *International Transactions in Operational Research*, n/a(n/a). <https://onlinelibrary.wiley.com/doi/abs/10.1111/itor.13106>. viii, 10, 18, 21
- [5] Weaving, Simon: *Evoke, don't show: Narration in cinematic virtual reality and the making of entangled*. *Virtual Creativity*, 2021. viii, 18
- [6] Young, R Michael: *Creating interactive narrative structures: The potential for ai approaches*. *Psychology*, 13:1–26, 2000. viii, 18
- [7] Nassar, Samah Gaber Mohamed: *Engaging by design: Utilization of vr interactive design tool in mise-en-scène design in filmmaking*. *International Design Journal*, 11(6):5, 2021. <https://digitalcommons.aaru.edu.jo/faa-design/vol11/iss6/5>. viii, 1, 18, 20, 21, 22
- [8] Pillai, Jayesh S. and Manvi Verma: *Grammar of vr storytelling: Analysis of perceptual cues in vr cinema*. In *Proceedings of the 16th ACM SIGGRAPH European Conference on Visual Media Production, CVMP '19*, New York, NY, USA, 2019.

- Association for Computing Machinery, ISBN 9781450370035. <https://doi.org/10.1145/3359998.3369402>. viii, 1, 18
- [9] Gödde, Michael, Frank Gabler, Dirk Siegmund, and Andreas Braun: *Cinematic narration in vr - rethinking film conventions for 360 degrees*. In *HCI*, 2018. viii, 19, 20, 21
- [10] Dambra, Savino, Giuseppe Samela, Lucile Sassatelli, Romaric Pighetti, Ramon Aparicio-Pardo, and Anne Marie Pinna-Déry: *Film editing: New levers to improve vr streaming*. In *Proceedings of the 9th ACM Multimedia Systems Conference*, pages 27–39, 2018. ix, xiii, xiv, 2, 9, 23, 25, 56
- [11] Farmani, Yasin and Robert J. Teather: *Evaluating discrete viewpoint control to reduce cybersickness in virtual reality*. pages 1–20, 2020. ix, x, 3, 22, 23, 26, 28, 36, 39, 78
- [12] Eftekharifar, Siavash, Anne Thaler, Adam O. Bebko, and Nikolaus F. Troje: *The role of binocular disparity and active motion parallax in cybersickness*. *Experimental brain research*, 2021. ix, 3, 28, 34
- [13] Ang, Samuel and John Quarles: *Gingervr: An open source repository of cybersickness reduction techniques for unity*. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pages 460–463, 2020. ix, 3
- [14] Fremerey, Stephan, Ashutosh Singla, Kay Meseberg, and Alexander Raake: *Av-track360: An open dataset and software recording people’s head rotations watching 360° videos on an hmd*. *MMSys ’18*, page 403–408, New York, NY, USA, 2018. Association for Computing Machinery, ISBN 9781450351928. <https://doi.org/10.1145/3204949.3208134>. x, 30
- [15] Pérez, Pablo and Javier Escobar: *Miro360: A tool for subjective assessment of 360 degree video for itu-t p.360-vr*. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3, 2019. x, 30
- [16] Pighetti, Romaric, wasp898, Savino D, sassatelli, and Joël CANCELA VAZ: *Uca4svr/toucan-vr: Toucan_vr*, March 2018. <https://doi.org/10.5281/zenodo.1204442>. x, 30
- [17] Rossi, Henrique Souza, Karan Mitra, Christer Åhlund, Irina Cotanis, Niclas Ögren, and Per Johansson: *Altruist: A multi-platform tool for conducting qoe subjective tests*. In *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, pages 99–102, 2023. x, 30
- [18] Gutiérrez, Jesús, Pablo Pérez, Marta Orduna, Ashutosh Singla, Carlos Cortés, Pramit Mazumdar, Irene Viola, Kjell Brunnström, Federica Battisti, Natalia Cieplinska, Dawid Juszka, Lucjan Janowski, Mikołaj Leszczuk, Anthony Olufemi Adeyemi-Ejeye, Yaosi Hu, Zhenzhong Chen, Glenn Van Wallendael, Peter Lambert, César Díaz, John Hedlund, Omar Hamsis, Stephan Fremerey, Frank Hofmeyer,

- Alexander Raake, Pablo César, Marco Carli, and Narciso García: *Subjective evaluation of visual quality and simulator sickness of short 360 videos: Itu-t rec. p.919*. IEEE Transactions on Multimedia, 24:3087–3100, 2022. xi, 15, 37, 38, 39, 43, 44
- [19] Pérez, Pablo, Nuria Oyaga, Jaime Jesus Ruiz, and Alvaro Villegas: *Towards systematic analysis of cybersickness in high motion omnidirectional video*. 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), pages 1–3, 2018. xi, 38, 39, 59
- [20] Bouchard, Stéphane, G. Robillard, Julie St-Jacques, Stéphane Dumoulin, M.J. Patry, and Patrice Renaud: *Reliability and validity of a single-item measure of presence in vr*. Proceedings. Second International Conference on Creating, Connecting and Collaborating through Computing, pages 59–61, 2004. xi, 38, 39, 59
- [21] Knorr, Sebastian, Cagri Ozcinar, Colm O Fearghail, and Aljosa Smolic: *Director’s cut: a combined dataset for visual attention analysis in cinematic vr content*. In *Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production*, pages 1–10, 2018. xii, 20, 34, 57
- [22] Nasrabadi, Afshin Taghavi, Alihsan Samiei, Anahita Mahzari, Mylène C. Q. Farias, Marcelo M. Carvalho, Ryan P. McMahan, and Ravi Prakash: *A taxonomy and dataset for 360° videos*. Proceedings of the 10th ACM Multimedia Systems Conference, 2019. xii, 35
- [23] Howard, Matt C. and Elise C. Van Zandt: *A meta-analysis of the virtual reality problem: Unequal effects of virtual reality sickness across individual differences*. Virtual Reality, 25:1221–1246, 2021. xiv, 50
- [24] Lee, Lik Hang, Tristan Braud, Pengyuan Zhou, Lin Wang, Dianlei Xu, Zijun Lin, Abhishek Kumar, Carlos Bermejo, and Pan Hui: *All one needs to know about meta-verse: A complete survey on technological singularity, virtual ecosystem, and research agenda*. ArXiv, abs/2110.05352, 2021. 1, 4, 11
- [25] Bremmers, J.: *Narrative cues within cinematic virtual reality: An exploratory study of narrative cues within the content and motives of virtual reality developers*. Master’s thesis, December 2017. <http://hdl.handle.net/2105/42782>. 1
- [26] Argyriou, Lemonia, Daphne Economou, and Vassiliki Bouki: *Design methodology for 360 immersive video applications: the case study of a cultural heritage virtual tour*. Personal and Ubiquitous Computing, pages 1–17, 2020. 1
- [27] Santos Althoff, Lucas dos, Henrique Domingues Garcia, Dário Daniel Ribeiro Morais, Sana Alamgeer, Myllena A. Prado, Gabriel C. Araujo, Ravi Prakash, Marcelo M. Carvalho, and Mylène C. Q. Farias: *Designing an user-centric framework for perceptually-efficient streaming of 360° edited videos*. Electronic Imaging, 2022. 1, 6
- [28] Ortega-Alvarez, Galo, Carlos Matheus-Chacin, Angel Garcia-Crespo, and Adrian Ruiz-Arroyo: *Evaluation of user response by using visual cues designed to direct*

- the viewer's attention to the main scene in an immersive environment.* *Multimedia Tools Appl.*, 82(1):573–599, jun 2022, ISSN 1380-7501. <https://doi.org/10.1007/s11042-022-13271-7>. 2
- [29] Speicher, Marco, Christoph Rosenberg, Donald Degraen, Florian Daiber, and Antonio Krüger: *Exploring visual guidance in 360-degree videos.* In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, pages 1–12, 2019. 2
- [30] Lin, Yen Chen, Yung Ju Chang, Hou Ning Hu, Hsien Tzu Cheng, Chi Wen Huang, and Min Sun: *Tell me where to look: Investigating ways for assisting focus in 360° video.* *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017. 2
- [31] Lankes, Michael and Argenis Ramirez Gomez: *Gazecues: Exploring the effects of gaze-based visual cues in virtual reality exploration games.* *Proc. ACM Hum.-Comput. Interact.*, 6(CHI PLAY), oct 2022. 2
- [32] Sassatelli, Lucile, Anne Marie Pinna-Déry, Marco Winckler, Savino Dambra, Giuseppe Samela, Romaric Pighetti, and Ramon Aparicio-Pardo: *Snap-changes: a dynamic editing strategy for directing viewer's attention in streaming virtual reality videos.* In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, pages 1–5, 2018. 2
- [33] Araújo, Gabriel De Castro, Henrique Domingues Garcia, Mylene Farias, Ravi Prakash, and Marcelo Carvalho: *360eavp: A 360-degree edition-aware video player.* In *Proceedings of the 15th International Workshop on Immersive Mixed and Virtual Environment Systems, MMVE '23*, page 18–23, New York, NY, USA, 2023. Association for Computing Machinery, ISBN 9798400701894. <https://doi.org/10.1145/3592834.3592879>. 2, 4, 9, 24
- [34] Serrano, Ana, Incheol Kim, Zhili Chen, Stephen DiVerdi, Diego Gutierrez, Aaron Hertzmann, and Belén Masiá: *Motion parallax for 360° rgbd video.* *IEEE Transactions on Visualization and Computer Graphics*, 25:1817–1827, 2019. <https://api.semanticscholar.org/CorpusID:73513463>. 3
- [35] Wolf, Dennis, Michael Rietzler, Laura Bottner, and Enrico Rukzio: *Augmenting teleportation in virtual reality with discrete rotation angles.* *ArXiv*, abs/2106.04257, 2021. 3
- [36] International Telecommunication Union: *ITU-T Recommendation BT.500-8: Methodology for the subjective assessment of the quality of television pictures*, 2023. <https://www.itu.int/rec/R-REC-BT.500-15-202305-I/en>. 3, 14, 15, 59
- [37] Pérez, Pablo, Ester González-Sosa, Jes'us Guti'errez, and Narciso García: *Emerging immersive communication systems: Overview, taxonomy, and good practices for qoe assessment.* In *Frontiers in Signal Processing*, 2022. 3, 4, 13

- [38] International Telecommunication Union: *P.800 : Methods for subjective determination of transmission quality*, 2019. <https://www.itu.int/rec/T-REC-P.800-199608-I.3>, 14, 15
- [39] Naderi, Babak, Sebastian Möller, and Ross Cutler: *Speech quality assessment in crowdsourcing: Comparison category rating method*. In *2021 Thirteenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2021. 3
- [40] Singla, Ashutosh, Werner Robitza, and Alexander Raake: *Comparison of subjective quality test methods for omnidirectional video quality evaluation*. In *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6, 2019. 3, 15, 25
- [41] Nehmé, Yana, Jean Philippe Farrugia, Florent Dupont, Patrick Le Callet, and Guillaume Lavoué: *Comparison of subjective methods for quality assessment of 3d graphics in virtual reality*. *ACM Transactions on Applied Perception (TAP)*, 18:1 – 23, 2021. 3
- [42] Pinson, Margaret H. and Stephen Wolf: *Comparing subjective video quality testing methodologies*. In *Visual Communications and Image Processing*, 2003. 4
- [43] Winkler, Stefan: *Analysis of public image and video databases for quality assessment*. *IEEE Journal of Selected Topics in Signal Processing*, 6:616–625, 2012. 4
- [44] Barman, Nabajeet and Maria G. Martini: *Qoe modeling for http adaptive video streaming—a survey and open challenges*. *IEEE Access*, 7:30831–30859, 2019. 4
- [45] Le Callet, Patrick, Sebastian Möller, Andrew Perkis, *et al.*: *Qualinet white paper on definitions of quality of experience*. European network on quality of experience in multimedia systems and services (COST Action IC 1003), 3(2012), 2012. 4
- [46] Perkis, Andrew, Christian Timmerer, Sabina Baraković, Jasmina Baraković Husić, Søren Bech, Sebastian Bosse, Jean Botev, Kjell Brunnström, Luis Cruz, Katrien De Moor, Andrea de Polo Saibanti, Wouter Durnez, Sebastian Egger-Lampl, Ulrich Engelke, Tiago H. Falk, Jesús Gutiérrez, Asim Hameed, Andrew Hines, Tanja Kojic, Dragan Kukolj, Eirini Liotou, Dragorad Milovanovic, Sebastian Möller, Niall Murray, Babak Naderi, Manuela Pereira, Stuart Perry, Antonio Pinheiro, Andres Pinilla, Alexander Raake, Sarvesh Rajesh Agrawal, Ulrich Reiter, Rafael Rodrigues, Raimund Schatz, Peter Schelkens, Steven Schmidt, Saeed Shafiee Sabet, Ashutosh Singla, Lea Skorin-Kapov, Mirko Suznjevic, Stefan Uhrig, Sara Vlahović, Jan Niklas Voigt-Antons, and Saman Zadtootaghaj: *Qualinet white paper on definitions of immersive media experience (imex)*, 2020. 4, 42
- [47] Dambra, Savino, Giuseppe Samela, Lucile Sassatelli, Romaric Pighetti, Ramon Aparicio-Pardo, and Anne Marie Pinna-Déry: *Film editing: New levers to improve vr streaming*. In *Proceedings of the 9th ACM Multimedia Systems Conference*, pages 27–39, 2018. 4

- [48] Serrano, Ana, Vincent Sitzmann, Jaime Ruiz-Borau, Gordon Wetzstein, Diego Gutierrez, and Belen Masia: *Movie editing and cognitive event segmentation in virtual reality video*. *ACM Transactions on Graphics (TOG)*, 36(4):1–12, 2017. 4, 18, 21, 28, 36, 57
- [49] Chiariotti, Federico: *A survey on 360-degree video: Coding, quality of experience and streaming*. *Comput. Commun.*, 177:133–155, 2021. 4
- [50] Santos, José, Jeroen van der Hooft, Maria Torres Vega, Tim Wauters, Bruno Volckaert, and Filip De Turck: *Efficient orchestration of service chains in fog computing for immersive media*. 2021 17th International Conference on Network and Service Management (CNSM), pages 139–145, 2021. 4
- [51] Lee, Lik Hang, Tristan Braud, Pengyuan Zhou, Lin Wang, Dianlei Xu, Zijun Lin, Abhishek Kumar, Carlos Bermejo, and Pan Hui: *All one needs to know about meta-verse: A complete survey on technological singularity, virtual ecosystem, and research agenda*. *ArXiv*, abs/2110.05352, 2021. 4
- [52] Guimard, Quentin, Florent Robert, Camille Bauge, Aldric Ducreux, Lucile Sassetelli, Hui Yin Wu, Marco Winckler, and Auriane Gros: *PEM360: A dataset of 360° videos with continuous Physiological measurements, subjective Emotional ratings and Motion traces*. In *MMSys 2022 - 13th ACM Multimedia Systems Conference*, Athlone, Ireland, June 2022. <https://hal.archives-ouvertes.fr/hal-03710323>. 4
- [53] Althoff, Lucas S., Mylène C. Q. Farias, Alessandro Rodrigues Silva, and Marcelo M. Carvalho: *Impact of alignment edits on the quality of experience of 360° videos*. *IEEE Access*, 11:108475–108492, 2023. 6
- [54] Althoff, Lucas S., Myllena A. Prado, Sana Alameer, Alessandro Silva, Ravi Prakash, Marcelo M. Carvalho, and Mylène C. Q. Farias: *360rat: A tool for annotating regions of interest in 360-degree videos*. *Proceedings of the Brazilian Symposium on Multimedia and the Web*, 2022. 6, 24
- [55] Morais, Dário Daniel Ribeiro, Lucas S. Althoff, Ravi Prakash, Marcelo M. Carvalho, and Mylène C. Q. Farias: *A content-based viewport prediction model*. *Electronic Imaging*, 2021. 6
- [56] Althoff, Lucas S., Mylène C. Q. Farias, and Li Weigang: *Once learning for looking and identifying based on yolo-v5 object detection*. *Proceedings of the Brazilian Symposium on Multimedia and the Web*, 2022. 7
- [57] Weigang, Li, Luiz Martins, Nikson Ferreira, Christian Miranda, Lucas Althoff, Walner Pessoa, Mylenè Farias, Ricardo Jacobi, and Mauricio Rincon: *Heuristic once learning for image & text duality information processing*. In *2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles*, pages 1353–1359, 2022. 7

- [58] Cerqueira, José Antonio Siqueira de, Lucas S. Althoff, Paulo Santos de Almeida, and Edna Dias Canedo: *Ethical perspectives in ai: A two-folded exploratory study from literature and active development projects - supplementary material*. In *Hawaii International Conference on System Sciences*, 2021. <http://hdl.handle.net/10125/71257>. 7
- [59] Ai, Hao, Zidong Cao, Jin Zhu, Haotian Bai, Yucheng Chen, and Ling Wang: *Deep learning for omnidirectional vision: A survey and new perspectives*. ArXiv, abs/2205.10468, 2022. 8
- [60] Martin, Daniel, Sandra Malpica, Diego Gutierrez, Belén Masiá, and Ana Serrano: *Multimodality in vr: A survey*. ACM Computing Surveys (CSUR), 2022. 8
- [61] Milgram, Paul and Fumio Kishino: *A taxonomy of mixed reality visual displays*. IEICE Transactions on Information and Systems, 77:1321–1329, 1994. 8, 9
- [62] Tong, Lingwei, Robert W. Lindeman, and Holger Regenbrecht: *Viewer’s role and viewer interaction in cinematic virtual reality*. Comput., 10:66, 2021. 8, 9, 18
- [63] Rossi, Silvia and Laura Toni: *Navigation-aware adaptive streaming strategies for omnidirectional video*. 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSp), pages 1–6, 2017. 10
- [64] Jerald, Jason, Tabitha C. Peck, Frank Steinicke, and Mary C. Whitton: *Sensitivity to scene motion for phases of head yaws*. In *APGV ’08*, 2008. 10
- [65] Litleskare, Sigbjørn and Giovanna Calogiuri: *Camera stabilization in 360° videos and its impact on cyber sickness, environmental perceptions, and psychophysiological responses to a simulated nature walk: A single-blinded randomized trial*. Frontiers in Psychology, 10, 2019. 10, 34
- [66] Serrano, Ana, Daniel Martin, Diego Gutierrez, Karol Myszkowski, and Belen Masia: *Imperceptible manipulation of lateral camera motion for improved virtual reality applications*. ACM Trans. Graph., 39(6), nov 2020, ISSN 0730-0301. <https://doi.org/10.1145/3414685.3417773>. 10
- [67] Hoßfeld, Tobias, Poul E. Heegaard, Martín Varela, Lea Skorin-Kapov, and Markus Fiedler: *From qos distributions to qoe distributions: a system’s perspective*. In *2020 6th IEEE Conference on Network Softwarization (NetSoft)*, pages 51–56, 2020. 11
- [68] YAMAZAKI, Tatsuya: *Quality of experience (qoe) studies: Present state and future prospect*. IEICE Transactions on Communications, E104.B(7):716–724, 2021. 11, 13
- [69] Shaikh, Junaid M., Markus Fiedler, and Denis Collange: *Quality of experience from user and network perspectives*. annals of telecommunications - annales des télécommunications, 65:47–57, 2010. 11
- [70] Kasteren, Anouk van, Kjell Brunnström, John Hedlund, and Chris Snijders: *Quality of experience of 360 video - subjective and eye-tracking assessment of encoding and freezing distortions*. Multim. Tools Appl., 81:9771–9802, 2022. 11

- [71] Anwar, Muhammad Shahid, Jing Wang, Wahab Khan, Asad Ullah, Sadique Ahmad, and Zesong Fei: *Subjective qoe of 360-degree virtual reality videos and machine learning predictions*. IEEE Access, 8:148084–148099, 2020, ISSN 2169-3536. <http://dx.doi.org/10.1109/ACCESS.2020.3015556>. 11
- [72] Hossfeld, Tobias, Anika Seufert, Frank Loh, Stefan Wunderer, and John Davies: *Industrial user experience index vs. quality of experience models*. IEEE Communications Magazine, 61(1):98–104, 2023. 11
- [73] Seufert, Anna Magdalena, Svenja Schröder, and Michael Seufert: *Delivering user experience over networks: Towards a quality of experience centered design cycle for improved design of networked applications*. SN Comput. Sci., 2:463, 2021. 11
- [74] Ookla: *Speedtest global index*. <https://www.speedtest.net/global-index>, visited on 2022-02-21. 12, 13
- [75] Cisco: *Cisco annual report 2018-2023*, March 2018. <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>. 12
- [76] Sun, Liyang, Fanyi Duanmu, Y. Liu, Yao Wang, Yinghua Ye, Hang Shi, and David H. Dai: *A two-tier system for on-demand streaming of 360 degree video over dynamic networks*. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 9:43–57, 2019. 12
- [77] Stockhammer, Thomas: *Dynamic adaptive streaming over http –: standards and design principles*. In *ACM SIGMM Conference on Multimedia Systems*, 2011. <https://api.semanticscholar.org/CorpusID:7097017>. 12
- [78] Zhang, Tong, Fengyuan Ren, Wenxue Cheng, Xiaohui Luo, Ran Shu, and Xiaolan Liu: *Modeling and analyzing the influence of chunk size variation on bitrate adaptation in DASH*. In *IEEE Conference on Computer Communications (INFOCOM)*, pages 1–9, 2017. 12
- [79] Graf, Mario, Christian Timmerer, and Christopher Mueller: *Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: Design, implementation, and evaluation*. In *Proc. ACM on Multimedia Systems Conference*, pages 261–271, 2017. 12
- [80] Chiariotti, Federico: *A survey on 360-degree video: Coding, quality of experience and streaming*. Comput. Commun., 177:133–155, 2021. 12
- [81] Shafi, Rabia, Wan Shuai, and Muhammad Usman Younus: *360-degree video streaming: A survey of the state of the art*. Symmetry, 12:1491, 2020. 12
- [82] Rondao Alface, Patrice, Jean François Macq, and Nico Verzijp: *Interactive omnidirectional video delivery: A bandwidth-effective approach*. Bell Labs Technical Journal, 16(4):135–147, 2012. 12

- [83] Corbillon, Xavier, Alisa Devlic, Gwendal Simon, and Jacob Chakareski: *Optimal set of 360-degree videos for viewport-adaptive streaming*. Proceedings of the 25th ACM international conference on Multimedia, 2017. 13
- [84] Netflix: *Internet connection speed recommendations*. <https://help.netflix.com/pt/node/306>, visited on 2018-10-22. 13
- [85] Konaszyński, Tomasz, Dawid Juszka, and Mikołaj Leszczuk: *Impact of the stimulus presentation structure on subjective video quality assessment*. Electronics, 12(22), 2023, ISSN 2079-9292. <https://www.mdpi.com/2079-9292/12/22/4593>. 13
- [86] Liu, Xiaochen, Wei Song, Qi He, Mario Di Mauro, and Antonio Liotta: *Speeding up subjective video quality assessment via hybrid active learning*. IEEE Transactions on Broadcasting, 69(1):165–178, 2023. 13
- [87] Huang, Xincheng, James Riddell, and Robert Xiao: *Virtual reality telepresence: 360-degree video streaming with edge-compute assisted static foveated compression*. IEEE Transactions on Visualization and Computer Graphics, 29(11):4525–4534, 2023. 13
- [88] Zhang, Junbin, Yixiao Wang, Hamidreza Tohidypour, Mahsa T. Pourazad, and Panos Nasiopoulos: *A generative adversarial network based tone mapping operator for 4k hdr images*. In *2023 International Conference on Computing, Networking and Communications (ICNC)*, pages 473–477, 2023. 13
- [89] T. T. Tran, Huyen, Nam P. Ngoc, Cuong T. Pham, Yong Ju Jung, and Truong Cong Thang: *A subjective study on user perception aspects in virtual reality*. Applied Sciences, 9(16), 2019, ISSN 2076-3417. <https://www.mdpi.com/2076-3417/9/16/3384>. 13
- [90] Zheng, Guoquan and Liang Yuan: *A review of qoe research progress in metaverse*. Displays, 77:102389, 2023, ISSN 0141-9382. <https://www.sciencedirect.com/science/article/pii/S0141938223000227>. 13
- [91] Islam, Md Tariqul, Christian Esteve Rothenberg, and Pedro Henrique Gomes: *Predicting xr services qoe with ml: Insights from in-band encrypted qos features in 360-vr*. In *2023 IEEE 9th International Conference on Network Softwarization (NetSoft)*, pages 80–88, 2023. 13
- [92] Zhai, Guangtao and Xiongkuo Min: *Perceptual image quality assessment: a survey*. Science China Information Sciences, 63(11), Apr 2020, ISSN 1869-1919. <http://dx.doi.org/10.1007/s11432-019-2757-1>. 14
- [93] International Telecommunication Union: *Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications*, 2023. <https://www.itu.int/rec/T-REC-P.910>. 14, 35, 39
- [94] International Telecommunication Union: *ITU-T Recommendation P.913: Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment*, 2021. <https://www.itu.int/rec/T-REC-P.913>. 14, 15

- [95] International Telecommunication Union: *ITU-T Recommendation P.919: Subjective test methodologies for 360° video on head-mounted displays*, 2020. <https://www.itu.int/rec/T-REC-P.919/en>. 14, 15, 38
- [96] International Telecommunication Union: *ITU-T Recommendation G.1011: Reference guide to quality of experience assessment methodologies*, 2016. <https://www.itu.int/rec/T-REC-G.1011/en>. 14, 15
- [97] International Telecommunication Union: *ITU-T Recommendation G.1035: Influencing factors on quality of experience for virtual reality services*, 2021. <https://www.itu.int/rec/T-REC-G.1035>. 15
- [98] Kozamernik, F., V. Steinmann, Paola Sunna, and Emmanuel Wyckens: *Samvig—a new ebu methodology for video quality evaluations in multimedia*. *Smpte Motion Imaging Journal*, 114:152–160, 2005. 15
- [99] Freitas, Pedro Garcia, Alexandre F. Silva, Judith Redi, and Mylène C. Q. Farias: *Performance analysis of a video quality ruler methodology for subjective quality assessment*. *Journal of Electronic Imaging*, 27, 2018. 17
- [100] Ostaszewska, A. and S. Żebrowska-Łucyk: *The Method of Increasing the Accuracy of Mean Opinion Score Estimation in Subjective Quality Evaluation*, pages 315–329. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, ISBN 978-3-642-15687-8. https://doi.org/10.1007/978-3-642-15687-8_16. 17
- [101] Tominaga, Toshiko, Takanori Hayashi, Jun Okamoto, and Akira Takahashi: *Performance comparisons of subjective quality assessment methods for mobile video*. 2010 Second International Workshop on Quality of Multimedia Experience (QoMEX), pages 82–87, 2010. 17
- [102] Janowski, Lucjan and Margaret H. Pinson: *The accuracy of subjects in a quality experiment: A theoretical subject model*. *IEEE Transactions on Multimedia*, 17:2210–2224, 2015. 17
- [103] Huynh-Thu, Quan and Mohammed Ghanbari: *Modelling of spatio-temporal interaction for video quality assessment*. *Signal Process. Image Commun.*, 25:535–546, 2010. 17
- [104] Rouse, David M., Romuald Pépion, Patrick Le Callet, and Sheila S. Hemami: *Trade-offs in subjective testing methods for image and video quality assessment*. In *Electronic Imaging*, 2010. 17
- [105] Seufert, Michael: *Statistical methods and models based on quality of experience distributions*. *Quality and User Experience*, 2020. 17
- [106] Brunnström, Kjell and Marcus Barkowsky: *Statistical quality of experience analysis: on planning the sample size and statistical significance testing*. *Journal of Electronic Imaging*, 27, 2018. 17

- [107] Ćmiel, Bogdan, Jakub Nawała, Lucjan Janowski, and Krzysztof Rusek: *Generalised score distribution: Underdispersed continuation of the beta-binomial distribution*. ArXiv, abs/2204.10565, 2022. 17
- [108] Sagot-Duvaouroux, Rémi, François Garnier, and Rémi Ronfard: *(re-)framing virtual reality*. In *WICED@Eurographics/EuroVis*, 2022. 18
- [109] Bordwell, David: *On the history of film style*. Harvard University Press, 1997. 18
- [110] Smith, Tim and John Henderson: *Attentional synchrony in static and dynamic scenes*. *Journal of Vision*, 8(6):773–773, 2008. 18
- [111] Kroma, Assem: *The technical dilemmas of creative design and rapid prototyping for immersive storytelling*. *Creativity and Cognition*, 2022. 18
- [112] Carpio, Rudy and James R. Birt: *The role of the embodiment director in virtual reality film production*. *Creative Industries Journal*, 15:189 – 198, 2022. 18
- [113] Pope, Vanessa C., Robert Dawes, Florian Schweiger, and Alia Sheikh: *The geometry of storytelling: Theatrical use of space for 360-degree videos and virtual reality*. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017. 19
- [114] Fearghail, Colm O., Emin Zerman, Sebastian B. Knorr, Fang Yi Chao, and Aljosa Smolic: *Use of saliency estimation in cinematic vr post-production to assist viewer guidance*. In *Proceedings of the 23rd Irish Machine Vision and Image Processing conference (IMVIP 2021)*, 2021. https://v-sense.scss.tcd.ie/wp-content/uploads/2021/09/IMVIP__saliency_post_production_guidance.pdf. 18, 20
- [115] Tong, Lingwei, Robert W. Lindeman, and Holger Regenbrecht: *Adaptive playback control: A framework for cinematic vr creators to embrace viewer interaction*. *Frontiers in Virtual Reality*, 2, 2022, ISSN 2673-4192. <https://www.frontiersin.org/articles/10.3389/frvir.2021.798306>. 20
- [116] Kjær, Tina, Christoffer B Lillelund, Mie Moth-Poulsen, Niels C Nilsson, Rolf Nordahl, and Stefania Serafin: *Can you cut it?: an exploration of the effects of editing in cinematic virtual reality*. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, page 4. ACM, 2017. 21
- [117] Speicher, Marco, Christoph Rosenberg, Donald Degraen, Florian Daiber, and Antonio Krüger: *Exploring visual guidance in 360-degree videos*. *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, 2019. 21
- [118] Montagud, Mario, Pilar Orero, and Anna Matamala: *Culture 4 all: accessibility-enabled cultural experiences through immersive vr360 content*. *Personal and Ubiquitous Computing*, 24(6):887–905, 2020. 21

- [119] Pavel, Amy, Björn Hartmann, and Maneesh Agrawala: *Shot orientation controls for interactive cinematography with 360 video*. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 289–297. ACM, 2017. 21
- [120] Rothe, Sylvia, Daniel Buschek, and Heinrich Hußmann: *Guidance in cinematic virtual reality-taxonomy, research status and challenges*. *Multimodal Technol. Interact.*, 3:19, 2019. 21, 22
- [121] Harley, Daniel, Aneesh P. Tarun, Daniel Germinario, and Ali Mazalek: *Tangible vr: Diegetic tangible objects for virtual reality narratives*. In *Proceedings of the 2017 Conference on Designing Interactive Systems, DIS '17*, page 1253–1263, New York, NY, USA, 2017. Association for Computing Machinery, ISBN 9781450349222. <https://doi.org/10.1145/3064663.3064680>. 21
- [122] Bork, Felix, Christian Schnelzer, Ulrich Eck, and Nassir Navab: *Towards efficient visual guidance in limited field-of-view head-mounted displays*. *IEEE Transactions on Visualization and Computer Graphics*, 24:2983–2992, 2018. 21
- [123] Harada, Yuki and Junji Ohyama: *Quantitative evaluation of visual guidance effects for 360-degree directions*. *Virtual Reality*, 2021. 22
- [124] Hu, Hou Ning, Yen Chen Lin, Ming Yu Liu, Hsien Tzu Cheng, Yung Ju Chang, and Min Sun: *Deep 360 pilot: Learning a deep agent for piloting through 360° sports videos*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1396–1405, 2017. 22
- [125] Wang, Miao, Yi Jun Li, Wenxuan Zhang, Christian Richardt, and Shimin Hu: *Transitioning360: Content-aware nfov virtual camera paths for 360° video playback*. 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pages 185–194, 2020. 22
- [126] Danieau, Fabien, Antoine Guillo, and Renaud Doré: *Attention guidance for immersive video content in head-mounted displays*. In *Virtual Reality (VR), 2017 IEEE*, pages 205–206. IEEE, 2017. 22
- [127] Mateer, John W.: *Directing for cinematic virtual reality: how the traditional film director’s craft applies to immersive environments and notions of presence*. *Journal of Media Practice*, 18:14 – 25, 2017. 22
- [128] Beck, Thomas and Sylvia Rothe: *Applying diegetic cues to an interactive virtual reality experience*. 2021 IEEE Conference on Games (CoG), pages 1–8, 2021. 22
- [129] Dorado, José L. and Pablo A. Figueroa: *Methods to reduce cybersickness and enhance presence for in-place navigation techniques*. 2015 IEEE Symposium on 3D User Interfaces (3DUI), pages 145–146, 2015. 22
- [130] Weech, Séamas, Sophie Kenny, and Michael Barnett-Cowan: *Presence and cybersickness in virtual reality are negatively related: A review*. *Frontiers in Psychology*, 10, 2019. 22

- [131] Tian, Feng, Minlei Hua, Tingting Zhang, and Wenrui Zhang: *Spatio-temporal editing method and application in virtual reality video*. 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 1:2290–2294, 2020. 22
- [132] Shi, Rongkai, Hai Ning Liang, Yuehua Wu, Difeng Yu, and Wenge Xu: *Virtual reality sickness mitigation methods*. Proceedings of the ACM on Computer Graphics and Interactive Techniques, 4:1 – 16, 2021. 22
- [133] Lin, Yun Xuan, Rohith Venkatakrishnan, Roshan Venkatakrishnan, Elham Ebrahimi, Wen Chieh Lin, and Sabarish V. Babu: *How the presence and size of static peripheral blur affects cybersickness in virtual reality*. ACM Transactions on Applied Perception (TAP), 17:1 – 18, 2020. 22
- [134] Yildirim, Caglar: *Don't make me sick: investigating the incidence of cybersickness in commercial virtual reality headsets*. Virtual Reality, 24:231–239, 2019. 22
- [135] Teixeira, Joel Anthony and Stephen A. Palmisano: *Effects of dynamic field-of-view restriction on cybersickness and presence in hmd-based virtual reality*. Virtual Reality, 25:433–445, 2020. 22
- [136] Wang, Miao, Yi Jun Li, Wenxuan Zhang, Christian Richardt, and Shimin Hu: *Transitioning360: Content-aware nfov virtual camera paths for 360° video playback*. 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pages 185–194, 2020. 22, 23
- [137] Su, Yu Chuan, Dinesh Jayaraman, and Kristen Grauman: *Pano2vid: Automatic cinematography for watching 360° videos*. ArXiv, abs/1612.02335, 2017. 22
- [138] Kang, Kyoungkook and Sunghyun Cho: *Interactive and automatic navigation for 360° video playback*. ACM Transactions on Graphics (TOG), 38:1 – 11, 2019. 22
- [139] Lee, Sangho, Jinyoung Sung, Youngjae Yu, and Gunhee Kim: *A memory network approach for story-based temporal summarization of 360° videos*. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 1410–1419, 2018. 22, 24
- [140] Lai, Wei Sheng, Yujia Huang, Neel Joshi, Chris Buehler, Ming Hsuan Yang, and Sing Bing Kang: *Semantic-driven generation of hyperlapse from 360 degree video*. IEEE Transactions on Visualization and Computer Graphics, 24:2610–2621, 2018. 22
- [141] Xu, Ran, Haoliang Wang, Stefano Petrangeli, Viswanathan Swaminathan, and Saurabh Bagchi: *Closing-the-loop: A data-driven framework for effective video summarization*. 2020 IEEE International Symposium on Multimedia (ISM), pages 201–205, 2020. 22, 24
- [142] Cao, Ruochen, James A. Walsh, Andrew Cunningham, Carolin Reichherzer, Subrata Dey, and B. Thomas: *A preliminary exploration of montage transitions in cinematic virtual reality*. 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pages 65–70, 2019. 23

- [143] Garrido, Luis Eduardo, Maite Frías-Hiciano, María del Pilar Moreno-Jiménez, Gabriella Nicole Cruz, Zoilo Emilio García-Batista, Kiero Guerra-Peña, and Leonardo Adrián Medrano: *Focusing on cybersickness: pervasiveness, latent trajectories, susceptibility, and effects on the virtual reality experience*. *Virtual Reality*, pages 1 – 25, 2022. 23
- [144] Tian, Nana, Phil Lopes, and Ronan Boulic: *A review of cybersickness in head-mounted displays: raising attention to individual susceptibility*. *Virtual Reality*, 2022. 23
- [145] Eom, Hyojung, Kwanguk Kim, Sungmi Lee, Yeon Ju Hong, Jiwoong Heo, Jae Jin Kim, and Eunjoo Kim: *Development of virtual reality continuous performance test utilizing social cues for children and adolescents with attention-deficit/hyperactivity disorder*. *Cyberpsychology, behavior and social networking*, 22 3:198–204, 2019. 23
- [146] Sassatelli, Lucile, Marco Winckler, Thomas Fisichella, Antoine Dezarnaud, Julien Lemaire, Ramon Aparicio-Pardo, and Daniela Trevisan: *New interactive strategies for virtual reality streaming in degraded context of use*. *Computers & Graphics*, 86:27–41, 2020. 23
- [147] Elwardy, Majed, Hans Jürgen Zepernick, and Thi My Chinh Chu: *On the number of subjects needed for 360° video quality experiments: An sos based analysis*. 2022 14th International Conference on Quality of Multimedia Experience (QoMEX), pages 1–4, 2022. 25
- [148] Orduna, Marta, Pablo Pérez, Jesús Gutiérrez, and Narciso García: *Content-immersive subjective quality assessment in long duration 360-degree videos*. In *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, pages 73–78, 2023. 25
- [149] Elwardy, Majed, Yan Hu, Hans Jürgen Zepernick, Thi My Chinh Chu, and Veronica Sundstedt: *Comparison of acr methods for 360° video quality assessment subject to participants’ experience with immersive media*. 2020 14th International Conference on Signal Processing and Communication Systems (ICSPCS), pages 1–10, 2020. 25
- [150] Nehmé, Yana, Jean Philippe Farrugia, Florent Dupont, Patrick Le Callet, and Guillaume Lavoué: *Comparison of subjective methods, with and without explicit reference, for quality assessment of 3d graphics*. *ACM Symposium on Applied Perception* 2019, 2019. 25
- [151] Borchert, Kathrin, Anika Seufert, Edwin Gamboa, Matthias Hirth, and Tobias Hofffeld: *In vitro vs in vivo: does the study’s interface design influence crowdsourced video qoe?* *Quality and User Experience*, 6(1):1, 11/02 2020, ISSN 2366-0147. <https://doi.org/10.1007/s41233-020-00041-2>. 39
- [152] Sevinç, Volkan and Mehmet Ilker Berkman: *Psychometric evaluation of simulator sickness questionnaire and its variants as a measure of cybersickness in consumer virtual environment*. *Applied ergonomics*, 82:102958, 2019. 39, 59

- [153] Rahimi, Kasra, Colin Banigan, and Eric D. Ragan: *Scene transitions and teleportation in virtual reality and the implications for spatial awareness and sickness*. IEEE Transactions on Visualization and Computer Graphics, 26:2273–2287, 2020. 42
- [154] Schober, Patrick, Christa Boer, and Lothar A. Schwarte: *Correlation coefficients: Appropriate use and interpretation*. Anesthesia & Analgesia, 126:1763–1768, 2018. 43
- [155] Rondon, Miguel Fabian Romero, Lucile Sassatelli, Ramon Aparicio-Pardo, and Frédéric Precioso: *Track: A new method from a re-examination of deep architectures for head motion prediction in 360-degree videos*. IEEE transactions on pattern analysis and machine intelligence, PP, 2021. 51
- [156] Rossi, Silvia, Francesca De Simone, Pascal Frossard, and Laura Toni: *Spherical clustering of users navigating 360° content*. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 4020–4024, 2019. 51, 54
- [157] LaValle, Steven M., Anna Yershova, Max Katsev, and Michael Antonov: *Head tracking for the oculus rift*. 2014 IEEE International Conference on Robotics and Automation (ICRA), pages 187–194, 2014. 51
- [158] Younis, Ola, Waleed Al-Nuaimy, Mohammad H. Alomari, and Fiona Rowe: *A hazard detection and tracking system for people with peripheral vision loss using smart glasses and augmented reality*. International Journal of Advanced Computer Science and Applications, 2019. 54
- [159] Kelly, JW, TA Doty, M Ambourn, and LA Cherep: *Distance perception in the oculus quest and oculus quest 2*. Frontiers in Virtual Reality, 3:850471, 2022. 54
- [160] Corbillon, Xavier, F. D. Simone, and G. Simon: *360-degree video head movement dataset*. Proceedings of the 8th ACM on Multimedia Systems Conference, 2017. 57
- [161] Hossfeld, Tobias, Raimund Schatz, and Sebastian Egger: *Sos: The mos is not enough!* 2011 Third International Workshop on Quality of Multimedia Experience, pages 131–136, 2011. 68

Appendix A

Free and Informed Consent Term

A.1 Free and Informed Consent Term

The Digital Signal Processing Group (GPDS) laboratory invites you to participate in the research entitled “Fade Rotation: Attention-Driving Transition Mechanism for User-Centric Content-Adaptive Virtual Reality Movies”. The expected benefit of this research is to understand the degrees of acceptability of a new attention-driving mechanism in 360-degree videos that in the future should integrate a content adaptation system for optimizing the experience of viewers of immersive videos. The survey is designed to be agile and completely safe for participants. You will always be accompanied by a researcher and the instructions intend to make your participation as simple as possible.

To participate, please read the information below carefully and check “Yes” to consent to your participation and start your session, or check “No” if you do not wish to participate.

1. **Procedure:** This experiment is scheduled to last 30 minutes, and you will be shown 36 videos of 30 seconds each, giving scores on the watching videos, and answer a pre- and post-questionnaire.
2. **Possible discomfort:** Eventually while watching the immersive videos, you may experience some initial discomfort that diminishes with time. If you need to stop at any moment, just call the researcher in charge. Since one of the measures taken will be the level of discomfort, we ask that you avoid pausing the video before the end, as this will mean losing data. However, should you wish to quit at any time, this will not cause any harm to you.
3. **Benefits and costs:** Your participation in this study will contribute important results to research in the areas of computer science and immersive media. You will

not incur any expenses or burdens from your participation in the study, nor will you receive any kind of reimbursement or gratuity for participating in the research, which is entirely voluntary. This is entirely voluntary, with the exception of those participants who request a transportation stipend.

4. **Privacy and confidentiality:** All information collected in this study is confidential and your name and that of your organization will not be identified in any way. Every effort will be made during data collection to ensure your privacy and anonymity. The data collected during the study is strictly for research activities, following the procedures and rules of the UnB's ethical committee.
5. **Safety protocols for performing subject experiments during the pandemic of Covid-19:** Our experiment will be conducted respecting the safety protocol of the GPDS/ENE/UnB laboratory.

The researchers responsible for the study can provide any clarification about the study by contacting the following e-mail addresses:

- Experimenter (contact): Lucas dos Santos Althoff, 190051612@aluno.unb.br - PPGI/UnB
- Supervisor: Mylène C. Q. Farias - PPGI/UnB

Do you think you are sufficiently informed about the research that will be carried out and do you freely and spontaneously agree to participate as a collaborator?

NO () YES ()

A.2 Laboratory setup of the experiments

The laboratory setup consisted by a swivel chair, a dedicated router, a server PC and the HMD. In the first experiment, the participants used the Oculus Rift S, while in the second experiment they used the Meta Quest 2. Figure A.1 shows two participants wearing the two devices. The safety protocols were carried carefully with participants wearing a face-mask and the sanitation of the complete equipment were applied at the beginning and at the end of each session.



(a) Setup with Oculus Quest 2. (b) Setup with Oculus Rift S.

Figure A.1: Participant wearing the HMD, and watching a experiment's video.



(a) Setup in room 1. (b) Setup in room 2.

Figure A.2: Participant wearing the HMD, and watching a experiment's video.

Appendix B

Mono360 Details

B.1 Survey interface

Thereafter we present the interface as like it is rendered in the HMD browser of the Mono360 software¹. Those pages are the interfaces for user interact during the stages of the experiment. All those pages are rendered for participants inside the HMD's browser, and users interact with it with the controller.

B.2 Recruitment Page

Figure B.8 shows the recruiting page that was hosted in a public domain, and prepared to facilitate the task of scheduling and recruiting participants.

¹The official repository of Mono360 can be found at: <https://gitlab.com/gpds-unb/mono360/-/wikis/Running-Mono360>

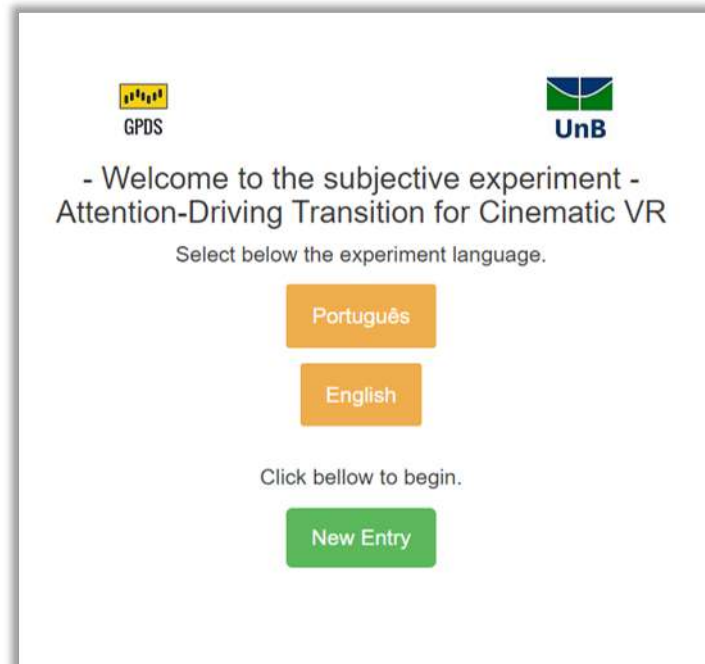


Figure B.1: Welcome page.

New Entry

Name *

Age *

Gender *

Select a gender

Nationality *

Nationality

Education Level *

Education Level

Profession *

Email *

Have you used any Head Mounted Display (HMD) before? *

No

If you marked yes in last question, how many times did you use HMD?

Select a value

Do you have normal, or corrected to normal, vision? *

No

Do you wear glasses or contact lenses? *

No

Next

Figure B.2: Pre-questionnaire.

Free and Informed Consent Term

The Digital Signal Processing Group (GPDS) laboratory invites you to participate in the research entitled "Fade Rotation: Attention-Driving Transition Mechanism for User-Centric Content-Adaptive Virtual Reality Movies". The expected benefit of this research is to understand the degrees of acceptability of a new attention-driving mechanism in 360-degree videos that in the future should integrate a content adaptation system for optimizing the experience of viewers of immersive videos. The survey is designed to be agile and completely safe for participants. You will always be accompanied by a researcher and the instructions intend to make your participation as simple as possible.

To participate, please read the information below carefully and check "Yes" to consent to your participation and start your session, or check "No" if you do not wish to participate.

1) Procedure
This experiment is scheduled to last 30 minutes, and you will be shown 36 videos of 30 seconds each, given notes on the watching videos, and answer a pre- and post-questionnaire.

2) Possible discomfort
Eventually while watching the immersive videos, you may experience some initial discomfort that diminishes with time. If you need to stop at any moment, just call the researcher in charge. Since one of the measures taken will be the level of discomfort, we ask that you avoid pausing the video before the end, as this will mean losing data. However, should you wish to quit at any time, this will not cause any harm to you.

3) Benefits and costs
Your participation in this study will contribute important results to research in the areas of computer science and immersive media. You will not incur any expenses or burdens from your participation in the study, nor will you receive any kind of reimbursement or gratuity for participating in the research, which is entirely voluntary. This is entirely voluntary, with the exception of those participants who request a transportation stipend.

4) Privacy and confidentiality
All information collected in this study is confidential and your name and that of your organization will not be identified in any way. Every effort will be made during data collection to ensure your privacy and anonymity. The data collected during the study is strictly for research activities, following the procedures and rules of the UnB's ethical committee.

5) Safety protocols for performing subject experiments during the pandemic of Covid-19
Our experiment will be conducted respecting the safety protocol of the GPDS/ENE/UnB laboratory.

The researchers responsible for the study can provide any clarification about the study by contacting the following e-mail addresses:
Responsible researcher: Lucas dos Santos Athor - 190051612@aluno.unb.br - PPGI/UnB
Advisor Professor: Mylene C. G. Farias - mylene@unb.com - PPGI/UnB

Do you think you are sufficiently informed about the research that will be carried out and do you freely and spontaneously agree to participate as a collaborator?

Figure B.3: Free and Informed Consent Term.

Welcome to your experiment session

Thank you for collaborating with our study.

The experiment is composed of seven phases with the following expected duration:

I	Instructions (3 min)
II	Training (2 min)
III	1st video sequence (9 min)
IV	Rest (5 min)
V	2nd video sequence (9 min)
VI	Post-questionnaire (2 min)

After starting the data collection, the researcher will give you all instructions needed to complete the experiment.

Figure B.4: Introduction of the training.

Instructions

At this point you are using a device (virtual reality head-mounted device) that allows you to control your viewing angle and interact with the virtual space. If you wear glasses, please wear them under the device to ensure your visual acuity. You will be required to watch the videos while seated on the subject chair.


Basic procedure

When you start your stimulus session you will watch a sequence of videos and give notes at the end of each video. To provide your scores you will have to point the beam to the desired score and click the button on the controller, to register the chosen score.

Assessing videos

After each video you will score three scales asking about your sensation of presence, your discomfort feeling, and your perceived experience. For each factor studied there will be five possible scores and each one has its specific description.

Equipment



Oculus Rift S

Giving Scores

Presence score
 "To which extent did you feel present in the virtual environment as if you were really there?"
 (1) Nothing (2) Very low (3) Reasonably (4) Very much (5) Entirely

Comfort score
 "Are you feeling any sickness or discomfort now?"
 (1) Unbearable (2) Unpleasant (3) Uncomfortable (4) Light effects (5) No Problem

Experience score
 "Give a score for your experience watching the previous video!"
 (1) Bad (2) Fair (3) Fair (+) (4) Good (5) Excellent

Caution while giving scores
 Please enter your score only for the last watched video, do not take into consideration the comparison with other previously watched videos. Note that in this experiment you will observe versions of the same video processed with different transition effects.

Training phase

Before you begin the experimental session, we will conduct a training where you will watch a single video and score it to make sure you understand the task of the experiment.
 The task to be performed in the training session will be the same as that performed in the main experiment. This training can be repeated as many times as you wish.

Start Training

Figure B.5: Instructions

Main Session

Now we will start the main experiment.
 Feel free to take a break in the middle of the experiment if you want to, the system will recover where you stopped if you type the same e-mail on the "New Entry" form.

After you evaluate the video, there will be no way to correct the score.
 The experiment has a duration of 30 minutes, without pauses.

Click on the button to start the experiment.

Start

Figure B.6: Session starting page.

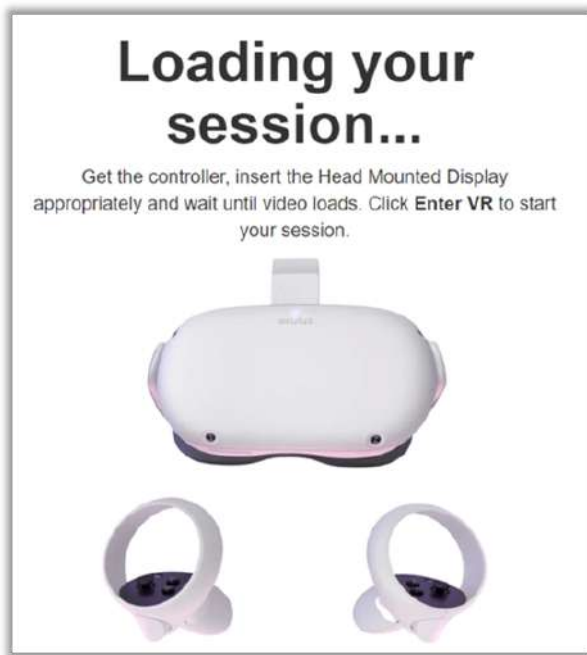


Figure B.7: Loading session.

PROJETO PPS-360 AGENDAMENTO

Bem vindo(a) à página de agendamento do experimento!
 Mecanismo de Condução da Atenção para Vídeos de Realidade Virtual Adaptáveis ao Conteúdo

Welcome to the experiment scheduling page!
 Attention Conduction Mechanism for Content-adaptive Virtual Reality Videos

-> **Onde?** Laboratório GPDS, prédio 565 11 - Entre a faculdade de Tecnologia e o Departamento de Artes (veja no mapa abaixo).
 -> **Nais informações?** (61)999965391 [Zap! ou Procurador Responsável](#)
 -> **Quando?** Entre as datas 07-29 de julho.
 -> **Como participar?** Agende a sua sessão no botãoável!

--- ENGLISH ---
 -> **Where?** GPDS, 565 11 building, Between Technology Faculty and the Theatre Department (Take a look at the map below).
 -> **Further informations?** (61)999965391 [Send a message in WhatsApp](#)
 -> **When?** From 07 to 29 de July.
 -> **How to participate?** Schedule your session at the violet button!

RESERVAR UM HORARIO (CLIQUE AQUI)

Descrição
 O laboratório de Processamento Digital de Sinais (GPDS), convida você a participar da pesquisa intitulada "Mecanismo de Condução da Atenção para Vídeos de Realidade Virtual Adaptáveis ao Conteúdo". O benefício esperado com esta pesquisa é o de compreender o nível de qualidade nos mecanismos de condução de atenção em vídeos 360-graus, que integrando um vídeo-player adaptativo para otimização da transmissão de de vídeos imersivos.
 A sua sessão vai durar menos de 50 minutos, onde você observará variados vídeos 360-graus.
 Colabore com o desenvolvimento de soluções em Realidade Virtual da Universidade de Brasília.

Description
 The Digital Signal Processing Laboratory (GPDS) invites you to participate in the research entitled "Attention Conduction Mechanism for Content-adaptive Virtual Reality Videos". The expected benefit of this research is to understand the degrees of quality of experience of a new arriving attention-driving mechanism for 360-degree videos, which will integrate a content adaptation system for optimizing the video streaming.
 Your session will mean 50 minutes, when you will watch several 360-degree videos.
 Come and support the development of VR solution at UNB!

Localização da aplicação do experimento (Experiment's localization)
 O experimento ocorrerá em nosso laboratório, localizado na Universidade de Brasília. The experiment will take place in our laboratory, located at the University of Brasília.

Motivação da pesquisa
 O ritmo de consumo global de conteúdo em vídeo aumentou 40% em 2016, e espera-se que os usuários tenham acesso a cada vez mais vídeos em todo o mundo. O objetivo desta pesquisa é propor um mecanismo de desenvolvimento de vídeos de realidade virtual de 360 graus adaptativos para aplicação em sessões de realidade virtual em ambientes de realidade virtual. A pesquisa será realizada em parceria com o laboratório de Realidade Virtual da Universidade de Brasília.

Research motivation
 We aim to face the growth of 360 degree videos consumption rate, reaching massive numbers of viewers worldwide. The aim of this study is to propose an attention-driving mechanism-based solution for the acceptance parameters to implement 360 degree videos adaptively. This will contribute with the research of University of Brasília in the field of Immersive Media. 2027. We will collaborate with our lab's partner and research in session with our team.

Responsabilidade pelo agendamento (responsibility for the equipment)
 Dr. Lucas B. de Aguiar | E-mail: lbas@unb.br | Telefone: (61) 3052-1010 | lbas@unb.br

Realidade Virtual
vídeos 360°
 30 Vídeos
1h de vídeos
 50 minutos
Tempo máximo

Figure B.8: Recruitment page