

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**AVALIAÇÃO DA QUALIDADE DE TÉCNICAS DE
INTERAÇÃO EM AMBIENTES VIRTUAIS
IMERSIVOS UTILIZANDO MEDIDAS
FISIOLÓGICAS**

Rafael Rieder

Tese apresentada como requisito parcial para a obtenção do grau de Doutor em Ciência da Computação na Pontifícia Universidade Católica do Rio Grande do Sul.

Orientador: Dr. Márcio Sarroglia Pinho

Porto Alegre

2011

Dados Internacionais de Catalogação na Publicação (CIP)

R549a Rieder, Rafael
Avaliação da qualidade de técnicas de interação em ambientes virtuais imersivos utilizando medidas fisiológicas / Rafael Rieder. – Porto Alegre, 2011.
178 f.

Tese (Doutorado) – Fac. de Informática, PUCRS.
Orientador: Prof. Dr. Márcio Serroglia Pinho.

1. Informática. 2. Realidade Virtual. 3. Fisiologia. I. Pinho, Márcio Serroglia. II. Título.

CDD 006.6

**Ficha Catalográfica elaborada pelo
Setor de Tratamento da Informação da BC-PUCRS**



Pontifícia Universidade Católica do Rio Grande do Sul
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

TERMO DE APRESENTAÇÃO DE TESE DE DOUTORADO

Tese intitulada "Avaliação da Qualidade de Técnicas de Interação em Ambientes Virtuais Imersivos Utilizando Medidas Fisiológicas", apresentada por Rafael Rieder, como parte dos requisitos para obtenção do grau de Doutor em Ciência da Computação, Sistemas Interativos e de Visualização, aprovada em 28/04/2011 pela Comissão Examinadora:

Prof. Dr. Márcio Sarroglia Pinho -
Orientador

PPGCC/PUCRS

Prof. Dr. Christian Haag Kristensen -

PPGP/PUCRS

Profa. Dra. Milene Selbach Silveira -

PPGCC/PUCRS

Profa. Dra. Fátima de Lourdes dos Santos Nunes Marques -

EACH/USP

Profa. Dra. Maria Helena Itaqui Lopes -

FAMED/PUCRS

Homologada em...../...../....., conforme Ata No. 10.... pela Comissão Coordenadora.

Prof. Dr. Fernando Luís Dotti
Coordenador.

PUCRS

Campus Central

Av. Ipiranga, 6681 - P. 32 - sala 507 - CEP: 90619-900

Fone: (51) 3320-3611 - Fax (51) 3320-3621

E-mail: ppgcc@pucrs.br

www.pucrs.br/facin/pos

AGRADECIMENTOS

Primeiramente, gostaria de agradecer a Deus por iluminar meu caminho durante a jornada da vida. Com fé e determinação, foi possível alcançar mais um objetivo.

Um agradecimento todo especial e carinhoso a minha esposa, amiga e companheira Dalise, pelo apoio e incentivo constante, tanto nas horas de bonança, como nos momentos difíceis.

Agradeço também aos meus pais, Adelar e Vera, pela educação, pelos ensinamentos de vida e por serem exemplos de caráter e dignidade. Aos meus irmãos, Rômulo e Rolf, por serem meus melhores amigos. Ao Rômulo, um agradecimento especial pelo suporte e apoio nestas constantes viagens a Porto Alegre. Ao Rolf, pelas longas e boas conversas. Aos meus sogros, Olavo e Maira, uma gratidão especial.

Agradeço ao meu orientador, Márcio Pinho, pela oportunidade e experiência em ser seu aluno de Mestrado e de Doutorado. É um profissional competente, que me ajudou a percorrer os caminhos e indicou os atalhos da vida acadêmica. Estendo também meus agradecimentos aos demais professores e funcionários do Programa de Pós-Graduação em Ciência da Computação (PPGCC) e Faculdade de Informática (FACIN) da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS).

Alguns professores merecem um agradecimento especial, por terem contribuído para a concepção deste trabalho. Ao professor Christian Kristensen, pelo apoio durante o projeto, com o qual compartilhamos dúvidas e acertos no uso de medidas fisiológicas. As professoras Maria Helena Lopes e Thais Russomano, pelas contribuições na área médica para interpretação dos sinais fisiológicos. Ao professor João Feliz, pelos ensinamentos e auxílio na parte de análise estatística. As professoras Milene Silveira e Fátima Nunes, pelas contribuições à pesquisa durante a apresentação da proposta de tese.

Agradeço também ao Grupo de Tecnologia em Computação Gráfica (Tecgraf), em especial a Antonio Scuri, pela amizade e pela confiança depositada em meu trabalho como desenvolvedor C/C++. Esta parceria foi fundamental para a conquista deste desafio.

Agradeço aos colegas do Grupo de Realidade Virtual (GRV), pelo aprendizado, amizade e alegria. São várias pessoas que contribuíram e continuam trabalhando para o avanço das pesquisas deste grupo. Em ordem cronológica, agradeço: Felipe Bacim, André Trombetta, Mauro Charão, Régis Kopper, Genilson Korogi, João Prauchner, Luciane Fraga, Pedro Xerxenesky, Fabrício Pretto, André Tomasi, Marcelo Donadussi, Laura Flach, Adonai Canêz, Jefferson Pinheiro, Diego Pereira, Tarso Hoffmeister, Daniel Camozzato, João Mossmann, Virgínia Andersson, Tales Bogoni, Pedro Vanzella, Lucas

Colnaghi, Filipi Vianna, Eduardo Fabris, Diogo de Lima, Otávio Gomes, Bruno Moreira, Silvair Frazão, Gustavo Goedert, Alberto Dietrich e Conrado Testa.

Agradeço aos amigos do PPGCC, Leandro Bogoni, Ana Winck, Karina Machado, Néelson Tenório Júnior, Eduardo Link, Fernando Marson e Thiago Silva, pela troca de ideias e ajuda mútua. Aos meus amigos da Psicologia da PUCRS, bolsistas e orientandos do professor Christian, agradeço pelo constante apoio. Um agradecimento especial também a todos os usuários que participaram dos testes, os quais contribuíram para o avanço da ciência.

Um agradecimento aos amigos que residem em Porto Alegre, os quais também me acolheram durante o período de doutoramento: Alexandre Zanatta, Gustavo Pereira e Matheus Sobiesiak. Da mesma forma, agradeço as instituições Faculdade Anglicana de Erechim (FAE), Faculdade Meridional (IMED), Universidade Regional Integrada do Alto Uruguai e das Missões (URI) e Universidade de Passo Fundo (UPF), pelas oportunidades como docente de Ensino Superior.

Finalmente, agradeço aos ensinamentos do Escotismo, os quais contribuíram na formação do meu caráter como cidadão.

AVALIAÇÃO DA QUALIDADE DE TÉCNICAS DE INTERAÇÃO EM AMBIENTES VIRTUAIS IMERSIVOS UTILIZANDO MEDIDAS FISIOLÓGICAS

RESUMO

Medidas fisiológicas, como frequência cardíaca (HR) e condutância de pele (SC), oferecem dados objetivos a respeito dos estados físico e emocional do usuário durante o processo interativo em ambientes virtuais. Isto pode ser útil, por exemplo, na avaliação de aspectos qualitativos de técnicas de interação tridimensionais, sem depender de dados subjetivos e considerando as diferentes reações do usuário. Com base nisto, o objetivo deste trabalho é apresentar uma nova metodologia para avaliar a qualidade de técnicas de interação em ambientes virtuais imersivos, fundamentada no estudo do relacionamento entre métricas de usabilidade e medidas fisiológicas utilizando análise multivariada de dados. Para tanto, esta metodologia define um protocolo de testes, um processo de normalização de medidas fisiológicas, e um conjunto de métodos estatísticos capazes de relacionar múltiplas medidas. Aplicando esta metodologia, um estudo de caso comparativo entre duas técnicas de interação (*ray-casting* e HOMER) mostra resultados fortemente significantes, apontando para a variabilidade da frequência cardíaca, medida pelo número de diferenças dos intervalos entre batimentos consecutivos superiores a 50ms (NN50), como um potencial índice para determinar o tempo gasto para desempenhar uma tarefa. O estudo também mostra que as medidas de HR e SC refletem o desempenho do usuário durante uma tarefa. Apesar desses resultados, o trabalho revela que medidas fisiológicas ainda não podem ser consideradas como substitutas de métricas de usabilidade, mas podem ser úteis no processo de interpretação e compreensão destas. Discussões também determinam a necessidade de aprofundar o estudo das medidas fisiológicas na predição de resultados, como forma de estabelecer diretrizes futuras baseadas numa associação bem definida entre comportamentos do organismo e ações desempenhadas nas interfaces tridimensionais.

Palavras-chave: avaliação de usabilidade, medidas fisiológicas, técnicas de interação.

QUALITY ASSESSMENT OF INTERACTION TECHNIQUES IN IMMERSIVE VIRTUAL ENVIRONMENTS USING PHYSIOLOGICAL MEASURES

ABSTRACT

Physiological measures, such as heart rate (HR) and skin conductance (SC), offer objective data about the user stress during the interaction process in virtual environments. This could be useful, for instance, to evaluate qualitative aspects of 3D interaction techniques without relying solely on subjective data and considering the different user responses. With this in mind, the goal of this project is to present a new methodology for quality assessment of the interaction techniques in immersive virtual environments, based on the study of the relationship between usability metrics and physiological measures using multivariate data analysis. Our methodology defines a testing protocol, a normalization procedure and statistical methods to correlate multiple measures. Applying this methodology, a case study comparison between two interaction techniques (ray-casting and HOMER) shows highly significant results, pointing to heart rate variability, as measured by the number of interval differences of successive inter-beat intervals greater than 50 ms (NN50), as a potential index to measure the time spent to perform a task. The study also shows that HR and SC measures reflect the user's performance during an interaction task. In spite of these results, this work reveals that physiological measures still cannot be considered as substitutes of usability metrics, but may be useful in the interpretation and understanding process of them. Discussions also indicate the need of further studies about the use of physiological measures in predicting outcomes, as a way to establish future guidelines based on a well-defined association between human body behaviors and the actions performed in 3D user interfaces.

Keywords: usability evaluation, physiological measures, interaction techniques.

LISTA DE FIGURAS

Figura 1. Circuito básico de um dispositivo que mede SC.....	21
Figura 2. Medição da SC.	22
Figura 3. Variantes de medição da HR utilizando um eletrocardiógrafo	23
Figura 4. Exemplo de fotopletismógrafo [KIR02].	24
Figura 5. Ciclo cardíaco normal, com seu conjunto de ondas	24
Figura 6. Dispositivo de monitoramento da frequência respiratória e formas de fixação. ...	27
Figura 7. Pupilômetro de baixo custo utilizando webcams [MIR10].	28
Figura 8. Exemplos de dispositivos de rastreamento de olhar [ASL10].	29
Figura 9. Dispositivo de monitoramento da temperatura da pele.	30
Figura 10. Visão lateral do AV e usuário vestindo dispositivos fisiológicos e de RV	34
Figura 11. Bar virtual e usuário interagindo na CAVE, usando dispositivos de <i>biofeedback</i>	37
Figura 12. Aplicação da rua virtual, à esquerda, e usuário interagindo na CAVE.....	39
Figura 13. Disposição dos sensores de <i>biofeedback</i> e <i>gamepad</i> utilizado no experimento	42
Figura 14. Jogo “Luxor: Amun Rising” e disposição dos equipamentos durante os testes	45
Figura 15. Exibição do cubo aramado, quando o objeto está sendo interceptado pelo raio.	55
Figura 16. Exibição do raio virtual para seleção, e da mão virtual para manipulação utilizando a técnica HOMER.	57
Figura 17. Posição inicial do sofá (à esquerda) e posição final desejada (à direita).....	58
Figura 18. Posições inicial dos livros (à esquerda) e final desejada (à direita).....	59
Figura 19. Equipamentos fisiológicos e de RV utilizados.	60
Figura 20. Disposição correta dos sensores para monitoramento fisiológico.	61
Figura 21. Usuário com equipamentos fisiológicos e de RV.....	63
Figura 22. Distribuição das respostas do questionário pré-teste.	66
Figura 23. Desempenho entre grupos com <i>ray-casting</i> , considerando as variáveis baseadas no tempo.	67
Figura 24. Desempenho entre grupos com <i>ray-casting</i> , considerando as variáveis baseadas no número de tentativas.....	67
Figura 25. Desempenho entre grupos com <i>ray-casting</i> , variável “precisão”.....	68
Figura 26. Desempenho entre grupos com HOMER, considerando as variáveis baseadas no tempo.....	69

Figura 27. Desempenho entre grupos com HOMER, considerando as variáveis baseadas no número de tentativas.	69
Figura 28. Desempenho entre grupos com HOMER, considerando a variável “precisão”.	70
Figura 29. Desempenho entre as técnicas, considerando as variáveis baseadas no tempo.	70
Figura 30. Desempenho entre as técnicas, considerando as variáveis baseadas no número de tentativas.	71
Figura 31. Desempenho entre as técnicas, considerando a variável “precisão”.	71
Figura 32. Distribuição das respostas nas questões que avaliam o uso dos dispositivos.	72
Figura 33. Distribuição das respostas nas questões que avaliam desconforto e descanso.	73
Figura 34. Distribuição das respostas nas questões que avaliam a interface visual do AV.	73
Figura 35. Distribuição das respostas nas questões que avaliam características da tarefa.	74
Figura 36. Distribuição das respostas nas questões que avaliam a interface visual do AV.	75
Figura 37. Distribuição das respostas nas questões que avaliam o protocolo proposto.	75
Figura 38. Desempenho do “Usuário 5” e seu traçado de SC ao longo do tempo, durante o uso da técnica <i>ray-casting</i>	81
Figura 39. Desempenho do “Usuário 5” e seu traçado de SC ao longo do tempo, durante o uso da técnica HOMER.	82
Figura 40. Exibição do novo recurso visual de seleção, em conjunto com a indicação de um cubo aramado ao redor do objeto.	85
Figura 41. Exibição do novo recurso visual para posicionamento dos objetos, representado por volumes transparentes.	85
Figura 42. Nova configuração dos dispositivos fisiológicos e de RV.	86
Figura 43. Distribuição das respostas do questionário pré-teste.	95
Figura 44. Desempenho entre grupos, considerando a variável “tempo” e as técnicas.	96
Figura 45. Desempenho entre grupos, considerando as variáveis “anexações” e “colisões” e as técnicas.	97
Figura 46. Desempenho entre grupos, considerando a variável “precisão” e as técnicas.	97
Figura 47. Desempenho entre técnicas, considerando a variável “tempo”.	98
Figura 48. Desempenho entre técnicas, considerando as variáveis “anexações” e “colisões”.	98

Figura 49. Desempenho entre técnicas, considerando a variável “precisão”.....	99
Figura 50. Avaliação da sensação de conforto durante o uso de dispositivos com fios. .	100
Figura 51. Avaliação do grau de irritabilidade (questão 2) e de influência do tempo limite (questão 3) durante a tarefa.	100
Figura 52. Avaliação dos graus de satisfação durante a interação com as técnicas <i>ray-casting</i> (questão 4) e HOMER (questão 5).....	101
Figura 53. Avaliação do grau de convicção na execução da tarefa (questão 6) e da qualidade da interface exibida (questão 7).	101
Figura 54. Exemplo de saída do <i>software</i> responsável em unir dados fisiológicos e de desempenho na tarefa.	112
Figura 55. Etapas do processo interativo e respectivas ações realizadas pelo usuário. .	113
Figura 56. Correlação entre as variáveis “tempo” e “NN50”, e respectivas tendências. ..	116
Figura 57. Correlação entre o período basal e as experiências, considerando a medida de HR.....	119
Figura 58. Correlação entre os experimentos, considerando a medida de SC.....	120
Figura 59. Visualização de sinais de ECG com registro de batimento faltante (à esquerda) e batimento extra (adaptado do manual CardioPro Infiniti [THO10]).	121
Figura 60. Dispositivo, em formato de cinto, que transmite os sinais da HR para um sensor de ECG.	122

LISTA DE TABELAS

Tabela 1. Medidas de domínio de tempo de HRV	25
Tabela 2. Medidas de domínio de frequência de HRV	26
Tabela 3. Comparativo entre trabalhos relacionados.	53
Tabela 4. Comentários dos usuários durante o questionário pós-teste.	76
Tabela 5. Sugestões dos usuários durante o questionário pós-teste.	77
Tabela 6. Equações utilizadas para normalizar medidas de HR.....	88
Tabela 7. Classificação dos sujeitos por idade, sexo e escolaridade.	95
Tabela 8. Dificuldades relatadas pelos usuários durante o questionário pós-teste.	103
Tabela 9. Comentários e sugestões dos usuários durante o questionário pós-teste.	104
Tabela 10. Dados descritivos das amostras, considerando as medidas de desempenho.	105
Tabela 11. Dados descritivos das amostras, considerando as medidas fisiológicas.	105
Tabela 12. Modelos de regressão para medidas fisiológicas e de desempenho com forte correlação.	106
Tabela 13. Dados descritivos das amostras, considerando as medidas subjetivas.....	108
Tabela 14. Modelos de regressão para as médias das medidas fisiológicas e de avaliação subjetiva do questionário pré-teste.	108
Tabela 15. Modelo de regressão para as médias das medidas fisiológicas e de avaliação subjetiva do questionário pós-teste.	109
Tabela 16. Modelos de regressão para as medidas fisiológicas na avaliação do uso da técnica <i>ray-casting</i>	109
Tabela 17. Resultado dos testes de confiabilidade dos questionários propostos.	117
Tabela 18. Resultado da análise fatorial para o questionário pré-teste.	117
Tabela 19. Resultado da análise fatorial para o questionário pós-teste.	118
Tabela 20. Custos para uso de medidas fisiológicas em testes de avaliação de usabilidade.....	123
Tabela 21. Comparativo entre trabalhos relacionados e as características desta pesquisa.	126

LISTA DE SIGLAS

2D – bidimensional

3D – tridimensional

3DUI – *3D User Interface*

ACM – *Association for Computing Machine*

ANOVA – Análise de Variância

AV – Ambiente Virtual (ou VE – *Virtual Environment*)

BF – *Breathing Frequency* (Taxa de Respiração)

BPM – Batimentos por minuto

BVP – *Blood Volume Pulse* (Pressão Sanguínea)

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior

CAVE – *Cave Automatic Virtual Environment*

ECG – Eletrocardiograma

EEG – Eletroencefalograma

EKG – Eletrogastrograma

EOG – Eletro-oculograma

FPS – *Frames Per Second* (Quadros Por Segundo)

H0 – Hipótese nula

H1 – Hipótese alternativa

HF – Alta frequência (*High Frequency*)

HMD – *Head Mounted Display* (Capacete de Realidade Virtual)

HOMER – *Hand-centered Object Manipulation Extending Ray-casting*

HR – *Heart Rate* (Frequência Cardíaca)

HRV – *Heart Rate Variability* (Variabilidade dos Batimentos Cardíacos)

IEEE – *Institute of Electrical and Electronics Engineers*

IHC – Interação Humano Computador

ITC-SOPI – *International Test Commission – Sense of Presence Inventory*

LF – Baixa frequência (*Low Frequency*)

LSD – *Least Significant Difference* (referente ao teste *post hoc* LSD de Fisher)

ms – milissegundos

ms² – milissegundos ao quadrado

NASA-TLX – *National Aeronautics and Space Administration – Task Load index*

NN – intervalo normalizado entre complexos QRS adjacentes

NN50 – Número de vezes em que os intervalos NN adjacentes foram superiores a 50 ms

pNN50 – Razão entre a contagem NN50 e o número total de intervalos NN

PSD – *Power Spectral Density* (Densidade Espectral de Potência)

PUCRS – Pontifícia Universidade Católica do Rio Grande do Sul

QRS – Complexo formado pelas ondas Q, R e S em um eletrocardiograma

RMSSD – *Root Mean Square of Successive Differences* (Raiz quadrada média da soma dos quadrados das sucessivas diferenças entre intervalos NN)

RR – intervalo entre duas ondas R, ou frequência ventricular

RV – Realidade Virtual (ou VR – *Virtual Reality*)

SC – *Skin Conductance* (Condutância de Pele)

SDNN – *Standard Deviation of the NN interval* (Desvio padrão de todos os intervalos NN)

SDANN – *Standard Deviation Average of the NN interval* (Desvio padrão da média de todos os intervalos NN)

SDRR – *Standard Deviation of the RR interval* (Desvio padrão de todos os intervalos RR)

SDSD – *Standard Deviation of the Successive Differences* (Desvio padrão das diferenças entre intervalos NN)

ST – *Skin Temperature* (Temperatura da Pele)

S-U-S – Slater, Usoh e Steed

TERV – Terapia de Exposição com Realidade Virtual (ou VRET – *Virtual Reality Exposure Therapy*)

TINN – *Triangular Interpolation of the NN interval* (Interpolação triangular dos intervalos NN)

ULF – Ultra baixa frequência (*Ultra Low Frequency*)

VIF – *Variance Inflation Factor* (Fator de Inflação da Variância)

VLF – Baixíssima frequência (*Very Low Frequency*)

μS – micro-Siemens

SUMÁRIO

1. INTRODUÇÃO	16
1.1 Contribuição do Trabalho	18
1.2 Organização do Documento	18
2. MEDIDAS FISIOLÓGICAS	20
2.1. Resposta da Condutância de Pele	20
2.2. Variabilidade da Frequência Cardíaca.....	22
2.3. Frequência Respiratória	27
2.4. Resposta Pupilar	28
2.5. Movimentos e Piscar dos Olhos	28
2.6. Temperatura da Pele	29
3. TRABALHOS RELACIONADOS	31
3.1 Metodologia de Análise dos Trabalhos.....	31
3.2 Avaliação de Presença em Ambientes Virtuais	33
3.2.1 Medidas fisiológicas e sensações de estresse.....	33
3.2.2 Medidas fisiológicas e quebras na sensação de presença	36
3.2.3 Medidas fisiológicas nas terapias de exposição com Realidade Virtual.....	40
3.3 Avaliação de Usabilidade em Ambientes Virtuais.....	41
3.3.1 Medidas fisiológicas e custo do usuário em jogos	41
3.3.2 Medidas fisiológicas e desconforto em ambientes virtuais.....	46
4. FUNDAMENTAÇÃO DA METODOLOGIA	49
4.1 Justificativa	49
4.2 Objetivos Específicos	54
4.3 Plataforma de Testes.....	54
4.3.1 Técnicas de interação	55
4.3.2 Ambientes virtuais.....	57
4.3.3 Dispositivos de Realidade Virtual.....	59
4.4 Princípios da Metodologia	60
4.4.1 Sensores e medidas fisiológicas	60
4.4.2 Medidas de desempenho do usuário	61
4.4.3 Concepção do protocolo de teste.....	62
4.5 Estudo Piloto	65
4.5.1 Caracterização do grupo	65
4.5.2 Desempenho entre técnicas.....	66

4.5.3	Avaliação subjetiva	72
4.5.4	Relação entre medidas fisiológicas e medidas de desempenho.....	77
4.5.5	Relação entre medidas fisiológicas e avaliações subjetivas	79
4.5.6	Medidas fisiológicas na comparação entre técnicas de interação.....	80
4.5.7	Medidas fisiológicas e alterações durante o processo interativo.....	81
4.6	Considerações sobre o Estudo Piloto.....	82
5.	DEFINIÇÃO DA METODOLOGIA	84
5.1	Adaptações da Plataforma de Testes.....	84
5.2	Definição das Medidas Fisiológicas.....	86
5.3	Definição das Medidas de Desempenho	88
5.4	Definição do Protocolo de Testes	88
5.5	Passos para Aplicação da Metodologia.....	90
5.5.1	Aplicação do protocolo de testes	91
5.5.2	Normalização de dados fisiológicos	92
5.5.3	Emprego de métodos estatísticos	92
6.	ESTUDO DE CASO.....	94
6.1	Experimento.....	94
6.2	Caracterização do grupo	94
6.3	Desempenho entre técnicas	96
6.4	Avaliação subjetiva	99
6.5	Relação entre medidas fisiológicas e medidas de desempenho	102
6.6	Relação entre medidas fisiológicas e avaliações subjetivas.....	107
6.7	Medidas fisiológicas na comparação entre técnicas de interação	110
6.8	Medidas fisiológicas na comparação entre experimentos	110
6.9	Medidas fisiológicas na comparação entre etapas do processo interativo	111
6.10	Medidas fisiológicas na comparação de desempenhos entre grupos.....	113
7.	DISCUSSÃO	115
7.1	Relacionamento entre medidas fisiológicas e métricas de usabilidade	115
7.2	Medidas fisiológicas como métrica de usabilidade	118
7.3	Normalização da Medida de Frequência Cardíaca.....	121
7.4	Custos e Benefícios do Uso do Método.....	122
7.5	Vantagens do Método.....	124
8.	CONCLUSÕES	127
	REFERÊNCIAS	129
	APÊNDICE A – Termo de Consentimento Livre e Esclarecido.....	138

APÊNDICE B – Questionário Pré-Teste (Estudo Piloto).....	140
APÊNDICE C – Instruções para o Experimento (Estudo Piloto).....	141
APÊNDICE D – Questionário Pós-Teste (Estudo Piloto).....	146
APÊNDICE E – Questionário Pré-Teste (Estudo de Caso)	149
APÊNDICE F – Instruções para o Experimento (Estudo de Caso).....	150
APÊNDICE G – Questionário Pós-Teste (Estudo de Caso)	154
APÊNDICE H – Glossário sobre Técnicas e Métodos Estatísticos.....	156
APÊNDICE I – Artigo aceito para publicação no INTERACT 2011.....	160

1. INTRODUÇÃO

Segundo Burdea e Coiffet [BUR03], Realidade Virtual (RV) é uma interface entre usuário e computador que envolve simulação em tempo real e interação por intermédio de múltiplos canais sensoriais, como visão, audição e tato. Para tanto, ambientes tridimensionais e dispositivos específicos podem proporcionar ao usuário a imersão em reproduções de situações reais ou imaginárias.

A quantidade de dispositivos, aliada à capacidade que esses têm de estimular os sentidos do usuário, determina o **nível de imersão** de um Ambiente Virtual (AV). Outro conceito que, junto à imersão, determina o sucesso de uma aplicação de RV é a **sensação de presença**, que é definida como a sensação de estar dentro do ambiente, como se este fosse real.

Para medir estas e outras características de interfaces tridimensionais (3D), métodos e ferramentas comumente empregados na avaliação de usabilidade em interfaces bidimensionais (2D), como protótipos, questionários e testes formativos e somativos podem ser utilizados. De acordo com Hix e Hartson [HIX93], estes instrumentos também são capazes de medir aspectos de usabilidade em interfaces 3D, como aprendizado, velocidade e precisão no desempenho de tarefas, latência do sistema e conforto do usuário.

Para facilitar a avaliação destas medidas de usabilidade nas interfaces 3D, Bowman *et al* [BOW04] sugere a divisão destas em três categorias: medidas de desempenho do sistema, medidas de desempenho na tarefa e medidas das preferências do usuário. As duas primeiras categorias se referem às medidas objetivas e buscam avaliar, respectivamente, a influência dos dispositivos e o rendimento do usuário durante uma tarefa específica. A terceira tem caráter subjetivo, e procura medir a satisfação do usuário na utilização do sistema.

No entanto, a adaptação de métodos e ferramentas para interfaces 3D pode acarretar em avaliações incompletas de características distintas destas aplicações, como o uso de dispositivos não-convencionais e técnicas de interação 3D [BOW04]. Estas características tendem a influenciar o desempenho e a satisfação do usuário nos experimentos e, portanto, necessitam de um processo que avalie seus diferentes recursos conforme o nível de experiência do usuário.

Uma das alternativas recentemente utilizadas para avaliar interfaces é o uso de medidas fisiológicas. De acordo com Malik *et al* [MAL96], o monitoramento fisiológico oferece informações sobre o balanço fisiológico do usuário, e suas medidas são

associadas ao estresse, baseadas no fato de que, se o estresse aumenta, o bem-estar diminui. Trabalhos na área de Computação Gráfica vêm utilizando este tipo de medida para avaliar o esforço físico e mental do usuário em jogos 2D [LIN05] [LIN06a] [LIN06b] [LIN08] e para avaliar a sensação de presença e conforto do usuário em AVs imersivos [MEE02] [MEE05] [BRO06] [BRO07] [KIM08]. No entanto, ainda não existem trabalhos que estudem o relacionamento entre estas medidas fisiológicas e as métricas que buscam avaliar a qualidade de técnicas de interação em interfaces 3D.

O emprego de medidas fisiológicas pode também ajudar na solução de dois problemas clássicos de avaliação de interfaces 3D. O primeiro deles diz respeito à computação de métricas objetivas, as quais podem exigir a alteração do código fonte dos programas – tarefa nem sempre possível ou desejável, por causa da complexidade do sistema ou da pouca disponibilidade de tempo para sua codificação. O segundo problema refere-se à confiabilidade das métricas subjetivas, que podem ter seus resultados influenciados por fatores externos ao processo de interação, como esforço físico e mental, ou pela mediação cognitiva do usuário, como omissão ou sumarização de informações.

Medidas fisiológicas, por sua vez, são respostas objetivas que independem da vontade do indivíduo, estão associadas a fatores como cansaço e irritação, e fornecem dados relacionados ao comportamento do organismo. Estas medidas podem indicar, por exemplo, os períodos de adaptação do usuário a um dispositivo ou técnica de interação 3D, uma vez que as alterações do nível de estresse podem ser visualizadas ao longo do tempo. Além disso, podem auxiliar na compreensão de resultados de desempenho e respostas de questionários. Logo, é possível que medidas fisiológicas sejam capazes de complementar os métodos de avaliação já existentes [HIX93] [ROS01] [BOW02], permitindo o entendimento do processo interativo como um todo, e contribuindo para o aumento da qualidade das aplicações de RV.

Uma forma de explorar a validade de medidas fisiológicas como critério de avaliação para interfaces 3D é por intermédio da comparação entre seus resultados e os resultados obtidos por meio de métricas convencionais de usabilidade. Para realizar este cruzamento de dados, métodos estatísticos de análise multivariada podem ser empregados. De acordo com Hair *et al* [HAI09], estes métodos permitem analisar, simultaneamente, a influência de múltiplas medidas sobre cada indivíduo ou variável sob investigação, independentemente da complexidade ou do contexto em que estas variações ocorrem. Desta forma, é possível verificar se os resultados obtidos com medidas fisiológicas são capazes de identificar os mesmos problemas apontados pelas métricas convencionais de usabilidade.

1.1 Contribuição do Trabalho

Este trabalho visa definir uma metodologia para avaliar a qualidade de técnicas de interação em ambientes virtuais imersivos utilizando medidas fisiológicas. Esta pesquisa se apresenta como inovadora, uma vez que pretende entender o relacionamento entre medidas fisiológicas e métricas de usabilidade comumente utilizadas para avaliar interfaces 3D, como, por exemplo, o tempo de interação, o número de tentativas realizadas para executar uma tarefa e os índices subjetivos de satisfação.

Para tanto, é apresentada **uma metodologia que considera a aplicação de um protocolo de teste e de métodos estatísticos destinados a comparar medidas fisiológicas e métricas convencionais de usabilidade.** Este formato de análise tem por objetivo descobrir relacionamentos que contribuam para o processo de avaliação da qualidade de técnicas de interação, complementando ou facilitando a interpretação dos resultados. Da mesma forma, a metodologia também determina quais medidas fisiológicas são capazes de indicar os mesmos resultados esperados por medidas convencionais de usabilidade, e se podem, eventualmente, atuarem como substitutas destas medidas convencionais em projetos nos quais a simplificação da etapa de testes é desejável.

Para validar esta comparação entre os dois tipos de medidas, este trabalho utiliza medidas fisiológicas como métrica de usabilidade, na comparação entre duas técnicas de interação de mesmo propósito, entre períodos de experimento e entre etapas do processo interativo. Nestes casos, a ideia é descobrir se a alteração fisiológica do usuário pode contribuir para o aperfeiçoamento da interface, destacando, por exemplo, eventuais características de técnicas, períodos de adaptação ao uso de dispositivos ou problemas que interfiram no desempenho e nas preferências do usuário. Da mesma forma, pretende-se também indicar qual das técnicas é geradora de um nível menor de estresse e, conseqüentemente, mais adequada para a realização de determinada tarefa.

1.2 Organização do Documento

O presente documento está organizado da seguinte forma: o Capítulo 2 apresenta uma caracterização sobre medidas fisiológicas, abordando seu uso e formas de coleta. O Capítulo 3 apresenta os trabalhos relacionados, com uma revisão dos métodos e avaliações de interfaces que utilizam medidas fisiológicas similares a abordagem

proposta. No Capítulo 4, destacam-se os passos realizados para a definição da metodologia, apresentando justificativas, princípios básicos adotados, técnicas estatísticas para análise e resultados preliminares alcançados por um estudo piloto. O Capítulo 5 apresenta uma revisão da metodologia, enquanto o Capítulo 6 apresenta um estudo de caso para avaliar a qualidade das técnicas de interação utilizando a nova metodologia. O Capítulo 7 apresenta uma discussão sobre os resultados alcançados no processo de avaliação, buscando apontar tendências e avaliar o uso desta nova abordagem. Por fim, o Capítulo 8 apresenta as conclusões do trabalho, assim como as perspectivas para sua continuidade.

2. MEDIDAS FISIOLÓGICAS

Durante a utilização de ambientes de RV, os usuários podem apresentar alterações fisiológicas ativadas pela ansiedade gerada na realização das tarefas [BOU08]. De acordo com a tarefa a ser executada e os recursos disponíveis pela aplicação, movimentos naturais e ações do usuário precisam ser adaptados para alcançar o resultado desejado, fato que pode causar momentos de ansiedade e nervosismo durante o período de interação no AV.

O uso de medidas fisiológicas na comunidade de RV é recente, e está principalmente ligado aos AVs de terapia expositiva para tratamento psicológico [WIE02a] [WIL05] [SLA06] [RIZ07]. Existem também alguns trabalhos que utilizam medidas fisiológicas voltados ao comportamento de usuários em jogos bidimensionais [LIN08] e à presença em AVs [MEE05].

Dentre as medidas utilizadas nestes trabalhos, as que apresentaram os melhores resultados são as medidas de variabilidade da frequência cardíaca e a resposta da condutância de pele. Outras medidas, como a frequência respiratória, o rastreamento dos olhos, a dilatação da pupila e a temperatura da pele, também são empregados, porém, sem apresentar resultados satisfatórios [MEE02] [WIL05] [LIN06b] ou relacionados às questões de ansiedade em AVs.

As próximas seções deste capítulo apresentam as diferentes medidas de monitoramento fisiológico não invasivo¹, utilizadas pelos trabalhos citados, com atenção especial às medidas de variabilidade da frequência cardíaca e resposta da condutância de pele.

2.1. Resposta da Condutância de Pele

A resposta da condutância de pele (SC, do inglês *Skin Conductance*) é um método que tem por objetivo medir a capacidade da pele de conduzir eletricidade [PRO03]. Também conhecida como resposta galvânica da pele, resposta eletrodermal ou reflexo psicogalvânico, é considerada pela comunidade científica como um indicador fisiológico que se correlaciona com os estados de emoção, excitação e atenção do indivíduo [PRO74] [DAW90].

¹ O monitoramento fisiológico não invasivo emprega apenas equipamentos que não rompem a pele e que não penetram fisicamente no corpo, evitando que um indivíduo fique propenso à exposição.

A pesquisa sobre resposta de SC é robusta e bem fundamentada, e apresenta trabalhos desde o final do século XIX [CON95]. A evolução destas pesquisas evidencia que as alterações dos níveis de suor, causadas pela ativação do sistema nervoso simpático², permitem quantificar o estado emocional de uma pessoa. Este sistema responde rapidamente e de forma involuntária a agitação psicológica ou mental do indivíduo, como sensação de medo, raiva e estresse.

Conforme Montagnu e Coles [MON66], quanto mais suor é produzido, maior é a capacidade da pele em conduzir corrente elétrica. Além disso, a SC apresenta um relacionamento linear com o aumento da atividade sudorípara, uma vez que as glândulas funcionam em paralelo, tornando cumulativo o princípio da condutividade no organismo.

Segundo Lykken e Venables [LYK71], é possível coletar a SC utilizando um dispositivo que mede a resistência elétrica entre dois pontos. Uma pequena tensão elétrica, inferior a 0,5 volts, pode ser aplicada em dois eletrodos, presos em dois dedos de uma das mãos do indivíduo, estabelecendo um circuito elétrico no qual o corpo atua como um resistor. Com isso, a variação em tempo real da condutância (que é o inverso da resistência) pode ser calculada [COR06] [BIO09a]. A Figura 1 ilustra o circuito básico para medir a SC.

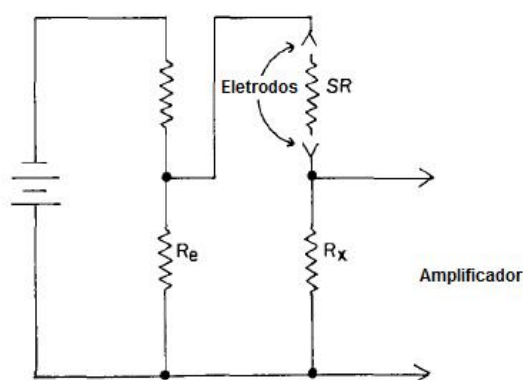


Figura 1. Circuito básico de um dispositivo que mede SC (adaptado de Lykken e Venables [LYK71]).

Segundo Dawson *et al* [DAW90] e Montagnu e Coles [MON66], as palmas das mãos e a as solas dos pés são utilizadas para medir SC por serem regiões que sofrem menor influência da temperatura do organismo, além de estarem associadas com a atividade mental da pessoa. McCleary [McC50] afirma ainda que para medir de forma eficaz a SC, os eletrodos devem ser fixados nas pontas dos dedos (falanges distais) e posicionados no centro das impressões digitais, pois estas são áreas de vasta

² Sistema responsável por estimular ações, como aceleração dos batimentos cardíacos e níveis de suor, que permitem ao organismo responder a situações de estresse, independente da vontade do indivíduo.

concentração de glândulas sudoríparas. A Figura 2 ilustra uma maneira de prender os eletrodos.



Figura 2. Medição da SC.

A unidade de medida padrão para SC é chamada de micro-Siemens (μS), podendo variar numa faixa entre 0 e 30 μS . Em estado de repouso, uma leitura normal de SC fica em torno de 2 μS , porém, esta medida pode variar conforme o tipo de pele da pessoa ou variáveis de ambiente, como umidade e temperatura.

2.2. Variabilidade da Frequência Cardíaca

Segundo Malik e Camm [MAL95], a variabilidade da frequência cardíaca (HRV, do inglês *heart rate variability*) é uma medida das variações dos batimentos na frequência cardíaca (HR, do inglês *heart rate*). Esta medida é considerada por especialistas da área médica como um importante indicador fisiológico e comportamental, que reflete a capacidade do indivíduo em adaptar-se a diferentes situações, principalmente àquelas ligadas aos estados emocionais e mentais ativados pelo sistema nervoso autônomo³. Normalmente, um ser humano apresenta uma HR entre 60 e 100 batimentos por minuto [GUY05].

A HRV tem relação direta com os sinais de atividade elétrica gerada pelo funcionamento do coração. Estes sinais, produzidos pelas células musculares cardíacas, apresentam informações que permitem observar o equilíbrio e funcionamento do organismo. Eles podem ser obtidos por meio de dispositivos de monitoramento pessoal, capazes de coletar e analisar uma série temporal de dados relativos aos intervalos dos batimentos cardíacos.

³ Sistema responsável pelo controle de funções vitais, como a respiração, a circulação sanguínea, a temperatura e a digestão. Está dividido em dois subsistemas: o parassimpático, que estimula principalmente atividades relaxantes, e o simpático, que estimula ações que mobilizam energia, permitindo ao organismo responder a situações de estresse.

Exemplos de equipamentos como estes são o eletrocardiógrafo e o fotopleletismógrafo, que possibilitam a medição e a gravação da HR em tempo real, de maneira não-invasiva.

De acordo com Nicolau *et al* [NIC03], o exame médico gerado pelo eletrocardiógrafo, o eletrocardiograma (ECG), é o registro mais conhecido e utilizado pelos médicos e fisiologistas, devido à eficiência na captação e avaliação dos sinais cardíacos, bem como à precisão dos dados.

Em equipamentos deste tipo, a medição da frequência cardíaca pode ser realizada por dois componentes básicos: um conjunto de eletrodos (transmissores) protegidos por uma capa plástica gelatinosa, presos ao peito da pessoa para capturar os batimentos cardíacos; e um equipamento receptor que estabelece uma rede para receber os dados.

Existem também equipamentos que possibilitam o monitoramento utilizando sensores que integram os eletrodos a uma cinta de tecido, a qual pode ser fixada no peito ou nos punhos do indivíduo. A Figura 3 apresenta três configurações de fixação de eletrodos e conexão com o eletrocardiógrafo.

No caso do fotopleletismógrafo (Figura 4), a medição utiliza um fotossensor preso em um dos dedos da mão do indivíduo. O processo considera a pulsação arterial (BVP, do inglês *blood volume pulse*), e consiste em medir a quantidade de luz refletida pela pele. A cada contração do coração, o sangue é forçado através de vasos periféricos, produzindo seu enchimento, o que, sob uma fonte de luz, altera a quantidade de luz refletida para o fotossensor [BIO09b]. Comparada ao eletrocardiógrafo, a precisão dos dados é menor, uma vez que o sensor é sensível a movimentos, que interferem no sinal capturado.

Os sinais capturados por estes aparelhos registram um gráfico com um determinado padrão de ondas. Esta característica é notável principalmente nos exames de ECG, por meio do conjunto de ondas denominadas P, Q, R, S e T, as quais correspondem aos eventos elétricos de ativação do músculo cardíaco (miocárdio).

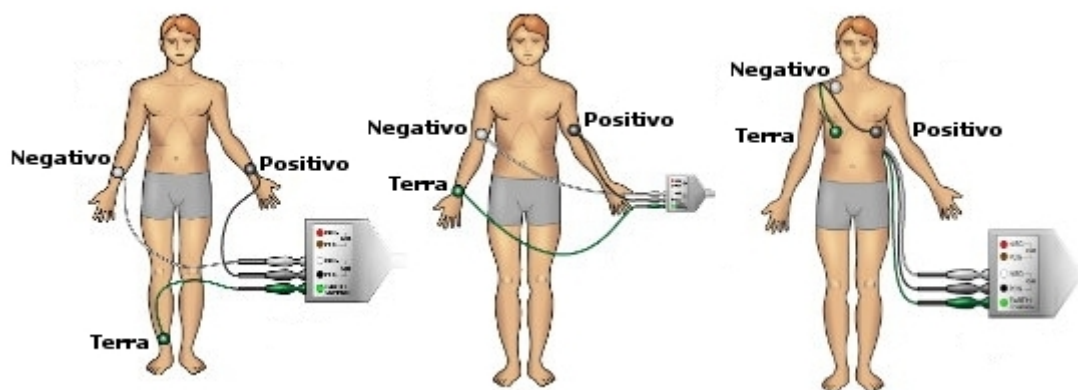


Figura 3. Variantes de medição da HR utilizando um eletrocardiógrafo (adaptado de AdInstruments [ADI10]).



Figura 4. Exemplo de fotopletismógrafo [KIR02].

A onda P é causada por correntes elétricas geradas pela despolarização⁴ dos átrios antes da sua contração. O complexo QRS é formado por três ondas, e é causado por correntes geradas pela despolarização dos ventrículos antes da contração. Já a onda T é causada por correntes geradas pela repolarização⁴ dos ventrículos após a contração. Ela encobre a repolarização dos átrios, que ocorre concomitantemente, porém com menor potência. A Figura 5 apresenta o gráfico característico de um ciclo cardíaco, com a apresentação do conjunto de ondas, intervalos e segmentos.

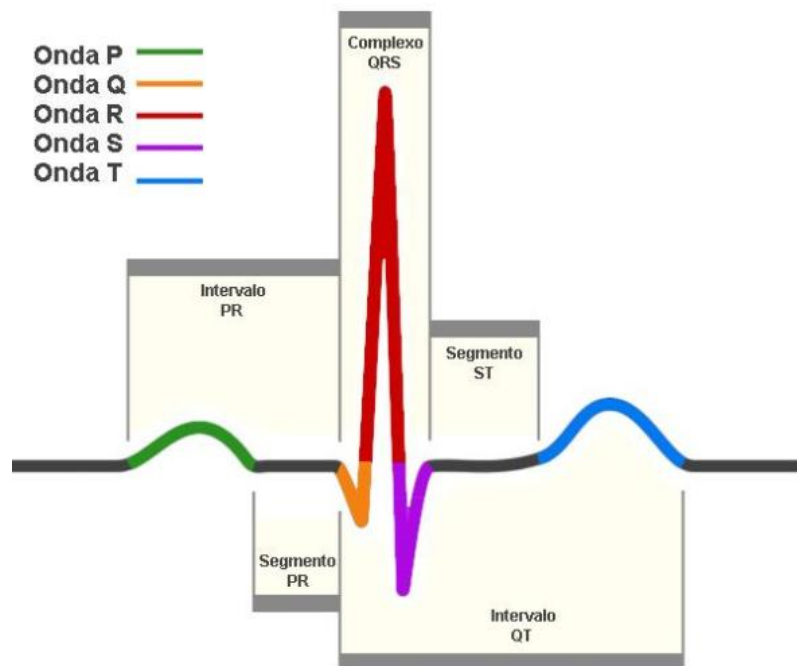


Figura 5. Ciclo cardíaco normal, com seu conjunto de ondas (adaptado de Skipping Hearts [SKI10]).

⁴ Em Biologia, despolarização é uma alteração no potencial de ação elétrico (impulso nervoso) de uma membrana celular, a qual se torna mais positiva ou menos negativa. Já repolarização é o processo oposto, que faz o potencial da membrana retornar ao valor anterior ao estímulo da despolarização.

A partir deste conjunto de ondas, é possível calcular a HRV. Conforme Malik e Camm [MAL95] e Madeiro *et al* [MAD06], várias medidas de HRV têm sido propostas, baseadas em modelos e fórmulas matemáticas que utilizam, principalmente, os intervalos NN (intervalo entre complexos QRS adjacentes) e RR (intervalo entre duas ondas R, ou frequência ventricular). Para alcançar estas medidas, métodos lineares de domínio de tempo e domínio de frequência, ou métodos não-lineares, podem ser aplicados.

De acordo com Malik *et al* [MAL96], as medidas lineares no domínio de tempo permitem determinar tanto a frequência cardíaca e alterações ao longo do tempo, como os intervalos entre sucessivos ciclos cardíacos normais (complexos QRS). Com base em métodos estatísticos ou geométricos, diferentes índices que refletem as flutuações da HR podem ser calculados. Como exemplos, podem ser citados a média dos intervalos NN, a média da HR, a diferença entre o maior e o menor intervalo NN, entre outros. A Tabela 1 apresenta uma relação das principais medidas de domínio de tempo.

Tabela 1. Medidas de domínio de tempo de HRV
(adaptado de Malik *et al* [MAL96]).

Variável	Unidade	Descrição
<i>Métodos Estatísticos</i>		
SDNN	ms	Desvio padrão de todos os intervalos NN
SDANN	ms	Desvio padrão das médias dos intervalos NN de todos os segmentos de 5 minutos do traçado de 24 horas
RMSSD	ms	Raiz quadrada média da soma dos quadrados das sucessivas diferenças entre intervalos NN
Índice SDNN	ms	Média dos desvios padrão dos intervalos NN calculados em intervalos de cinco minutos
SDSD	ms	Desvio padrão das diferenças entre intervalos NN adjacentes.
NN50		Número de pares de intervalos NN adjacentes, os quais a diferença excede 50 ms
pNN50	%	A contagem NN50 dividido pelo número total de todos os intervalos NN
<i>Métodos Geométricos</i>		
Índice triangular de HRV		Número total de intervalos NN (integral da distribuição de densidade) dividido pelo valor máximo da distribuição de densidade
TINN	ms	Largura base do triângulo, obtido por interpolação matemática, a partir do histograma de distribuição dos intervalos NN
Gráfico de Poincaré		Mapa de pontos em um plano cartesiano, no qual cada intervalo NN é correlacionado com o intervalo antecedente, definindo um ponto no gráfico

Já no caso das medidas lineares de domínio de frequência, oscilações periódicas do sinal são analisadas, fornecendo informações sobre a quantidade de variação do ritmo cardíaco em diferentes frequências. Conforme Vanderlei *et al* [VAN09], a análise da

densidade espectral de potência (PSD, do inglês *Power Spectral Density*) é o método mais utilizado, determinando a distribuição da variância do sinal em função da frequência. Para calculá-la, o método comumente utilizado é a Transformada Rápida de Fourier, devido à sua simplicidade, eficiência e boa apresentação gráfica [AUB03]. Desta forma, a HRV pode ser decomposta em componentes oscilatórios fundamentais, analisados de acordo com o período de tempo para ter algum significado [COM10]. Os dois períodos de tempo (também chamado de épocas) frequentemente utilizados são de 24 horas (longo período) e 5 minutos (curto período). A Tabela 2 expressa esses componentes de medida.

Tabela 2. Medidas de domínio de frequência de HRV
(adaptado de Malik *et al* [MAL96]).

Variável	Unidade	Descrição	Faixa de Frequência
<i>Análises de registros em períodos curtos (padrão: 5 minutos)</i>			
Potência Total 5 min.	ms ²	Variância de intervalos NN sobre o segmento temporal	aproxim. ≤ 0,4Hz
VLF	ms ²	Potência em frequência muito baixa	≤ 0,4Hz
LF	ms ²	Potência em frequência baixa	0,04 – 0,15Hz
LF norm.	u.n.	Potência em frequência baixa, em unidade normalizada LF/ (Potência Total – VLF) x 100	
HF	ms ²	Potência em frequência alta	0,15 – 0,40Hz
HF norm.	u.n.	Potência em frequência alta, em unidade normalizada HF/ (Potência Total – VLF) x 100	
LF/ HF		Razão LF (ms ²) sobre HF (em ms ²)	
<i>Análises de registros em períodos longos (padrão: 24 horas)</i>			
Potência Total	ms ²	Variância de todos os intervalos NN	aproxim. ≤ 0,4Hz
ULF	ms ²	Potência em frequência ultra baixa	≤ 0,003Hz
VLF	ms ²	Potência em frequência muito baixa	0,003 – 0,04Hz
LF	ms ²	Potência em frequência baixa	0,04 – 0,15Hz
HF	ms ²	Potência em frequência alta	0,15 – 0,40Hz

Por fim, medidas não-lineares são utilizadas para interpretar, explicar e prever o comportamento dos fenômenos fisiológicos dos sistemas humanos, em razão de sua natureza dinâmica complexa, como interações hemodinâmicas, eletrofisiológicas e de humor [VAN09]. Entretanto, de acordo com Malik *et al* [MAL96] e Godoy *et al* [GOD05], a pesquisa ainda carece de aprofundamento científico, com amostras expressivas e acompanhamento prolongado para análise de dados biomédicos, incluindo a análise da HRV.

2.3. Frequência Respiratória

A frequência respiratória (BF, do inglês *Breathing Frequency*) é uma medida que computa o número de vezes que um indivíduo respira dentro de um determinado período de tempo, geralmente expresso em minutos. Também conhecida como taxa de respiração ou de ventilação pulmonar, é considerada pela comunidade médica como um dos principais indicadores de instabilidade fisiológica, capaz de ajudar na identificação de eventos adversos graves, como paradas cardiorrespiratórias [RES10]. De acordo com Tortora e Derrickson [TOR08], um adulto saudável, em estado de repouso, apresenta uma BF entre 12 e 20 respirações por minuto.

A BF é geralmente tomada em repouso, e utilizada como medida de controle quando outras medidas fisiológicas são capturadas. O equipamento que coleta esta frequência é um sensor sensível à expansão dos pulmões, o qual é amarrado ao tronco do indivíduo por cima da roupa, durante uma expiração completa. Este sensor pode ser fixado em torno do peito, para coletar a respiração torácica, ou em torno do abdômen, para coletar a respiração diafragmática. A Figura 6 apresenta o dispositivo e as respectivas formas de fixação.

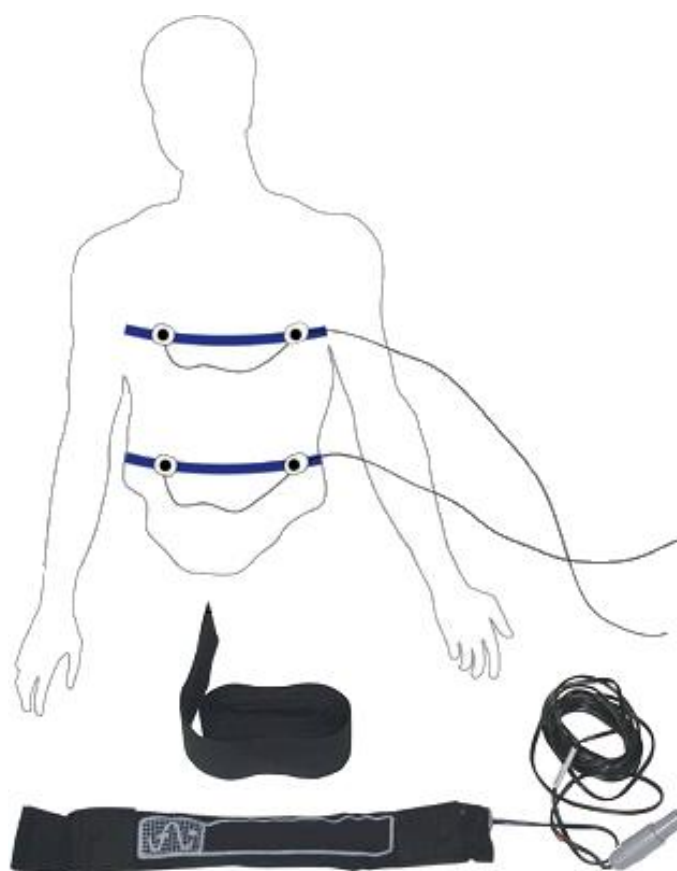


Figura 6. Dispositivo de monitoramento da frequência respiratória e formas de fixação.

2.4. Resposta Pupilar

A resposta pupilar é um parâmetro que mede alterações dos diâmetros das pupilas nos olhos de uma pessoa. O tamanho da pupila é afetado pela constrição e dilatação involuntária da íris, para controlar a intensidade de luz que passa por este orifício, ou para determinar a nitidez de um objeto em foco. De acordo com Hess e Polt [HES60] [HES64], as variações do tamanho da pupila também refletem alterações no comportamento emocional de uma pessoa, as quais podem indicar, por exemplo, o esforço cognitivo gasto para executar uma tarefa ou o interesse por um estímulo visual.

Para monitoramento fisiológico, pupilômetros são geralmente utilizados, os quais medem as variações do tamanho da pupila em milímetros. Andreassi [AND07] destaca que o método mais popular de medir o diâmetro da pupila é utilizando câmeras de vídeo sob iluminação infravermelha. Uma luz infravermelha de baixa intensidade é utilizada para iluminar o olho, possibilitando que o sensor do pupilômetro determine precisamente onde começa a pupila, e meça sua distância transversal. Equipamentos que utilizam este método utilizam câmeras para filmar os olhos e um sistema de processamento do sinal. A Figura 7 apresenta um exemplo deste tipo de dispositivo.

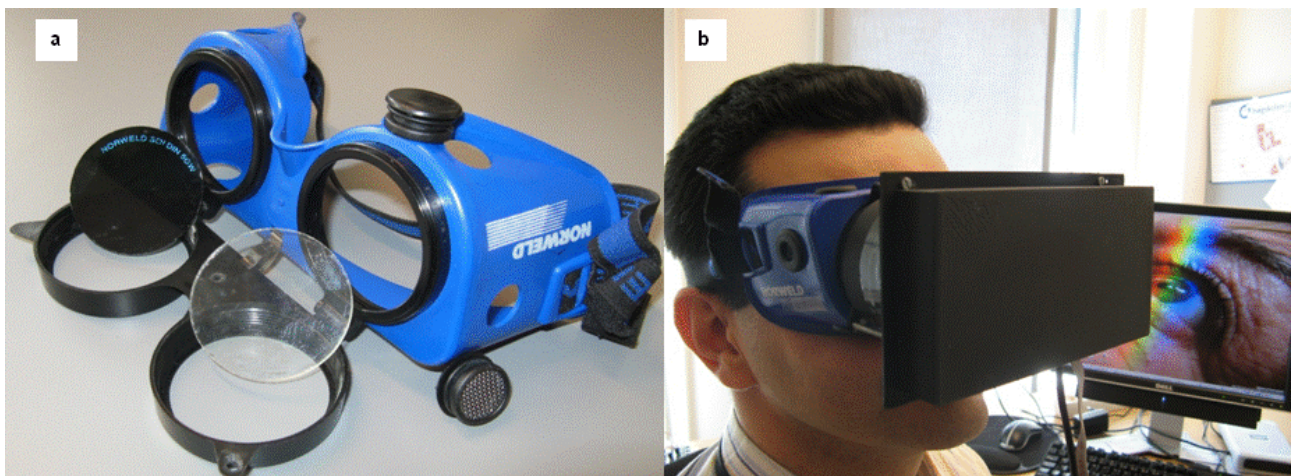


Figura 7. Pupilômetro de baixo custo utilizando webcams [MIR10].

2.5. Movimentos e Piscar dos Olhos

O rastreamento do olhar é um processo de medição dos pontos para onde se está olhando, ou do movimento dos olhos em relação à cabeça. De acordo com Duchowski [DUC07], a captura destas ações possibilita entender como o sistema visual age na identificação de imagens e quais são as regiões de interesse desta. Assim, pode-se compreender o funcionamento das atividades mentais de uma pessoa, como o

esforço para o aprendizado e a solução de problemas. Por outro lado, o piscar dos olhos permite descobrir momentos de fadiga, lapsos de atenção e estresse durante uma determinada atividade visual [AND07].

De acordo com Andreassi [AND07], os quatro principais métodos para medição dos movimentos dos olhos são a técnica de lentes de contato, o método de reflexo corneal, a varredura por imagens de vídeo (video-oculografia) e a eletro-oculografia⁵ (EOG). Os equipamentos que empregam estas técnicas coletam, basicamente, a posição e rotação dos olhos, os registros de piscar dos olhos e os pontos de observação.

A Figura 8 apresenta três exemplos de equipamentos utilizados para rastreamento de olhar. O primeiro, à esquerda, é um dispositivo para ambientes *desktop*, que deve ser utilizado em conjunto com um sistema de rastreamento da cabeça. O segundo é um óculos *mobile* que permite o rastreamento do olhar e a gravação destes em fitas de vídeo. Já o terceiro é um capacete de RV ou HMD (do inglês *head-mounted display*), na qual um módulo de video-oculografia está acoplado.



Figura 8. Exemplos de dispositivos de rastreamento de olhar [ASL10].

2.6. Temperatura da Pele

A temperatura da pele (ST, do inglês *Skin Temperature*) é um parâmetro físico que está associado ao conceito de equilíbrio térmico do corpo. Sua medição varia de acordo com o local do corpo onde é feita a coleta, as condições climáticas e o nível de atividade da pessoa [GUY05]. Apesar disto, de acordo com Clark [CLA84], existem características comuns de padrões de ST para grupos similares de pessoas que possibilitam seu emprego para fins diagnósticos.

⁵ Equipamento que registra o potencial elétrico córneo-retiniano, por meio de eletrodos fixados na pele, localizados nos cantos nasal e temporal da face próximos dos olhos.

Para o monitoramento psicofisiológico, a ST é utilizada como medida de controle, coletada em período de descanso, ou então antes e depois da realização do experimento. O sensor de temperatura (termistor) é normalmente anexado à parte dorsal ou palmar de um dos dedos da mão ou do pé, na parte dorsal ou palmar. As variações da temperatura são convertidas para sinais de corrente elétrica, e estão relacionadas à quantidade de sangue que irriga a pele nas extremidades do corpo. A Figura 9 apresenta o dispositivo.

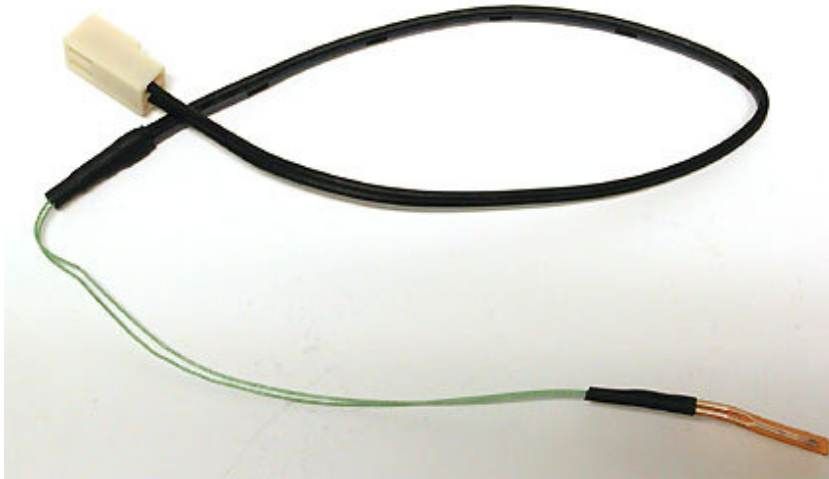


Figura 9. Dispositivo de monitoramento da temperatura da pele.

Com base na fundamentação teórica realizada nesta seção e nas seções anteriores, o próximo capítulo apresenta trabalhos e projetos de pesquisa que empregam medidas fisiológicas na avaliação de características distintas de AVs, como sensação de presença e desempenho do usuário na realização de tarefas interativas.

3. TRABALHOS RELACIONADOS

Este capítulo apresenta uma revisão dos trabalhos de Computação com o objetivo de mostrar as pesquisas que utilizam medidas fisiológicas para avaliação de AVs.

3.1 Metodologia de Análise dos Trabalhos

Como forma de organizar a análise dos trabalhos, e destacar as medidas fisiológicas utilizadas em projetos de RV, uma metodologia de revisão foi criada. Esta metodologia busca identificar, selecionar e apresentar os estudos existentes, a partir de artigos publicados em conferências e revistas científicas da área de Computação.

Para tanto, os seguintes passos foram executados:

- Definir as palavras-chave e acrônimos da pesquisa, com base no tema central de pesquisa;
- Fixar o período de busca do material desejado;
- Buscar artigos publicados em revistas e conferências utilizando a combinação de, ao menos, duas palavras-chave geradas;
- Selecionar os artigos encontrados, separando-os de acordo com a combinação de palavras-chave utilizadas;
- Refinar o material relevante, por intermédio da leitura dos artigos.

A definição das palavras-chave e acrônimos tomou como base os contextos nos quais este trabalho está concentrado, quais sejam: “medidas fisiológicas”, “interfaces de RV” e “métodos de avaliação”. Considerando estes três contextos, as seguintes palavras-chave e acrônimos foram especificados: *physiological measures*, *physiological signals* e *biofeedback* para medidas fisiológicas; *virtual reality*, *VR*, *virtual environment*, *VE*, *immersive*, *3D user interface*, *3DUI* e *3D interaction technique* para interfaces de RV; *evaluation*, *validation* e *assessment* para métodos de avaliação.

Para condução da pesquisa, foram considerados os materiais publicados em conferências e revistas eletrônicas no período de 01 de janeiro de 1990 até 31 de dezembro de 2010.

Com relação às buscas propriamente ditas, estas se concentraram nas bases de dados disponibilizadas pelo Portal de Periódicos do CAPES (Coordenação de

Aperfeiçoamento de Pessoal de Ensino Superior)⁶, nas bases de periódicos disponíveis na biblioteca da PUCRS, sendo estas a Elsevier Science Direct⁷, CiteseerX⁸, ACM (*Association for Computing Machine*)⁹, Springer¹⁰ e IEEE (*Institute of Electrical and Electronics Engineers*)¹¹, e no mecanismo de busca Google Scholar¹². As palavras-chave e acrônimos eram combinados de duas formas, considerando os contextos apresentados: “medidas fisiológicas” e “interfaces de RV”; “medidas fisiológicas”, “interfaces de RV” e “métodos de avaliação”. Ao todo, utilizaram-se 96 combinações diferentes de pesquisa.

O procedimento de seleção de trabalhos se utilizou da combinação destas palavras-chave para organizar o material encontrado. Os artigos foram organizados, internamente, de acordo com o ano da publicação e nome da revista ou conferência. Para evitar redundância de material, artigos já filtrados por outras buscas eram descartados.

Em seguida, um refinamento foi realizado, após leitura prévia dos documentos coletados. Para tornar-se relevante, a análise destes materiais considerou a existência dos seguintes aspectos nas publicações: descrição dos recursos computacionais e fisiológicos utilizados; métodos e protocolos de avaliação aplicados; características dos grupos de estudo e resultados alcançados.

Ao final, este processo resultou na seleção de 10 artigos que tratavam essencialmente da utilização de medidas fisiológicas como métrica para avaliar AVs.

Devido à natureza dos trabalhos selecionados, as pesquisas foram classificadas em dois grupos, conforme os objetivos da avaliação realizada. O primeiro grupo continha os trabalhos que utilizam medidas fisiológicas para **avaliar presença**, e o segundo, **avaliar usabilidade**.

As próximas subseções apresentam o detalhamento destes trabalhos.

⁶ <http://periodicos.capes.gov.br/>

⁷ <http://www.sciencedirect.com/>

⁸ <http://citeseerx.ist.psu.edu/>

⁹ <http://portal.acm.org/>

¹⁰ <http://www.springerlink.com/>

¹¹ <http://ieeexplore.ieee.org/>

¹² <http://scholar.google.com>

3.2 Avaliação de Presença em Ambientes Virtuais

Para facilitar a compreensão, os estudos desta subseção foram divididos em três categorias, baseadas na relação das medidas fisiológicas com as sensações de estresse, as quebras de presença e as terapias de exposição.

3.2.1 Medidas fisiológicas e sensações de estresse

O trabalho de Meehan *et al* [MEE02] apresenta o uso de medidas fisiológicas como alternativa para avaliação de presença em AVs capazes de provocar sensações de estresse ao usuário. Para tanto, são consideradas medidas de HR, SC e ST durante a interação em um ambiente que simula uma caminhada ao redor de um fosso.

Nesta pesquisa, os autores tentam provar se as medidas fisiológicas relacionadas a cada usuário podem resultar em dados confiáveis, válidos, sensíveis e objetivos, capazes de determinar quanta presença o AV evoca nos usuários em situações estressantes.

De acordo com os autores, são considerados dados confiáveis aqueles que produzem resultados iguais ou compatíveis em diferentes experimentos. Por sua vez, dados válidos são aqueles que se correlacionam com medidas subjetivas já bem estabelecidas. Já dados sensíveis se caracterizam por diferenciarem vários níveis de presença, enquanto que dados objetivos estão protegidos de tendências externas.

A ideia é que, dado um AV realista, medidas fisiológicas semelhantes às aquelas evocadas em um ambiente real possam ser reproduzidas, passando sensação de presença ao sujeito.

Para comprovar isto, os pesquisadores expuseram 69 usuários a um AV que simulava a caminhada ao redor de um fosso virtual, com o objetivo de provocar sensações de estresse no indivíduo. Neste ambiente, o usuário realizava ações de pegar e transportar objetos de uma sala virtual para outra, tendo que obrigatoriamente passar pelo fosso virtual para largar o objeto em sua nova posição.

Para simular a caminhada, os pesquisadores montaram um tablado de madeira com saliência nas bordas, com o intuito de gerar ao usuário a sensação tátil de estar à beira de um fosso quando estes limites eram alcançados. Esta área foi subdividida e mapeada para a aplicação, representando as duas salas virtuais do AV. Os dispositivos utilizados para interação foram um HMD com visão estereoscópica e rastreador de cabeça, e um rastreador de posição, preso a um *mouse*, o qual era controlado pela mão

direita do usuário. A Figura 10 apresenta o AV e um usuário vestindo os sensores e dispositivos de RV durante uma sessão de teste.



Figura 10. Visão lateral do AV e usuário vestindo dispositivos fisiológicos e de RV (adaptado de Meehan *et al* [MEE02]).

As medidas de SC e ST foram coletadas utilizando sensores presos aos dedos da mão esquerda do usuário. Para medir a HR, um eletrocardiógrafo foi utilizado, com sensores presos ao peito de cada participante.

Além do uso dos dispositivos de RV e aparelhos de medição fisiológica, um questionário da *University College London* [SLA95] [USO99] foi aplicado ao final de cada sessão. Este documento, composto de 13 questões, avaliou a sensação de presença, o comportamento do usuário como se estivesse presente no AV, e a facilidade de locomoção, de maneira subjetiva.

Por meio desta configuração, três experimentos foram realizados visando comprovar a eficiência no uso de dados fisiológicos como medida de presença. O primeiro experimento buscava analisar os efeitos de múltiplas exposições, observando se a presença evocada pelo AV declinava a cada novo teste. No segundo experimento, foi realizada uma comparação entre AVs com e sem estímulo tátil, procurando detectar se a existência desse artifício aumenta a sensação de presença. Já o terceiro experimento examinava o efeito de se usar diferentes taxas de atualização da imagem (10, 15, 20 e 30 quadros por segundo – FPS), investigando se, quanto mais rápida a atualização das imagens, mais presença era evocada.

Em todos os experimentos, foram realizados testes que coletavam as reações fisiológicas em situações consideradas “não-ameaçadoras”, nas quais o usuário interagiu apenas na sala que não continha o fosso, denominada “sala de treinamento” e em situações “estressantes”, nas quais o usuário executava toda a tarefa interagindo também com a “sala do fosso”.

Para múltiplas exposições, participaram 10 pessoas executando tarefas no AV três vezes por dia. No experimento, um teste em situação “estressante” foi intercalado por dois testes em situação “não-ameaçadora” durante quatro dias.

O experimento com estímulo tátil contou com dois grupos de 26 pessoas realizando tarefas distintas. Enquanto um grupo testava com tato, o outro testava sem, durante dois dias.

Já para os testes que investigaram o efeito de diferentes taxas de exibição, 33 usuários foram divididos em quatro grupos, testando o AV durante quatro dias.

Os resultados da pesquisa mostraram que medidas fisiológicas podem ser utilizadas como medida de presença em AVs geradores de estresse, e podem servir como substitutas a determinadas abordagens subjetivas comumente utilizadas.

Dentre as medidas testadas, a HR apresentou resultados estatisticamente significantes na comparação entre os ensaios em situações “não-ameaçadoras” e “estressantes”, bem como mostrou forte correlação com o questionário de presença. A HR também indicou variações distintas no ritmo cardíaco entre AVs com e sem estímulo tátil, e na comparação entre aplicações com diferentes taxas de atualização das imagens.

Neste trabalho, a SC também apresentou resultado satisfatório, porém não tão consistente como a HR. Apenas os experimentos de múltiplas exposições e *feedback* tátil forneceram dados para uma medida de presença confiável, válida, sensível e objetiva. Já as respostas de ST não apresentaram resultados capazes de determinar quanta sensação de presença o AV evoca no usuário.

Os autores também realizaram uma comparação entre usuários, considerando apenas as primeiras exposições no AV de cada experimento. O objetivo era avaliar o efeito das múltiplas exposições na presença, uma vez que decréscimos significativos entre as medidas fisiológicas da primeira e as experiências subsequentes foram encontrados. Desta forma, foi possível evitar uma possível variância devido às diferenças naturais entre os indivíduos.

Nesta análise, foi constatado que a sensação de presença diminuiu entre um experimento e o seguinte, sem chegar, entretanto, a um quadro sem sensação de presença, no experimento de múltiplas exposições. Resultados similares também foram

encontrados no segundo experimento, quando a interação contava com a inclusão do estímulo tátil, e no terceiro experimento, quando ocorria uma variação expressiva da sensação de presença, comparando sessões de 15 para 20 FPS e de 15 para 30 FPS.

Uma continuidade deste trabalho, também apresentado por Meehan *et al* [MEE05], acrescentou o efeito da latência na sensação de presença. Neste caso, os autores definem latência como o período de tempo entre um movimento da cabeça, no mundo real, e sua exibição através do HMD, no mundo virtual. O experimento considerava que, quanto menor o período de latência, mais presença era evocada.

Para estes testes, duas configurações foram utilizadas: 50ms e 90ms. O experimento contou com a participação de 64 usuários, divididos em dois grupos. No primeiro grupo, foram coletadas as respostas da HR, enquanto no segundo as respostas da SC. Como cada participante participava de apenas um teste, o experimento foi balanceado entre testes com 50ms e 90ms.

Da mesma forma que o trabalho anterior, a HR apresentou resultados consistentes e correlação com o questionário de presença, constatando uma maior sensação de presença quando a latência era de 50ms.

No entanto, os autores ressaltam que apesar de medidas fisiológicas poderem ser utilizadas na avaliação de presença em AVs, é preciso continuar investigando a relação entre estas respostas e os questionários em outras abordagens, como a interação em ambientes com avatares ou que exploram recursos de iluminação e física.

3.2.2 Medidas fisiológicas e quebras na sensação de presença

Slater *et al* [SLA06] relataram a dificuldade em elaborar e conduzir questionários que medem presença subjetivamente. Em função disto, conduziram um experimento que procurava entender a relação entre medidas fisiológicas, quebras na sensação de presença e os efeitos das interações das pessoas com personagens virtuais em um AV que simula um cenário de bar (*pub*).

Segundo Slater e Steed [SLA00], as quebras de presença (*breaks in presence*) são fenômenos percebidos pelo usuário que interrompem sua sensação de presença durante a experiência no AV, devolvendo sua consciência à situação real que o cerca. Exemplos disso contemplam tanto obstáculos reais, como se enroscar em cabos ou se bater em paredes, como efeitos do próprio sistema, como perceber falhas na aplicação de texturas em objetos do AV.

No caso do experimento, as quebras eram causadas por pausas na projeção do AV, fazendo com que o cenário virtual “desaparecesse” por alguns instantes.

O experimento analisou 20 usuários interagindo em um sistema de CAVE (do inglês *Cave Automatic Virtual Environment*), com telas de projeção para o chão e três paredes. Um óculos *shutter glasses* com rastreador de cabeça ofereceu visão estereoscópica do cenário. Neste AV, a tarefa do usuário era deslocar-se por um bar virtual, prestando atenção ao que acontecia no ambiente, e interagir livremente com as personagens virtuais.

As medidas fisiológicas eram processadas pelo equipamento Procomp Infiniti [PRO03], da Thought Technology¹³, o qual realizava a leitura das respostas de SC e HR. A Figura 11 apresenta uma parte do AV visualizado pelo usuário, bem como a disposição dos equipamentos de RV e *biofeedback* durante o experimento.



Figura 11. Bar virtual e usuário interagindo na CAVE, usando dispositivos de *biofeedback*. Adaptado de Slater *et al* [SLA06].

Antes de cada teste, cada participante preencheu dois questionários. O primeiro questionário, de natureza demográfica, coletava informações sobre experiências com AVs e o uso de jogos de computador. Já o segundo questionário, de 27 questões objetivas, visava obter informações sobre ansiedade social do participante.

O experimento dividiu-se em três etapas, denominadas basal, treinamento e experimental, nas quais a coleta das medidas fisiológicas sempre era realizada. Na etapa basal, os usuários entravam na CAVE e ficavam parados por, aproximadamente, dois minutos, sem visualizar qualquer imagem. O objetivo era coletar dados fisiológicos como base para posterior análise. Na etapa de treinamento, os usuários interagiam em uma sala virtual repleta de números, procurando adaptar-se aos equipamentos de RV. Para isto, moviam-se pelo cenário procurando os números, em ordem crescente.

¹³ <http://www.thoughttechnology.com>

Na etapa experimental, iniciava-se a interação com o bar virtual, no qual cinco avatares conversavam com os participantes, repassando informações sobre o AV. Durante esta etapa, a interação era interrompida por quatro pausas de dois segundos na projeção do AV, a cada intervalo de cinco minutos.

Ao final de cada experimento, os usuários eram entrevistados, durante aproximadamente 15 minutos, para relatar sobre as experiências vivenciadas na sessão experimental. Eles também preenchiam um questionário que descrevia suas ações durante a interação com os personagens virtuais.

O estudo mostrou que a HR e os parâmetros de HRV, calculados com base nos intervalos entre batimentos cardíacos, variaram significativamente entre as etapas de treinamento e experimental. Os autores destacam que isto acontece devido ao fato dos usuários estarem mais relaxados e atentos na fase experimental, uma vez que já haviam se familiarizado com os equipamentos e a experiência de RV.

A pesquisa também mostrou que participantes que se consideraram socialmente mais ansiosos apresentaram um aumento na HR em torno de nove batimentos cardíacos, em comparação com aqueles que se consideraram socialmente mais confiantes. Isto mostrou que o AV retratou o esperado em uma situação real equivalente.

Com base nas medidas de SC e HRV, foi possível visualizar os momentos em que ocorriam as quebras na sensação de presença, associados aos repentinos “apagões” durante a projeção do AV.

Porém, de acordo com os autores, o resultado mais surpreendente está relacionado às alterações na HR quando um avatar repassa informações ao usuário. Em dois participantes, a HR chegou a apresentar um aumento de quase 3% quando estes prestavam atenção à fala de personagens virtuais.

Apesar dos resultados promissores, os autores também advertem que a metodologia empregada pode não distinguir quebras de presença de situações inesperadas que acarretam em mudanças de estado do usuário, como, por exemplo, alteração na atenção e desorientação.

Dando continuidade a esta pesquisa, Brogni *et al* [BRO06] [BRO07] utilizaram uma configuração similar de equipamentos e métodos de análise de dados para avaliar o impacto de realismo visual no nível de estresse do usuário. Neste estudo, os autores analisam se dados fisiológicos podem ajudar na interpretação de questionários e de dados de desempenho do usuário.

Para tanto, os autores utilizaram um AV que simulava uma rua comercial, onde 16 avatares circulavam, transmitindo um aspecto de cenário movimentado, como mostra a

Figura 12. Diferentemente do experimento anterior, não foram provocadas pausas na projeção. Neste caso, a diminuição da sensação de presença era induzida pela redução da qualidade visual das imagens do AV, a qual estava condicionada à repetição de texturas e à mudança na aparência dos avatares.

O experimento contou com a participação de 40 pessoas, e tinha como objetivo a exploração do AV até encontrar uma porta de saída. Para navegar no AV, o usuário utilizava um *joystick*.



Figura 12. Aplicação da rua virtual, à esquerda, e usuário interagindo na CAVE (adaptado de Brogni *et al* [BRO07]).

Como medidas fisiológicas foram utilizados apenas os parâmetros de HRV. Antes de cada teste, cada participante preenchia um questionário de natureza demográfica, e era treinado para detectar as quebras de presença. Após cada experimento, os usuários respondiam dois questionários de medição de presença, disponíveis na literatura (S-U-S [USO00] e ITC-SOPI [LES01]).

As etapas basal e de treinamento foram similares ao trabalho de Slater *et al* [SLA06], com exceção de que o usuário comunicava os avaliadores por meio de um botão no *joystick*, nos momentos em que sentia quebras de presença. Já a etapa experimental teve duração de três minutos, e foi dividida em duas fases de 90 segundos, para fins de análise estatística.

Na etapa experimental, os usuários foram divididos em quatro diferentes grupos, com o objetivo de avaliar o impacto nas variações da cena [VIN04], de acordo com qualidade das texturas utilizadas e o realismo visual dos personagens virtuais. Para tanto, os pesquisadores combinaram condições sem repetição e com repetição de 50% das texturas, e avatares com aparência foto-realística ou aparência de *cartoon*.

Nesta pesquisa, os parâmetros de HRV indicaram que o usuário sentiu um nível de estresse maior quando estava iniciando a realização de uma nova tarefa no AV. Da mesma forma, o nível de estresse diminuiu gradualmente à medida que o tempo passa, uma vez que o usuário ia se familiarizando com o AV, e se sentindo mais confortável.

Considerando a análise da HRV e os resultados dos questionários, também foi possível observar que situações nas quais texturas não eram repetidas representaram um fator de influência maior no nível de estresse. O trabalho sugere que os efeitos da fidelidade visual podem ter um impacto positivo nas respostas do usuário.

Os autores também destacam que é preciso avaliar a influência do tempo de interação do usuário no AV como indutor de estresse, pois existe a chance do usuário acostumar-se com os problemas ou limitações do sistema, acarretando em alterações no seu comportamento fisiológico.

3.2.3 Medidas fisiológicas nas terapias de exposição com Realidade Virtual

Nesta seção, são apresentados trabalhos que tratam do uso de medidas fisiológicas como parâmetro do estado emocional do usuário, sem uma relação direta com avaliação de usabilidade. De qualquer maneira, serve como um exemplo do uso deste recurso e para fins de entendimento do poder desta ferramenta.

Conforme Bush [BUS08], a Terapia de Exposição com RV (TERV ou VRET, do inglês *Virtual Reality Exposure Therapy*) é uma técnica de tratamento de fobias e transtornos de ansiedade, que combina recursos de *hardware* e *software* em um sistema computacional capaz de imergir o paciente em AVs que simulam situações geradoras de ansiedade.

Hodges *et al* [HOD01] destacam ainda que a TERV é uma maneira de oferecer um cenário de tratamento capaz de desestruturar lembranças traumáticas em pacientes com fobias e transtornos de ansiedade, de maneira gradual, prolongada e sistemática.

Revisões bibliográficas apresentadas por Krijn *et al* [KRI04a] e Powers e Emmelkamp [POW08] demonstram que estudos de caso têm apresentado resultados satisfatórios no tratamento de fobias específicas, como medo de voar [ROT00] [ROT02] [MÜH03] e medo de altura (acrofobia) [EMM02] [KRI04b].

Para observar o comportamento dos pacientes durante as sessões de exposição aos AVs causadores de estresse, os experimentos citados utilizam equipamentos de monitoramento fisiológico, oferecendo ao terapeuta informações que o ajudam a treinar o paciente na auto-regulação do medo, e avaliar sua evolução clínica [WIE02b]. Parsons e Rizzo [PAR08] destacam que aspectos qualitativos do AV, como nível de presença e

imersão, são avaliados por outros instrumentos, como questionários e entrevistas com os pacientes, após a sessão de terapia.

Por este motivo, este trabalho opta em não abordar as pesquisas que envolvem o uso de medidas fisiológicas em AVs destinados às terapias de exposição.

3.3 Avaliação de Usabilidade em Ambientes Virtuais

As subseções a seguir apresentam os trabalhos que utilizam medidas fisiológicas para avaliar aspectos usabilidade de jogos de computador, e para minimizar o desconforto em AVs.

3.3.1 Medidas fisiológicas e custo do usuário em jogos

Os trabalhos de Lin *et al* [LIN05] [LIN06a] apresentam um estudo sobre o uso de medidas fisiológicas como métrica de usabilidade para avaliação de videogames, investigando o relacionamento do *biofeedback* com medidas de avaliação subjetiva e de desempenho em uma tarefa. Para tanto, são consideradas medidas de HR, SC e BVP.

Estes trabalhos destacam que os métodos de usabilidade estão preocupados em avaliar a produtividade e a satisfação do usuário, não medindo questões importantes como seu esforço cognitivo, que tem relação com a qualidade da experiência. Pesquisas da área de Interação Humano Computador (IHC) [SWE93] [WIL00] denominam este esforço como “custo do usuário”, referindo-se ao empenho afetivo, mental e físico que o usuário realiza para se adaptar a um sistema, com vistas de manter um mesmo nível de desempenho e satisfação durante o processo interativo. Os autores defendem também o uso de medidas fisiológicas para compreender os comportamentos do usuário durante a interação, uma vez que profissionais da área clínica, como fisiologistas e psicólogos, vêm utilizando estes parâmetros para determinar estados emocionais, esforço cognitivo e estresse.

Para investigar a correlação entre as diferentes medidas de usabilidade, os trabalhos analisaram sujeitos interagindo com o vídeo game Super Mario 64 da Nintendo¹⁴. Dezoito universitários na faixa de 19 a 31 anos participaram dos testes, sendo 14 deles do sexo masculino. Os jogos foram disputados utilizando um equipamento Nintendo 64 e uma televisão de 25 polegadas.

¹⁴ <http://www.nintendo.com/>

Nesta experiência, cada pessoa foi avaliada enquanto jogava em três diferentes níveis de dificuldade do jogo, tendo por objetivo realizar as tarefas corretamente, no menor tempo possível. Na primeira tarefa, o objetivo era alcançar o topo de uma montanha, derrotando um inimigo durante o caminho. Na segunda tarefa, era preciso bater em um poste de madeira, para evitar o ataque do inimigo. Já na terceira tarefa, o objetivo era percorrer um trajeto esquiando, evitando quedas e desviando de obstáculos.

Para avaliar esta experiência, o trabalho utilizou duas formas de avaliação: uma subjetiva, utilizando questionários, e outra objetiva, utilizando dados de desempenho na tarefa e medidas fisiológicas.

Como nos experimentos de Slater *et al* [SLA06], apresentado na Seção 3.2.2, a coleta dos parâmetros fisiológicos foi realizada utilizando-se dos recursos do dispositivo ProComp Infinity. Os sensores de SC foram fixados nos dedos da mão esquerda do usuário, enquanto os sensores de HR e BVP, em dedos da mão direita, conforme ilustra a Figura 13.



Figura 13. Disposição dos sensores de *biofeedback* e *gamepad* utilizado no experimento (adaptado de Lin *et al* [LIN06a]).

Na avaliação subjetiva, foram utilizados dois questionários. O primeiro, aplicado antes da sessão de testes, tinha por objetivo captar informações demográficas e sobre a experiência em jogos dos participantes. Já o segundo, aplicado após a experiência, avaliava os níveis de dificuldade da tarefa e o estresse causado pelo jogo, utilizando a escala de avaliação de esforço mental [ZIJ93].

O experimento dividiu-se em quatro fases. Na primeira fase, os usuários tomavam ciência dos testes, e respondiam ao primeiro questionário. Na segunda fase, eram repassadas instruções sobre as regras e as tarefas do jogo, para leitura individual. Na terceira fase, os participantes jogavam os três níveis de jogo, com tempo máximo de 10 minutos para completar cada tarefa. Na última fase, cada usuário descansava por 15 minutos, respondia a um questionário e relatava suas impressões sobre a experiência.

Segundo os autores, uma comparação entre as medidas de SC coletadas em cada tarefa apresentou diferenças significantes entre as médias normalizadas de SC (ANOVA, $F = 11,6$, $p < 0,001$), apontando a terceira tarefa como a maior causadora de estresse, seguida pela segunda e pela primeira, respectivamente. Este mesmo comportamento se refletiu na avaliação subjetiva do nível de estresse, indicando uma correlação entre estas duas medidas.

Para comparar a medida de SC e o desempenho em cada tarefa, primeiramente, os autores optaram por classificar os participantes em três grupos distintos, baseados no número de acertos: o grupo entre 0 e 7 acertos (baixo), o grupo entre 8 e 15 acertos (médio) e o grupo com mais de 16 acertos (alto). Neste caso, o trabalho apontou diferenças significantes entre os grupos. O grupo “baixo” apresentou um nível maior de estresse, seguido dos grupos “médio” e “alto”, sugerindo que jogadores com maior habilidade apresentavam variações fisiológicas menores do que jogadores inexperientes.

Uma terceira análise dos dados fisiológicos avaliou a evolução do nível de estresse dentro de cada grupo, mostrando que o comportamento dentro dos grupos seguia a mesma tendência intergrupos, ou seja, à medida que o número de acertos aumentava dentro do grupo, a SC decrescia, indicando uma diminuição do estresse, sendo mais evidente no grupo “baixo”.

O estudo também avaliou a relação entre medidas fisiológicas e eventos de frustração. Para tanto, vídeos da primeira e da segunda tarefas, contendo eventos como ser atacado por inimigos, cair acidentalmente de uma montanha ou de uma ponte, e ser inesperadamente atingido por uma bomba, foram apresentados aos usuários. A avaliação foi realizada considerando-se os 5 segundos anteriores e os 10 segundos posteriores ao evento. Dos 355 eventos de frustração identificados, 256 indicaram um aumento de mais de 5% no sinal de SC.

Também foi investigada a relação entre desempenho na tarefa e frustração. Os resultados sugeriram uma tendência de decréscimo na média de eventos frustrantes com o aumento no número de acertos.

Referente às análises utilizando medidas de HR e BVP, os autores destacam que estas medidas não apresentaram diferenças significantes entre as tarefas e os níveis de desempenho. No entanto, o trabalho relata que esta avaliação foi prejudicada pelo excesso de ruídos no sinal coletado, uma vez que o dispositivo de BVP é sensível a movimentos.

Como conclusão, os autores citam que apesar de terem encontrado fortes correlações entre medidas fisiológicas, dados subjetivos e de desempenho, faz-se necessário uma avaliação com um número maior de participantes, separando-os por nível de experiência, e utilizando outros indicadores, como a HRV e o rastreamento do olhar. Com referência a estes indicadores, os mesmos autores apresentam trabalhos posteriores nos quais inserem parâmetros de HRV e dados sobre o rastreamento do olhar para avaliar o esforço cognitivo [LIN06b] [LIN08]. Estes trabalhos apresentam resultados comparativos entre estas medidas e métricas de usabilidade que possibilitam compreender mudanças de desempenho, efeitos de aprendizagem ao longo do tempo e diferenças no esforço cognitivo de cada usuário.

Para tanto, uma nova análise com 10 participantes (nove homens e uma mulher) foi realizada, utilizando uma das fases do jogo “Luxor: Amun Rising”¹⁵. O objetivo do jogo era destruir conjuntos de três ou mais esferas, da mesma cor, antes destas alcançarem uma pirâmide. Para interagir com a aplicação, o usuário utilizava apenas o *mouse*. Recursos sonoros foram desabilitados para evitar a influência das músicas nas reações fisiológicas. Os participantes jogaram os níveis “iniciante”, “intermediário” e “experiente”.

Neste experimento o movimento dos olhos e o tamanho da pupila foram gravados por uma câmera de rastreamento, presa à cabeça do usuário. Como nos trabalhos anteriores de Lin *et al* [LIN05] [LIN06a], utilizou-se de um sensor de ECG, conectado dispositivo ProComp Infinity para coletar dados de HR. Para este experimento, os eletrodos do sensor foram presos no peito do usuário. A interpretação do sinal e os cálculos para os parâmetros de HRV utilizaram recursos do *software* CardioPro, que acompanha o equipamento utilizado. A Figura 14 apresenta a disposição dos equipamentos, bem como o cenário de jogo utilizado.

Outra medida capturada foi o movimento das mãos do usuário. Para tanto, foi realizado um registro da distância percorrida pelo cursor, considerando tempo e localização dos cliques de *mouse* efetuados.

Para avaliação subjetiva, primeiramente foi utilizado um questionário pré-teste, de caráter demográfico, para avaliação de conhecimentos prévios, e um segundo, pós-teste, para avaliar a carga de trabalho mental do usuário. Para esta avaliação, os autores utilizaram o questionário NASA-TLX [HAR88] [HAR06], um instrumento capaz de avaliar a carga de trabalho a partir de uma escala multidimensional, baseada nas avaliações de

¹⁵ <http://luxor-amun-rising.com/>

seis sub-escalas: exigência mental, exigência física, exigência temporal, nível de realização, nível de esforço e nível de frustração.



Figura 14. Jogo “Luxor: Amun Rising” e disposição dos equipamentos durante os testes (adaptado de Lin *et al* [LIN08]).

O experimento também se dividiu nas mesmas quatro fases do trabalho anterior, quais sejam: responder o questionário pré-teste, receber instruções sobre o jogo, interagir com o jogo e responder o questionário pós-teste. Como complemento a esta metodologia, antes de iniciar o jogo eram coletadas informações sobre o tamanho da pupila e sinais de ECG do usuário para servirem de base para posterior análise.

Neste experimento, os autores encontraram diferenças significantes entre os três níveis de tarefa, tanto na avaliação subjetiva, utilizando as medidas do NASA-TLX, como na avaliação fisiológica, utilizando os parâmetros de HRV. As comparações também mostraram correlação entre estas medidas, além de, em ambos os casos, indicarem o nível de jogo “experiente” como o que leva ao maior desgaste do usuário, seguido dos níveis “intermediário” e “iniciante”. As comparações entre a medida de tamanho da pupila e a dificuldade na tarefa não apresentaram diferença significativa, e apenas correlacionaram-se com a medida subjetiva no nível de jogo “iniciante”.

Procurando entender a relação entre o custo de utilizar uma interface e o comportamento dos usuários nos níveis de dificuldade, o trabalho comparou o movimento das mãos e dos olhos do usuário com as medidas anteriores, com base em outro trabalho anterior de Lin *et al* [LIN04]. Para tanto, foram considerados os parâmetros de número de fixações do olhar, duração destas fixações, distância dos movimentos oculares,

velocidade de movimentos sacádicos¹⁶, número de cliques e distância percorrida pelo cursor.

Os resultados apresentaram o movimento dos olhos como uma medida sensível às variações de esforço cognitivo, diminuindo à medida que o nível de dificuldade do jogo aumentava. No entanto, isto somente ficou evidente em comparações com o nível “iniciante”. Além disso, o trabalho relatou que a velocidade de movimentos sacádicos é um parâmetro que apresenta correlação com a medida de HRV.

A partir destes estudos de caso, os autores propõem duas novas medidas. A primeira, criada para determinar o custo do usuário, combina a soma da distância dos movimentos oculares com a distância percorrida pelo cursor. Já a segunda, para medir seu desempenho, utiliza a razão entre o número de fixações do olhar sobre o número de cliques do *mouse*.

Para a primeira medida, denominada “medida olho-mão para custo do usuário”, foram aplicados testes de análise de variância e comparações *post hoc*, que apontaram diferenças significantes entre os três níveis do jogo (ANOVA, $p = 0,02$; Fisher LSD test, $p < 0,01$ entre iniciante e experiente e $p < 0,05$ nos demais casos). Os autores também encontraram forte correlação com o questionário NASA-TLX ($r = 0,87$, $p < 0,01$). Já para a segunda medida, denominada “índice de desempenho olho-mão”, foi examinada a correlação com o número de conjuntos de esferas destruídos, a qual também indicou que conforme decrescia a razão, aumentava o desempenho na tarefa ($r = -0,78$, $p < 0,01$).

Os autores concluem que todos estes estudos de caso ofereceram subsídios para entender o impacto de ambientes interativos, como jogos de computador, e avaliar o esforço e desempenho dos usuários por meio de múltiplas e variadas fontes de dados. Isto permitiu também o aperfeiçoamento de diretrizes para projetos de interfaces.

3.3.2 Medidas fisiológicas e desconforto em ambientes virtuais

Dentro do escopo dos artigos selecionados para este estudo, apenas Kim *et al* [KIM08] apresentam um estudo sobre o uso de parâmetros fisiológicos para reduzir o desconforto (do inglês *cybersickness*¹⁷) durante uma exposição a AVs imersivos. O trabalho propõe um sistema para detecção em tempo real de *cybersickness*, que utiliza

¹⁶ Movimentos oculares rápidos e pequenos, que direcionam ambos os olhos de uma região de interesse para outra a ser fixada.

¹⁷ Termo utilizado para descrever o desconforto, como fadiga e náusea, que pode ocorrer durante a utilização de AVs, enquanto o usuário visualiza ou interage com cenas em movimento, e permanece fisicamente parado.

uma rede neural artificial com entradas para diferentes sinais fisiológicos, entre eles ECG, SC, BVP, ST, BF, EOG, eletrogastrograma¹⁸ (EGG) e eletroencefalograma¹⁹ (EEG). O objetivo é minimizar o desconforto, sem afetar a sensação de presença, por intermédio da redução automática do campo de visão e da desaceleração do deslocamento do usuário no AV, quando o *cybersickness* é detectado pelo sistema.

Para tanto, foi realizado um estudo de caso com 47 participantes (23 homens e 24 mulheres), em condições saudáveis, sem disfunções do sistema vestibular e livres de efeitos de medicamentos. O AV simulava uma área universitária, e era projetado em uma CAVE de três telas. Nesta aplicação, o objetivo do usuário era encontrar 10 lixeiras, espalhadas pelo cenário, utilizando um *joystick* em forma de volante para o deslocamento no cenário. O usuário interagiu neste ambiente duas vezes, com e sem o sistema de detecção, dentro do intervalo de um mês.

Para coletar os sinais fisiológicos, um equipamento do tipo polígrafo²⁰ foi utilizado, permitindo conectar diferentes sensores de medição. A leitura e a geração dos dados foram realizadas pelo *software Acknowledge*, da empresa Biopac²¹. Estes dados alimentavam o sistema de detecção, informando 13 tipos de sinais fisiológicos, os quais eram normalizados para servir de entrada à rede neural. Esta rede, ao final, informava um índice de probabilidade de ocorrer ou não *cybersickness*, o qual era enviado à aplicação para que esta pudesse controlar a exibição do AV.

Além disso, quatro questionários eram aplicados aos participantes. Antes da experiência virtual, os questionários de tendência imersiva (*Immersive Tendency Questionnaire* [WIT98]) e de susceptibilidade a mal-estar causado pela movimentação (*Motion Sickness Susceptibility Questionnaire* [GOL98]) eram preenchidos. Após os testes, os participantes respondiam aos questionários de presença (*Presence Questionnaire* [WIT98]) e de simulação de mal-estar (*Simulator Sickness Questionnaire* [KEN93]).

Ao iniciar o procedimento de teste, o participante recebia explicações sobre a tarefa a ser executada e instruções sobre como manipular objetos no AV. Em seguida, os

¹⁸ Equipamento que registra o potencial elétrico produzido pela musculatura gástrica, por meio de eletrodos fixados na região abdominal do paciente.

¹⁹ Equipamento que registra o potencial elétrico produzido pelo encéfalo, por meio de eletrodos aplicados no couro cabeludo, na superfície encefálica, ou até mesmo dentro da substância encefálica.

²⁰ Aparelho capaz de medir e gravar registros de diversas variáveis fisiológicas, utilizado pela polícia como detector de mentiras durante um depoimento.

²¹ <http://www.biopac.com/>

questionários pré-teste eram preenchidos, e os equipamentos de medição e interação, fixados.

Logo após, o usuário passava por um período basal de 5 minutos, visualizando uma imagem estática. O período de exposição era iniciado, com duração máxima de 9 minutos e meio. A experiência também poderia ser encerrada antes do tempo previsto, pelo próprio participante, nos casos de enjôos, vômitos ou solicitação de término.

Posteriormente, uma imagem estática novamente era visualizada pelo usuário durante 1 minuto, antes de preencher aos questionários pós-teste. Os dados fisiológicos eram adquiridos durante a visualização das imagens estáticas e durante a interação com o AV.

A análise dos dados apontou para uma frequência de *cybersickness* significativamente menor quando os usuários utilizavam o sistema de detecção, em comparação à situação na qual este sistema estava ausente. O mesmo comportamento foi observado na comparação com o questionário de simulação de *cybersickness*. Ainda relacionado a este assunto, os resultados indicaram que a alteração do campo de visão influenciou na severidade do *cybersickness*.

Já a análise do questionário de presença indicou que não existe diferença significativa entre as condições experimentais, mostrando que a existência de um sistema de detecção não interfere na sensação de presença do usuário.

Uma diferença significativa na medida de taquiarritmia gástrica foi encontrada na comparação entre os parâmetros fisiológicos e o ambiente que usava o sistema de detecção automática. Obtida pelo EGG, a taquiarritmia apresentou um elevado aumento entre os períodos basal e experimental, quando o usuário utilizou o sistema proposto.

Como conclusão, os autores destacam a eficácia do método na redução do *cybersickness* induzido por dispositivos de RV, uma vez que o uso de medicamentos para o mal-estar e uma sessão de adaptação aos dispositivos de *biofeedback* são desnecessários. Conforme os resultados apresentados, o método possibilitou a apresentação de um *feedback* adequado ao comportamento fisiológico de cada usuário durante a exposição ao AV.

Observando aspectos importantes dos trabalhos apresentados nesta seção e nas seções anteriores, os próximos dois capítulos apresentam a fundamentação e a definição de uma metodologia destinada a avaliar a qualidade de interfaces 3D, usando medidas fisiológicas em conjunto com métricas convencionais de usabilidade.

4. FUNDAMENTAÇÃO DA METODOLOGIA

O objetivo principal deste trabalho é definir uma metodologia para avaliar a qualidade de técnicas de interação em ambientes virtuais imersivos utilizando medidas fisiológicas. Com base nos trabalhos anteriormente citados, esta proposta se apresenta como inovadora, uma vez que pretende determinar o relacionamento entre métricas de usabilidade e medidas fisiológicas em interfaces 3D, por intermédio de um protocolo de testes e de técnicas estatísticas, específicos para esta análise. Da mesma forma, também pretende verificar medidas fisiológicas como métrica de usabilidade, por meio da comparação entre duas técnicas de interação, entre períodos do experimento e entre etapas do processo interativo utilizando testes de hipóteses.

Em linhas gerais, a ideia é permitir o uso de medidas fisiológicas para identificar momentos do processo interativo que apresentam falhas de projeto, peculiaridades passíveis de melhoramentos ou características que apontam a vantagem de uma técnica sobre a outra. Como exemplo, pode-se citar a detecção de problemas de uma nova técnica de interação, e a definição do período necessário para adaptar-se a um novo dispositivo.

Além disso, o uso das medidas fisiológicas pode simplificar o processo de testes de aplicações de RV, reduzindo a dependência de dados subjetivos dos usuários e evitando a necessidade de modificação de *software* legado, para coleta de dados objetivos.

O restante deste Capítulo detalha a base da metodologia, justificando sua importância no aperfeiçoamento de interfaces de RV e descrevendo os procedimentos adotados na criação de um protocolo de teste e dos passos para análise estatística. Um estudo piloto também é apresentado, com o intuito de levantar subsídios para revisão dos procedimentos para avaliações com um grupo maior de usuários. Um estudo de caso mais detalhado é apresentado pelo Capítulo seguinte, compreendendo a definição dos procedimentos e os testes de validação.

4.1 Justificativa

O uso de medidas fisiológicas como métrica de usabilidade de AVs ainda é um assunto de pesquisa pouco explorado. Conforme apresentado no Capítulo 3, foram encontrados apenas 10 trabalhos publicados, sendo o primeiro no ano de 2002, os quais buscam relacionar algumas métricas de avaliação, como presença e conforto, com

medidas fisiológicas do usuário. Além disso, apenas metade dos trabalhos envolve o estudo destas medidas em AVs imersivos.

Logo, dentro do escopo dos trabalhos pesquisados, concluiu-se não existirem estudos que explorem a relação entre medidas fisiológicas e métricas de usabilidade utilizadas na avaliação de interfaces 3D.

Considerando os trabalhos relacionados, a proposta desta pesquisa se assemelha às abordagens de Lin *et al* [LIN05] [LIN06a] [LIN06b] [LIN08], uma vez que também pretende explorar a correlação entre medidas de desempenho, respostas subjetivas e medidas fisiológicas para avaliar usabilidade. No entanto, este trabalho busca estudar o relacionamento destas medidas em interfaces 3D, nas quais diferentes dispositivos e técnicas de interação estão envolvidos. Outra diferença com relação aos trabalhos de Lin *et al* reside no fato de que enquanto aqueles são voltados à avaliação de jogos 2D, este trabalho discute um novo método de avaliação de usabilidade que objetiva melhorar o desenvolvimento de técnicas de interação 3D.

Em resumo, pretende-se descobrir se métricas objetivas e subjetivas, destinadas à avaliação do sistema, do desempenho e das preferências do usuário em interfaces 3D, podem ser complementadas ou, eventualmente, substituídas por medidas fisiológicas.

Métricas subjetivas estão sujeitas, por exemplo, a problemas como diferentes interpretações de dados, omissão de informações e dissociação com fatores como cansaço e estresse. Neste caso, medidas fisiológicas podem contribuir na avaliação da qualidade de técnicas de interação, uma vez que apresentam dados objetivos, sofrem influência de fatores externos como esforço físico e mental, e são respostas independentes e precisas relacionadas ao comportamento do indivíduo. Além disso, podem servir de ajuda na compreensão das preferências de cada indivíduo.

Já as métricas objetivas requerem, em alguns casos, alterações no código fonte do *software*, podendo tornar o sistema complexo, aumentar os custos de projeto e consumir tempo de desenvolvimento. No caso de *softwares* legados, por exemplo, a inserção destas métricas pode exigir um considerável esforço de programação para adaptação do código, o que torna o processo suscetível à introdução de erros ou ao prolongamento de prazos para compreensão dos detalhes do sistema [BIS02]. No caso de *softwares* em geral, de acordo com Tullis e Albert [TUL08], a inclusão de métricas em código pode influenciar no tempo destinado para análise de dados e na previsão de lançamento de uma versão de sistema. Medidas fisiológicas, por sua vez, podem reduzir a inserção de métricas em código, pela capacidade de refletir as ações e os comportamentos ligados ao estado físico e emocional de um indivíduo. Além disso, os resultados fisiológicos

permitem a combinação com outros dados quantitativos e qualitativos em um estudo de usabilidade. A análise do traçado gráfico de uma resposta fisiológica, por exemplo, pode apresentar diretamente as alterações relevantes no nível de estresse do usuário, e facilitar a avaliação de uma técnica de interação, ou da própria interface.

Uma das razões para este trabalho optar pelo estudo de medidas fisiológicas é que estas medidas fornecem dados numéricos relacionados ao comportamento natural de uma pessoa. Como apresentado no Capítulo 2, medidas fisiológicas têm ligação com o funcionamento do sistema nervoso autônomo, que é responsável pelas funções vitais do organismo frente às modificações do ambiente, e sobre as quais o indivíduo não tem controle e independem da sua vontade. Assim, espera-se que estas medidas permitam avaliar a qualidade de interfaces 3D com base no comportamento do usuário, durante o processo interativo e na comparação entre técnicas de interação.

Outra motivação para o uso de medidas fisiológicas é a possibilidade de melhor interpretar métricas subjetivas e de desempenho em tarefas interativas, servindo como complemento às medidas utilizadas por ferramentas e métodos de avaliação de usabilidade já bem fundamentados na literatura [HIX93] [ROS01] [BOW02], como protótipos, questionários, testes formativos e somativos. Desta forma, determinadas situações do processo interativo, como a ocorrência de falhas em uma técnica, podem ser corroboradas com precisão e confiabilidade, com base em dados objetivos relativos ao próprio usuário. Assim, pode-se reduzir o tempo de análise de dados e, provavelmente, o número de sujeitos de um teste.

Além disto, as medidas fisiológicas podem ser vistas como métricas objetivas que permitem detectar situações que, eventualmente, possam ter sido omitidas durante as diferentes etapas de coleta de dados. Respostas subjetivas em questionários, por exemplo, estão sujeitas à mediação cognitiva de cada sujeito e suscetíveis à influência de fatores externos, como cansaço físico e mental, irritação e, inclusive, esquecimento. Tratando-se de AVs tridimensionais, existe uma tendência a ocorrer situações desta natureza, uma vez que o usuário está envolvido com o aprendizado de novas técnicas, dispositivos e metáforas de interface 3D, o que acaba aumentando a carga cognitiva.

Como exemplo, pode-se citar também uma provável entrevista na qual o usuário relata a sensação de incômodo durante a realização de uma tarefa, causada pela exibição imperfeita de imagens. Neste caso, não é possível apontar o momento em que este defeito ocorreu, o quanto isto influenciou seu desempenho, e o quanto alterou seu comportamento natural.

Outra razão para o uso de medidas fisiológicas é que estas podem apontar problemas comuns aos usuários durante o processo interativo. O traçado fisiológico pode destacar alterações relevantes no nível de estresse do usuário, o que pode ter relação, por exemplo, com os momentos de dificuldade durante a experiência. Além disso, este traçado pode mostrar os períodos de adaptação à interface e de aprendizado de uma técnica de interação. Para tanto, os sinais coletados devem ser normalizados, uma vez que as respostas variam de pessoa para pessoa, e são particulares de cada indivíduo.

Em relação à escolha de medidas fisiológicas para avaliar a qualidade de técnicas de interação, este trabalho opta pelo estudo da HR, HRV e SC. Conforme descrito no Capítulo 2, estas medidas já apresentam resultados satisfatórios em pesquisas que envolvem as áreas clínica e de RV, e são não-invasivas. Além disso, elas também apontam para resultados relevantes, em comparação as demais medidas fisiológicas apresentadas no Capítulo 3, como BVP, ST e BF.

Já para avaliar se medidas fisiológicas podem ou não serem substituídas de métricas de usabilidade, este trabalho sugere uma comparação entre duas técnicas de interação, de mesmo propósito, utilizando métricas objetivas e subjetivas de usabilidade. Durante o processo interativo, medidas fisiológicas também são coletadas. Após este procedimento, os resultados obtidos com as primeiras medidas são comparados com os dados fisiológicos utilizando métodos estatísticos de análise multivariada de dados. De acordo com Hair *et al* [HAI09], estes métodos permitem avaliar, simultaneamente, a influência de múltiplas medidas sobre cada indivíduo ou variável sob investigação, independente da complexidade ou do contexto em que estas variações ocorrem. Assim, é possível verificar se os resultados obtidos com medidas fisiológicas são capazes de complementar ou indicar os mesmos problemas apontados pelas métricas de usabilidade.

Ressalte-se, por fim, que esta pesquisa não pretende estudar a relação entre medidas fisiológicas e sensações de presença e conforto, uma vez que os trabalhos relacionados já apresentam resultados significantes, considerando avaliações subjetivas e medidas fisiológicas em AVs imersivos [MEE02] [MEE05] [BRO06] [BRO07] [KIM08]. No entanto, utiliza-se a experiência destes trabalhos para elaborar o protocolo de testes, que considera a coleta de dados em três etapas: basal, treinamento e experimental.

A seguir, a Tabela 3 apresenta detalhes das abordagens similares a esta proposta. Neste comparativo, as revisões que apresentam o uso de medidas fisiológicas nas terapias de exposição com RV são descartadas, devido ao foco do estudo destes trabalhos estarem ligados à relação entre medidas fisiológicas e o estado de estresse do paciente, e não a questões de usabilidade da aplicação.

Tabela 3. Comparativo entre trabalhos relacionados.

Características	MEEHAN et al	SLATER et al	BROGNI et al	LIN et al	KIM et al
<i>Interface</i>	3D	3D	3D	2D	3D
<i>Ambiente Virtual</i>	Imersivo	Imersivo	Imersivo	–	Imersivo
<i>Aplicação</i>	Simulação	Simulação	Simulação	Jogo	Simulação
<i>Visualização</i>	Estéreo	Estéreo	Estéreo	Mono	Estéreo
<i>Dispositivos</i>	HMD, rastreador, mouse e tablado	CAVE e shutter	CAVE, shutter e joystick	Monitor e mouse	CAVE, shutter e joystick
<i>Tarefas de Interação</i>	Seleção, Manipulação e Navegação	Navegação	Navegação	–	Navegação
<i>Estímulos explorados</i>	Visual e Tátil	Visual e Sonoro	Visual e Sonoro	Visual	Visual e Sonoro
<i>Medidas Fisiológicas utilizadas</i>	HR, SC e ST	HR, HRV e SC	HRV	HR, HRV, SC e BVP	ECG, EOG, SC, BVP, ST, EGG, BF e EEG
<i>Nº de experimentos</i>	04	01	01	02	01
<i>Nº de usuários e grupos</i>	10; 01 26; 02 33; 04 64; 02	20; 01	40; 04	18; 01 10; 01	47; 01
<i>Medidas objetivas (desempenho)</i>	–	–	Tempo	Tempo, número de acertos, rastreamento do olhar, movimento das mãos, tamanho da pupila	–
<i>Medidas subjetivas (questionários)</i>	<i>Pós-teste:</i> presença, comportamento e locomoção	<i>Pré-teste:</i> demográfico, experiência e ansiedade social <i>Pós-teste:</i> ações no AV	<i>Pré-teste:</i> demográfico, experiência e treinamento de presença <i>Pós-teste:</i> presença e inventário de presença (ITC-SOPI)	<i>Pré-teste:</i> demográfico e experiência em jogos <i>Pós-teste:</i> dificuldade na tarefa, nível de estresse e carga de trabalho (NASA-TLX)	<i>Pré-teste:</i> tendência imersiva e de mal-estar <i>Pós-teste:</i> presença e simulador de desconforto
<i>Protocolo</i>	Treinamento – Experimento	Basal – Treinamento – Experimento	Basal – Treinamento – Experimento	Basal – Experimento	Basal – Experimento – Basal
<i>Foco do trabalho</i>	Medir presença sob estresse	Medir interrupções de presença	Medir o impacto do realismo visual	Medir a usabilidade de jogos de computador	Prever o desconforto em AVs
<i>Medidas Fisiológicas significantes</i>	HR e SC	HR, HRV e SC	HRV	HRV e SC	Taquiarritmia gástrica no EGG
<i>Referências</i>	[MEE02] [MEE05]	[SLA06]	[BRO06] [BRO07]	[LIN05] [LIN06a] [LIN06b] [LIN08]	[KIM08]

4.2 Objetivos Específicos

Com base na justificativa apresentada na Seção 4.1, e considerando a comparação entre duas técnicas de interação de mesmo propósito, os objetivos específicos deste trabalho são:

- Identificar se um ou mais parâmetros fisiológicos são capazes de avaliar a qualidade de uma técnica de interação, realizando uma comparação entre estes parâmetros e medidas de desempenho;
- Identificar se um ou mais parâmetros fisiológicos são capazes de avaliar a qualidade de uma técnica de interação, realizando uma comparação entre estes parâmetros e avaliações subjetivas;
- Identificar se as medidas de SC e HR apresentam diferenças significativas na comparação entre técnicas de interação, possibilitando afirmar qual delas é geradora de um nível menor de estresse e, conseqüentemente, mais adequada para a realização de determinada tarefa;
- Identificar se as medidas de SC e HR apresentam alterações significativas associadas a momentos do processo interativo na qual um determinado tipo de tarefa (seleção ou manipulação) ou ação está acontecendo.

Para atingir estes objetivos, inicialmente, as próximas seções descrevem a plataforma de testes, os procedimentos para concepção de uma nova metodologia de avaliação, e um estudo piloto para avaliação preliminar destas. Posteriormente, o Capítulo 5 define a metodologia, por meio da revisão da plataforma de testes, do protocolo de testes e das técnicas estatísticas, e realiza uma avaliação final com um grupo maior de usuários, aplicando este novo método.

4.3 Plataforma de Testes

Para que seja possível ilustrar o uso da metodologia, primeiramente são descritas as técnicas de interação escolhidas para avaliação usando medidas fisiológicas. Da mesma forma, são apresentados os AVs que servirão como plataforma de testes, e os dispositivos de RV utilizados.

4.3.1 Técnicas de interação

De acordo com Bowman *et al* [BOW04], uma técnica de interação é um método que permite a concretização de tarefas em uma interface gráfica, utilizando tanto componentes de *hardware*, como de *software*. A parte de *software* é responsável em mapear a informação recebida de dispositivos de entrada para alguma ação compreensível pelo sistema. Além disso, o resultado dessa operação é mapeado pela aplicação para um dispositivo de saída, que procurará representar as ações do usuário de forma tão natural quanto possível.

Duas técnicas de interação tradicionais na área de RV foram escolhidas para serem avaliadas utilizando medidas fisiológicas: apontamento e manuseio por raio [BOL80], e apontamento por raio e manuseio com a mão [BOW97]. Ambas as técnicas oferecem recursos para seleção e manipulação de objetos em AVs, e são de simples aprendizado. Além disso, são comumente utilizadas como parâmetro em interfaces que visam avaliar novas técnicas de interação.

A técnica de apontamento e manuseio por raio (*ray-casting*) é representada no AV pela figura de um “raio-laser”, emitido a partir da mão do usuário. Este raio tem por objetivo estender o alcance da mão para que o objeto alvo seja selecionado mesmo à distância e, posteriormente, manipulado.

Para selecionar um objeto, o raio deve ser apontado na direção deste, procurando interceptá-lo. A fim de facilitar a percepção do usuário, um *feedback* visual pode ser utilizado para realçar o objeto indicado. A Figura 15 apresenta um exemplo na qual a exibição de um cubo aramado, na cor vermelha, destaca o objeto selecionado.

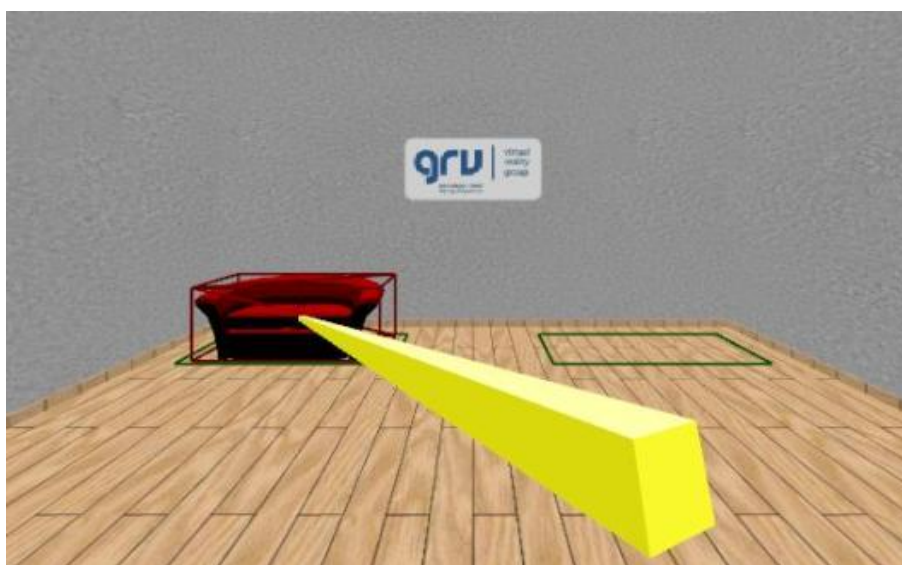


Figura 15. Exibição do cubo aramado, quando o objeto está sendo interceptado pelo raio.

Para confirmar a seleção, o usuário pressiona um botão que habilita a manipulação do objeto indicado. Posteriormente, um estímulo sensorial também pode ser utilizado para confirmar a realização desta tarefa.

Para facilitar a compreensão da técnica, o objeto é anexado (grudado) ao raio. Desta forma, entende-se que o objeto seguirá os movimentos do raio, podendo ser reposicionado para qualquer lugar do ambiente por intermédio de movimentos de punho ou braço.

Para liberar o objeto após manipulá-lo, o usuário pressiona um botão que solta o objeto em sua nova posição, seguido ou não de um estímulo sensorial que confirme esta ação. Após isto, um novo objeto pode ser selecionado.

De acordo com Bowman e Hodges [BOW97], o uso do “raio-laser” facilita as tarefas de seleção e manipulação, uma vez que basta apontar para o objeto desejado. No entanto, cabe ressaltar algumas desvantagens desta técnica. Operações de giro requerem atenção e precisão, uma vez que o manuseio não é centrado na mão do usuário. Além disso, a técnica carece de um método que controle a distância entre o objeto e o usuário.

Para solucionar estas deficiências, a técnica de apontamento por raio e manuseio com a mão, ou HOMER (*Hand-centered Object Manipulation Extending Ray-casting*) foi desenvolvida [BOW97]. Ela mantém as facilidades de apontamento e seleção da técnica anterior. No entanto, no momento da manipulação, ao invés do objeto ser preso ao raio, a mão virtual do usuário se move até a posição do objeto, e este é anexado à mão. Quando o objeto é solto, a mão retorna a sua posição natural. Desta forma, é possível manipular o objeto com movimentos centrados na mão virtual – ou seja, pode-se posicionar e girar o objeto como se ele estivesse na palma da mão do usuário, reduzindo o esforço físico e cognitivo do usuário.

O processo de apontamento de objetos é o mesmo da técnica anteriormente apresentada, bem como o processo de pegar e soltar os objetos.

Para diferenciar a técnica HOMER da técnica anterior, o “raio-laser” é ocultado durante o processo de manipulação, e substituído pela exibição de uma mão virtual, que controla a posição e a orientação do objeto. Somente após soltar o objeto, o raio torna a ser visível e a mão, ocultada. A Figura 16 mostra este contraste.

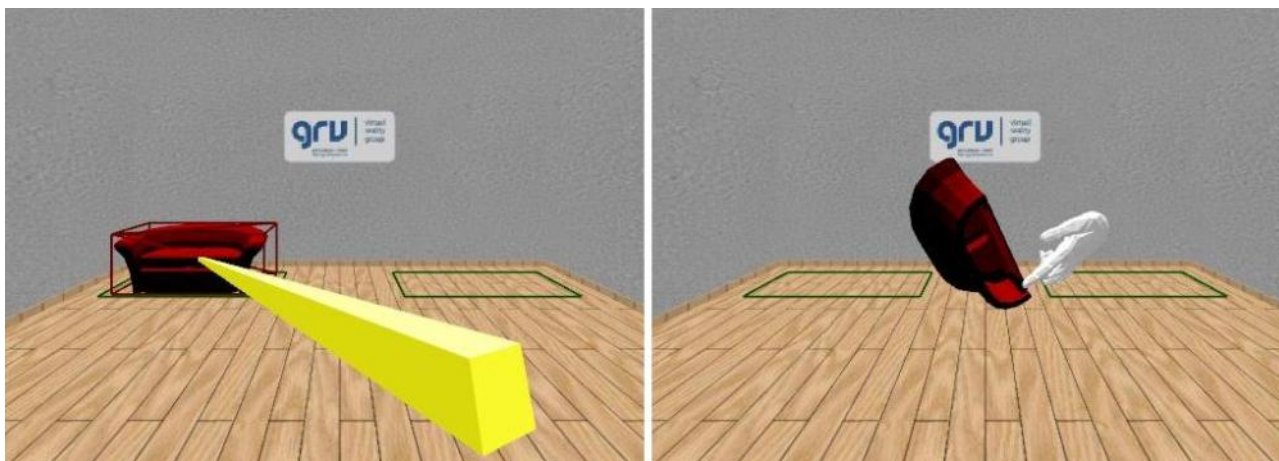


Figura 16. Exibição do raio virtual para seleção, e da mão virtual para manipulação utilizando a técnica HOMER.

4.3.2 Ambientes virtuais

De acordo com Bowman *et al* [BOW04], um AV é um cenário sintético tridimensional, gerado por computador em tempo real, na qual o usuário pode interagir utilizando múltiplos canais sensoriais, como visão, audição e tato.

Duas aplicações foram desenvolvidas para este projeto: um AV para treinamento das técnicas de interação, e outro destinado para avaliação do experimento. Estas aplicações dispõem de uma tela inicial de configuração, na qual se registra o sujeito participante e a técnica a ser avaliada. Em seguida, verifica-se a conexão com os dispositivos de RV – se estão habilitados, o cenário virtual é liberado para exibição e interação. As aplicações são automaticamente encerradas após o término do tempo estipulado, ou podem ser finalizadas de maneira manual, caso haja necessidade. Em caso de problemas na conexão com os dispositivos, ou falha no preenchimento das opções de configuração, as aplicações exibem na tela os erros detectados e aguardam o encerramento manual.

Estes *softwares* foram desenvolvidos na linguagem C/C++, em plataforma Windows, utilizando recursos das bibliotecas OpenGL²², GLUT²³ e SmallVR²⁴. Esta última é uma biblioteca gráfica que facilita o desenvolvimento de aplicações de RV, abstraindo aspectos de implementação como o controle de dispositivos não convencionais, como

²² <http://www.opengl.org>

²³ <http://www.opengl.org/resources/libraries/glut/>

²⁴ <http://grv.inf.pucrs.br/projects/SmallVR/Tutorial/CursoRV.html>

rastreadores e HMDs, e permitindo o gerenciamento de um grafo de cena, mantendo a estrutura básica de um programa construído com a GLUT.

O AV de treinamento tem por propósito oferecer uma prática inicial com a interface gráfica 3D, os diferentes equipamentos e as duas técnicas de interação para seleção e manipulação de objetos. A intenção é oferecer ao usuário um período de familiarização e aprendizado dos recursos disponíveis no ambiente.

O cenário virtual simula uma sala simples, com o chão representado por uma textura de madeira, e paredes e teto com textura de concreto bruto. O ambiente é decorado por um quadro, ao fundo da sala, e um sofá vermelho, posicionado no canto esquerdo e com os assentos virados de frente para o observador. Além disso, existem duas áreas demarcadas em verde – uma onde está o sofá, e a outra vazia, como mostra a Figura 17.

A tarefa proposta para treinamento consiste em reposicionar o sofá, girando-o em 180° graus e trocando-o de lado (mover da esquerda para a direita). Ao final da manipulação, o sofá deverá estar do lado direito da sala, virado de costas, e dentro da área demarcada. O treinamento é feito usando as técnicas de interação apresentadas na Seção 4.3.1. A Figura 17 também mostra os estados inicial e final do sofá dentro do AV.

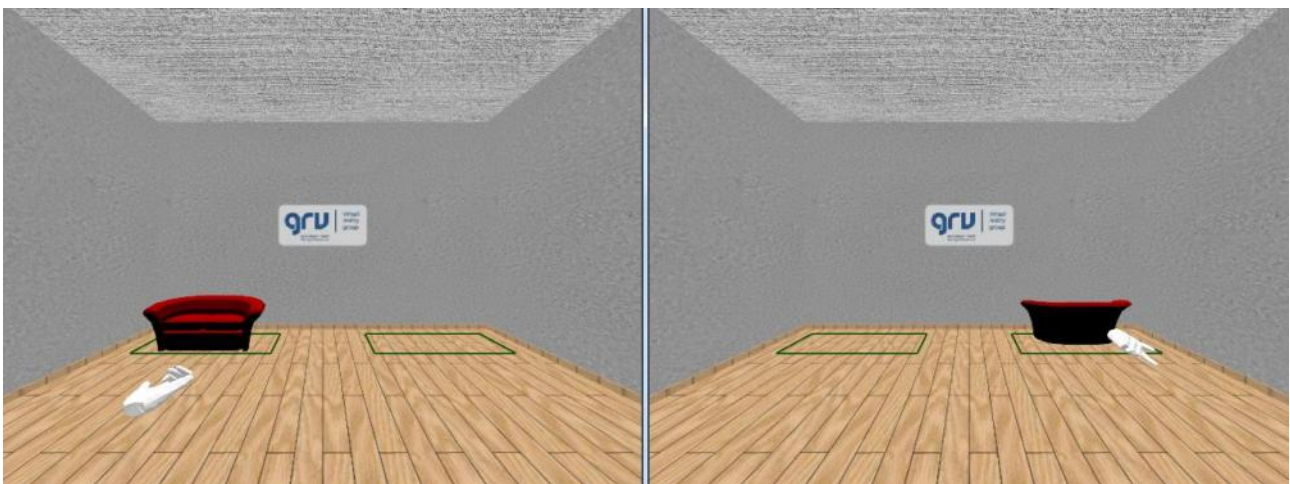


Figura 17. Posição inicial do sofá (à esquerda) e posição final desejada (à direita).

Já o AV de experimento tem como propósito a seleção e manipulação de diferentes objetos, com base na experiência adquirida no AV de treinamento. O cenário virtual desta aplicação também é uma sala, na qual quatro livros numerados, de tamanhos e cores diferentes, estão espalhados pelo chão do cenário. Existem também quatro áreas demarcadas em cores transparentes no chão, similares a cor dos livros, como mostra a Figura 18.

A tarefa do experimento consiste em pegar os livros, girá-los o quanto for necessário, e posicioná-los nas áreas demarcadas. Para realizar a tarefa corretamente, é preciso organizar os livros de forma que a capa numerada esteja visível ao usuário, e a lombada do livro esteja voltada para o lado esquerdo da sala. Ao final, os livros deverão estar posicionados no chão da sala, em ordem crescente (da esquerda para a direita).

Para fins de ilustração, a Figura 18 também mostra os estados inicial e final dos livros dentro do AV.

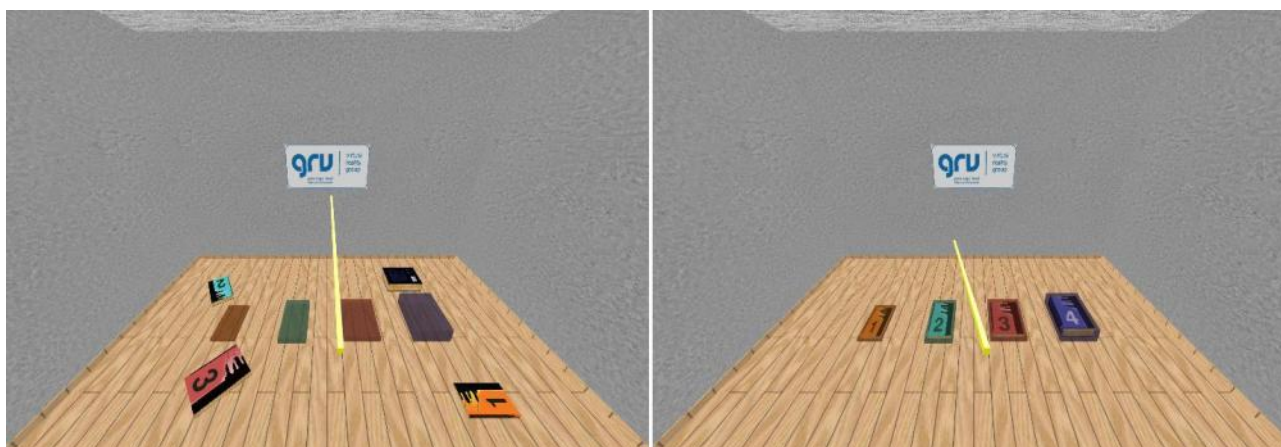


Figura 18. Posições inicial dos livros (à esquerda) e final desejada (à direita).

Outra característica presente nos dois AVs é a liberação automática do objeto em manipulação a cada colisão com o cenário, com o intuito de estressar e provocar reações fisiológicas no usuário.

4.3.3 Dispositivos de Realidade Virtual

Para interagir nos AVs utilizando as técnicas de interação descritas na Seção 4.3.1, a seguinte configuração de equipamentos de RV é utilizada:

- Um HMD i-Glasses, com rastreador de movimentos da cabeça, o qual permite ao usuário a visualização do AV;
- Um rastreador magnético de movimentos, modelo Polhemus FastTrak, com um sensor habilitado, o qual possibilita controlar um apontador para indicação e manipulação de objetos;
- Um botão de pressão (“alavanca”), acoplado ao sensor de rastreamento e conectado por fios a um *mouse*, o qual oferece ao usuário uma forma de confirmar a seleção e a liberação de objetos.

Além destes dispositivos, sensores fisiológicos descritos na Seção 4.4.1 também foram utilizados. A Figura 19 mostra os equipamentos dispostos na sala do experimento.



Figura 19. Equipamentos fisiológicos e de RV utilizados.

4.4 Princípios da Metodologia

As subseções a seguir apresentam os passos iniciais adotados para estabelecer um novo método de avaliação de interfaces 3D, considerando a comparação entre duas técnicas de interação de mesmo propósito. Para tanto, são apresentadas as medidas fisiológicas e de desempenho do usuário utilizadas no experimento, e uma sequência de passos que procura englobar, em um único protocolo, a aplicação dos métodos de avaliação já fundamentados na literatura, e o novo método, baseado na coleta de medidas fisiológicas.

4.4.1 Sensores e medidas fisiológicas

Para esta avaliação, um sensor de ECG e um sensor de SC são utilizados para coletar, respectivamente, as medidas de HR e SC. Esses dispositivos são ligados ao codificador de sinais Procomp Infiniti, e têm seus dados capturados e processados pelo *software* Biograph Infiniti, ambos fornecidos pela Thought Technology. Este *software*

também possui recursos que permitem a geração da análise estatística de uma sessão de experimento, bem como a exportação de dados em formato de relatório ou arquivo texto.

Com relação às medidas de HRV, estas são obtidas a partir do processamento dos dados de HR, e resultam em 39 diferentes medidas de domínio de tempo e frequência, como visto no Capítulo 2, Seção 2.2.

Ao sensor de ECG estão conectados três eletrodos, os quais são fixados nos punhos do usuário com pulseiras de borracha, de acordo com a posição apresentada pela Figura 20. Já o sensor de SC está conectado a dois eletrodos, os quais são fixados nas pontas nos dedos do usuário utilizando fitas de velcro, como também mostra a Figura 20.



Figura 20. Disposição correta dos sensores para monitoramento fisiológico.

4.4.2 Medidas de desempenho do usuário

Durante a experiência com as técnicas de interação, as seguintes medidas de usabilidade relacionadas ao desempenho do usuário são coletadas:

- Tempo: período total de interação;
- Precisão: distância entre o centro do objeto e o centro da posição final desejada;
- Colisões: total de colisões do objeto anexado com o cenário virtual;
- Anexações: número de vezes em que o usuário pegou o objeto;
- Liberações: número de vezes em que o usuário soltou o objeto.

Das três primeiras medidas, também são derivadas outras variáveis, quais sejam:

- Tempo em seleção: soma total dos períodos em que o usuário esteve tentando pegar um objeto;
- Tempo em manipulação: soma total dos períodos em que o usuário esteve manuseando um objeto;
- Sensibilidade: número de vezes em que ocorreram duas ou mais colisões num tempo inferior a 1s;

- Maior tempo em seleção: período máximo em que o usuário esteve tentando pegar um objeto;
- Maior tempo em manipulação: período máximo em que o usuário esteve manuseando um objeto.

4.4.3 Concepção do protocolo de teste

Para a execução dos testes foi elaborado um protocolo com base nas pesquisas apresentadas no Capítulo 3 [MEE02] [MEE05] [SLA06] [BRO06] [BRO07] [KIM08] [LIN05] [LIN06a] [LIN06b] [LIN08]. A sequência de passos proposta permite que dados de desempenho e avaliações subjetivas sejam tomados, em tempos distintos, oferecendo uma variabilidade de informações para avaliação da interface 3D. Da mesma forma, possibilita que medidas fisiológicas sejam coletadas de maneira contínua ao usuário durante o período de interação no AV.

O protocolo desenvolvido é composto de 9 etapas, as quais foram executadas dentro de um tempo estimado de 60 minutos, na seguinte ordem:

- Aplicação dos Questionários Pré-teste;
- Fixação dos Dispositivos;
- Coleta de Dados Basal Pré-teste;
- Realização do Treinamento – técnica “A”;
- Realização do Experimento – técnica “A”;
- Realização do Treinamento – técnica “B”;
- Realização do Experimento – técnica “B”;
- Retirada dos Dispositivos;
- Aplicação dos Questionários Pós-Teste.

Na etapa de “Aplicação dos Questionários Pré-teste”, o protocolo prevê uma breve apresentação do instrutor e do experimento ao usuário, a entrega de um termo de consentimento (Apêndice A) para leitura, e o preenchimento de um questionário (Apêndice B) com perguntas relacionadas à experiência em AVs. Uma escala progressiva de 1 a 7 é utilizada para avaliar cada questão solicitada. Por fim, é entregue também um documento com instruções que apresenta o AV, explicando como realizar tarefas utilizando duas diferentes técnicas de interação (Apêndice C). A duração desta etapa é de aproximadamente 10 minutos, e o usuário, sentado junto a uma mesa, realiza a leitura

dos documentos. Conforme Kim *et al* [KIM08], este tempo também é importante para a estabilização dos sinais fisiológicos antes da coleta de dados basal.

Na etapa de “Fixação de Dispositivos”, o usuário é convidado a sentar-se em outra cadeira para a fixação dos sensores fisiológicos e dos dispositivos de RV. Primeiramente, solicita-se o desligamento de aparelhos eletrônicos, a retirada de relógio ou pulseiras e a acomodação de maneira confortável. Em seguida, os sensores fisiológicos são fixados na mão esquerda (SC) e punhos (ECG) do sujeito, além de vestir o HMD. Antes de entregar o dispositivo a ser usado para apontamento, pergunta-se ao usuário se este se sente bem com a posição e o ajuste dos equipamentos. Com base nos trabalhos de Meehan *et al* [MEE02] [MEE05], o tempo estimado para esta etapa é de, aproximadamente, 3 minutos.

A etapa de “Coleta de Dados Basal Pré-Teste” é iniciada após o usuário estar bem acomodado com os dispositivos. Solicita-se para que encontre uma posição confortável na cadeira, deixando seus braços em posição de descanso sobre as pernas. A Figura 21 mostra o usuário já vestido com os equipamentos de teste. Com base nas pesquisas de Slater *et al* [SLA06] e Brogni *et al* [BRO07], neste momento os dispositivos de medição de dados fisiológicos são ligados, e uma coleta de 3 minutos é realizada. O objetivo é utilizar este conjunto de dados como base de comparação para as etapas em que o usuário estiver interagindo nos AVs.



Figura 21. Usuário com equipamentos fisiológicos e de RV.

Em seguida, o sujeito inicia a etapa de “Realização do Treinamento – técnica A” em contato com o AV de treinamento, para aprender a utilizar os equipamentos de RV e a primeira técnica de interação. Tarefas de seleção e manipulação podem ser executadas neste ambiente, dentro de um tempo limite de 5 minutos que também engloba o monitoramento fisiológico do sujeito. Ao final deste período, é concedido um tempo de 1 minuto e 30 segundos para descanso, antes de iniciar-se a próxima etapa.

Na etapa “Realização do Experimento – técnica A”, o usuário interage em um AV diferente, realizando um novo conjunto de tarefas dentro de um tempo limite de 7 minutos. Em seguida, é oferecido um período de 2 minutos para descanso.

O mesmo procedimento é aplicado para as etapas subsequentes, nas quais a segunda técnica de interação é avaliada. Cabe ressaltar que, para uma análise não tendenciosa, é importante que a ordem das técnicas avaliadas seja balanceada de acordo com o número de participantes do experimento.

Os tempos estipulados para a realização das etapas de treinamento e experimento são fundamentados pelas pesquisas de Lin *et al* [LIN05] [LIN06a] [LIN06b] [LIN08] e Brogni *et al* [BRO06] [BRO07], enquanto que a inserção de um intervalo de descanso entre as etapas é baseada no trabalho de Kim *et al* [KIM08].

Em seguida, na etapa de “Retirada dos Dispositivos”, os dispositivos de RV (HMD e rastreador) são retirados do usuário, e faz-se uma nova coleta de dados fisiológicos. O tempo estimado para esta etapa é de 3 minutos, o qual toma por base as pesquisas de Meehan *et al* [MEE02] [MEE05]. Este procedimento tem por objetivo capturar informações sobre o comportamento do usuário na readaptação ao ambiente natural após o período de imersão nos AVs. Assim como realizado em etapa anterior, pede-se ao sujeito que encontre uma posição confortável, mantendo os braços descansados.

Na etapa subjetiva de “Aplicação dos Questionários Pós-teste”, os sensores fisiológicos são retirados, e o sujeito é convidado a sentar-se novamente junto à mesa para responder um questionário (Apêndice D) que avalia aspectos de usabilidade, como a interface, as técnicas de interação e os procedimentos adotados durante toda a sessão de testes. Para este questionário, também é utilizada uma escala progressiva de avaliação, de 1 a 7.

Ao final, é aberta uma sessão para dúvidas, críticas e sugestões verbais por parte do usuário, concluindo com o agradecimento pela participação. O tempo destinado a esta etapa é de 10 minutos, com base nos trabalhos de Slater *et al* [SLA06] e Lin *et al* [LIN08].

4.5 Estudo Piloto

Para avaliar o protocolo, o AV de treinamento e as técnicas de interação propostas, considerando suas diferentes etapas de execução, um conjunto de testes preliminares foi elaborado. O experimento contou com a participação de 10 usuários, estudantes de graduação e pós-graduação, saudáveis, sendo 8 homens e 2 mulheres com idade entre 20 e 44 anos, e teve a duração de 2 dias.

Para estes testes, os sujeitos foram divididos em dois grupos, de forma a equilibrar a ordem de utilização das técnicas de interação. Um grupo iniciou os experimentos utilizando a técnica *ray-casting*, enquanto o outro iniciou com a técnica HOMER.

O estudo piloto também permitiu a definição de uma metodologia estatística para estudo da relação entre medidas fisiológicas, de desempenho e de avaliação subjetiva, sugerindo-se o uso de testes de hipóteses e de análise multivariada, considerando-se um nível de significância de 5%. Um glossário de termos técnicos foi elaborado (Apêndice H), o qual apresenta, de forma sucinta, os métodos e as técnicas estatísticas aplicados nesta proposta, com base no trabalho de Hair *et al* [HAI09] e de manuais de diferentes *softwares* estatísticos [EAS97][AYR07].

Como forma de certificar o protocolo e adequar tempos e procedimentos para testes posteriores, que reúnam um grupo maior de usuários, esta avaliação optou por testar as duas técnicas de interação apresentadas na Seção 4.3.1 apenas no AV de treinamento. Para tanto, foram suprimidas as instruções (Apêndice C) e as questões 7, 9 e 14 do questionário pós-teste (Apêndice D) relacionadas ao AV do experimento.

4.5.1 Caracterização do grupo

O questionário pré-teste (Apêndice B) tinha como objetivo fazer a caracterização do grupo que estava sendo avaliado, e apresentou a distribuição de respostas apresentada pela Figura 22. As questões abordadas referiam-se, respectivamente, aos conhecimentos sobre informática (Q. 01) e RV (Q. 02), ao uso de dispositivos não-convencionais (Q. 03) e de medição fisiológica (Q. 04), e as sensações de incômodo (Q. 05) e mal-estar (Q. 06) na adaptação a uma nova interface.

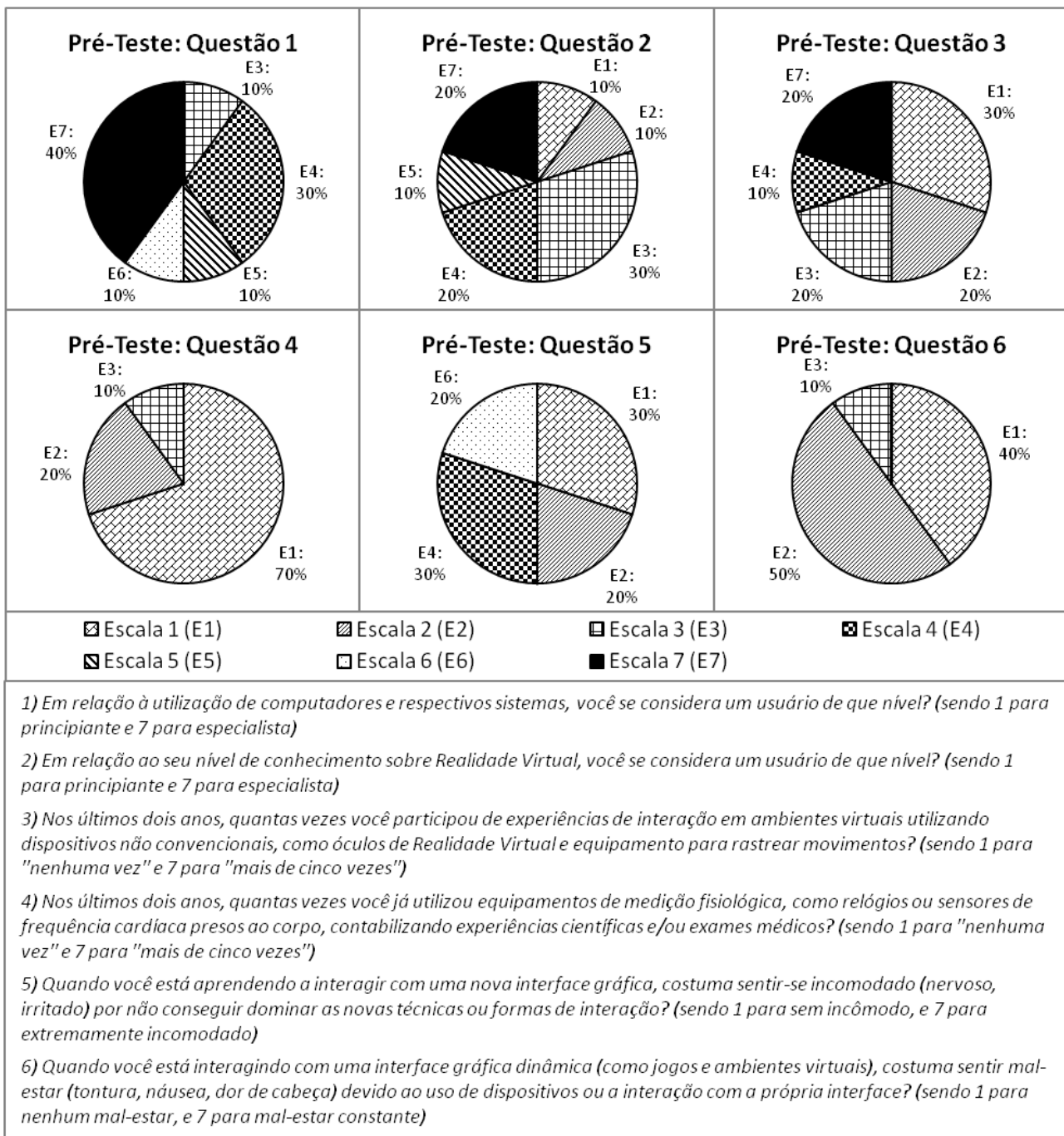


Figura 22. Distribuição das respostas do questionário pré-teste.

4.5.2 Desempenho entre técnicas

Uma comparação entre grupos considerando apenas a técnica *ray-casting* mostra que o desempenho melhora após a primeira experiência. As médias das variáveis baseadas no tempo apresentam valores menores para o grupo que experimentou a técnica *ray-casting* na segunda experiência com o AV de treinamento, como mostra a Figura 23.

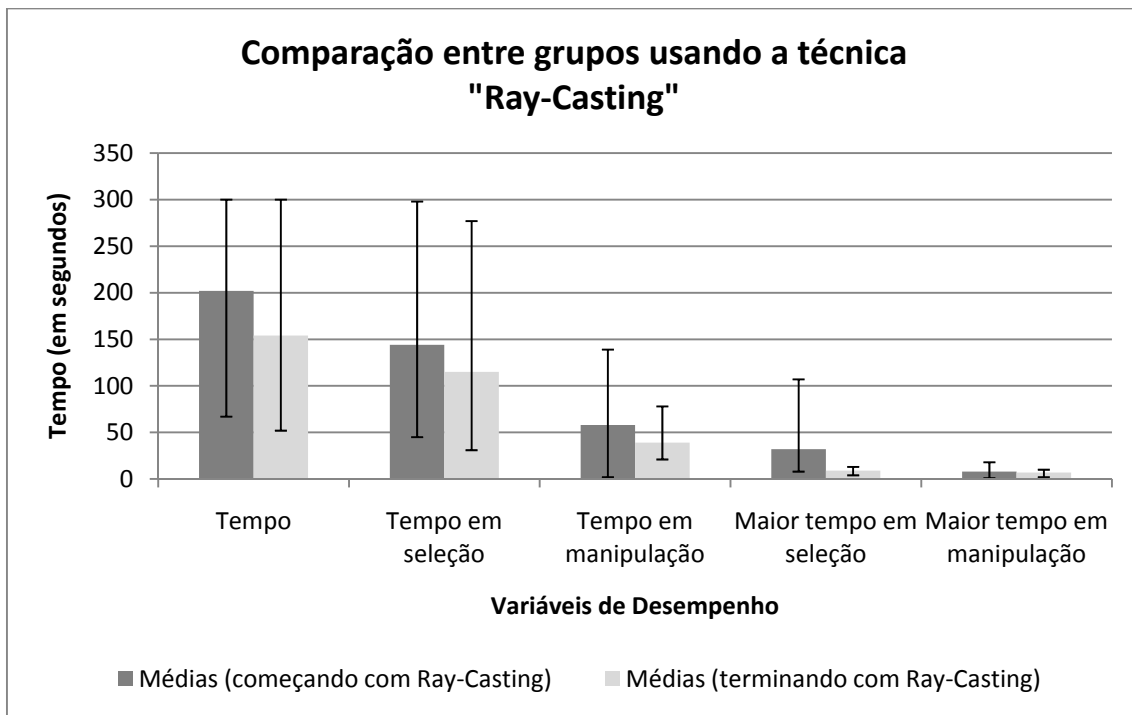


Figura 23. Desempenho entre grupos com *ray-casting*, considerando as variáveis baseadas no tempo.

A Figura 24 apresenta uma comparação entre as médias das variáveis baseadas no número de tentativas, a qual aponta para valores maiores na segunda experiência com o AV de treinamento, com exceção da variável “liberações”. Este nível maior de dificuldade pode ser justificado pela diferença na forma de interação das técnicas, uma vez que este grupo testou, primeiramente, com uma técnica que permitia uma interação mais natural que a *ray-casting*.

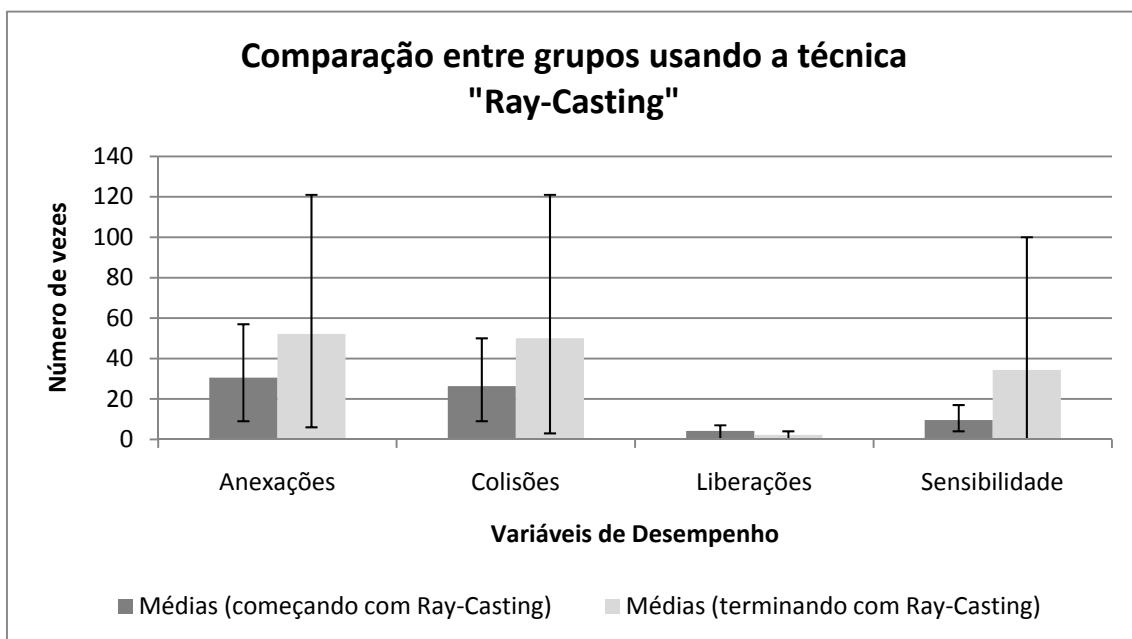


Figura 24. Desempenho entre grupos com *ray-casting*, considerando as variáveis baseadas no número de tentativas.

Já a Figura 25 mostra que as médias da variável “precisão” não apresentam diferença significativa entre a ordem de utilização da técnica.

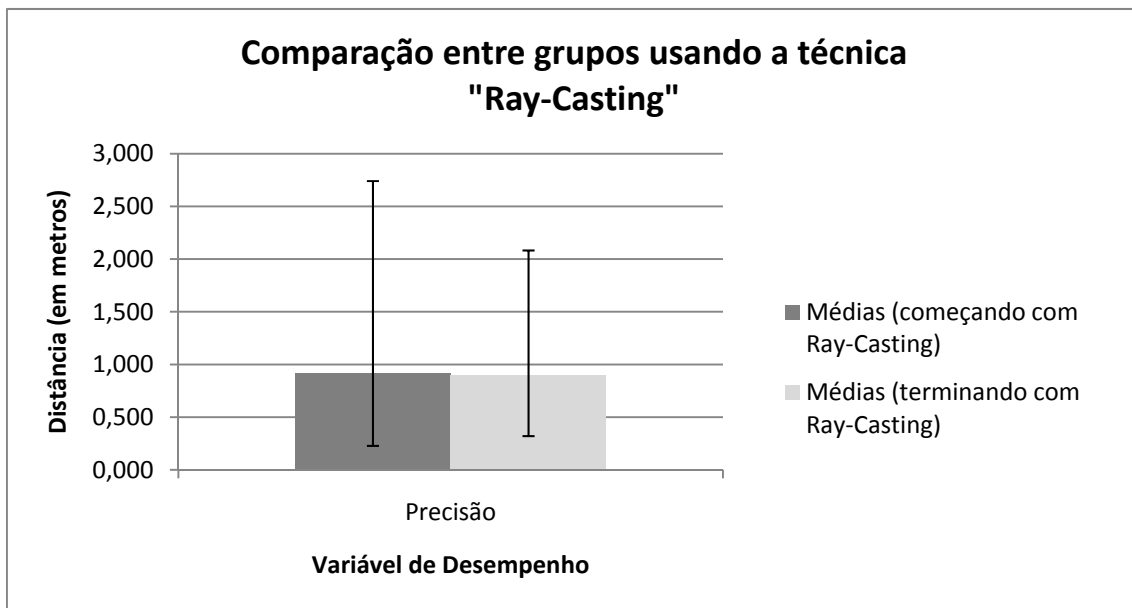


Figura 25. Desempenho entre grupos com *ray-casting*, variável “precisão”.

Considerando as observações explanadas, pode-se concluir que o usuário já estava acostumado ao AV de treinamento no momento da segunda interação, e que a precisão em realizar a tarefa continua a mesma, independente da ordem das técnicas. Os mesmos resultados foram apontados por uma análise de variância (ANOVA).

Este procedimento de comparação entre grupos também foi realizado considerando apenas a técnica HOMER. Nesta análise, o desempenho também melhora após a primeira experiência, como mostram as variáveis “Tempo”, “Tempo em Seleção” e “Tempo em Manipulação” da Figura 26. Cabe destacar também que o tempo da segunda experiência é, praticamente, a metade do tempo gasto na primeira interação com o AV de treinamento.

Já a Figura 27 aponta para uma média menor no número de tentativas durante a segunda experiência com o AV. Neste caso, cabe a mesma justificativa relacionada à Figura 24 – porém, interpretada de forma inversa: a facilidade encontrada na segunda experiência ocorreu devido ao fato de que grupo testou, primeiramente, com uma técnica que permitia uma interação menos natural que a HOMER.

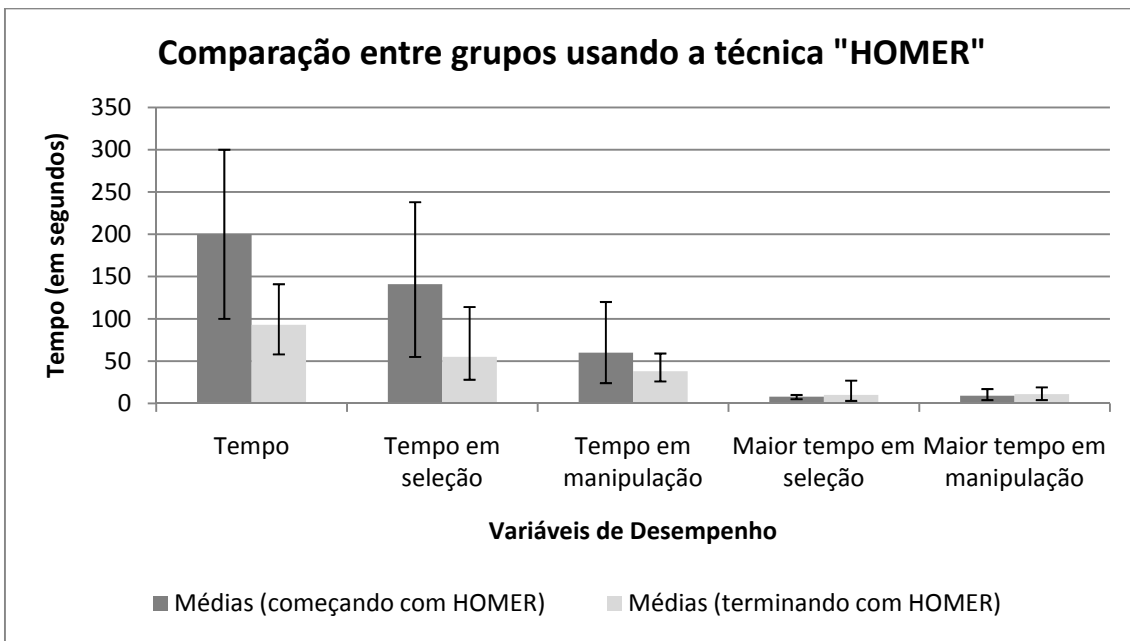


Figura 26. Desempenho entre grupos com HOMER, considerando as variáveis baseadas no tempo.

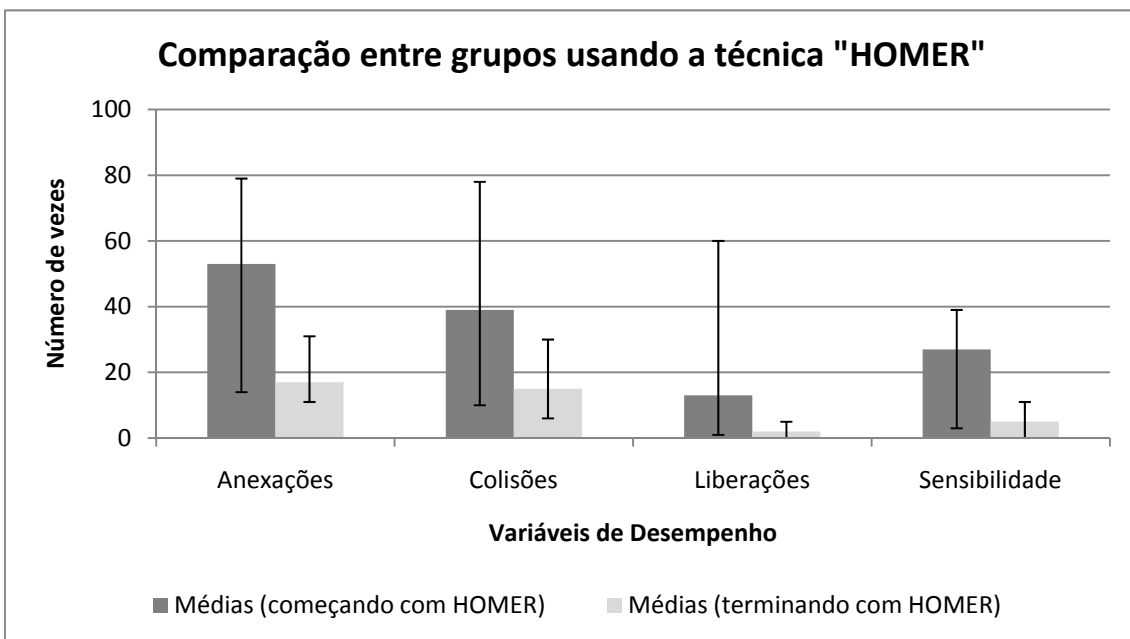


Figura 27. Desempenho entre grupos com HOMER, considerando as variáveis baseadas no número de tentativas.

A mesma analogia, neste caso, pode ser feita para a variável “precisão”, que aponta para uma pequena diferença ao grupo que terminou a experiência utilizando a técnica HOMER, como mostra a Figura 28. Uma ANOVA foi realizada para esta comparação, e também não apontou diferenças significativas entre os grupos.

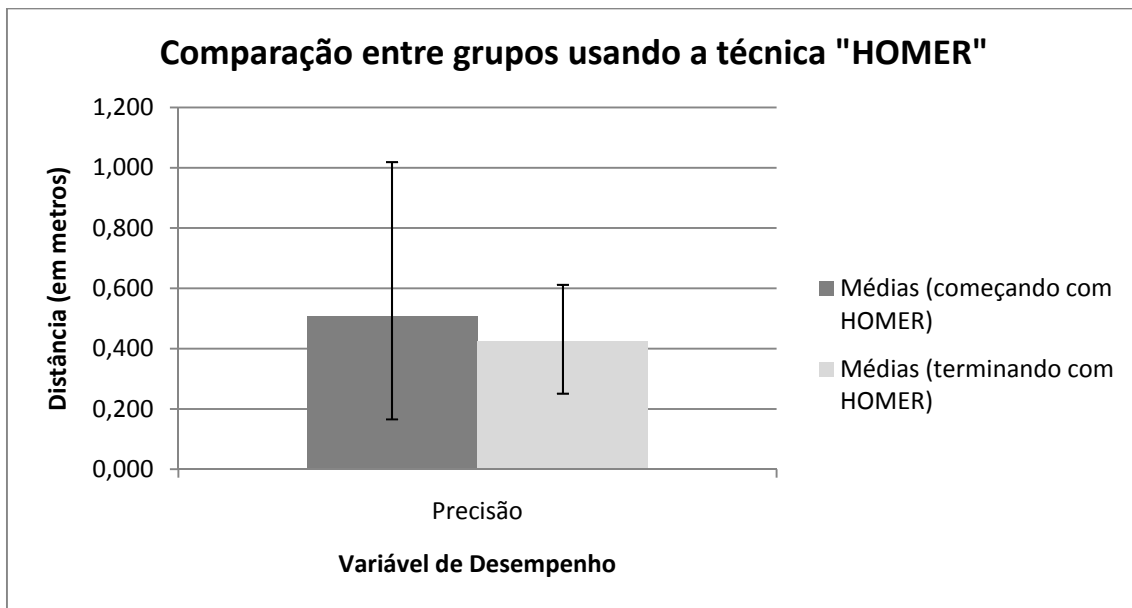


Figura 28. Desempenho entre grupos com HOMER, considerando a variável “precisão”.

Uma simples comparação entre as duas técnicas torna evidente a vantagem de usar a técnica HOMER para completar a tarefa deste AV. A Figura 29 mostra que o tempo total gasto na tarefa foi menor, enquanto a Figura 30 mostra que uma quantidade menor de tentativas foi necessária para manipular o objeto, com uma redução no número de colisões. Já a Figura 31 destaca que, com HOMER, a posição final do objeto ficou, em média, mais próxima do alvo desejado.

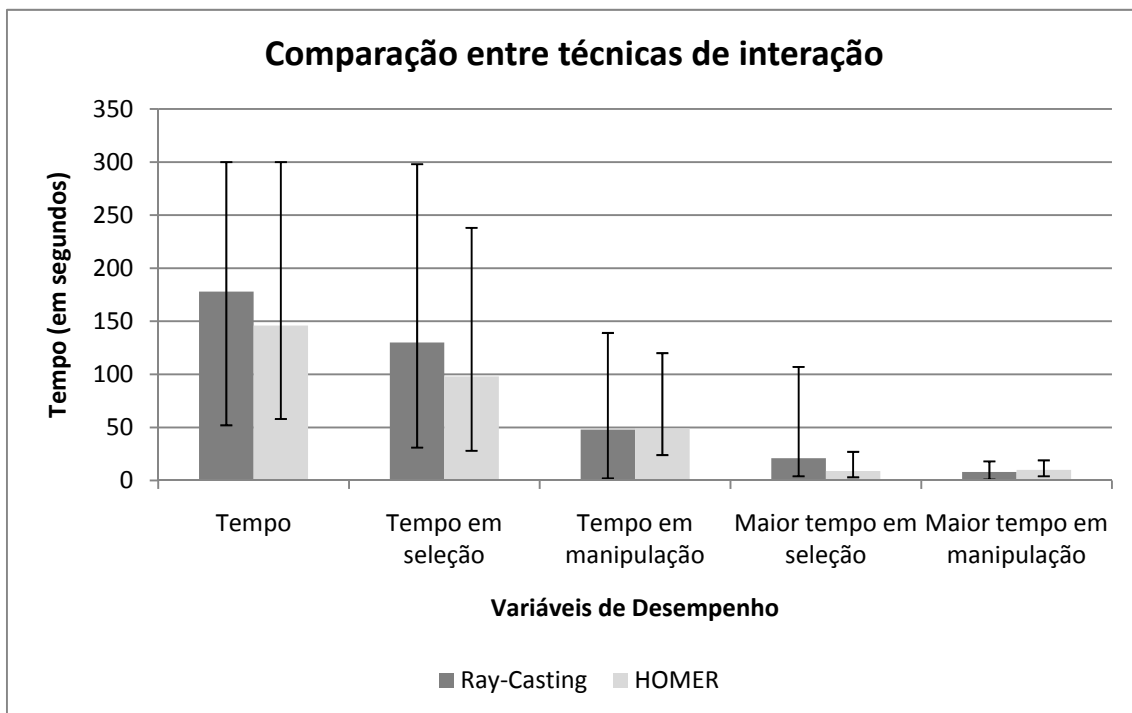


Figura 29. Desempenho entre as técnicas, considerando as variáveis baseadas no tempo.

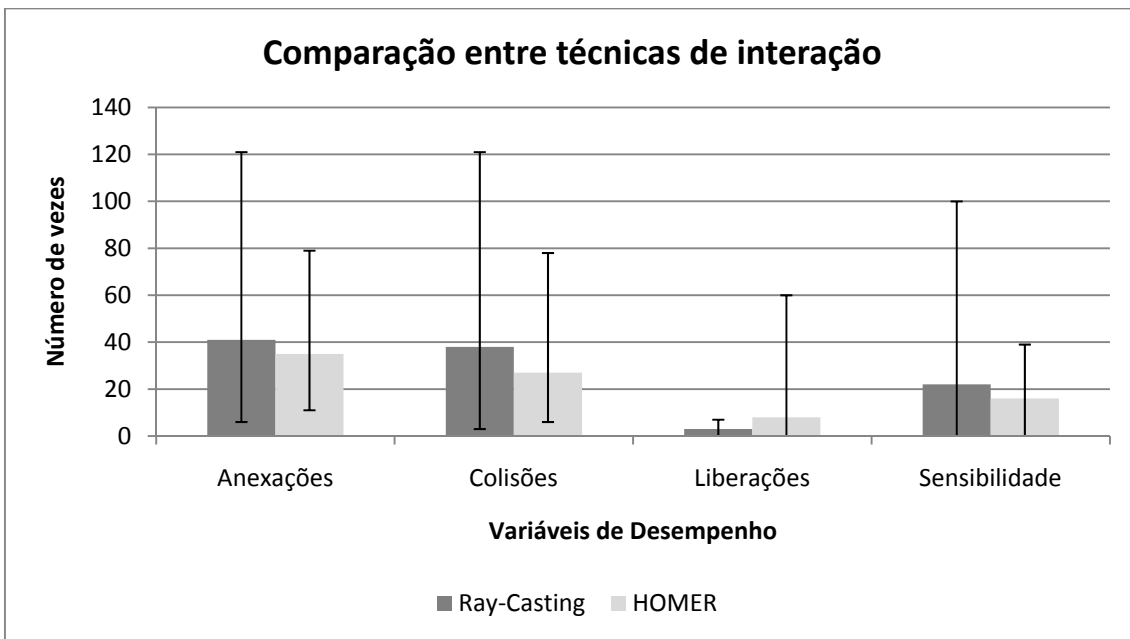


Figura 30. Desempenho entre as técnicas, considerando as variáveis baseadas no número de tentativas.

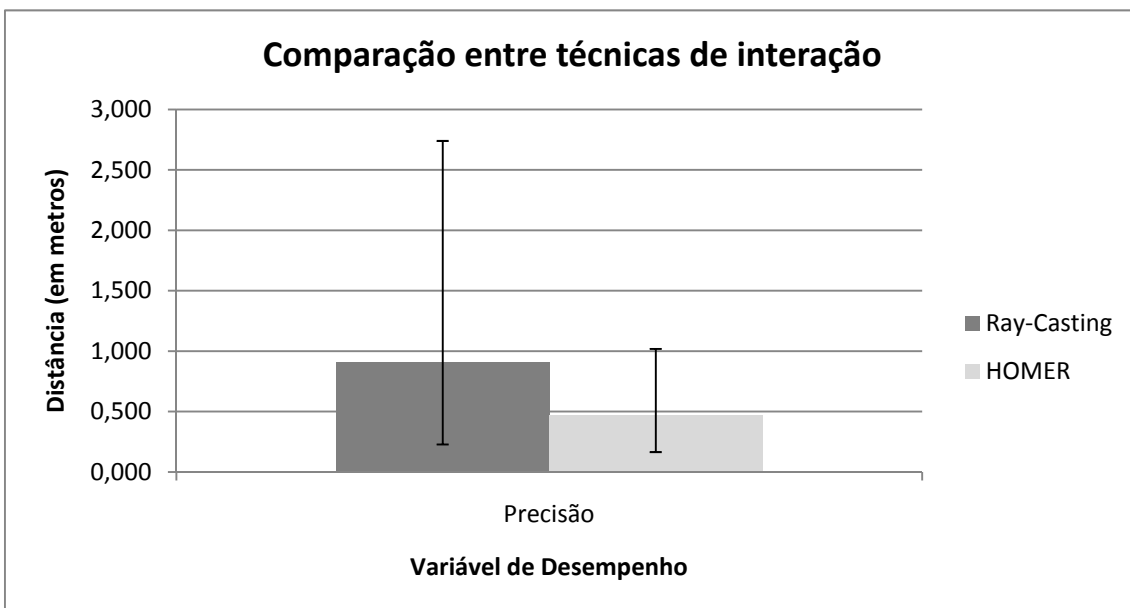


Figura 31. Desempenho entre as técnicas, considerando a variável “precisão”.

Por fim, observando os gráficos que apresentam o desempenho com base no número de tentativas, nota-se que os usuários quase não utilizaram o recurso de liberar o objeto, acionado pela alavanca presa ao dispositivo de rastreamento. Este fato justifica-se pelo elevado número de colisões, uma vez que o objeto é liberado automaticamente a cada colisão no AV, como destacado na Seção 4.3.2.

4.5.3 Avaliação subjetiva

O questionário pós-teste (Apêndice D) tinha como objetivo avaliar as técnicas de interação, a tarefa, os dispositivos, a interface visual e o protocolo utilizado.

As perguntas 1, 2 e 3 faziam referência à sensação de conforto no uso de diferentes dispositivos, as quais abordaram, respectivamente, a presença de fios presos ao corpo (Q. 01), o tempo de adaptação aos dispositivos não-convencionais (Q. 02) e a comodidade em vestir estes equipamentos. As respostas apontaram uma sensação moderada para alta de conforto à maioria dos usuários, conforme mostra a Figura 32.

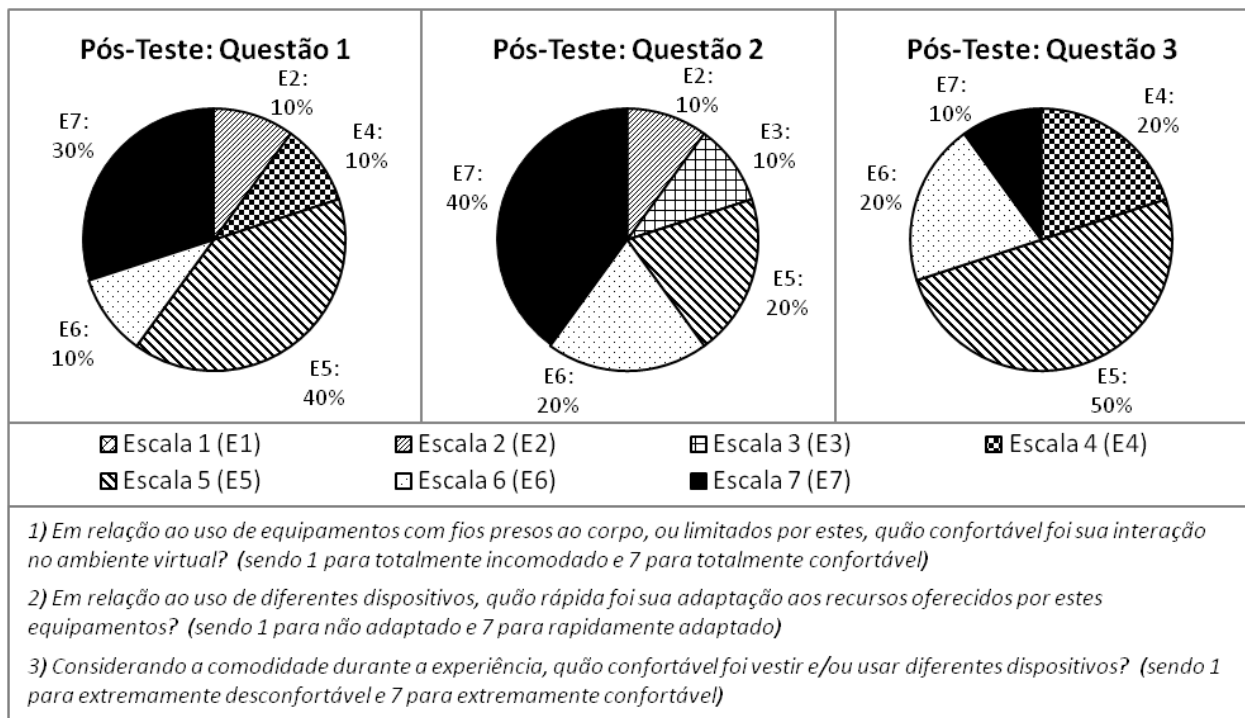


Figura 32. Distribuição das respostas nas questões que avaliam o uso dos dispositivos.

Resultados satisfatórios foram coletados nas questões 4 e 5, responsáveis pela avaliação de sensações de desconforto (Q. 04) e do tempo para descanso entre as tarefas (Q. 05). As respostas indicaram que sensações de mal-estar estiveram praticamente ausentes durante a experiência, e que o tempo previsto para descanso entre os testes das técnicas foi suficiente, como mostra a Figura 33.

As questões 6 e 11 procuraram avaliar a interface visual do AV. A primeira avaliava o *feedback* visual (Q. 06), na qual metade dos usuários caracterizaram como totalmente satisfatório. Já as respostas da segunda questão (Q. 11), que avaliava a convicção de ter posicionado corretamente o objeto no local desejado, indicaram que o cenário apresentou uma sensação de profundidade aceitável, apesar da ausência de recursos de estereoscopia. A Figura 34 apresenta os resultados referentes a estas questões.

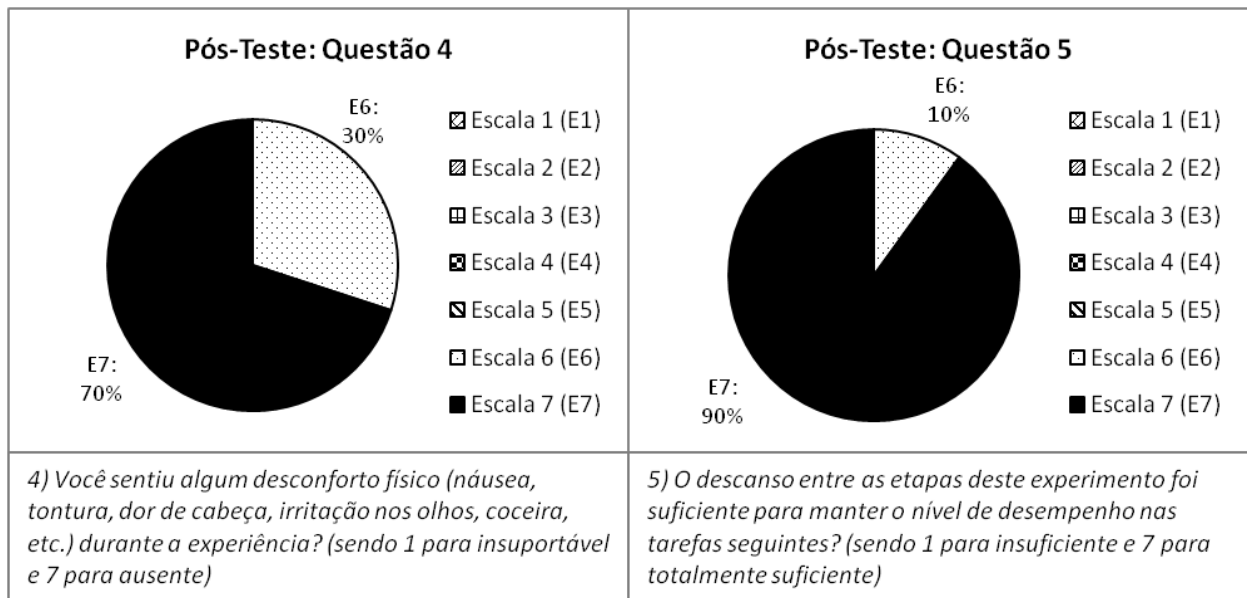


Figura 33. Distribuição das respostas nas questões que avaliam desconforto e descanso.

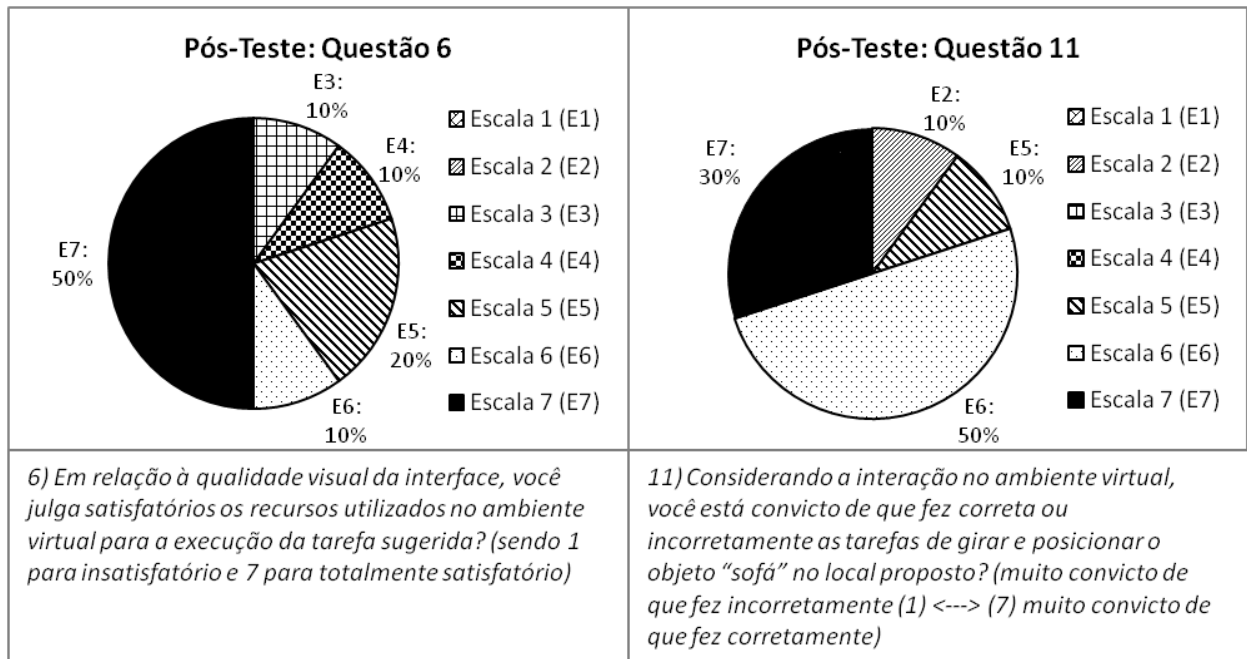


Figura 34. Distribuição das respostas nas questões que avaliam a interface visual do AV.

Características da tarefa foram avaliadas pelas questões 7, 12, 13 e 14. A primeira, relacionada ao nível de dificuldade da tarefa (Q. 07), apresentou uma variabilidade de respostas, as quais tiveram relação com o nível de conhecimento de RV do participante. Já a questão 12 (Q. 12) apontou como fator de influência no desempenho o fato de que, a cada colisão no cenário virtual, era necessário pegar novamente o objeto desejado. Por sua vez, a questão 13 (Q. 13) estabeleceu entre baixa e moderada impaciência o grau de irritabilidade causado pelas constantes colisões. Por fim, a questão 14 (Q. 14), avaliou

como pouco relevante a pressão exercida pelo tempo limite existente para realizar a tarefa no AV. Os resultados destas questões podem ser visualizados na Figura 35.

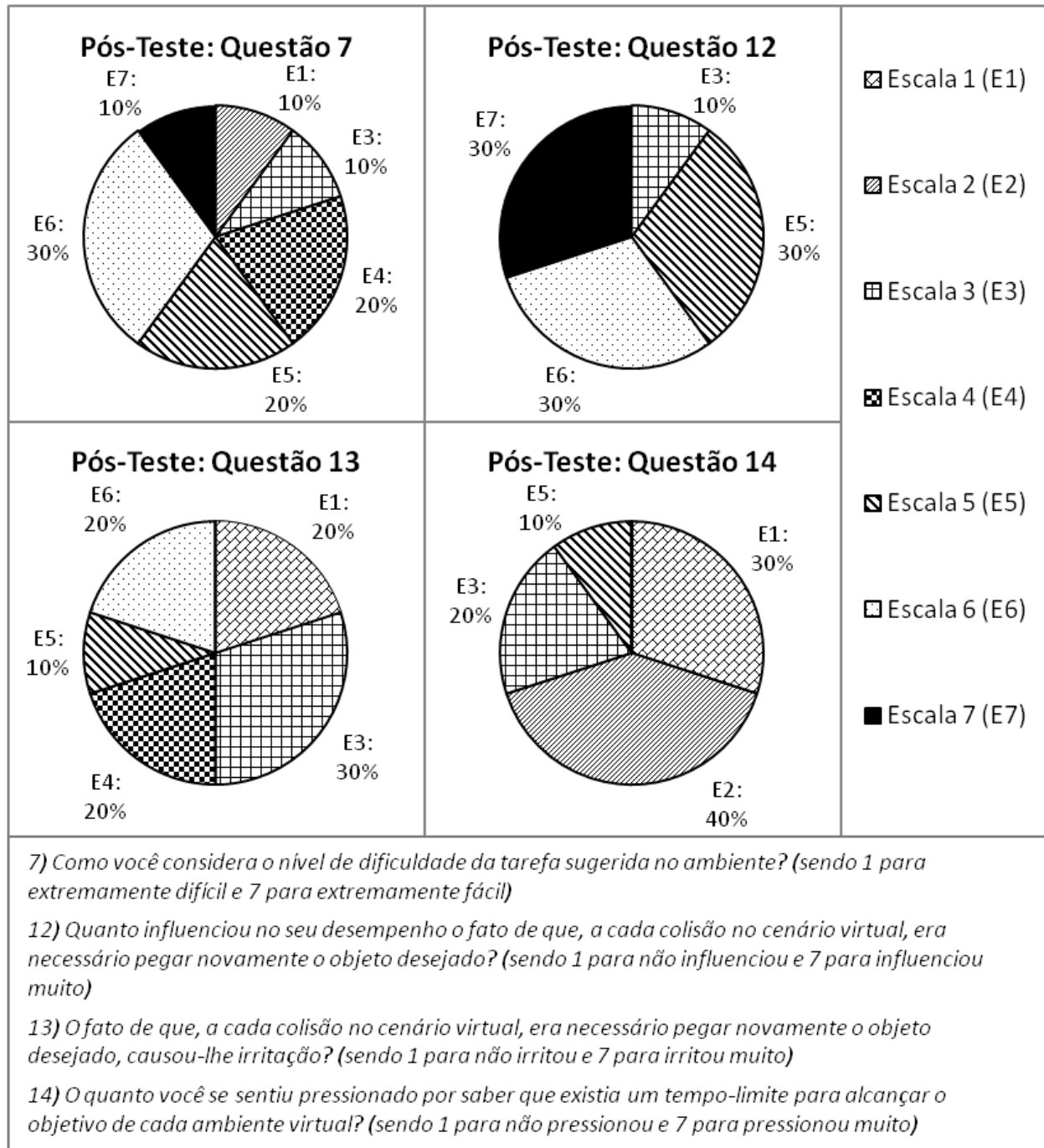


Figura 35. Distribuição das respostas nas questões que avaliam características da tarefa.

Uma avaliação entre as técnicas de interação foi contemplada pelas questões 9 e 10, as quais buscavam avaliar o grau de satisfação dos recursos oferecidos pelas técnicas. A Figura 36 apresenta uma comparação com base nas respostas dos sujeitos, as quais indicam uma leve vantagem no uso da técnica HOMER (Q. 10) sobre a *ray-casting* (Q. 09), corroborando com a avaliação de desempenho da Seção 4.5.2.

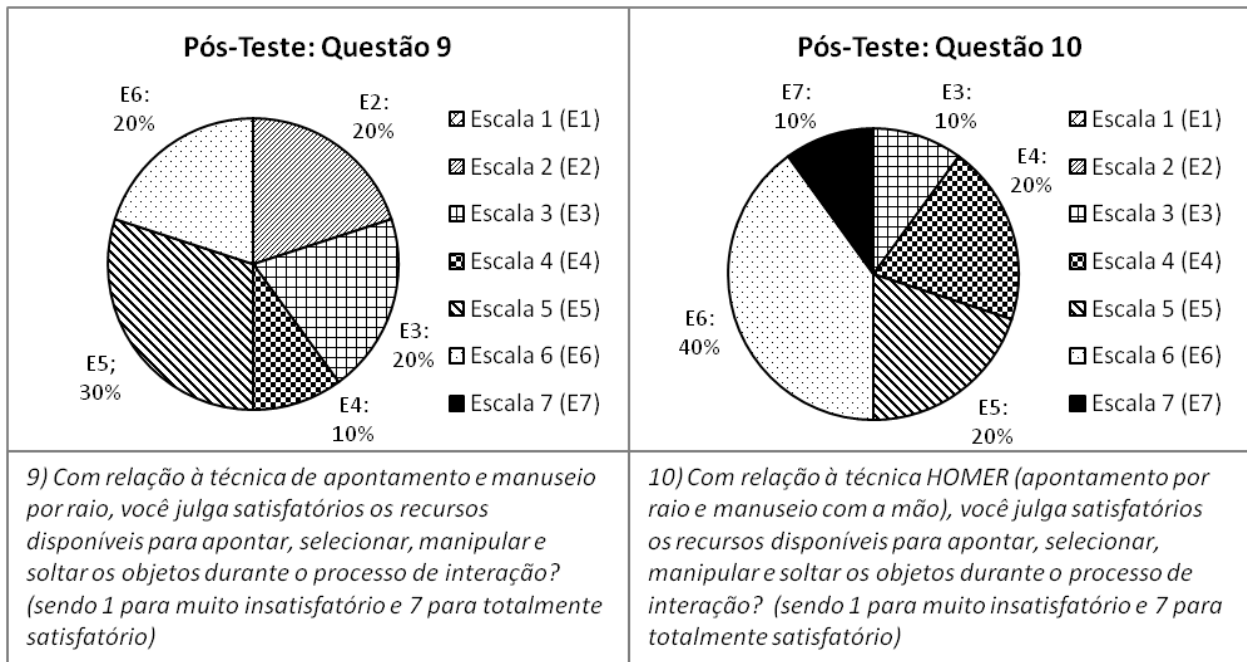


Figura 36. Distribuição das respostas nas questões que avaliam a interface visual do AV.

Por fim, as questões 8 e 15 avaliaram pontos importantes do protocolo utilizado. A questão 8 (Q. 08) analisou as instruções do experimento (Apêndice C), consideradas como esclarecedoras, enquanto que a questão 15 (Q. 15) avaliou a sequência de passos do protocolo, apontada como totalmente satisfatória pelos participantes. Os resultados podem ser visualizados na Figura 37.

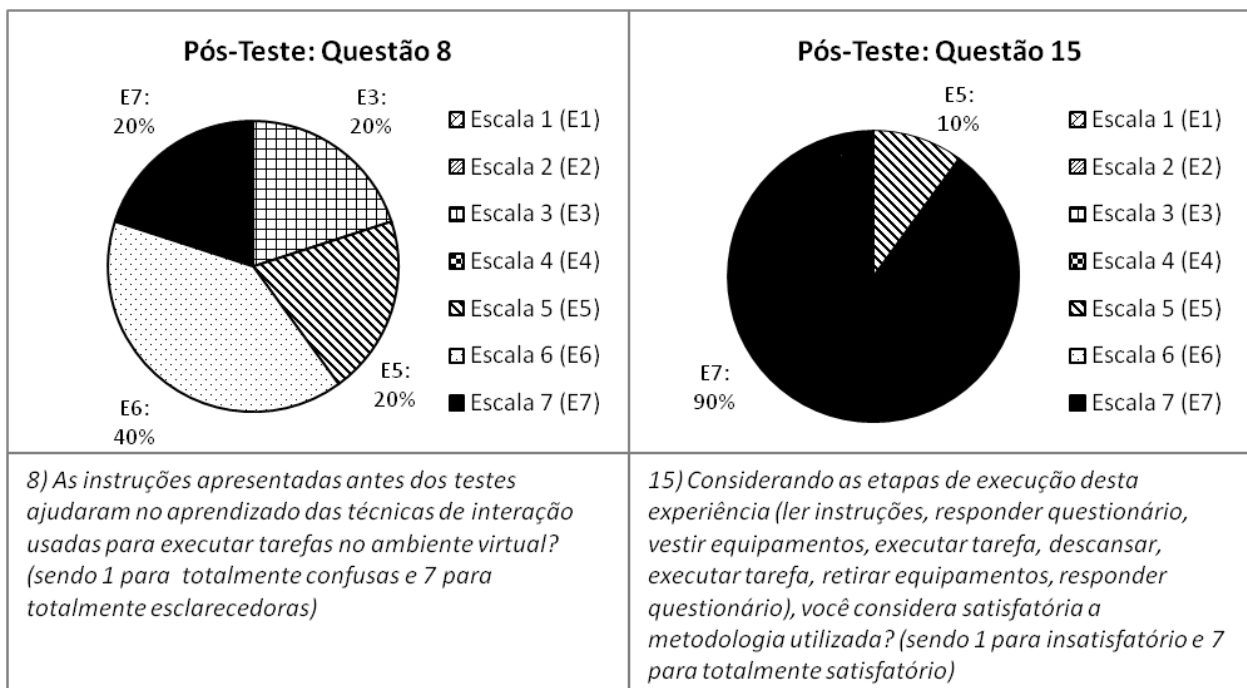


Figura 37. Distribuição das respostas nas questões que avaliam o protocolo proposto.

Nesta etapa também foi solicitada a colaboração dos participantes por meio de comentários descritivos no questionário pós-teste. A Tabela 4 e a Tabela 5 apresentam uma compilação das contribuições relatadas pelos sujeitos durante o estudo piloto.

Tabela 4. Comentários dos usuários durante o questionário pós-teste.

<i>Natureza da Dificuldade</i>	<i>Momentos de irritação, desconforto ou dificuldade</i>
Em relação aos dispositivos	Cabo do HMD ficou preso algumas vezes, restringindo um pouco o movimento da cabeça.
	A resolução do HMD é baixa (640x480).
	Os cabos dificultaram alguns movimentos. O dispositivo de manuseio era um pouco largo, e causou um pouco de desconforto.
Em relação ao ambiente virtual	Nos momentos de colisão nos cantos do cenário, foi difícil determinar a superfície onde o objeto colidiu.
	A ausência de estereoscopia causou um leve incômodo, tornando complicado “imaginar” a posição do sofá, e acostumar-se com o fato.
	Faltam sombras nos objetos para se ter noção de distâncias e evitar colisões com o cenário.
Em relação às técnicas de interação	Girar o objeto ao redor do eixo vertical é desconfortável, pois é necessário girar o punho além do limite.
	A técnica por raio é muito mais frustrante pela dificuldade em girar o objeto.
Em relação à satisfação do usuário	A primeira tarefa foi mais desconfortável, pois não havia entendido muito bem como mover o objeto, e demorei a perceber que eram as colisões que me obrigavam a pegar novamente o objeto.
	Realmente, é frustrante ter que pegar o objeto toda vez que ocorre uma colisão. Atrapalhou na primeira tarefa, porém a partir da segunda já estava acostumado e cuidava para não colidir.
	Quando me lembrei do limite de tempo, fiquei ansioso para completar a tarefa.
	Houve algum desconforto devido à entrada de luminosidade externa no meu campo de visão.

Tabela 5. Sugestões dos usuários durante o questionário pós-teste.

<i>Natureza da Dificuldade</i>	<i>Dicas, críticas ou elogios</i>
Em relação aos dispositivos	Uma melhor forma de pressionar o botão, evitando ter que procurá-lo durante a interação.
	Melhorar a posição dos fios para não “prender” o usuário.
Em relação ao ambiente virtual	Uma sinalização visual ou sonora ajudaria na identificação das colisões.
	Incluir sombra nos objetos, para melhorar a sensação de profundidade, ou usar estereoscopia.
Em relação à satisfação do usuário	A tentativa de irritar o usuário foi válida, e teve relação com a técnica que eu usava no momento.
	Dependendo do usuário, o número excessivo de colisões poderia levar a desistência da experiência.
Em relação ao protocolo e a metodologia proposta	Gravar a sessão em vídeo.
	Em pesquisas futuras, comparar o uso de dispositivos usando medidas fisiológicas, como, por exemplo, exibir um vídeo no HMD e no monitor.
	Limpar sensores fisiológicos e HMD perante o usuário, antes de cada teste.
	Utilizar uma imagem de fundo no HMD entre um teste e outro, ou desligá-lo. As imagens “fora de sincronismo” cansam os olhos.
	Durante a sessão basal, definir se os olhos devem ficar abertos ou fechados. Uma música de fundo poderia também relaxar mais o participante.
	Testar o mesmo usuário em horários diferentes.
	Escurecer mais a sala de testes.

4.5.4 Relação entre medidas fisiológicas e medidas de desempenho

Como um dos objetivos do trabalho é identificar parâmetros fisiológicos capazes de avaliar a qualidade de técnicas de interação, decidiu-se realizar uma análise para verificar a correlação das 10 variáveis de desempenho descritas na Seção 4.4.2, com as 41 medidas fisiológicas geradas pelo *software* Biograph Infiniti a partir dos dados do equipamento Procomp Infiniti.

Para realizar estas comparações utilizou-se a técnica de análise multivariada de dados. De acordo com Hair *et al* [HAI09], análise multivariada de dados refere-se a todos os métodos estatísticos que simultaneamente analisam múltiplas medidas sobre cada indivíduo ou objeto sob investigação.

Para definir os métodos estatísticos passíveis de serem aplicados à análise de correlação entre medidas fisiológicas e medidas de desempenho, foi necessário ampliar artificialmente a amostra obtida nos testes preliminares, por intermédio da geração de dados aleatórios. A geração deste tipo de dados é uma técnica comumente usada em estatística para, a partir de uma amostra simples, validar a forma de análise dos dados. De acordo com Morettin e Bussab [MOR05], uma amostra aleatória simples pode definir uma referência empírica para medir a eficiência de um teste estatístico em relação a outros tipos de amostras.

É importante deixar claro que os dados aleatórios não servem para validar a tese de que existe relação entre medidas fisiológicas e medidas de desempenho, mas sim para validar se a análise multivariada, neste caso, pode ser usada para analisar a correlação existente entre estas variáveis, com outros tipos de amostras.

Com base nos dados existentes da amostra de 10 sujeitos do teste preliminar, foram produzidos dados aleatórios para uma população de 100 sujeitos. Estes dados fictícios foram gerados utilizando o *software* Microsoft Excel, e ficam compreendidos dentro do intervalo entre os valores mínimo e máximo de cada variável.

A partir deste referencial, diferentes testes estatísticos, também descritos no Apêndice H, foram aplicados utilizando o *software* SPSS²⁵, um programa para análise estatística, mineração de dados e análise preditiva. Em seguida, chegou-se a sugestão da seguinte sequência de passos analíticos destinadas para esta proposta:

- Realizar uma **análise exploratória** dos dados para verificar a existência de valores *outlier*²⁶, a normalidade da distribuição e a multicolinearidade das variáveis. Esta análise permite identificar a consistência dos dados, a forma de distribuição dos dados, e evitar a redundância de variáveis. Testes como o Gráfico de Box-Plot [TUK77], o Kolmogorov-Smirnov [KOL33] e o Fator de Inflação da Variância [HAI09] (ou VIF, do inglês *Variance Inflation Factor*) podem ser utilizados;

²⁵ <http://www.spss.com/>

²⁶ *Outlier* é uma observação numérica distante do resto dos dados. A identificação de *outliers* permite preservar a forma de uma distribuição [BAR94].

- Aplicar um processo de **regressão** para descobrir as variáveis que estão associadas ao desempenho do usuário [HOS00]. Um procedimento de seleção de variáveis pode ser utilizado, primeiramente, para definir o conjunto de medidas de interesse. Em seguida, pode-se aplicar uma **regressão múltipla**, para identificar os fatores independentes (medidas fisiológicas) associados a uma medida de desempenho e, conseqüentemente, indicar quais medidas fisiológicas podem ser utilizadas para avaliar a qualidade da interação, na comparação entre técnicas de mesmo propósito.

Para reduzir a complexidade da análise, decidiu-se ainda pela diminuição do número de variáveis, tanto de desempenho, como de medidas fisiológicas. No primeiro caso, podem ser excluídas as medidas derivadas apresentadas na Seção 4.4.2, uma vez que estas tendem a apresentar relação uma com a outra. Já para variáveis fisiológicas, o conjunto pode ser reduzido apenas para as medidas de domínio de tempo e frequência para curtos períodos, uma vez que as interações duram, no máximo, 5 minutos.

4.5.5 Relação entre medidas fisiológicas e avaliações subjetivas

Seguindo a mesma abordagem da Seção 4.5.4, também é preciso efetuar uma análise multivariada de dados para a relação entre medidas subjetivas e medidas fisiológicas. Neste caso, é necessário comparar 15 questões do questionário pós-teste com as 41 medidas fisiológicas disponibilizadas pelo *software* Biograph Infiniti.

Para identificar os métodos estatísticos para esta análise, também foram gerados dados aleatórios para uma população de 100 sujeitos, a partir da base inicial. Após diferentes testes, chegou-se a esta sequência de passos analíticos:

- Realizar uma **análise exploratória** dos dados, conforme já descrita na Seção 4.5.4, a fim de se obter uma sumarização das respostas coletadas;
- Aplicar um processo de **regressão** para descobrir as variáveis que estão associadas à avaliação subjetiva do usuário. Como apresentado na Seção 4.5.4, faz-se uma seleção de variáveis preditoras a partir do conjunto de questões. E, após, uma **regressão múltipla** para identificar os fatores independentes (medidas fisiológicas) associados a uma medida subjetiva, determinando quais medidas fisiológicas podem ser utilizadas para avaliar a qualidade da interação, na comparação entre técnicas de mesmo propósito.

Para o comparativo com dados subjetivos, a redução do número de perguntas do questionário também é recomendável. De acordo com Hair *et al* [HAI09], para validar um questionário é preciso de uma amostra de, pelo menos, 10 sujeitos por questionamento para a análise estatística ser confiável – o que pode ser um agravante, devido a dificuldade de mobilizar um grupo de 150 pessoas para os testes desta proposta.

4.5.6 Medidas fisiológicas na comparação entre técnicas de interação

Considerando apenas os dados da amostra original e as medidas de HR e SC, foram realizadas comparações entre as técnicas de interação utilizando testes *t* de Student para duas amostras pareadas [SNE89]. Para tanto, as médias dos sujeitos foram geradas a partir de dados normalizados de seus sinais fisiológicos, coletados durante a experiência com cada técnica.

Para a comparação usando medidas de HR não foi encontrada diferença significativa entre as técnicas. Uma justificativa para este resultado é que, durante a análise, foi detectada a presença de ruídos no sinal de HR de todos os sujeitos. Isto provavelmente ocorreu devido à ausência de gel condutor nos sensores de ECG durante o experimento, o qual tem por objetivo reduzir a resistência elétrica da pele, e oferecer uma melhor condutância.

Com base nesta constatação, as medidas fisiológicas derivadas deste sinal, e utilizadas nas análises anteriores, também foram prejudicadas, e mereceria novo estudo.

Já para a comparação usando medidas de SC, foi detectada diferença significativa entre as técnicas, permitindo concluir que a técnica de *ray-casting* causa mais estresse ao usuário, uma vez que exige mais esforço do que a HOMER ($p = 0,02$, $t = 2,42$). Em média, os valores de SC da técnica *ray-casting* foram maiores que os da técnica HOMER.

Também foi realizada uma comparação entre os sinais de SC coletados nos períodos pré- e pós-basal. Na primeira situação, foram comparadas as médias de SC entre o período basal pré-teste e a primeira experiência no AV, a qual apontou para uma forte tendência do usuário estar mais tenso durante a interação ($p < 0,01$, $t = -5,36$). Em média, os valores de SC em período basal (descanso) foram menores do que durante o experimento.

Já na segunda situação, as médias de SC entre a segunda experiência e o período basal pós-teste foram comparadas. Neste caso, a análise não apontou para diferenças significantes, o que determina a necessidade de um tempo maior para descanso após a interação com o AV. Da mesma forma, também é possível concluir que esta coleta é

desnecessária, tornando a retirada dos sensores um fator que pode ajudar no processo de relaxamento do usuário.

4.5.7 Medidas fisiológicas e alterações durante o processo interativo

Para medir a relação entre medidas fisiológicas e as etapas do processo interativo, foram comparados os estados de seleção e manipulação de cada técnica, também utilizando testes t de Student para duas amostras pareadas. Para esta análise, apenas os sinais de SC foram considerados, devido aos problemas encontrados com o sinal de HR.

Visualmente, observando o comportamento do sinal de SC durante o processo interativo, não é possível apontar uma relação entre o sinal e as etapas de seleção e manipulação que estão sendo desempenhadas. A Figura 38 e a Figura 39 apresentam exemplos dos gráficos de desempenho gerados para o “Usuário 5”, mostrando o traçado do sinal e as etapas realizadas ao longo de toda a interação.

Por intermédio do teste t de Student, foram encontradas fortes correlações entre os sinais de SC coletados e as etapas do processo interativo para ambas as técnicas (*ray-casting*: $r = 0,971$; *HOMER*: $r = 0,998$). Porém, foi detectada diferença significativa apenas na comparação entre os sinais de SC e as etapas do processo interativo durante o uso da técnica *HOMER* ($p = 0,02$, $t = -2,36$). Em média, os valores de SC durante a etapa de seleção foram menores do que durante a etapa de manipulação.

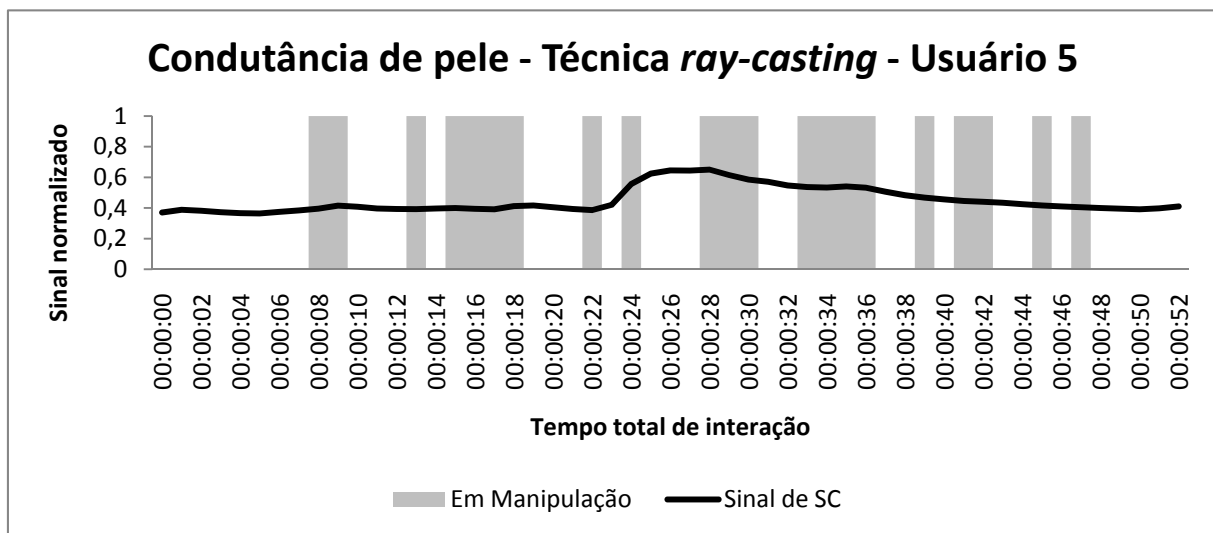


Figura 38. Desempenho do “Usuário 5” e seu traçado de SC ao longo do tempo, durante o uso da técnica *ray-casting*.

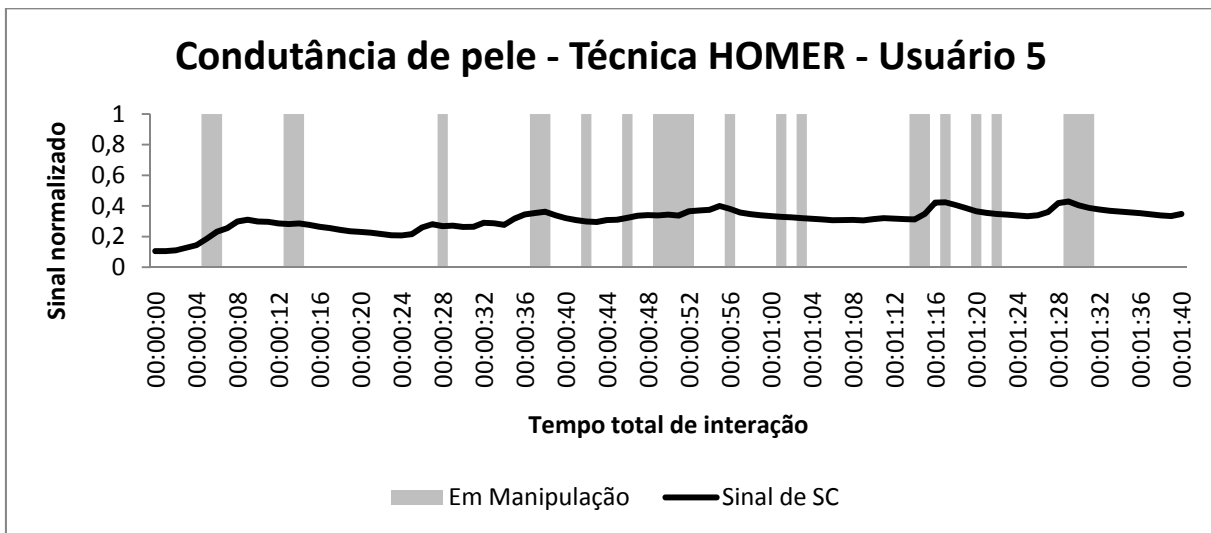


Figura 39. Desempenho do “Usuário 5” e seu traçado de SC ao longo do tempo, durante o uso da técnica HOMER.

4.6 Considerações sobre o Estudo Piloto

Este estudo permitiu uma avaliação preliminar sobre a nova metodologia, antes de utilizá-la em um processo avaliativo com um grupo maior de usuários. Para tanto, apontou itens do protocolo, da plataforma de teste e do processo de análise estatística que careciam de revisão, adaptação e definição.

De acordo com os resultados apresentados pela Seção 4.5, os seguintes itens deveriam ser considerados:

- Aperfeiçoar a AV dos livros, com novos recursos visuais como a adição de um *feedback* visual para seleção de objetos, e volumes transparentes nos locais de destino dos livros para melhorar a noção de distância em relação ao chão, além de recursos sonoros para informar sobre as colisões;
- Escurecer a sala de testes durante o experimento;
- Limpar os eletrodos e o HMD antes de cada teste;
- Aplicar o gel condutor nos sensores de ECG;
- Deixar o HMD desligado durante o período basal pré-teste, no qual o usuário deverá permanecer com os olhos abertos;
- Rever a posição dos fios e equipamentos dentro da sala de testes;
- Determinar quais perguntas do questionário pós-teste são relevantes, reduzindo o tamanho do questionário;
- Determinar as medidas fisiológicas de estudo, considerando apenas medidas de curto período;

- Aumentar a amostra de observação;
- Utilizar apenas o AV dos livros para as fases de treinamento e experimental;
- Estudar a redução do tempo de protocolo, retirando a medição fisiológica pós-teste, e diminuindo os tempos da etapa basal pré-teste, treinamento e experimental.

O Capítulo 5 apresenta um detalhamento sobre estes tópicos, como forma de revisar e definir o uso da metodologia para estudos de caso com um número maior de usuários.

5. DEFINIÇÃO DA METODOLOGIA

Conforme destacado no Capítulo 4, a avaliação preliminar apresentou itens da metodologia proposta que necessitavam de ajustes, os quais poderiam contribuir na adequação e definição da metodologia para estudos de caso com grupos maiores de usuários.

Para tanto, esta revisão considera as sugestões dos participantes, os resultados obtidos e as considerações sobre o estudo piloto (Seções 4.5 e 4.6) para aperfeiçoar o AV e os procedimentos de análise estatística. As próximas seções apresentam as adaptações da plataforma de testes e as definições das medidas fisiológicas, das medidas de desempenho e do protocolo de testes propostos para este trabalho. Ao final, são apresentados os passos para aplicação da metodologia desenvolvida.

5.1 Adaptações da Plataforma de Testes

Tomando como base o estudo piloto da Seção 4.5 e os comentários dos usuários no questionário pós-teste, foi necessário executar algumas alterações no AV de experimento, antes de sua utilização na etapa de testes finais. Para tanto, o AV recebeu três novos recursos, com o objetivo de sinalizar mais claramente o momento da indicação de objetos, melhorar a noção de distância em relação ao chão, e informar sobre a ocorrência de colisões. São eles:

- Um *feedback* visual adicional para a etapa de seleção de objetos que, quando habilitado, muda a textura do quadro localizado no fundo do cenário, como mostra a Figura 40;
- Volumes transparentes nos locais de destino dos livros, com o objetivo de facilitar o posicionamento dos objetos e melhorar a noção de distância em relação ao chão (Figura 41);
- Um sinal sonoro, disparado na ocorrência de colisões no AV.

Devido a essas alterações, também foi preciso atualizar o documento de instruções sobre o AV, como mostra o Apêndice F.

Uma mudança na configuração de equipamentos de RV também foi necessária, devido a um problema no sensor de rastreamento do HMD. Para corrigir este defeito, foi adicionado um segundo sensor ao dispositivo de rastreamento magnético, o qual passou a ter a função de capturar os movimentos da cabeça do usuário. A Figura 42 apresenta a

nova configuração de equipamentos, com destaque para sensor de rastreamento preso a uma tiara colocada na cabeça do usuário, durante a realização de um teste.

Para suporte ao recurso sonoro adicionado ao AV, foi utilizada a biblioteca irrKlang²⁷, destinada à reprodução de som 2D e 3D para aplicações multimídia, como jogos e AVs.

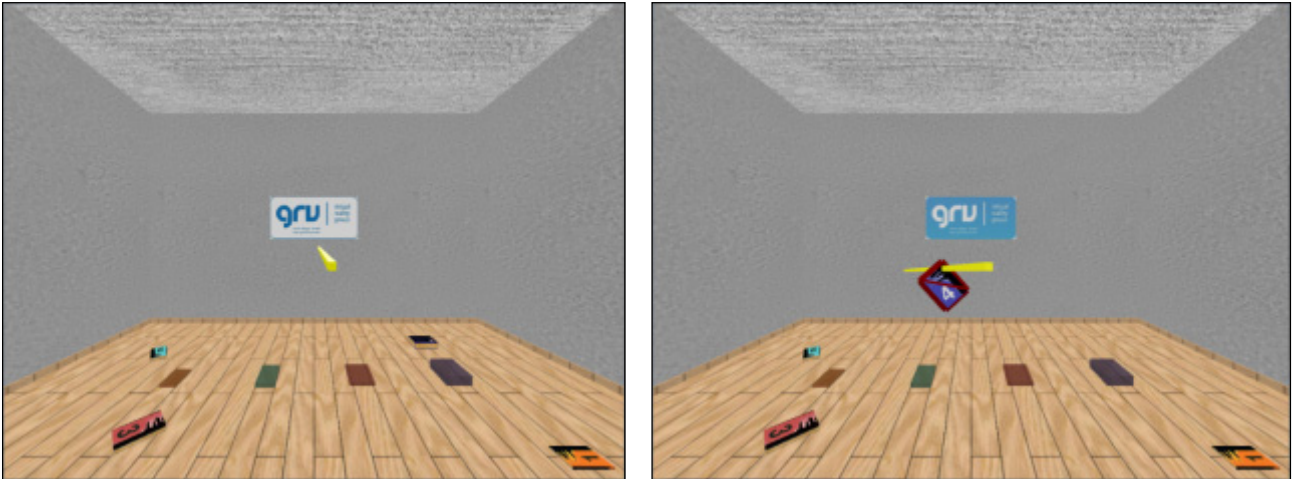


Figura 40. Exibição do novo recurso visual de seleção, em conjunto com a indicação de um cubo aramado ao redor do objeto.



Figura 41. Exibição do novo recurso visual para posicionamento dos objetos, representado por volumes transparentes.

²⁷ <http://www.ambiera.com/irrklang/>



Figura 42. Nova configuração dos dispositivos fisiológicos e de RV.

5.2 Definição das Medidas Fisiológicas

Como apresentado pela Seção 4.5.4, para reduzir a complexidade da análise estatística e facilitar a interpretação dos resultados, é preciso escolher medidas fisiológicas que permitam a aplicação de métodos lineares de domínio de tempo e de domínio de frequência sobre o tempo total destinado à interação no AV.

Para tanto, esta metodologia opta em utilizar apenas medidas fisiológicas de domínio de tempo e de domínio de frequência para curtos períodos, geradas pelo *software* Biograph Infiniti a partir dos dados do equipamento Procomp Infiniti. Conforme destacado pela Seção 2.2, estas medidas podem ser analisadas considerando períodos padrão de 5 minutos, semelhantes aos tempos destinados para interação com cada técnica de interação.

Assim, as seguintes medidas fisiológicas foram selecionadas para a etapa de testes finais:

- Média da SC, expressa em micro-Siemens (μs);
- Média da HR, expressa em batimentos por minuto (bpm);
- Desvio padrão de todos os intervalos RR (SDRR), expresso em milissegundos (ms);
- Contagem do número de vezes em que os intervalos NN adjacentes foram superiores a 50 ms (NN50);

- Razão entre a contagem NN50 e o número total de intervalos NN, expressa em percentual (pNN50);
- Média da potência total do sinal de HR, considerando a banda de baixíssima frequência (VLF), expressa em milissegundos ao quadrado (ms^2);
- Média da potência total do sinal de HR, considerando a banda de baixa frequência (LF), expressa em ms^2 ;
- Média da potência total do sinal de HR, considerando a banda de alta frequência (HF), expressa em ms^2 ;
- Razão entre as medidas de LF e HF (LF/ HF).

O estudo piloto também apontou para a necessidade de definir métodos padronizados de normalização e exclusão de medidas de HR e SC, visando à correta comparação estatística entre as diferentes medidas de avaliação estabelecidas para este trabalho.

Para normalizar a medida de SC, foi adotada uma escala de 0 a 1, aplicada para cada sinal de SC coletado, a qual atribuía o valor mínimo 0 à menor medida de SC, e o valor máximo 1 à maior. Desta forma, os novos valores gerados se tornavam uniformes, e mantinham as características de cada indivíduo.

Para a medida de HR, foi necessário aplicar um procedimento de exclusão de sujeitos da amostra, devido à forma de coleta utilizada ser suscetível a geração de ruídos no sinal de HR (vide Seção 4.5.6). O procedimento eliminava os participantes que apresentassem valores de HR fora dos critérios a seguir apresentados, os quais foram executados na seguinte ordem:

- HR em repouso: com base na coleta basal, eliminação de casos nos quais as medidas de HR foram inferiores a 60 bpm ou superiores a 100 bpm. De acordo com Guyton e Hall [GUY05], convencionou-se como normal no ser humano uma HR entre 60 e 100 batimentos por minuto;
- HR máxima: eliminação de casos nos quais a medida de HR coletada durante a experiência é maior do que a HR máxima, estimada a partir da equação linear de Tanaka [TAN01], conforme mostra a Tabela 6;
- HR alvo: eliminação de casos nos quais a medida de HR coletada durante a experiência é maior do que a HR alvo, estimada a partir do método de Karvonen [KAR57], calculado conforme a Tabela 6. Para aplicação deste método, determinou-se como padrão um nível de intensidade de 50% para o exercício proposto, uma vez que o esforço físico exigido pela tarefa pode ser

considerado dentro de uma zona de atividade moderada [FER03] [AME05], já que o sujeito realiza o teste sentado e executa apenas movimentos espaciais utilizando os braços e a cabeça;

- HR mínima: eliminação dos casos na qual a medida coletada durante a experiência foi inferior a 60 bpm.

Tabela 6. Equações utilizadas para normalizar medidas de HR.

Método	Fórmula
Tanaka	$HR_{\text{máxima}} = 208 - (0,7 \times \text{idade})$
Karvonen	$HR_{\text{alvo}} = ((HR_{\text{máxima}} - HR_{\text{repouso}}) \times \text{intensidade}\%) + HR_{\text{repouso}}$

5.3 Definição das Medidas de Desempenho

Com base no estudo piloto, e conforme apresentado pela Seção 4.5.4, é preciso também diminuir o número de medidas de usabilidade, relacionadas ao desempenho do usuário, visando simplificar o processo de análise estatística.

Para tanto, as variáveis derivadas apresentadas na Seção 4.4.2 foram descartadas devido à forte correlação entre elas e sua medida matriz – fato que, em um processo de avaliação, tende a apontar para mesmos resultados ou conclusões. Já no caso da medida do número de liberações, esta também foi descartada por derivar da diferença entre o número de anexações e colisões, e também por não ter apresentado relevância durante o estudo piloto.

As seguintes medidas de desempenho, também definidas na Seção 4.4.2, foram selecionadas para a coleta e posterior análise estatística:

- Tempo;
- Anexações;
- Colisões;
- Precisão.

5.4 Definição do Protocolo de Testes

O estudo piloto também mostrou a necessidade de revisar o protocolo de testes, reduzindo seu tempo de execução. Durante os testes, observou-se que os usuários que levaram mais tempo para finalizar o teste acabaram apresentando um rendimento inferior

aos demais. Além disso, estes apresentaram dificuldades no aprendizado da interface e seus recursos.

Para facilitar o aprendizado do usuário, e oferecer uma maior satisfação ao longo de sua participação no teste, optou-se em utilizar apenas um dos AVs tanto para as etapas de treinamento, como para as etapas de experimento. O AV dos livros foi escolhido por apresentar mais de um objeto para interação, por oferecer diferentes graus de dificuldade na seleção e manuseio dos objetos, por oportunizar as colisões entre os próprios objetos e o cenário virtual, e por manter o mesmo contexto ao longo de todo o experimento.

Esta decisão resultou em uma pequena alteração na aplicação, a qual recebeu uma nova opção de registro na tela inicial de configuração, destinada à escolha da etapa a ser executada (treinamento ou experimento).

Para reduzir o tempo do teste, foi necessária a revisão das etapas do protocolo apresentado na seção 4.4.3. Para esta tarefa, questionários foram simplificados, estimativas de tempo foram reduzidas e novos procedimentos para coleta de dados fisiológicos foram incluídos, como forma de oferecer um protocolo que considerava tanto o conforto do usuário durante o teste, como uma coleta de dados adequada e confiável.

Na etapa de “Aplicação dos Questionários Pré-teste”, a questão referente aos conhecimentos sobre computadores e respectivos sistemas foi excluída. O Apêndice E apresenta o novo questionário. Para esta etapa, foi estimado um período de 8 minutos para sua conclusão.

Na etapa de “Fixação dos Dispositivos”, foram incluídos os procedimentos de limpeza dos punhos do usuário, com álcool gel, e a aplicação de gel condutor nos sensores de ECG. Adotou-se este último procedimento como forma de reduzir os ruídos causados pela resistência elétrica da pele, conforme destacado na Seção 4.5.6. Para esta etapa, foi mantida a estimativa de tempo de 3 minutos.

Para a “Coleta de Dados Basal Pré-teste”, definiu-se um padrão de uso do HMD, o qual permanecia desligado para evitar o cansaço visual antes do início das experiências. Neste período, o usuário também permanecia com os olhos abertos para acostumar-se com a presença dos equipamentos fixados a seu corpo. Um tempo de 2 minutos foi destinado para a coleta de dados fisiológicos.

As etapas de “Realização do Treinamento” com as técnicas “A” e “B” tiveram uma redução no tempo de treinamento de 5 para 3 minutos, devido ao fato de que a maioria dos usuários ter solicitado, no estudo piloto, o encerramento do treinamento antes do tempo anteriormente estipulado.

Já para as etapas de “Realização do Experimento” com as técnicas “A” e “B”, foi mantido o tempo de duração de 7 minutos para o processo interativo. No entanto, ao final destas etapas foi estabelecido um tempo adicional de 3 minutos para descanso do participante, para a qual o HMD era novamente desligado.

Na etapa de “Retirada dos Dispositivos”, incluiu-se o procedimento de limpeza dos punhos e de higienização dos dispositivos fisiológicos e de RV com lenços umedecidos e flanelas secas, após a retirada dos equipamentos. Além disso, foi retirado o procedimento de coleta de dados pós-teste, conforme justificativa apontada na Seção 4.5.6. Para esta etapa, foi estimado um tempo de 4 minutos.

O questionário aplicado na etapa de “Aplicação dos Questionários Pós-Teste” foi revisado, passando de 15 para 7 questões. Perguntas que serviram de base para avaliar os procedimentos adotados para o experimento foram suprimidas, bem como questões que avaliavam aspectos similares foram agrupadas. Assim, o novo questionário pós-teste, apresentado no Apêndice G, contemplava a avaliação de usabilidade de dispositivos, da tarefa, das técnicas de interação e da interface gráfica, de maneira simplificada. Para esta etapa, foi estimado um período de 5 minutos.

Considerando essas alterações e os novos tempos estimados para as etapas do protocolo, previu-se a duração de cada teste em um tempo de, aproximadamente, 45 minutos.

Por fim, foram tomados cuidados com relação ao escurecimento da sala durante o experimento. Para tanto, adotou-se o fechamento das persianas e a manutenção da iluminação artificial da sala. Da mesma forma, uma revisão na posição dos fios e equipamentos antes de cada experimento também foi determinada, com vistas a manter a disposição inicial apontada pela Figura 19.

5.5 Passos para Aplicação da Metodologia

Baseado nas definições apresentadas pelas Seções 5.2, 5.3 e 5.4, e pelos Apêndices E e G, as próximas subseções apresentam, de maneira resumida e em ordem, os passos necessários para aplicar a metodologia proposta por este trabalho. Sua definição considera o uso das medidas fisiológicas de HR e SC para avaliar a qualidade de técnicas de interação em AVs imersivos. No entanto, a metodologia não impede que outras medidas fisiológicas sejam agregadas ao processo, como forma de aprimorar a avaliação de interfaces 3D.

5.5.1 Aplicação do protocolo de testes

Esta etapa é responsável pela apresentação do experimento, utilização de dispositivos de RV e fisiológicos, exposição ao AV e técnicas de interação, e coleta de dados, tanto fisiológicos, como de desempenho e de avaliação da interface. Segue a ordem de execução do protocolo:

- Aplicação de Questionário Pré-teste: apresentação do instrutor, preenchimento do termo de consentimento (Apêndice A) e do questionário para coleta de dados demográficos e de experiência em AVs (Apêndice E), leitura do documento de instruções do experimento (Apêndice F), e estabilização dos sinais fisiológicos. Duração estimada de 8 minutos;
- Fixação dos Dispositivos: limpeza dos punhos do sujeito, aplicação de gel condutor aos sensores de ECG, acomodação para início do experimento, fixação dos dispositivos, e ajustes técnicos. Duração estimada de 3 minutos;
- Coleta de Dados Basal Pré-teste: captura das medidas fisiológicas em período de descanso, na qual o HMD está desligado e o sujeito de olhos abertos. Duração de 2 minutos;
- Realização do Treinamento – técnica “A”: interação no AV para treinar a seleção e manipulação de objetos, utilizando a primeira técnica. Duração de 3 minutos;
- Realização do Experimento – técnica “A”: interação no AV, utilizando a primeira técnica determinada para avaliação. Duração de 7 minutos;
- Intervalo entre experimentos: período para descanso, com equipamentos desligados. Duração de 3 minutos;
- Realização do Treinamento – técnica “B”: interação no AV para treinar a seleção e manipulação de objetos, utilizando a segunda técnica. Duração de 3 minutos;
- Realização do Experimento – técnica “B”: interação no AV, utilizando a segunda técnica determinada para avaliação. Duração de 7 minutos;
- Retirada dos Dispositivos: limpeza dos punhos do sujeito, higienização dos equipamentos, e período para descanso. Duração estimada de 4 minutos;
- Aplicação dos Questionários Pós-Teste: preenchimento do questionário de avaliação da interface e do experimento (Apêndice G), conversas informais, e agradecimento. Duração estimada de 5 minutos.

Maiores detalhes podem ser obtidos nas Seções 4.4.3 e 5.4.

5.5.2 Normalização de dados fisiológicos

Esta etapa é destinada à aplicação de métodos de padronização de dados fisiológicos para posterior comparação de resultados. A normalização considera as medidas fisiológicas de SC e HR de cada sujeito:

- Medida de SC: conversão de dados para uma escala de 0 a 1, na qual o valor mínimo 0 refere-se à menor medida coletada, enquanto que o valor máximo 1 refere-se à maior medida;
- Medida de HR: para dados coletados no período basal, eliminação de medidas inferiores a 60 bpm ou superiores a 100 bpm; para dados coletados durante as experiências, aplicação da equação linear de Tanaka para encontrar a HR máxima e eliminar dados superiores a esta medida, aplicação do método de Karvonen para determinar a HR alvo e eliminar dados superiores a esta medida, e eliminação de medidas inferiores a 60 bpm.

A fundamentação e as fórmulas utilizadas nesta etapa podem ser contempladas na Seção 5.2.

5.5.3 Emprego de métodos estatísticos

Esta etapa se refere à aplicação de métodos estatísticos capazes de comparar o relacionamento entre medidas fisiológicas e métricas de usabilidade. O objetivo é identificar resultados que permitam indicar desempenho e preferências do sujeito em AVs, a partir de sua própria resposta fisiológica.

Para tanto, recomenda-se executar os seguintes passos:

- Aplicação de métodos de análise exploratória de dados, considerando técnicas de sumarização de dados, testes de normalização, testes para detecção de valores extremos (*outliers*) e técnicas que verificam as correlações entre variáveis;
- Aplicação de técnicas de regressão múltipla para geração de modelos de predição, considerando métodos que selecionem as variáveis relevantes e seus respectivos coeficientes de determinação.

Já para a comparação de desempenho entre duas técnicas de interação de mesmo propósito, e o comportamento do sujeito em diferentes períodos do experimento e etapas do processo interativo, utilizando medidas fisiológicas, recomenda-se:

- Aplicação de testes t de Student, e/ou;
- Aplicação de testes ANOVA.

Como forma de exemplificar o uso desta metodologia, o Capítulo 6 apresenta um estudo de caso que utiliza os passos anteriormente apresentados na comparação entre duas técnicas de interação, de mesmo propósito. Para tanto, são consideradas as medidas fisiológicas e de desempenho, apresentadas pelas Seções 5.2 e 5.3, e os questionários apresentados nos Apêndices E e G. Testes estatísticos também são apresentados no próximo capítulo, como sugestão para emprego na avaliação de interfaces 3D.

6. ESTUDO DE CASO

Com o intuito de avaliar a eficiência do uso de medidas fisiológicas como preditoras de qualidade de interfaces 3D, este capítulo mostra um estudo de caso baseado nos testes preliminares apresentados no Capítulo 4, e nas definições metodológicas exibidas no Capítulo 5.

As próximas seções expõem os resultados alcançados com o grupo de usuários avaliado, por intermédio do uso da metodologia apresentada pela Seção 5.5.

6.1 Experimento

O experimento contou com a participação de 54 usuários, saudáveis, sendo 28 homens e 26 mulheres com idade entre 17 e 57 anos. Os mesmos participaram dos testes em horários pré-agendados, dentro de um período de duas semanas.

Para os testes, os sujeitos foram divididos em dois grupos iguais com 14 homens e 13 mulheres. Para equilibrar a ordem de utilização das técnicas de interação, um grupo iniciou os testes utilizando a técnica *ray-casting*, enquanto o outro iniciou com a técnica HOMER, no AV de experimento (Seção 4.3.2).

Para este estudo também foi definido um nível de significância de 5% para todos os testes estatísticos aplicados.

6.2 Caracterização do grupo

A Figura 43 apresenta a distribuição de respostas, com base no questionário pré-teste (Apêndice E). De acordo com o gráfico e a escala definida pelo questionário, a maioria dos sujeitos avaliados tem pouco conhecimento sobre RV (questão 1), está participando pela primeira vez de uma experiência envolvendo dispositivos não convencionais (questão 2) e fisiológicos (questão 3), e não é propensa a situações de mal-estar durante a interação com interfaces gráficas dinâmicas, como jogos e ambientes virtuais (questão 5). Nota-se também que os sujeitos indicam ter tendência a exprimir pouca irritabilidade em contato com uma nova interface gráfica (questão 4).

Por sua vez, a Tabela 7 exhibe a distribuição dos sujeitos por faixa etária, sexo e nível de escolaridade, na qual nota-se o equilíbrio entre sexos e a maior concentração de participantes na faixa etária dos 20 aos 29 anos. No que diz respeito ao nível de escolaridade, percebe-se uma presença maior de sujeitos com formação de ensino superior e pós-graduação.

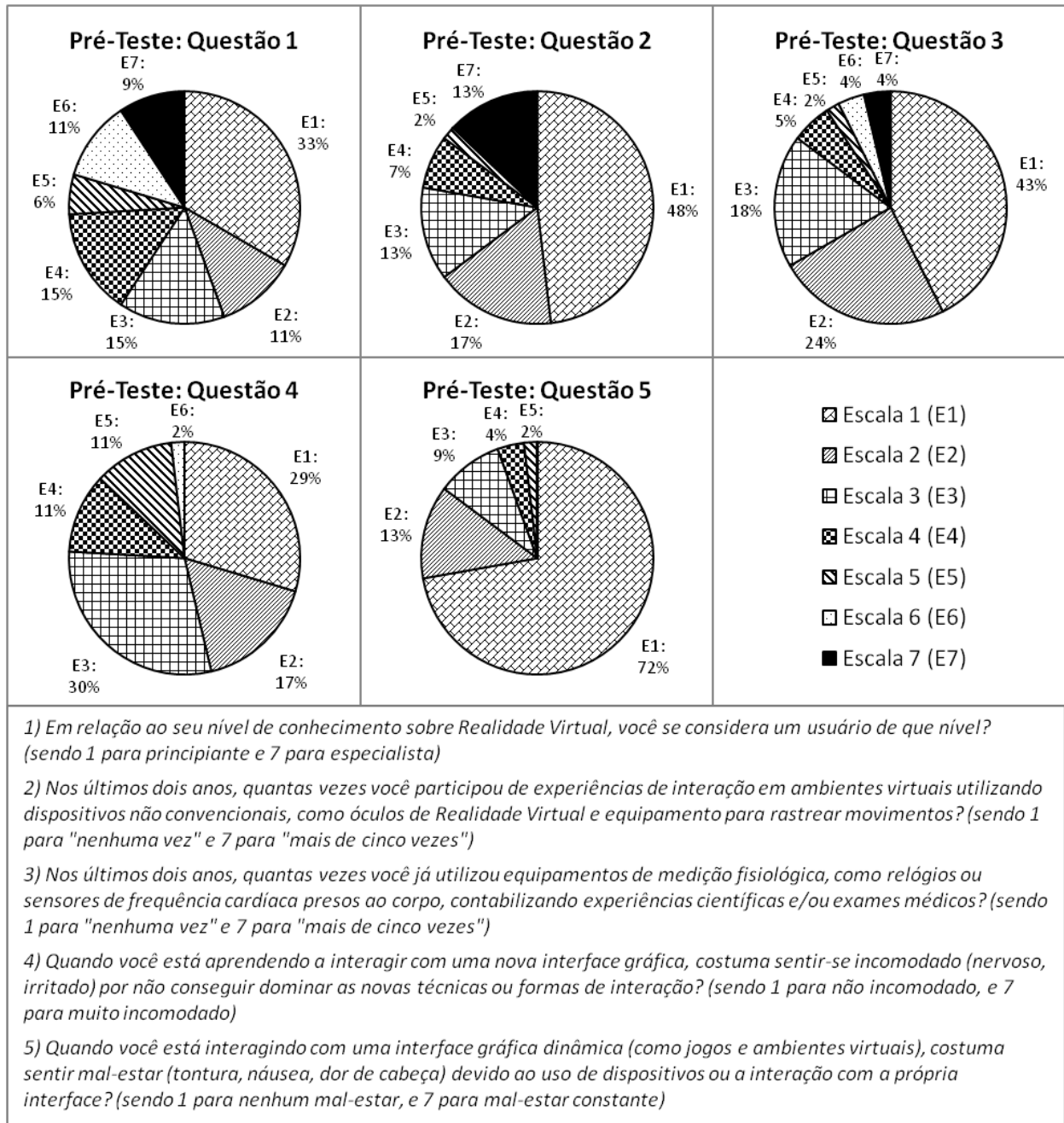


Figura 43. Distribuição das respostas do questionário pré-teste.

Tabela 7. Classificação dos sujeitos por idade, sexo e escolaridade.

GRUPOS DE IDADE	SEXO		ESCOLARIDADE		
	Masculino	Feminino	Ensino Médio	Ensino Superior	Pós-Graduação
19 anos ou menos	1	1	2	0	0
20 a 29 anos	18	16	11	16	7
30 a 39 anos	7	6	0	1	12
40 anos ou mais	2	3	0	0	5

6.3 Desempenho entre técnicas

Seguindo o mesmo método de análise aplicado na Seção 4.5.2, novas comparações entre grupos e técnicas foram realizadas, considerando os dados de desempenho coletados durante a avaliação.

Na comparação entre grupos, pode-se observar na Figura 44 que, para a variável, “tempo”, o desempenho melhora após a primeira experiência para ambas as técnicas, como já fora apontado na Seção 4.5.2. Por intermédio de testes t de Student para duas amostras independentes [SNE89], nota-se que a diferença é significativa apenas para a técnica HOMER ($p < 0,01$, $t = 2,71$), indicando que os usuários que usaram esta técnica na segunda experiência levaram, em média, menos tempo para completar a tarefa.

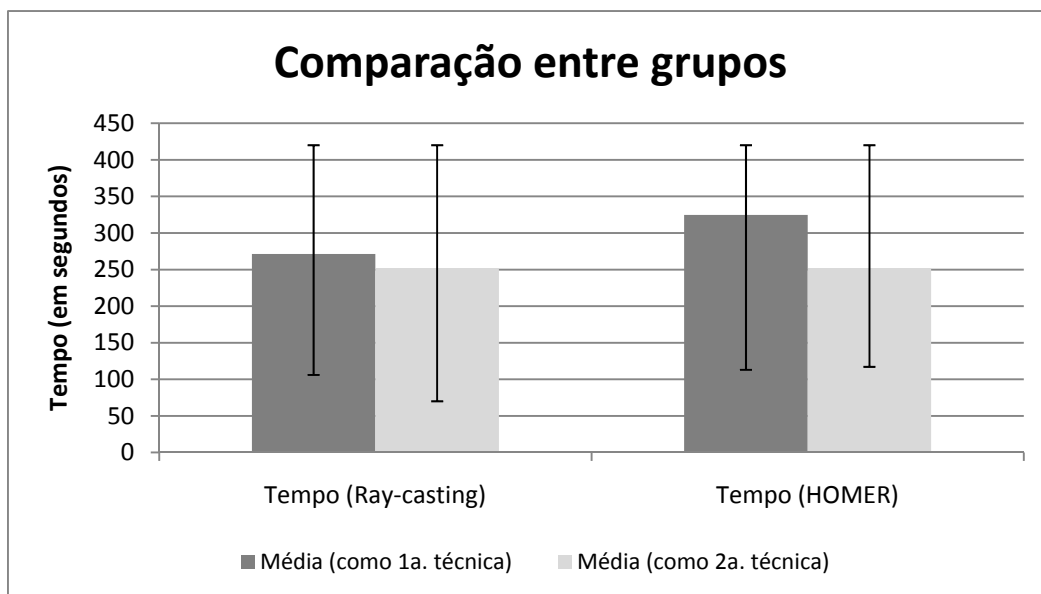


Figura 44. Desempenho entre grupos, considerando a variável “tempo” e as técnicas.

Por outro lado, para as variáveis “anexações” e “colisões”, o desempenho entre as experiências é similar para as duas técnicas, conforme mostra a Figura 45. Testes t de Student para duas amostras independentes também não apontaram diferenças significativas entre os grupos.

Já na comparação que considera a variável “precisão”, nota-se que o desempenho também melhora após a primeira experiência (Figura 46). Porém, neste caso, as diferenças foram significativas tanto para a técnica *ray-casting* ($p = 0,02$, $t = 2,05$), como para a técnica HOMER ($p = 0,01$, $t = 2,37$), revelados por testes t de Student para duas amostras independentes, considerando variâncias diferentes. Estes resultados indicam que os usuários posicionaram, em média, os objetos mais próximos do alvo na segunda experiência.

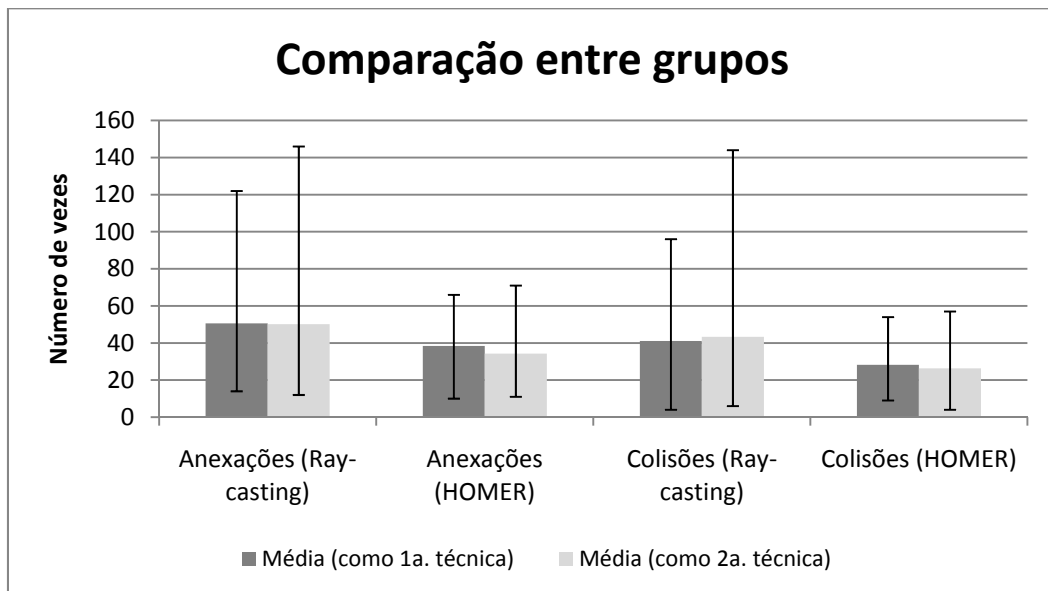


Figura 45. Desempenho entre grupos, considerando as variáveis “anexações” e “colisões” e as técnicas.

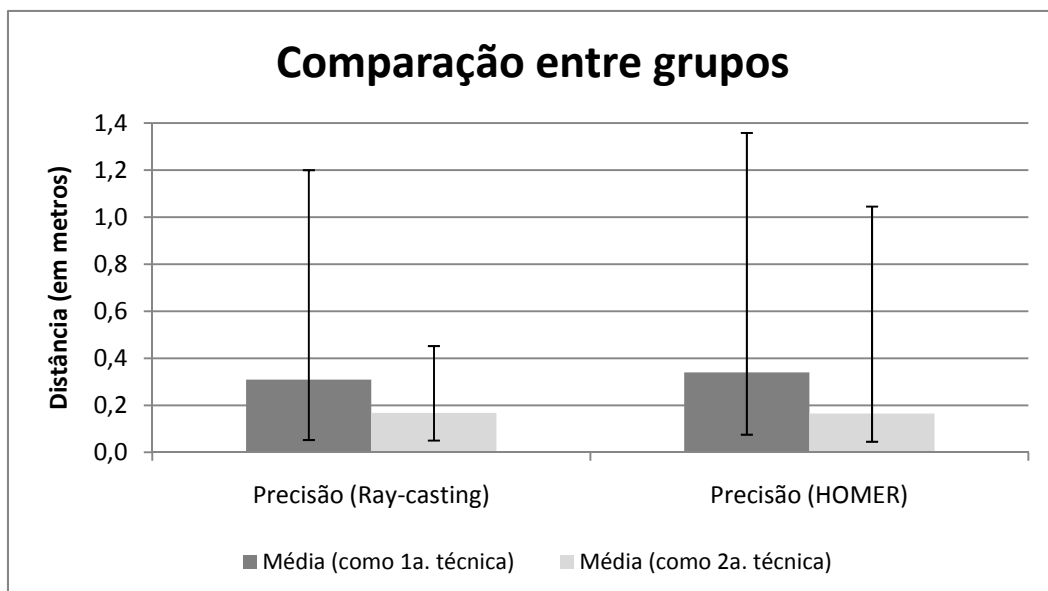


Figura 46. Desempenho entre grupos, considerando a variável “precisão” e as técnicas.

Na comparação entre técnicas, pode-se observar na Figura 47 o melhor desempenho da técnica *ray-casting*, considerando a variável “tempo” – diferentemente do resultado apresentado pela Seção 4.5.2, o qual apontava para uma leve vantagem da técnica HOMER. Uma análise de dados utilizando teste t de Student para duas amostras pareadas mostrou uma diferença significativa entre as técnicas ($p = 0,04$, $t = -1,79$). Isto indica que os usuários levaram menos tempo para executar as tarefas de interação utilizando *ray-casting*.

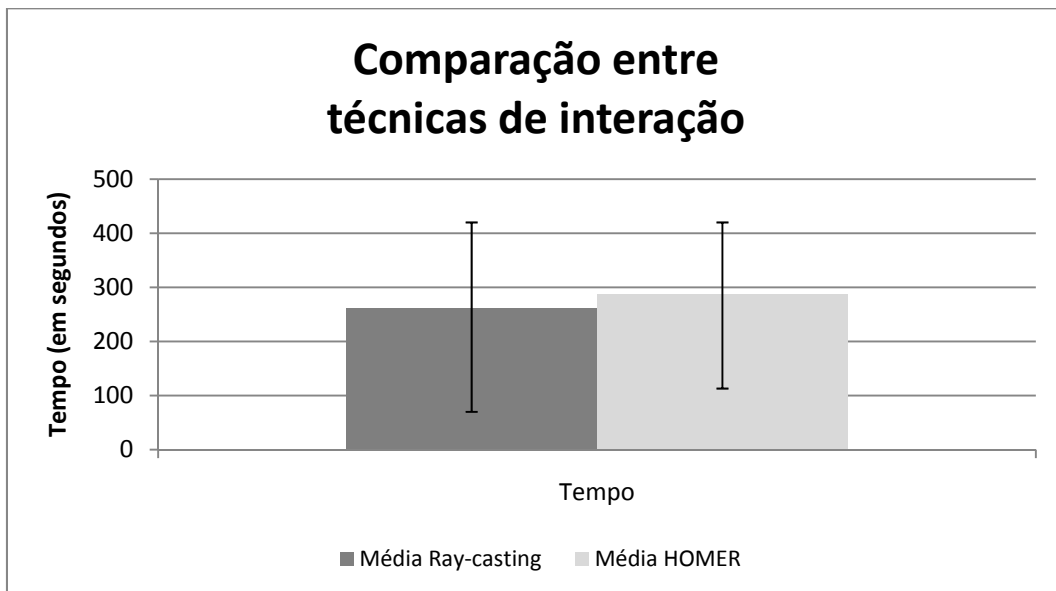


Figura 47. Desempenho entre técnicas, considerando a variável “tempo”.

No entanto, a técnica *ray-casting* apresenta um número maior de tentativas para manipular o objeto (anexações) e também um maior número de colisões com o AV, como mostra a Figura 48. Logo, a técnica HOMER se apresenta mais ajustada para as tarefas propostas, com diferenças fortemente significantes apontadas em testes t de Student para duas amostras pareadas. Os resultados mostram, em média, um número maior de tentativas de pegar o objeto ($p < 0,01$, $t = 3,74$) e, conseqüentemente, um número maior de colisões ($p < 0,01$, $t = 4,12$) durante interações com *ray-casting* do que com HOMER.

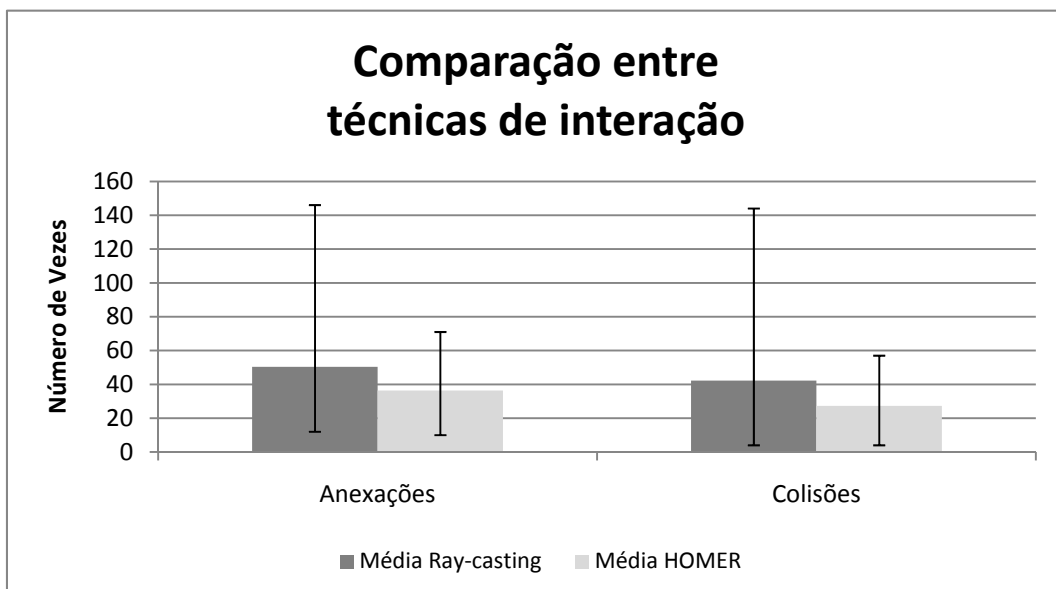


Figura 48. Desempenho entre técnicas, considerando as variáveis “anexações” e “colisões”.

Por fim, observa-se na Figura 49 que a comparação entre técnicas apresenta um desempenho similar, no que diz respeito à “precisão” em posicionar o objeto no alvo

especificado pela tarefa. Testes t de Student para duas amostras em par para médias comprovam este quadro, não apontando diferenças significativas entre as técnicas.

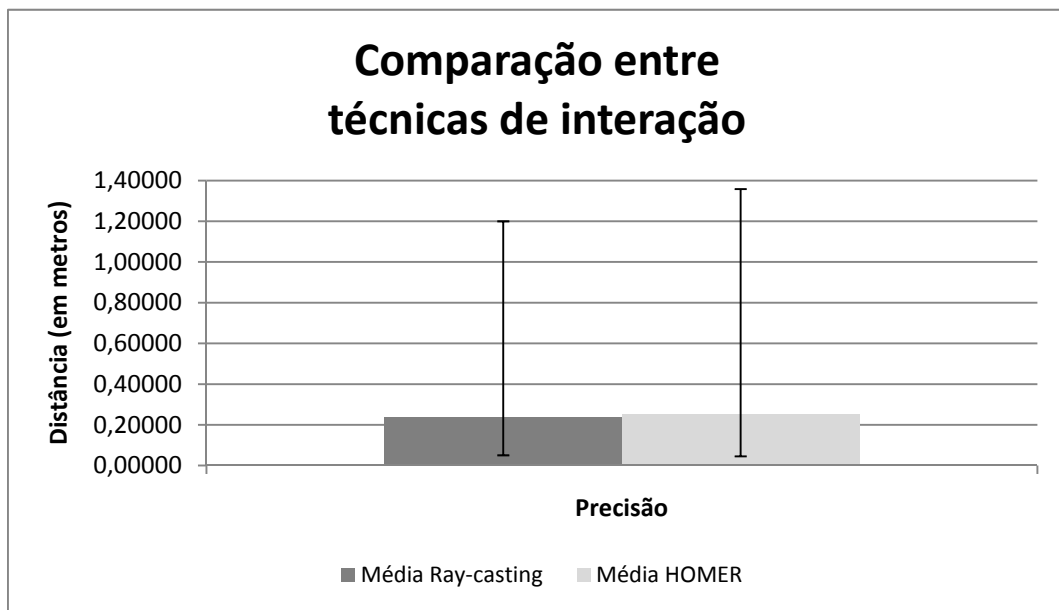


Figura 49. Desempenho entre técnicas, considerando a variável “precisão”.

6.4 Avaliação subjetiva

O questionário pós-teste (Apêndice G) tinha por objetivo avaliar as preferências dos usuários em relação ao uso das técnicas de interação e dos dispositivos, quanto à realização da tarefa e do contato com a interface gráfica. Para este questionário, manteve-se a escala progressiva de 1 a 7, definida na Seção 4.4.3.

A questão 1 fazia referência à sensação de conforto no uso de diferentes dispositivos com fios presos ao corpo. As respostas apontaram uma sensação moderada de conforto à maioria dos usuários, representada pela distribuição das respostas entre as escalas 5 e 6, como mostra a Figura 50.

Características da tarefa foram avaliadas pelas Questões 2 e 3. Para a primeira pergunta, a maioria dos usuários classificou como “baixo” o grau de irritabilidade causado pelas constantes colisões, representado pela respostas distribuídas principalmente na escala 2. Da mesma forma, para a segunda pergunta, os usuários avaliaram a informação de que havia um tempo limite para executar as tarefas como “pouco relevante”, representado pela distribuição das respostas entre as escalas 1 e 2. A Figura 51 apresenta os resultados obtidos.

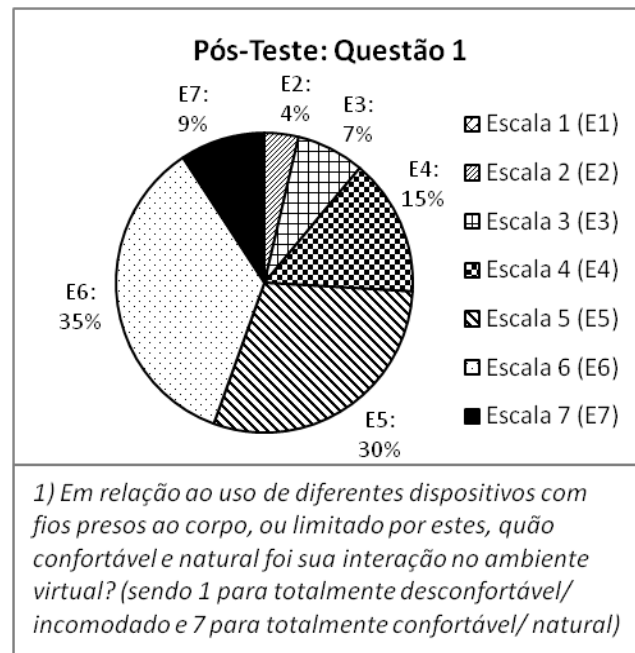


Figura 50. Avaliação da sensação de conforto durante o uso de dispositivos com fios.

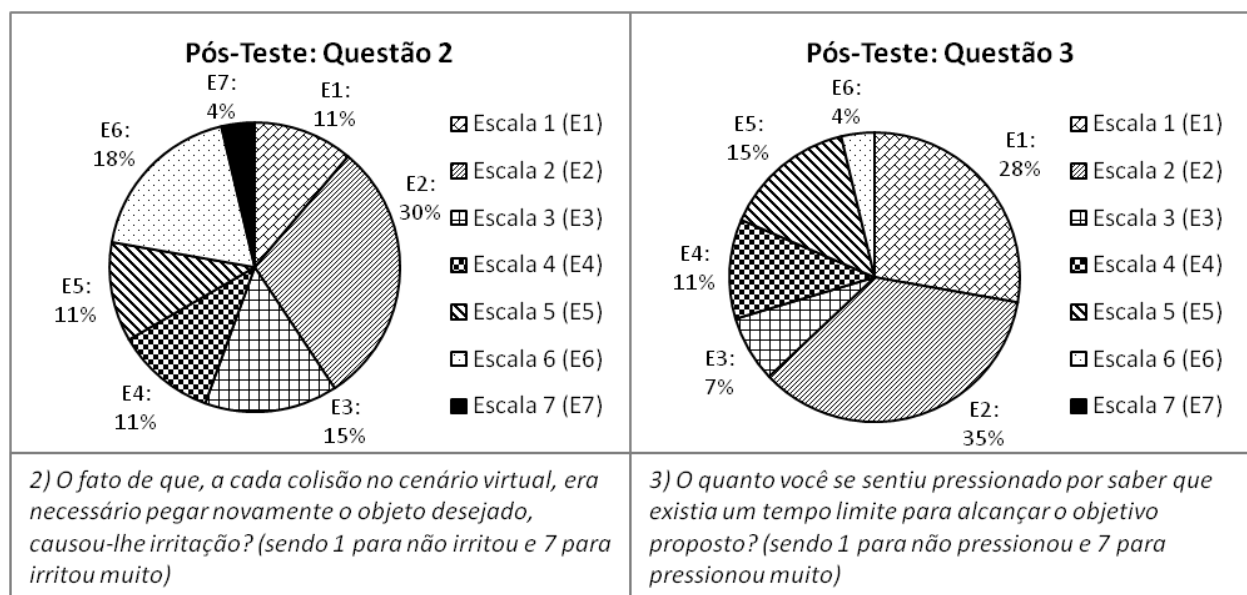


Figura 51. Avaliação do grau de irritabilidade (questão 2) e de influência do tempo limite (questão 3) durante a tarefa.

Buscando realizar uma comparação entre as técnicas de interação, as Questões 4 e 5 avaliaram o grau de satisfação do usuário com os recursos oferecidos pelas técnicas. Uma comparação entre as respostas utilizando teste t de Student para duas amostras pareadas indicou não existir diferença significativa entre as técnicas. Isto comprova a distribuição equilibrada e similar das respostas para estas questões, como mostra o gráfico da Figura 52.

As Questões 6 e 7, por sua vez, procuraram avaliar a interface gráfica do AV. A primeira avaliava o grau de convicção na realização correta da tarefa, com base nos

elementos gráficos disponíveis. Neste caso, a maioria dos usuários classificou ter manipulado e posicionado os objetos de maneira correta ou bem próxima desta, representado pela distribuição das respostas entre as escalas 6 e 7. A segunda questão, que avaliava a qualidade dos *feedbacks* visuais e sonoros oferecidos pela aplicação, também apresentou respostas com um índice de aceitação satisfatório, distribuídas principalmente entre as escalas 6 e 7. A Figura 53 apresenta estes resultados.

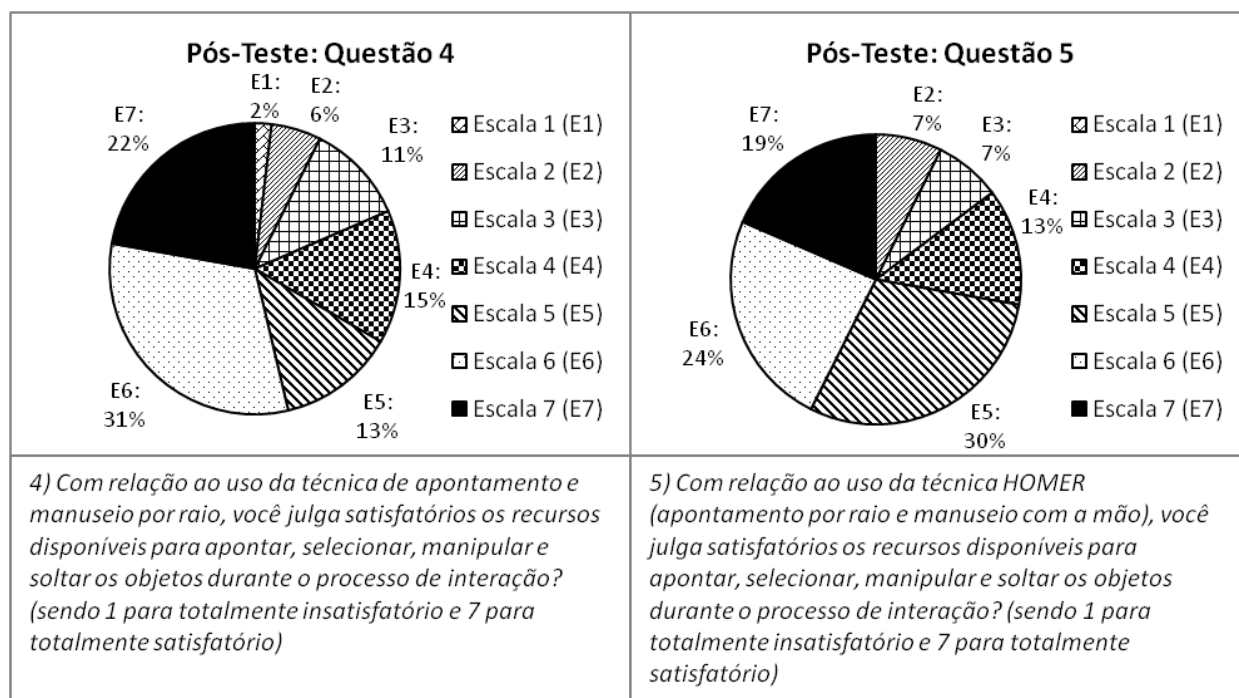


Figura 52. Avaliação dos graus de satisfação durante a interação com as técnicas *ray-casting* (questão 4) e HOMER (questão 5).

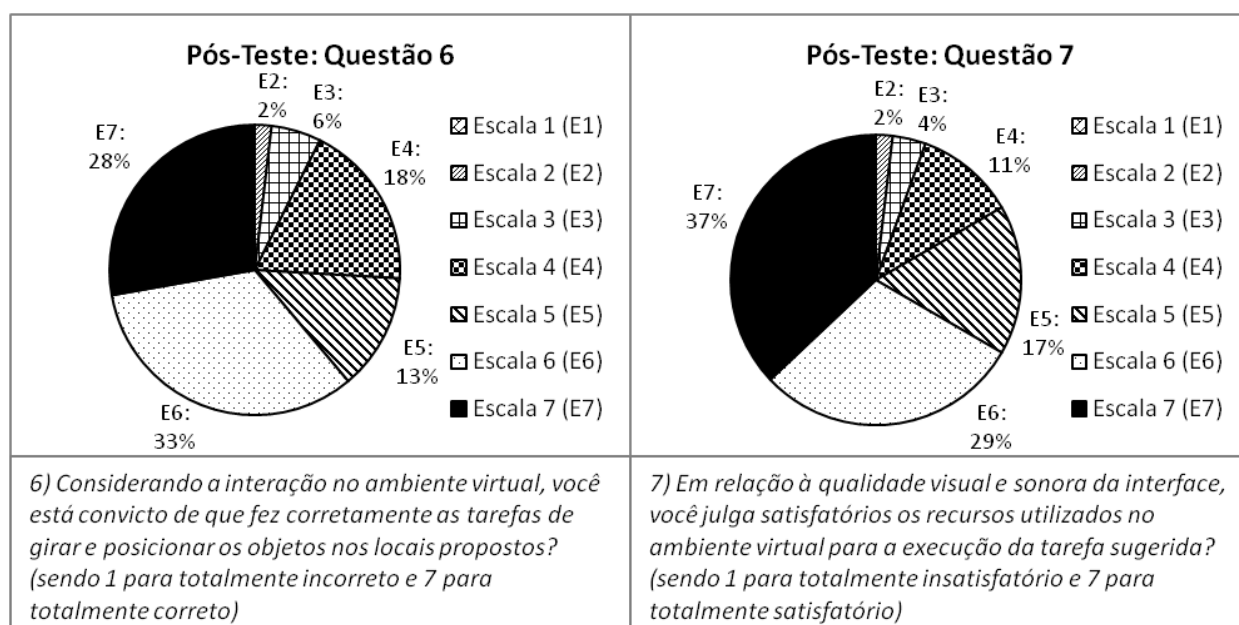


Figura 53. Avaliação do grau de convicção na execução da tarefa (questão 6) e da qualidade da interface exibida (questão 7).

Nesta etapa também foi solicitada a colaboração dos participantes, por meio de comentários descritivos no questionário pós-teste. A Tabela 8 apresenta uma compilação das dificuldades encontradas durante o experimento, destacando os momentos de desconforto ou irritação vivenciados. Já a Tabela 9 apresenta relatos referentes à satisfação do usuário e sugestões para o aperfeiçoamento da aplicação. Ao lado de cada comentário, entre parênteses, também é apresentado o número total de vezes em que as citações foram descritas no questionário pelos usuários.

6.5 Relação entre medidas fisiológicas e medidas de desempenho

Conforme apresentado na Seção 4.5.4, é necessário empregar um conjunto de métodos de análise multivariada de dados para avaliar a relação entre medidas fisiológicas e medidas de desempenho. Desta forma, é possível identificar quais medidas fisiológicas são capazes de indicar o desempenho na tarefa, ou se estas podem, ao menos, auxiliar na interpretação dos resultados. A análise envolveu quatro variáveis de desempenho, definidas na Seção 5.3, e nove variáveis fisiológicas, definidas na Seção 5.2.

Inicialmente, aplicou-se ao conjunto de dados o processo de normalização apresentado na Seção 5.2, ajustando as respostas coletadas para a devida comparação estatística. Este procedimento acabou detectando a presença de ruídos em algumas medidas de HR coletadas de 22 usuários, levando ao descarte destes dados. Assim, a amostra original precisou ser subdividida em dois novos conjuntos: uma amostra que contemplava apenas a medida de SC, para toda a população avaliada (54 indivíduos); e outra amostra que considerava as demais medidas, baseadas na HR, para uma população de 32 sujeitos.

Em seguida, uma série de testes estatísticos foi aplicada, considerando a sequência de passos definida na Seção 4.5.4, a qual determina a execução das etapas de análise exploratória de dados e regressão múltipla para identificar variáveis fisiológicas preditoras de desempenho. Uma breve descrição das técnicas e métodos estatísticos utilizados pode ser visualizada no Apêndice H.

Tabela 8. Dificuldades relatadas pelos usuários durante o questionário pós-teste.

<i>Natureza da Dificuldade</i>	<i>Momentos de desconforto</i>	<i>Momentos de irritação</i>
Em relação aos dispositivos	Leve irritação nos olhos ao final do experimento (8)	Pequena demora dos eventos de clique do <i>mouse</i> durante ações de pegar ou soltar objetos (6)
	Pressão do HMD na testa ou nariz (7)	Disposição da alavanca junto ao sensor de rastreamento (4)
	Peso do HMD (7)	
	Baixa resolução (640x480) do HMD (5)	
	Restrição de movimentos devido ao cabo do HMD (2)	
	Pressão da tiara que prendia o sensor de rastreamento (2)	
	Disposição dos fios dos sensores fisiológicos nos pés (2)	
	Cocleira fraca nos olhos (2)	
	Cansaço no braço (1)	
	Dor de cabeça após término do experimento (1)	
Muitos giros para visualizar o AV (1)		
Suor no dedal da alavanca (1)		
Em relação ao ambiente virtual	Falta de gravidade (1)	Percepção das posições corretas dos objetos, em relação aos alvos (12)
		Liberação dos objetos quando ocorriam colisões (12)
		Determinar a superfície de colisão do objeto (4)
		Transição entre técnicas (2)
		<i>Feedback</i> sonoro das colisões (1)
Em relação às técnicas de interação	Girar o objeto ao redor do eixo vertical é desconfortável, pois é necessário girar o punho além do limite (11)	Mover objetos por grandes distâncias (4)
		Excesso de giros com HOMER (4)
		Dificuldades em entender a forma de girar objetos com <i>ray-casting</i> (2)
		Tamanho grande do raio, em comparação aos livros (2)

Tabela 9. Comentários e sugestões dos usuários durante o questionário pós-teste.

<i>Natureza da Dificuldade</i>	<i>Comentários de satisfação</i>	<i>Sugestões</i>
Em relação aos dispositivos		Melhorar a disposição da alavanca junto ao sensor de rastreamento (4)
		Utilizar um HMD com maior resolução (3)
		Dar preferência a equipamentos sem fios (1)
Em relação ao ambiente virtual	Divertido, por apresentar diferentes ângulos de visualização (2)	Incluir um <i>feedback</i> visual que informe quando o objeto está corretamente posicionado (3)
	Desafiador, por exigir habilidade e atenção (2)	Escurecer um pouco mais a sala (2)
	Sensação de imersão e esquecimento do tempo (2)	Incluir outros <i>feedbacks</i> sonoros para diferenciar tipos de colisão (1)
		Adicionar sombras e efeitos de iluminação (1)
		Adicionar recursos de física (1)
Em relação às técnicas de interação	HOMER se apresentou como a técnica com interação mais próxima do natural (1)	Tempo maior de treinamento (1)
	Preferência por <i>ray-casting</i> , uma vez que utiliza a mesma metáfora para seleção e manipulação (1)	
	Claras instruções de ensino (1)	

Para a etapa de análise exploratória, foram aplicados os seguintes testes para cada estágio de análise:

- Sumarização dos dados: Estatística Descritiva;
- Normalização dos dados: teste Kolmogorov-Smirnov;
- Detecção de valores extremos (*outliers*): gráficos *stem-and-leaf* e *box-plots*;
- Correlações entre variáveis: correlação linear de Pearson, para as distribuições normais, e correlação de Spearman, para as distribuições não-normais.

Já para a etapa de regressão múltipla, foi considerado o método passo a passo (*stepwise*) para a construção dos modelos de regressão e para a seleção das variáveis preditoras. Neste processo, cada modelo gerado pode apresentar uma ou mais variáveis

preditoras, e seus respectivos coeficientes de determinação (r^2), os quais informam o poder de explicação que estas medidas têm sobre a variabilidade dos resultados alcançados. No caso deste trabalho, este coeficiente indica se uma medida fisiológica pode ser substituída ou não de uma medida de desempenho. Para testar se a regressão realizada é estatisticamente significativa, uma ANOVA também é gerada para cada modelo.

Para efeito de geração dos modelos de regressão, a análise considerou apenas as medidas fisiológicas que apresentaram resultados estatisticamente significativos ($p < 0,05$) nos testes de correlação, para ambas as técnicas.

Como forma de ilustrar as amostras utilizadas nesta análise, a Tabela 10 e a Tabela 11 apresentam, respectivamente, dados descritivos referentes às medidas de desempenho e as medidas fisiológicas.

Tabela 10. Dados descritivos das amostras, considerando as medidas de desempenho.

MEDIDAS DE DESEMPENHO	TÉCNICAS DE INTERAÇÃO			
	<i>Ray-Casting</i>		HOMER	
	<i>Média ± Desvio Padrão</i>	<i>Mínimo; Máximo</i>	<i>Média ± Desvio Padrão</i>	<i>Mínimo; Máximo</i>
Tempo	261,6667 ± 123,7783	70; 420	288,0556 ± 104,7838	113; 420
Anexações	50,3704 ± 29,7456	12; 146	36,3519 ± 16,4937	10; 71
Colisões	42,2222 ± 28,7905	4; 144	27,2778 ± 14,7283	4; 57
Precisão	0,2387 ± 0,2613	0,0503; 1,2000	0,2523 ± 0,2831	0,0453; 1,3582

Tabela 11. Dados descritivos das amostras, considerando as medidas fisiológicas.

MEDIDAS FISIOLÓGICAS	TÉCNICAS DE INTERAÇÃO			
	<i>Ray-Casting</i>		HOMER	
	<i>Média ± Desvio Padrão</i>	<i>Mínimo; Máximo</i>	<i>Média ± Desvio Padrão</i>	<i>Mínimo; Máximo</i>
SC**	0,4215 ± 0,1184	0,2296; 0,7235	0,4058 ± 0,1216	0,1988; 0,7999
HR*	93,3856 ± 18,9545	63,39; 145,04	91,4125 ± 16,1623	61,86; 127,41
SDRR*	62,4206 ± 38,9310	16,48; 214,82	68,9909 ± 45,3213	23,41; 164,79
NN50*	19,4375 ± 19,8021	1; 81	19,9688 ± 15,47	1; 66
pNN50*	7,2894 ± 4,0089	1,03; 18,77	7,4653 ± 5,5108	1,15; 29,86
VLF*	313,0234 ± 468,4144	15,67; 2127,71	285,1353 ± 374,7727	8,94; 1739,19
LF*	177,4950 ± 125,9665	0,28; 573,00	198,7619 ± 136,0865	30,45; 649,88
HF*	61,4022 ± 43,0470	0,23; 205,45	68,8969 ± 51,3895	9,37; 283,72
LF/ HF*	3,4872 ± 2,8075	0,83; 14,99	3,4378 ± 2,2932	0,88; 12,15

Nota: * amostra normalizada de 32 indivíduos ** amostra normalizada de 54 indivíduos

Após a aplicação dos testes especificados, duas medidas fisiológicas (NN50 e HF) apresentaram relação significativa com duas medidas de desempenho (“tempo” e “precisão”). No entanto, apenas uma destas relações indicou, no modelo de regressão, resultados fortemente significantes para ambas as técnicas (“tempo” x NN50, $p < 0,01$), conforme mostra a Tabela 12.

Pela análise dos resultados da Tabela 12 pode-se observar que a medida de desempenho “precisão” apresenta relação estatisticamente significativa ($p < 0,01$) com medidas fisiológicas NN50 e HF apenas na experiência com a técnica *ray-casting*.

Por outro lado, verifica-se que somente a medida NN50 pode ser considerada como a variável com grau de influência mais significativo para a medida de “tempo”, pois resultados fortemente significantes foram encontrados em ambas as técnicas ($p < 0,01$). Ainda com base na Tabela 12, pode-se dizer que a medida NN50 é capaz de indicar o desempenho do usuário, para a medida de “tempo”, com um poder explicativo de 61,98% para experiências com a técnica *ray-casting*, e 28,83% para experiências com a técnica HOMER.

Tabela 12. Modelos de regressão para medidas fisiológicas e de desempenho com forte correlação.

TÉCNICAS DE INTERAÇÃO	MEDIDAS DE DESEMPENHO	MEDIDAS FISIOLÓGICAS	REGRESSÃO	ANOVA	
			r^2	Teste F	p(valor)
HOMER	Tempo	NN50	28,83%	12,15	0,00**
	Precisão	NN50	7,26%	2,35	0,13
	Precisão	HF	2,35%	0,72	0,59
<i>Ray-casting</i>	Tempo	NN50	61,98%	48,91	0,00**
	Precisão	NN50	43,28%	22,89	0,00**
	Precisão	HF	31,33%	13,69	0,00**

Nota: * $p < 0,05$ ** $p < 0,01$

Uma regressão para as médias das medidas NN50 e “tempo” também foi gerada, como forma de agrupar o poder explicativo (r^2) da resposta fisiológica selecionada em um único modelo, independente da técnica de interação. Como resultado, obteve-se um coeficiente de determinação de 45,16% (ANOVA, $p < 0,01$, $F = 24,70$).

No entanto, os resultados de predição para ambas as análises apresentam percentuais de médio poder explicativo. Os coeficientes de determinação obtidos apresentam valores distantes do índice próximo à correlação perfeita ($r^2 = 100\%$), mostrando que a variância da medida NN50 não pode explicar sozinha a variância da medida “tempo”. Em outras palavras, pode-se afirmar que a medida fisiológica NN50 não

é capaz de substituir, por completo, a medida de desempenho “tempo”, ou estimar seu resultado com alta precisão.

6.6 Relação entre medidas fisiológicas e avaliações subjetivas

Seguindo a abordagem apresentada na Seção 6.5, também é preciso efetuar uma análise multivariada de dados para a relação entre medidas subjetivas e medidas fisiológicas. Este estudo envolveu as cinco questões do questionário pré-teste (Apêndice E), as sete questões do questionário pós-teste (Apêndice G), e as nove variáveis fisiológicas, apresentadas na Seção 5.2.

Com relação às variáveis fisiológicas, considerou-se o processo de normalização apresentado pela Seção 6.5, o qual definiu a subdivisão da amostra em dois conjuntos distintos (54 indivíduos para a medida de SC, e 32 para as medidas baseadas na HR).

Para seleção de testes estatísticos, seguiu-se a definição apresentada pela Seção 4.5.5, a qual determinou a execução de uma sequência de passos para análise exploratória e regressão múltipla dos dados, em busca das variáveis fisiológicas preditoras de preferências do usuário. Os testes aplicados para estas etapas seguiram o mesmo padrão apresentado pela Seção 6.5.

Para comparar as respostas dos questionários com as medidas fisiológicas, fez-se uma média geral das medidas fisiológicas coletadas durante a experiência. E, para determinar a geração dos modelos de regressão, foram consideradas as medidas fisiológicas que apresentaram resultados estatisticamente significativos ($p < 0,05$) nos testes de correlação.

A Tabela 13 apresenta os dados descritivos referentes às medidas de avaliação subjetiva utilizados nesta avaliação. Os dados descritivos referentes às medidas fisiológicas são os mesmos apresentados pela Tabela 11, na Seção 6.5.

Primeiramente, os testes especificados foram aplicados para verificar a relação entre as respostas do questionário pré-teste e as medidas fisiológicas. Neste estudo, duas medidas fisiológicas (NN50 e SC) apresentaram relação significativa com questões que abordavam o nível de conhecimento de RV (questão 1), a experiência com dispositivos não convencionais (questão 2), e a tendência a exprimir um leve grau de irritabilidade no aprendizado com novas interfaces (questão 4). Os resultados alcançados podem ser visualizados na Tabela 14. Pela análise destes resultados, nota-se que os modelos de regressão apresentaram resultados significantes ($p < 0,05$) para as medidas de SC e NN50 como preditoras de avaliação, porém com baixo poder explicativo para a questão 1 (SC, $r^2 = 11,44\%$; NN50, $r^2 = 15,26\%$), a questão 2 (NN50, $r^2 = 13,82\%$) e a

questão 4 (SC, $r^2 = 10,49\%$). Desta forma, pode-se afirmar que estas medidas fisiológicas não apresentam um coeficiente de determinação capaz de prever o nível de conhecimento de RV, a experiência do usuário com dispositivos não convencionais, e a tendência a sentir-se irritado com o aprendizado de novas interfaces.

Tabela 13. Dados descritivos das amostras, considerando as medidas subjetivas.

QUESTÕES	Amostra	QUESTIONÁRIOS			
		Pré-Teste		Pós-Teste	
		Média \pm Desvio Padrão	Mínimo; Máximo	Média \pm Desvio Padrão	Mínimo; Máximo
1	**	3,1852 \pm 2,0839	1; 7	5,1296 \pm 1,2293	2; 7
	*	3,6250 \pm 1,9960	1; 7	4,7188 \pm 1,2504	2; 7
2	**	2,500 \pm 2,0351	1; 7	3,5185 \pm 1,8300	1; 7
	*	2,6563 \pm 2,1039	1; 7	3,5938 \pm 1,9652	1; 7
3	**	2,2593 \pm 1,5804	1; 7	2,6111 \pm 1,5470	1; 6
	*	2,4688 \pm 1,7411	1; 7	2,500 \pm 1,5450	1; 6
4	**	2,6296 \pm 1,4048	1; 6	5,1482 \pm 1,6184	1; 7
	*	2,7500 \pm 1,5658	1; 6	5,1250 \pm 1,5606	2; 7
5	**	1,500 \pm 0,9467	1; 5	5,1111 \pm 1,4494	2; 7
	*	1,5625 \pm 1,0140	1; 5	4,9375 \pm 1,4128	2; 7
6	**			5,5370 \pm 1,3277	2; 7
	*			5,7188 \pm 1,1705	3; 7
7	**			5,7963 \pm 1,2646	2; 7
	*			5,7813 \pm 1,3616	2; 7

Nota: * amostra normalizada de 32 indivíduos ** amostra normalizada de 54 indivíduos

Tabela 14. Modelos de regressão para as médias das medidas fisiológicas e de avaliação subjetiva do questionário pré-teste.

AVALIAÇÕES SUBJETIVAS	MEDIDAS FISIOLÓGICAS	REGRESSÃO	ANOVA	
		r^2	Teste F	p(valor)
Questão 1	SC	11,44%	6,71	0,01*
	NN50	15,26%	5,40	0,03*
Questão 2	NN50	13,82%	4,81	0,03*
	LF	10,65%	3,57	0,07
	HF	7,50%	2,43	0,13
Questão 4	SC	10,49%	6,09	0,02*

Nota: * $p < 0,05$ ** $p < 0,01$

Também foram aplicados testes para averiguar a relação entre as respostas do questionário pós-teste e as medidas fisiológicas. Para esta análise, excluiu-se a comparação com as médias das medidas fisiológicas para as Questões 4 e 5, uma vez que se tratavam de perguntas específicas de avaliação de cada uma das técnicas de interação utilizadas. Neste caso, as respostas foram comparadas com as medidas fisiológicas da experiência com a técnica *ray-casting*, para a questão 4, e da experiência com a técnica HOMER, para a questão 5.

A análise baseada nas médias apresentou apenas uma medida fisiológica (LF/HF) com relação significativa ($p < 0,05$) para a questão que avaliava a qualidade visual e sonora da interface (questão 7). No entanto, como mostra o modelo da Tabela 15, apenas 13,72% da variância da medida LF/HF pode explicar as respostas da questão 7. Assim, pode-se dizer que LF/HF não é capaz de substituir este quesito do questionário pós-teste.

Tabela 15. Modelo de regressão para as médias das medidas fisiológicas e de avaliação subjetiva do questionário pós-teste.

AVALIAÇÕES SUBJETIVAS	MEDIDAS FISIOLÓGICAS	REGRESSÃO	ANOVA	
		r^2	Teste F	p(valor)
Questão 7	LF/HF	13,72%	4,77	0,03*

Nota: * $p < 0,05$ ** $p < 0,01$

Já a análise que considerou o grau de satisfação do uso das técnicas de interação apresentou relações significativas ($p < 0,05$) somente entre as medidas fisiológicas (HR, NN50 e HF) e as respostas que avaliaram a técnica *ray-casting* (questão 4), conforme mostra a Tabela 16. Neste caso, os modelos de regressão gerados também apontam para coeficientes de determinação de baixo poder explicativo (HR, $r^2 = 12,83\%$; NN50, $r^2 = 17,02\%$; HF, $r^2 = 23,39\%$), o que impossibilita prever a satisfação do usuário no uso da técnica *ray-casting* por meio das medidas fisiológicas HR, NN50 e HF. Logo, não é possível substituir o quesito proposto pela questão 4 no questionário pós-teste.

Tabela 16. Modelos de regressão para as medidas fisiológicas na avaliação do uso da técnica *ray-casting*.

AVALIAÇÕES SUBJETIVAS	MEDIDAS FISIOLÓGICAS	REGRESSÃO	ANOVA	
		r^2	Teste F	p(valor)
Questão 4 (<i>ray-casting</i>)	HR	12,83%	4,42	0,04*
	NN50	17,02%	6,15	0,02*
	HF	23,39%	9,16	0,01*

Nota: * $p < 0,05$ ** $p < 0,01$

6.7 Medidas fisiológicas na comparação entre técnicas de interação

Considerando o processo de normalização e as nove medidas fisiológicas apresentadas pela Seção 5.2, foram realizadas comparações entre as técnicas de interação utilizando testes t de Student para duas amostras pareadas e ANOVA. Para tanto, utilizou-se as médias dos sinais fisiológicos coletados durante a experiência, para cada uma das técnicas.

Para esta análise, foi definida como hipótese nula (H_0) o fato de não existir diferença entre o uso das técnicas, e como hipótese alternativa (H_1), o uso da técnica *ray-casting* causar maior estresse ao usuário do que o uso da técnica HOMER. Em dados absolutos, isto significa dizer que os valores de medição fisiológica coletados durante a experiência com *ray-casting* devem ser superiores aos coletados para a experiência com HOMER.

Após empregar os testes para cada medida fisiológica, não foram encontrados resultados significantes, rejeitando a hipótese H_1 . Logo, isto indica que, para este estudo de caso, não existe diferença entre usar a técnica *ray-casting* ou a técnica HOMER.

6.8 Medidas fisiológicas na comparação entre experimentos

Para destacar as diferenças fisiológicas existentes entre períodos de interação e de descanso (basal), foram realizadas comparações entre as medidas coletadas para cada uma das técnicas, bem como entre o período basal e os experimentos. Para análise estatística, foram utilizados os testes t de Student para duas amostras pareadas e ANOVA.

A primeira análise buscou identificar se o processo interativo altera os níveis de ansiedade (medido pela SC) ou esforço físico (medido pela HR) de uma pessoa, considerando os dados basais e os dados coletados nas experiências. Para tanto, definiu-se como hipótese H_0 o fato de não existir diferença significativa de ansiedade ou esforço físico entre os estados basal e interativo, e como hipótese H_1 , o fato do período basal apresentar valores de medição inferiores ao dos períodos em experimento – ou seja, o usuário estar de fato relaxado e tranquilo antes de desenvolver as tarefas no AV.

Na comparação entre os períodos basal e da primeira experiência, foram encontradas diferenças fortemente significativas entre as medidas normalizadas de HR, para ambos os testes ($p < 0,01$, $t = -3,97$, $F = 15,74$). Resultados similares com medidas de HR também foram obtidos na comparação entre os períodos basal e da segunda experiência ($p < 0,01$, $t = -4,86$, $F = 23,65$), e considerando a média dos experimentos

($p < 0,01$, $t = -5,32$, $F = 28,27$). Assim, pode-se afirmar que, durante o período basal, em média, ocorrem menos batimentos cardíacos do que durante os experimentos. Isto comprova que os usuários, neste estudo, passam de um estado de relaxamento para um estado de esforço físico, caracterizado pelas experiências com o processo interativo.

Por outro lado, na comparação entre estes mesmos períodos com medidas normalizadas de SC não foram encontrados resultados significantes. Logo, a hipótese H0 é aceita, indicando que não existe diferença significativa entre os níveis de ansiedade gerados pelo período basal e os experimentos. Isto pode ser justificado pelas características do grupo avaliado, conforme mostra a Seção 6.2, com maioria de sujeitos participando pela primeira vez de uma experiência de RV.

A segunda análise comparou a alteração dos níveis de ansiedade e de esforço físico entre as experiências, independente da técnica. Para tanto, definiu-se como hipótese H0 o fato de não existir diferença significativa de ansiedade ou esforço físico entre a primeira e a segunda experiência, e como hipótese H1, o fato da primeira experiência apresentar valores de medição superiores ao da segunda experiência. Isto pode ser um indicativo de que, na segunda experiência, o usuário está mais acostumado ao processo interativo.

Inicialmente, a comparação entre experiências considerando as medidas de HR não apontou diferenças significantes, indicando que o esforço físico exigido pelas experiências é o mesmo, rejeitando-se a hipótese H1.

Já na comparação que considerou as medidas de SC, foram encontradas diferenças significantes entre as experiências, em ambos os testes aplicados ($p < 0,01$, $t = 3,16$, $F = 10,00$). Neste caso, pode-se dizer que, em média, os usuários estão mais ansiosos na primeira do que na segunda experiência. Isto também comprova que a carga cognitiva exigida numa primeira experiência é maior, pois o usuário está aprendendo a interagir no AV e acostumando-se às etapas do processo interativo.

6.9 Medidas fisiológicas na comparação entre etapas do processo interativo

Para medir a relação entre medidas fisiológicas e as etapas do processo interativo, foram comparadas as etapas de seleção e manipulação de cada técnica, também utilizando testes t de Student para duas amostras pareadas e ANOVA. Para esta análise, apenas a medida normalizada de SC foi considerada, devido às características de sua coleta, a qual é processada segundo a segundo pelo *software* Biograph Infiniti. Cada técnica de interação foi avaliada individualmente, com base nas médias de SC das etapas interativas.

Para facilitar esta comparação, foi também desenvolvido um *software* capaz de juntar, em um único arquivo, dados referentes às medidas de desempenho e medidas fisiológicas de cada usuário. A ferramenta, desenvolvida na linguagem C, tinha por objetivo mostrar a evolução do usuário durante o processo interativo, considerando a técnica de interação utilizada, e agrupando as medidas fisiológicas e as ações desempenhadas no AV ao longo do tempo. A Figura 54 apresenta um exemplo de saída em arquivo após processamento dos dados.

```
Tempo;SC;Tarefa;Ação;ÁreaColidida
001;6.410750;0;0;0
002;7.269375;0;0;0
003;7.262375;0;0;0
004;7.503375;0;0;0
005;7.719625;4;1;0
006;7.779875;4;2;5
007;7.644751;0;0;0
008;7.510375;0;0;0
009;7.746500;4;1;0
010;7.908249;4;0;0
```

Figura 54. Exemplo de saída do *software* responsável em unir dados fisiológicos e de desempenho na tarefa.

Como hipótese H0, foi definido que inexistia diferença significativa de ansiedade entre as etapas interativas (seleção = manipulação). Por outro lado, como hipótese H1, definiu-se que a etapa de manipulação apresenta níveis de ansiedade maiores que a etapa de seleção.

Como resultados, não foram encontrados diferenças significantes para nenhuma das técnicas de interação avaliadas para os dois testes estatísticos aplicados, os quais confirmam a hipótese H0.

Procurando analisar o processo interativo com um maior nível de detalhe, foi também realizada uma comparação entre as médias das etapas de seleção e manipulação e das ações desempenhadas pelos usuários (anexar, liberar ou colidir um objeto) no AV. Para tanto, seguiu-se a ordem de ações executadas pelo usuário, como mostra a Figura 55. Como exemplo, comparou-se as médias da etapa de seleção com as médias das ações de “anexar”, estas com a etapa de manipulação, e assim por diante.

A análise realizada apontou para dois resultados distintos entre etapas e ações interativas, os quais não permitem identificar uma característica que seja independente da

técnica utilizada. Na primeira situação, considerando apenas a técnica HOMER, foi encontrada uma forte diferença significativa entre a etapa de manipulação e a ação de “colidir” ($p < 0,01$, $t = 3,13$, $F = 9,77$). Em média, a etapa de manipulação gera mais ansiedade ao usuário do que nos momentos de colisão no AV. No segundo caso, considerando apenas o uso da técnica *ray-casting*, existe uma diferença significativa entre a ação de “colidir” e a etapa de seleção ($p < 0,05$, $t = -1,99$, $F = 3,96$). Em média, a ação de “colidir” gera menos ansiedade ao usuário do que durante a etapa de seleção.

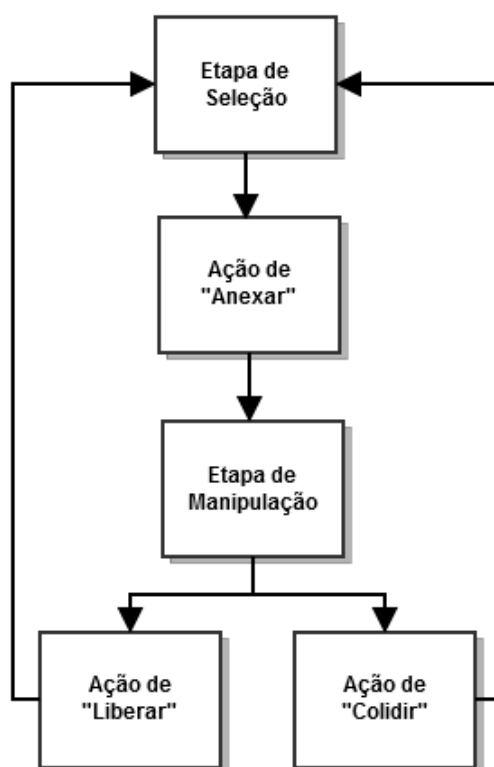


Figura 55. Etapas do processo interativo e respectivas ações realizadas pelo usuário.

6.10 Medidas fisiológicas na comparação de desempenhos entre grupos

Outras comparações também foram realizadas sobre o conjunto de dados normalizados, considerando uma avaliação baseada no desempenho entre grupos, e de acordo com a técnica de interação utilizada.

Para classificar os sujeitos em grupos, foram consideradas as quatro variáveis de desempenho: tempo, anexações, colisões e precisão. A classificação foi obtida pela soma destas variáveis, ordenadas do menor para o maior resultado. Os 10 menores resultados foram classificados como o grupo do “Melhor Desempenho”, enquanto os 10 maiores resultados como o grupo do “Pior Desempenho”. Ao final, quatro grupos foram gerados.

Com relação às medidas fisiológicas, foram selecionadas para avaliação as medidas de SC e HR, por serem indicadoras dos níveis de ansiedade e esforço físico, e a

NN50, a qual apresentou os resultados mais significantes nas avaliações apresentadas nas Seções 6.5 e 6.6.

Para analisar os grupos, definiu-se o teste t de Student para duas amostras independentes, uma vez que os dados já estavam normalizados.

A comparação entre grupos de “Melhor Desempenho” mostrou, para a medida NN50, uma diferença significativa entre as técnicas utilizadas ($p = 0,01$, $t = -2,77$) na aplicação do teste t de Student para variâncias diferentes. Isto mostrou que os usuários com melhor desempenho que utilizaram a técnica *ray-casting* tiveram uma menor HRV do que o grupo que utilizou a técnica HOMER. Pode-se dizer que o grupo que utilizou *ray-casting* conseguiu se adaptar melhor aos recursos oferecidos por esta técnica durante o processo interativo, uma vez que apresentou um valor inferior da medida de variabilidade NN50, o que reflete um ritmo cardíaco mais cadenciado e menos oscilatório [STE94].

Já na comparação entre grupos de “Melhor Desempenho” e “Pior Desempenho” utilizando *ray-casting*, diferenças significativas foram encontradas para as medidas SC ($p = 0,02$, $t = 2,13$) e HR ($p = 0,03$, $t = 2,21$). Estes resultados mostraram que, em média, os usuários com melhor desempenho ficaram mais ansiosos e tiveram um desgaste físico maior do que o grupo com pior desempenho.

Além disso, uma forte diferença significativa foi encontrada para a medida NN50, por intermédio do teste t de Student para variâncias diferentes ($p < 0,01$, $t = -7,17$). Este resultado indicou que, em média, o grupo de “Melhor Desempenho” acabou se adaptando melhor à técnica *ray-casting* do que o grupo de “Pior Desempenho”, mesmo tendo um resultado significativo maior de batimentos cardíacos.

Uma análise entre os de “Melhor Desempenho” e “Pior Desempenho” para a técnica HOMER também foi realizada, a qual apontou para uma diferença significativa apenas para a medida NN50, considerando variâncias diferentes no teste t de Student ($p = 0,02$, $t = -2,89$). Neste caso, o resultado indicou também que o grupo de “Melhor Desempenho” acabou se adaptando melhor à técnica HOMER do que o grupo de “Pior Desempenho”.

7. DISCUSSÃO

O objetivo desta seção é discutir os resultados alcançados no processo de avaliação, buscando apontar tendências e avaliar o uso da nova metodologia. O propósito é comentar os resultados obtidos com este estudo, os quais merecem atenção em futuros estudos de caso, e abordar o custo de uso deste método como instrumento de avaliação.

7.1 Relacionamento entre medidas fisiológicas e métricas de usabilidade

Como visto nas Seções 6.5 e 6.6, as medidas fisiológicas coletadas por este trabalho não podem ser consideradas como substitutas de medidas de desempenho e de avaliação subjetiva do usuário em AVs. Pelos resultados obtidos, as medidas fisiológicas apenas indicam tendências de comportamento do organismo do usuário durante o processo interativo – o que pode servir como recurso complementar na interpretação de métricas de usabilidade comumente utilizadas em processos avaliativos.

Na comparação entre medidas fisiológicas e de desempenho, pode-se destacar apenas o comportamento da medida NN50, a qual apresentou para ambas as técnicas, diferenças significantes e poder explicativo próximo de um nível aceitável para a variável “tempo”. Conforme apresentado na Seção 2.2, a medida NN50 diz respeito ao número total de intervalos NN adjacentes, superiores a 50ms – a qual é um indicativo de estabilização da HR.

A Figura 56 apresenta um gráfico de correlação entre as variáveis “tempo” e “NN50”, considerando a experiência dos sujeitos com as duas técnicas de interação. O gráfico indica a tendência da medida NN50 aumentar conforme cresce o tempo de execução da tarefa.

Este resultado pode ser interpretado sobre dois pontos de vista diferentes. Primeiramente, sujeitos que terminam as experiências em menor tempo estão mais concentrados e, com isso, empregam mais esforço físico para desempenhar as tarefas, em comparação com aqueles que despendem mais tempo. Sob outro aspecto, pode-se dizer que os sujeitos que levam mais tempo podem se acostumar com a interface, o que acaba deixando-os mais tranquilos e estabiliza suas HRs para níveis próximos de uma medida basal. De qualquer forma, é uma medida que merece ser explorada e reavaliada em pesquisas futuras.

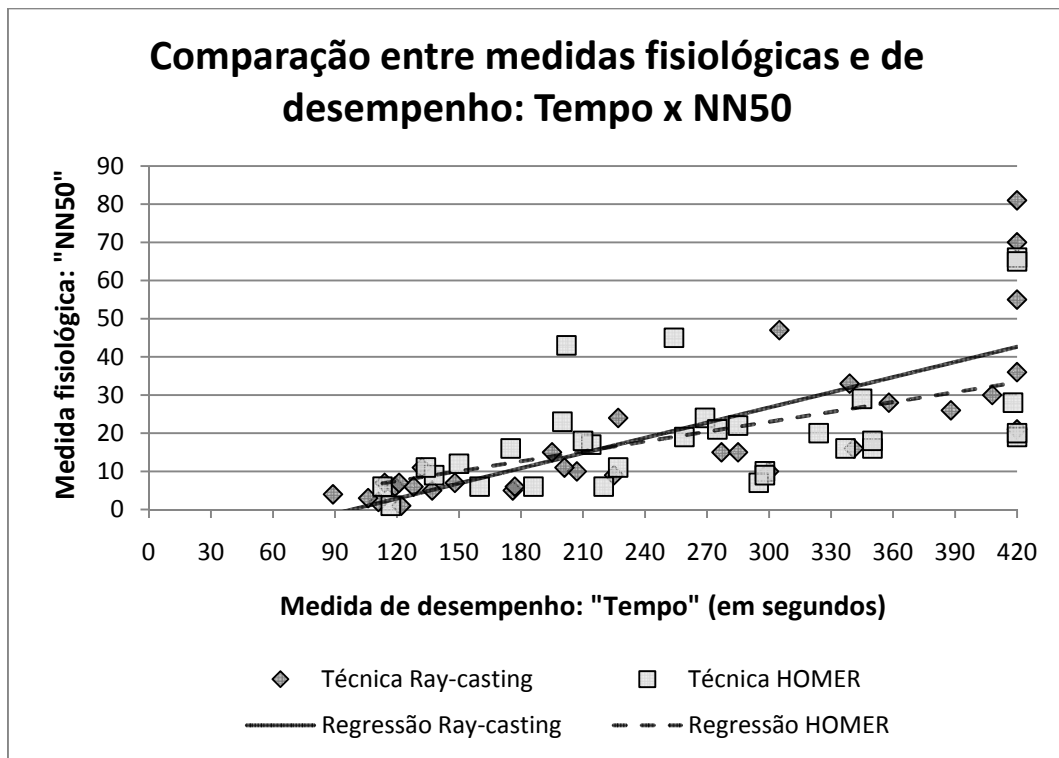


Figura 56. Correlação entre as variáveis “tempo” e “NN50”, e respectivas tendências.

Já na comparação entre medidas fisiológicas e de avaliação subjetiva, algumas medidas apresentaram resultados estatisticamente significantes, como a SC, a NN50 e a LF/HF – porém, nenhuma delas apresentou poder explicativo relevante, ou se aproximou de um nível aceitável. Esta baixa relação pode ter origem na configuração dos questionários pré- e pós-teste, os quais estavam sendo utilizados pela primeira vez. Da mesma forma, as medidas fisiológicas e as questões subjetivas empregadas podem não ter avaliado os mesmos construtos²⁸, o que acabou gerando esta baixa relação entre os resultados.

Para tanto, foram realizados alguns testes estatísticos para verificar a confiabilidade dos questionários pré- e pós-teste. Primeiramente, foram aplicados os testes alfa de Cronbach [CRO51] e método das metades [GUT45], cujos resultados são apresentados pela Tabela 17. De acordo com Hair *et al* [HAI09], os coeficientes de confiabilidade que indicam um bom instrumento de avaliação devem ser superiores a 0,7 – fato não observado nesta análise. Uma das possíveis causas para este resultado é o tamanho da amostra utilizada (54 sujeitos), que ficou abaixo da média recomendada pela literatura, que considera a dimensão da escala utilizada no instrumento. Como a

²⁸ Na Estatística, construto é um conceito teórico que atua como “bloco de construção”, usado para definir relações. De acordo com Hair *et al* [HAI09], ele pode representar um conceito tão simples, como idade, renda, sexo, ou tão complexo quanto *status* econômico, conhecimento, preferência ou atitude.

escala variava de 1 a 7, o ideal seria aplicar os questionários a um grupo de, no mínimo, 70 pessoas, dando margem à obtenção de resultados mais confiáveis.

Tabela 17. Resultado dos testes de confiabilidade dos questionários propostos.

QUESTIONÁRIOS	ALFA DE CRONBACH	MÉTODO DAS METADES	ITENS
Pré-Teste	0,50	0,10	5
Pós-Teste	0,35	0,48	7

Uma análise fatorial também foi aplicada aos questionários para verificar suas relevâncias e determinar seus construtos. De acordo com Malhotra [MAL06], recomenda-se para este tipo de análise a escolha do maior número possível de fatores (gerados de acordo com o número de questões) que correspondam a, no mínimo, 60% da variância total. Como mostra a Tabela 18, para o questionário pré-teste, 63,54% do total da variância observada pode ser explicado por apenas dois fatores, considerando autovalores²⁹ totais maiores que 1. Já para o questionário pós-teste, a Tabela 19 mostra que 67,60% do total da variância pode ser explicado por três fatores, porém apenas dois deles possuem autovalores maiores que 1. Este resultado também é corroborado pelos baixos índices KMO de cada questionário, os quais avaliam a adequação da análise fatorial.

Para tanto, recomenda-se revisar a escala utilizada por estes questionários, antes de uma nova aplicação. É provável que, a partir desta validação, seja possível realizar uma análise mais apurada das relações entre medidas de avaliação subjetiva e medidas fisiológicas.

Tabela 18. Resultado da análise fatorial para o questionário pré-teste.

FATORES	AUTOVALORES INICIAIS		
	TOTAL	% DA VARIÂNCIA	% VAR ACUMULADA
1	1,92	38,30	38,30
2	1,26	25,24	63,54
3	0,92	18,48	82,02
4	0,60	12,05	94,07
5	0,30	5,93	100,00

Medida KMO = 0,51

²⁹ De acordo com Hair *et al* [HAI09], o critério de autovalores representa a quantia de variância explicada por um fator, o qual, para ser relevante, deve ser maior ou igual a 1.

Tabela 19. Resultado da análise fatorial para o questionário pós-teste.

FATORES	AUTOVALORES INICIAIS		
	TOTAL	% DA VARIÂNCIA	% VAR ACUMULADA
1	2,24	32,04	32,04
2	1,50	21,36	53,40
3	0,99	14,20	67,60
4	0,73	10,46	78,06
5	0,70	10,05	88,11
6	0,49	6,93	95,04
7	0,35	4,96	100,00

Medida KMO = 0,60

7.2 Medidas fisiológicas como métrica de usabilidade

Conforme apresentado nas Seções 6.7, 6.8, 6.9 e 6.10, as medidas fisiológicas apresentam resultados satisfatórios na avaliação de técnicas de interação, períodos de experimento e grupos de desempenho. Por outro lado, ainda necessitam de um estudo mais aprofundado em relação às etapas do processo interativo.

No caso da comparação entre técnicas de interação, apesar dos resultados da Seção 6.7 não indicarem diferenças significantes entre as técnicas, a análise traduz o panorama obtido na comparação utilizando somente medidas de desempenho, abordadas pela Seção 6.3. Esta análise também mostrou um equilíbrio entre as técnicas, no qual a técnica *ray-casting* leva vantagem em relação ao tempo de execução da tarefa, a técnica HOMER em relação ao número de anexações e colisões no AV, e um desempenho similar em relação a medida de precisão. Neste caso, pode-se dizer que as medidas fisiológicas refletiram, em nível geral, o desempenho das técnicas. Isto dá margem para o uso de medidas fisiológicas num processo inicial de avaliação, como critério para tomada de decisão, e o uso de medidas de desempenho na tarefa em avaliações posteriores, para estudar em mais detalhes a técnica ou o processo interativo.

Com relação aos períodos do experimento, apresentados pela Seção 6.8, as medidas fisiológicas apresentaram resultados satisfatórios na distinção entre as etapas previstas em protocolo. Primeiramente, as comparações utilizando a medida de HR mostram que os experimentos exigem certo esforço físico do usuário para desempenhar as tarefas interativas, e que este esforço não se altera entre os experimentos, conforme mostra a Figura 57. Estes fatos comprovam a importância dos tempos estabelecidos em protocolo para os períodos basal e de descanso entre os testes com as técnicas.

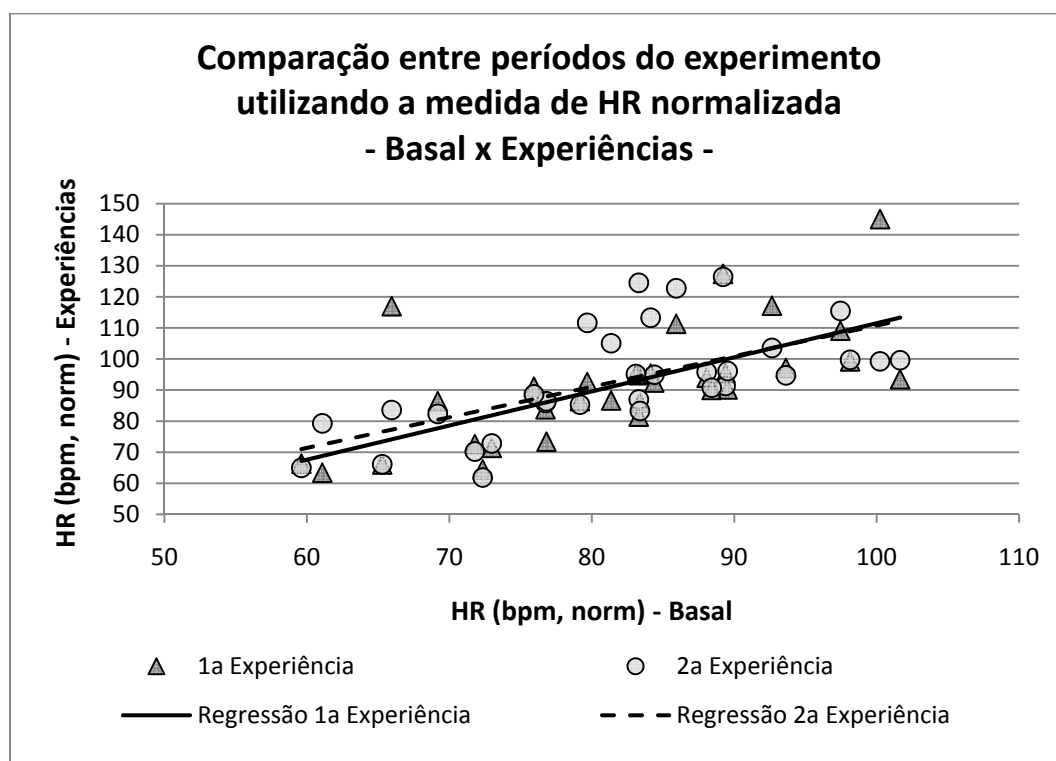


Figura 57. Correlação entre o período basal e as experiências, considerando a medida de HR.

Referente à medida de SC, esta mostra que o nível de ansiedade diminui significativamente entre os experimentos, comprovando que o usuário se acostuma aos recursos e funcionalidades à medida que interage no AV. Logo, a SC pode ser utilizada como métrica indicativa do nível de adaptação do usuário à interface 3D. A Figura 58 apresenta esta tendência.

Por outro lado, a medida de SC não identificou claramente mudanças no nível de ansiedade entre os períodos basal e dos experimentos. Uma das razões para a ocorrência deste fato pode ter relação com as características do grupo avaliado, o qual era formado por sujeitos que estavam participando pela primeira vez de experiências com AV e dispositivos de RV. Para avaliar este aspecto, sugere-se a realização de análises que considerem a classificação de grupos de acordo com a experiência, e a inserção de um período para coleta basal sem o uso dos dispositivos de RV, para verificar o grau de influência do uso destes equipamentos no grau de ansiedade do sujeito. Outras sugestões são o emprego de avaliações que considerem os períodos do dia ou horários de teste, e a classificação por habilidade física ou cognitiva.

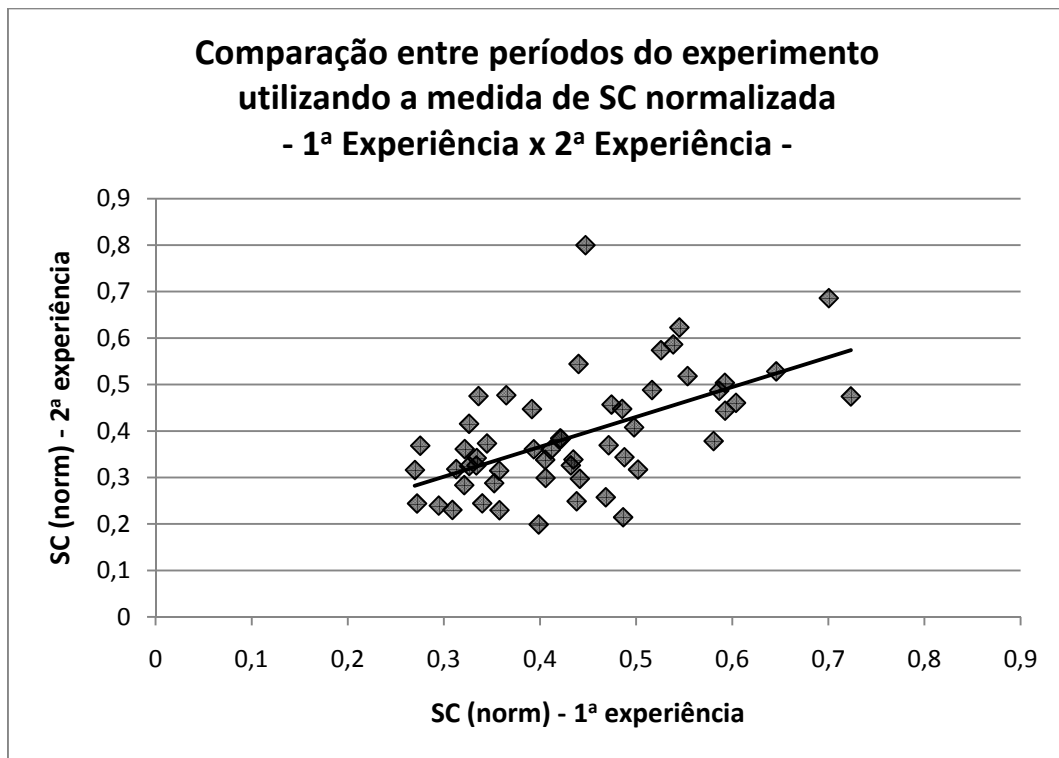


Figura 58. Correlação entre os experimentos, considerando a medida de SC.

Na comparação entre grupos de desempenho, apresentada pela Seção 6.10, a medida NN50 novamente mostrou-se relevante, indicando sua importância como uma métrica associada ao desempenho das técnicas, já observado pela Seção 7.1. Esta análise comprova a necessidade de um estudo mais detalhado sobre seu comportamento na avaliação da qualidade de técnicas de interação em interfaces 3D.

Em relação à comparação entre medidas fisiológicas e etapas do processo interativo, não foi possível identificar quais momentos deste processo demandaram maior esforço físico ou acarretaram maiores alterações no nível de ansiedade do usuário. Assim, também não foi possível associar eventuais alterações fisiológicas a problemas ou falhas na aplicação, indicados pelos usuários no questionário pós-teste. Provavelmente, isto ocorreu devido às próprias características dos AVs utilizados, as quais não foram suficientemente estressantes a ponto de deixar os sujeitos irritados por constantes colisões ou ansiosos em completar a tarefa no tempo estabelecido.

Com este panorama, sugere-se que o estudo seja refeito, considerando o uso de AVs capazes de gerar um nível maior de ansiedade e esforço físico ao usuário. Da mesma forma, recomenda-se decompor o processo interativo em um nível maior de detalhe, e identificar na literatura estudos sobre o comportamento das medidas fisiológicas em eventos de frustração ou satisfação, o que poderá contribuir para uma avaliação mais adequada das etapas do processo interativo.

7.3 Normalização da Medida de Frequência Cardíaca

Apesar dos resultados interessantes obtidos na comparação de medidas fisiológicas com outras medidas de avaliação de usabilidade, o processo de análise da medida de HR, e seus parâmetros de HRV, ficou prejudicado, em parte, pela necessidade de normalizar os dados coletados. Para as comparações que envolveram medidas oriundas da HR a amostra foi reduzida de 54 para 32 sujeitos, em face da necessidade de eliminar os dados dos sujeitos que apresentaram problemas na leitura do sinal, com base em uma metodologia de normalização apresentada pela Seção 5.2.

De acordo com Combatalade [COM10], apesar da precaução tomada em relação à preparação da pele, aplicação do gel condutor, colocação dos eletrodos e instruções ao usuário, é muito difícil gravar dados de HR absolutamente livres de ruídos. Isto torna o processo de normalização do sinal de HR essencial, principalmente na detecção de dois tipos de artefatos³⁰: batimentos faltantes ou batimentos extras.

Batimentos faltantes podem ocorrer quando o sinal fica distorcido a ponto de impossibilitar ao algoritmo de processamento de sinal a identificação do padrão de batimentos, resultando em longos intervalos entre batimentos. Por outro lado, batimentos extras ocorrem quando o programa analisa uma distorção de sinal e detecta dois padrões de batimentos quando deveria existir apenas um, resultando em resposta anormal, com intervalos muito curtos entre batimentos. Graficamente, estes artefatos podem ser observados na Figura 59 que apresenta o registro de intervalos entre batimentos.

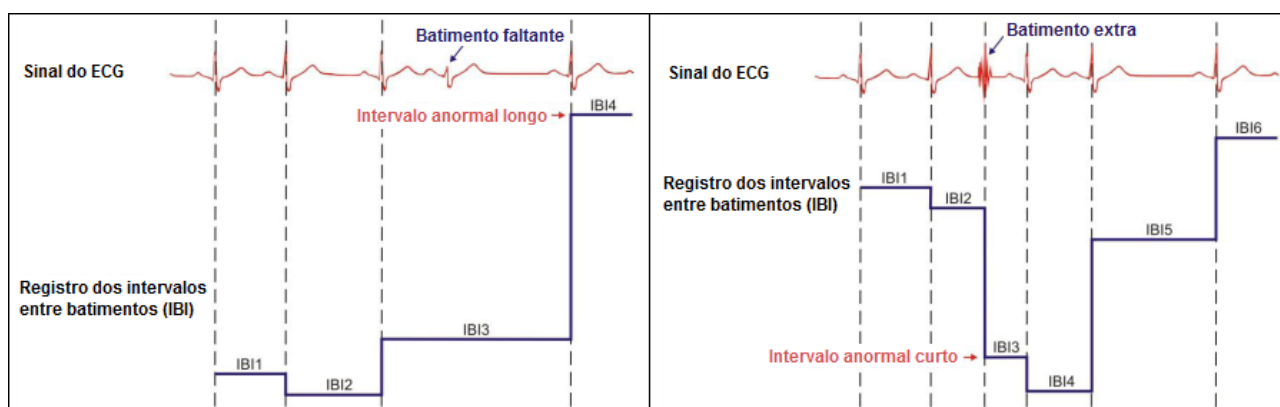


Figura 59. Visualização de sinais de ECG com registro de batimento faltante (à esquerda) e batimento extra (adaptado do manual CardioPro Infiniti [THO10]).

³⁰ Artefatos são informações indesejáveis contidas em um sinal fisiológico, geralmente provenientes de outras fontes. Eles podem ocultar ou distorcer o verdadeiro sinal, impossibilitando a realização de análises significativas.

Uma forma de eliminar estes artefatos é utilizar um *software* especializado na análise da HRV. Neste caso, é recomendável também o acompanhamento ou auxílio de um profissional da área médica, para garantir que a limpeza dos dados não interfira nos resultados posteriores.

É importante mencionar também que existem eventos fisiológicos naturais que se assemelham a estes artefatos, como contrações atriais ou ventriculares prematuras. Em função disto, recomenda-se seguir um protocolo adequado no caso de presença de pessoas com problemas cardíacos e/ou que estejam sob medicação.

Outra sugestão para minimizar a ocorrência de ruídos é a utilização de pulseiras auto-ajustáveis ou de velcro, para evitar que os eletrodos fiquem apertados ou afrouxados nos punhos do usuário. Pode-se ainda utilizar um dispositivo não invasivo que fique preso no peito do usuário, mais próximo ao coração, e que transmita a HR para um sensor de ECG sem o uso de fios, como mostra o exemplo da Figura 60.



Figura 60. Dispositivo³¹, em formato de cinto, que transmite os sinais da HR para um sensor de ECG.

Por fim, recomenda-se também o trabalho conjunto entre especialistas das áreas de Computação e Medicina, visando traçar diretrizes que permitam associar, de forma mais clara, o comportamento do organismo e as ações desempenhadas nas interfaces 3D, considerando o uso das medidas fisiológicas.

7.4 Custos e Benefícios do Uso do Método

A abordagem apresentada por este trabalho pode ser avaliada como custosa devido ao processo de aprendizado envolvido para interpretação e análise de uma medida fisiológica. Para entendê-la, é preciso compreender bem as diferentes teorias, fórmulas e conceitos técnicos relacionados às áreas de Medicina, Psicologia e Estatística.

³¹ <http://www.thoughttechnology.com/polar.htm>

Para tanto, um período de estudos deve ser considerado pelo projeto, preferencialmente com a ajuda de profissionais destas áreas.

O uso de medidas fisiológicas também exige um investimento inicial para aquisição de equipamentos, *software* e produtos que permitam sua utilização em testes de avaliação. Este custo pode variar, pois é dependente do número de medidas fisiológicas a serem coletadas, e dos tipos de dispositivos e programas para leitura e interpretação dos dados. Dependendo das expectativas e da previsão do projeto, este investimento pode valer a pena, pois permitirá o uso da aparelhagem em diferentes testes, fornecerá resultados que complementam ou apresentam novas tendências aos processos avaliativos, e poderá ser empregado em novos projetos.

Para efeitos de estimativa de custos, a Tabela 20 apresenta o investimento realizado para o estudo de caso deste trabalho, o qual utilizou soluções de *hardware* e *software* de uma mesma empresa (Thought Technology). Neste caso, foram utilizados sensores para captura de dois sinais fisiológicos (SC e ECG), uma unidade de leitura e conversão de sinais e um programa para visualização e análise dos dados.

Tabela 20. Custos para uso de medidas fisiológicas em testes de avaliação de usabilidade.

EQUIPAMENTOS	CUSTO	Taxa de CONVERSÃO***	TOTAL
Sensor de ECG com pulseiras*	US\$ 310,00	R\$ 1,80	R\$ 558,00
Sensor de SC*	US\$ 275,00	R\$ 1,80	R\$ 495,00
Procomp5 Infiniti – sistema de 5 canais*	US\$ 2.720,00	R\$ 1,80	R\$ 4.896,00
Procomp Infiniti Physiology Suite v5.0*	US\$ 275,00	R\$ 1,80	R\$ 495,00
Acessórios**	R\$ 50,00	–	R\$ 50,00
Despesas de Importação	R\$ 2.259,82	–	R\$ 2.259,82
TOTAL			R\$ 8.753,82

* Fonte: <http://bio-medical.com/>

** Gel condutor, pilhas alcalinas, lenços umedecidos, álcool gel, papel toalha e flanelas

*** Cotação do dólar em Novembro/2010

Outros custos associados ao uso deste método envolvem a definição das variáveis a serem analisadas e o tempo de protocolo gasto para executar todas as etapas do experimento. Geralmente, avaliações de usabilidade de interfaces 3D já abrangem a comparação entre múltiplas variáveis, fato que pode agravar-se com a inclusão de medidas fisiológicas. Para tanto, o uso dos métodos estatísticos sugeridos por este trabalho visa auxiliar na avaliação das técnicas de interação, reduzindo a complexidade do processo analítico. Já o tempo de protocolo, apesar de dificultar a participação de um

número maior e mais variado de sujeitos para os testes, possibilita estudar detalhadamente o comportamento das medidas fisiológicas de um grupo em sucessivas avaliações, o que contribui para o aperfeiçoamento da interface, associando a evolução de desempenho às medidas fisiológicas. Isto também pode resultar em maior satisfação do usuário, com melhor aceitação da interface, maior produtividade e um número reduzido de erros.

Uma sugestão para facilitar a interpretação e análise de uma medida fisiológica é automatizar o processo de aplicação da metodologia. Por intermédio de ferramentas computacionais, pode-se gerar uma biblioteca contendo os recursos utilizados pela metodologia para normalizar dados fisiológicos e empregar diferentes técnicas estatísticas sobre todo o conjunto de dados. Além disso, é possível agregar à biblioteca diferentes formatos de visualização dos resultados, como gráficos e relatórios pré-definidos.

7.5 Vantagens do Método

Com relação à aplicação do novo método de avaliação proposto por este trabalho, cabe destacar a definição de uma metodologia estatística para análise dos dados fisiológicos, principalmente na comparação entre medidas fisiológicas e métricas comumente utilizadas na avaliação de interfaces 3D, como tempo, número de tentativas, precisão e os índices subjetivos de satisfação presentes em questionários. Esta série de técnicas estatísticas, estabelecidas pelas Seções 4.5.4, 4.5.5, 5.5, 6.5 e 6.6, permitiu a compreensão e o estudo detalhado das relações entre as medidas estudadas, fomentando a elaboração de conclusões refinadas e a indicação de tendências para futuros estudos na área. Como exemplo, pode-se citar o uso da regressão múltipla para comparar e combinar diferentes variáveis dependentes e independentes na predição de resultados.

Da mesma forma, a definição de um formato padrão de protocolo possibilitou a coleta de dados fisiológicos sem que isto interferisse no desempenho do usuário durante o processo interativo. Conforme os comentários dos participantes, apresentados na Seção 4.5.3, apenas foram citados sugestões de melhorias ligadas ao aperfeiçoamento do AV e a adequação dos dispositivos de RV.

Para fortalecer o uso desta metodologia e das medidas fisiológicas como métrica de usabilidade, é recomendável a realização de novas comparações com outras técnicas de interação 3D, tanto para tarefas de seleção, manipulação ou navegação. Da mesma forma, ainda cabe avaliar outras características inerentes ao processo interativo utilizando esta abordagem, como o impacto do uso de visualização em estéreo, dos graus de

liberdade de movimentos oferecidos pelas técnicas, e de diferentes estímulos sensoriais na execução de tarefas.

A seguir, a Tabela 21 apresenta novamente o quadro comparativo entre as pesquisas que usam medidas fisiológicas para avaliação de interfaces 3D, mostrado pela Tabela 3, com a inclusão da abordagem proposta por este trabalho. Como pode ser visto, esta pesquisa manteve a maioria das características utilizadas e fundamentadas em trabalhos anteriores, porém com foco voltado a avaliação da qualidade de interação em interfaces 3D.

Tabela 21. Comparativo entre trabalhos relacionados e as características desta pesquisa.

Características	MEEHAN	SLATER	BROGNI	LIN	KIM	RIEDER
<i>Interface</i>	3D	3D	3D	2D	3D	3D
<i>Ambiente Virtual</i>	Imersivo	Imersivo	Imersivo	–	Imersivo	Imersivo
<i>Aplicação</i>	Simulação	Simulação	Simulação	Jogo	Simulação	Simulação
<i>Visualização</i>	Estéreo	Estéreo	Estéreo	Mono	Estéreo	Mono
<i>Dispositivos</i>	HMD, rastreador, mouse e tablado	CAVE e shutter	CAVE, shutter e joystick	Monitor e mouse	CAVE, shutter e joystick	HMD, rastreador e mouse
<i>Tarefas de Interação</i>	Seleção, Manipulação e Navegação	Navegação	Navegação	–	Navegação	Seleção e Manipulação
<i>Estímulos explorados</i>	Visual e Tátil	Visual e Sonoro	Visual e Sonoro	Visual	Visual e Sonoro	Visual e Sonoro
<i>Medidas Fisiológicas utilizadas</i>	HR, SC e ST	HR, HRV e SC	HRV	HR, HRV, SC e BVP	ECG, EOG, SC, BVP, ST, EGG, BF e EEG	HR, HRV e SC
<i>Nº de experimentos</i>	04	01	01	02	01	01
<i>Nº de usuários e grupos</i>	10; 01 26; 02 33; 04 64; 02	20; 01	40; 04	18; 01 10; 01	47; 01	54; 02
<i>Medidas objetivas (desempenho)</i>	–	–	Tempo	Tempo, número de acertos, rastreamento do olhar, movimento das mãos, tamanho da pupila	–	Tempo, Precisão, Anexações e Colisões
<i>Medidas subjetivas (questionários)</i>	<i>Pós-teste:</i> presença, comportamento e locomoção	<i>Pré-teste:</i> demográfico, experiência e ansiedade social <i>Pós-teste:</i> ações no AV	<i>Pré-teste:</i> demográfico, experiência e treinamento de presença <i>Pós-teste:</i> presença e inventário de presença (ITC-SOPI)	<i>Pré-teste:</i> demográfico e experiência em jogos <i>Pós-teste:</i> dificuldade na tarefa, nível de estresse e carga de trabalho (NASA-TLX)	<i>Pré-teste:</i> tendência imersiva e de mal-estar <i>Pós-teste:</i> presença e simulador de desconforto	<i>Pré-teste:</i> demográfico e experiência (questionário próprio) <i>Pós-teste:</i> Avaliação subjetiva (questionário próprio)
<i>Protocolo</i>	Treinamento e Experimento	Basal, Treinamento e Experimento	Basal, Treinamento e Experimento	Basal e Experimento	Basal, Experimento e Basal	Basal, Treinamento e Experimento
<i>Foco do trabalho</i>	Medir presença sob estresse	Medir interrupções de presença	Medir o impacto do realismo visual	Medir a usabilidade de jogos de computador	Prever o desconforto em AVs	Avaliar a qualidade de interação utilizando uma nova metodologia
<i>Medidas Fisiológicas significantes</i>	HR e SC	HR, HRV e SC	HRV	HRV e SC	Taquiarritmia gástrica no EGG	HR, SC e HRV (NN50)
<i>Referências</i>	[MEE02] [MEE05]	[SLA06]	[BRO06] [BRO07]	[LIN05] [LIN06a] [LIN06b] [LIN08]	[KIM08]	

8. CONCLUSÕES

Este trabalho definiu uma metodologia para avaliar a qualidade de técnicas de interação em interfaces 3D, tendo por base o emprego de medidas fisiológicas durante o processo avaliativo. Um estudo de caso foi realizado, como forma de certificar o uso deste método, que engloba a comparação entre duas técnicas de interação utilizando um protocolo padronizado de teste, um procedimento de normalização de dados fisiológicos, e o uso de técnicas estatísticas para análise detalhada dos relacionamentos existentes entre métricas de usabilidade e medidas fisiológicas.

Os resultados alcançados apontam para tendências promissoras. A expectativa é que esta metodologia evolua e seja fortalecida com a realização de novas comparações com outras técnicas de interação 3D, tanto para tarefas de seleção, manipulação ou navegação, e também considerando outras características de interfaces 3D, como estereoscopia, diferentes graus de liberdade e exploração de múltiplos estímulos sensoriais.

Medidas fisiológicas ainda não podem ser consideradas como substitutas de medidas de desempenho e de avaliação subjetiva. Porém, são medidas que podem ser utilizadas como recurso complementar na interpretação de métricas de usabilidade comumente utilizadas em processos avaliativos, com destaque para a medida NN50, a qual apresentou uma boa relação com medidas de tempo de execução das tarefas. Os resultados obtidos neste estudo foram aceitos para publicação no *13th IFIP TC13 Conference on Human-Computer Interaction (INTERACT 2011)*, os quais podem ser visualizados no Apêndice I deste documento. Como trabalho futuro, recomenda-se uma investigação mais detalhada da medida NN50, visando contribuir de forma objetiva ao processo de avaliação de técnicas de interação em interfaces 3D.

A pesquisa também apresenta um formato de validação dos questionários propostos, como forma de estabelecer melhores relacionamentos entre medidas fisiológicas e métricas de preferência do usuário. Esta análise apontou a importância em adequar à escala de avaliação subjetiva antes dos testes, conforme o tamanho da amostra de sujeitos disponível – fato que pode ter contribuído para que medidas com diferenças estatisticamente significativas, como a SC e a NN50, terem apresentado baixo poder explicativo neste trabalho.

Para a comparação entre métricas de usabilidade e medidas fisiológicas, cabe destacar a importância do protocolo de testes e da metodologia estatística utilizada no processo avaliativo. Estes instrumentos possibilitaram a compreensão e o estudo

detalhado das relações entre as diferentes medidas estudadas, fomentando a elaboração de conclusões mais refinadas e a indicação de tendências para futuros estudos na área. Um conjunto de técnicas estatísticas, capazes de analisar múltiplas variáveis, é sugerido pelo trabalho com o objetivo de avaliar características distintas de interfaces 3D considerando diferentes relacionamentos entre medidas de usabilidade.

As medidas fisiológicas como métrica de usabilidade também apresentaram resultados interessantes. Na comparação entre experimentos, as medidas de HR e SC foram capazes de comprovar a importância dos tempos estabelecidos no protocolo de testes para a coleta basal e o descanso entre as experiências, indicaram o nível de esforço físico exigido para execução de tarefas, e mostraram que os usuários se acostumam com a interação em um AV, independente do seu nível de experiência. Já na comparação entre técnicas de interação, as medidas de HR e SC refletiram, em nível geral, os mesmos resultados alcançados pela análise entre medidas de desempenho, indicando que medidas fisiológicas podem ser úteis para decidir quais técnicas de interação melhor se adaptam a uma determinada interface 3D.

No entanto, é prudente aplicar novas avaliações para formalizar o uso de medidas fisiológicas como métrica de usabilidade. Exemplos de trabalhos futuros são análises que considerem a classificação de grupos de acordo com a experiência, a verificação do grau de influência dos dispositivos de RV, o emprego de avaliações que considerem os períodos do dia ou horários de teste, e a classificação por habilidade física ou cognitiva.

Em relação à comparação entre medidas fisiológicas e etapas do processo interativo, não foi possível identificar quais momentos deste processo demandaram maior esforço físico ou acarretaram alterações no nível de ansiedade do usuário. Para tanto, recomenda-se um novo estudo, considerando a decomposição do processo interativo em um nível maior de detalhe.

Por fim, cabe destacar que este projeto é um primeiro passo para a definição de um processo avaliativo, específico para interfaces 3D, que considera o uso de medidas fisiológicas. Para tanto, é importante que um trabalho em conjunto com a área médica seja estabelecido, visando formalizar diretrizes que permitam associar, de forma mais clara, o comportamento fisiológico e as ações desempenhadas em interfaces 3D.

REFERÊNCIAS

- [ADI10] ADINSTRUMENTS. "ECG, EKG - Research". Data Acquisition & Analysis for Life Science - ADInstruments, 2010. Capturado em: <http://www.adinstruments.com/solutions/research/Psychophysiology/ECG,-EKG/>, Maio 2010.
- [AME05] AMERICAN COLLEGE OF SPORTS MEDICINE. "ACSM's Advanced Exercise Physiology". 1. ed., Philadelphia: Lippincott Williams & Wilkins, 2005, 704p.
- [AND07] ANDREASSI, J. L. "Psychophysiology: human behavior and physiological response". 5. ed., Mahwah: LEA, 2007, 538p.
- [ASE04] ASENDORPF, J. "Psychologie der Persönlichkeit". Berlin: Springer, 2004, 508p.
- [ASL10] ASL. "Applied Science Laboratories - Eye Tracking Expertise". Capturado em: <http://www.asleyetracking.com/>, Novembro 2010.
- [AUB03] AUBERT, A. E.; SEPS, B.; BECKERS, F. "Heart Rate Variability in Athletes". *Sports Medicine*, vol. 33-12, 2003, pp. 889-919.
- [AYR07] AYRES, M. *et al.* "BioEstat 5.0: aplicações estatísticas nas áreas de ciências biológicas e médicas", Sociedade Civil Mamirauá, Belém, 2007, 364p.
- [BAR94] BARNETT, V.; LEWIS, T. "Outliers in Statistical Data". New York: Wiley, 1994.
- [BIO09a] BIO-MEDICAL.COM. "Skin Conductance Sensor: how it works". Bio-Medical.com. Capturado em: http://www.bio-medical.com/product_info.cfm?inventory__imodel=SA9309M, Maio 2009.
- [BIO09b] BIO-MEDICAL.COM. "Heart Rate/ BVP Sensor: how it works". Bio-Medical.com. Capturado em: http://www.bio-medical.com/product_info.cfm?inventory__imodel=SA9308M, Maio 2009.
- [BIS02] BISBAL, J. *et al.* "Legacy Information Systems: Issues and directions". *IEEE Software*, vol. 16-5, 2002, pp. 103-111.
- [BOL80] BOLT, R. A. "'Put-That-There': Voice and Gesture at the Graphics Interface". In: SIGGRAPH '80: Proceedings of the 7th annual conference on Computer graphics and interactive techniques, 1980, pp. 262-270.
- [BOU08] BOUCHARD, S. *et al.* "Anxiety Increases the Feeling of Presence in Virtual Reality". *Presence: Teleoperators and Virtual Environments*, vol. 17-4, Agosto

2008, pp. 376-391.

- [BOW02] BOWMAN, D. A.; GABBARD, J. L.; HIX, D. "A Survey of Usability Evaluation in Virtual Environments: classification and comparison of methods". *Presence*, vol. 11-4, 2002, pp. 404-424.
- [BOW04] BOWMAN, D. A. *et al.* "3D User Interfaces: theory and practice". Boston: Addison-Wesley, 2004, 512p.
- [BOW97] BOWMAN, D. A.; HODGES, L. F. "An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments". In: *I3D '97: Proceedings of the 1997 symposium on Interactive 3D graphics*, 1997, pp. 35-38.
- [BRO06] BROGNI, A. *et al.* "Variations in Physiological Responses of Participants During Different Stages of an Immersive Virtual Environment Experiment". In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 2006, pp. 376-382.
- [BRO07] BROGNI, A. *et al.* "Responses of Participants During an Immersive Virtual Environment Experience". *The International Journal of Virtual Reality*, vol. 6-2, 2007, pp. 1-10.
- [BUR03] BURDEA, G. C.; COIFFET, P. "Virtual Reality Technology". 2. ed., Hobokon: Wiley-Interscience, 2003, 464p.
- [BUS08] BUSH, J. "Viability of Virtual Reality Exposure Therapy as a Treatment". *Computers in Human Behavior*, vol. 24-3, May 2008, pp. 1032-1040.
- [CLA84] CLARK, R. P. "Human Skin Temperature and Its Relevance in Physiology and Clinical Assessment". *Recent Advances in Medical Thermology*, New York, 1984, pp. 5-15.
- [COM10] COMBATALADE, D. "Basics of Heart Rate Variability Applied to Psychophysiology", MAR953-00, Thought Technology Ltd., Montreal, 2010, 31p.
- [CON95] CONESA, J. "Electrodermal palmar asymmetry and nostril dominance". *Perceptual and Motor Skills*, vol. 80, 1995, pp. 211-216.
- [COR06] EDA Sensor: Designing with microcontrollers, Final Projects. Electrodermal Activity Sensor, 2006. Capturado em: http://courses.cit.cornell.edu/ee476/FinalProjects/s2006/hmm32_pjw32/index.html, Maio 2009.

- [CRO51] CRONBACH, L. J. "Coefficient alpha and the internal structure of tests". *Psychometrika*, vol. 16-3, 1951, pp. 297-334.
- [DAW90] DAWSON, M. E.; SCHELL, A. M.; FILION, D. L. "The Electrodermal System". In: CACIOPPO, J. T.; TASSINARY, L. G. *Principles of Psychophysiology: physical, social and inferential elements*. Cambridge: Cambridge University Press, 1990. Cap. 10, pp. 295-324.
- [DUC07] DUCHOWSKI, A. T. "Eye Tracking Methodology: Theory and Practice". 2. ed., New York: Springer-Verlag, 2007, 334p.
- [EAS97] EASTON, V. J.; MCCOLL, J. H. "STEPS - Statistics Glossary", 1997. Capturado em: <http://www.stats.gla.ac.uk/steps/glossary/>, Novembro 2010.
- [EMM02] EMMELKAMP, P. M. G. *et al.* "Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia". *Behaviour Research and Therapy*, vol. 40-5, 2002, pp. 509-516.
- [FER03] FERNANDES FILHO, J. "A Prática da Avaliação Física". 2. ed., Rio de Janeiro: Shape, 2003, 216p.
- [GOD05] GODOY, M. F.; TAKAKURA, I. T.; CORREA, P. R. "Relevância da análise do comportamento dinâmico não-linear (Teoria do Caos) como elemento prognóstico de morbidade e mortalidade em pacientes submetidos à cirurgia de revascularização miocárdica". *Arquivos de Ciência e Saúde*, vol. 12-4, 2005, pp. 167-171.
- [GOL98] GOLDING, J. F. "Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness". *Brain Research Bulletin*, vol. 47-5, Novembro 1998, pp. 507-516.
- [GUT45] GUTTMAN, L. "A Basis for Analyzing Test-Retest Reliability". *Psychometrika*, vol. 10-4, 1945, pp. 255-282.
- [GUY05] GUYTON, A. C.; HALL, J. E. "Textbook of Medical Physiology". Philadelphia: Saunders, vol. 11, 2005, 1104p.
- [HAI09] HAIR JR., J. F. *et al.* "Análise Multivariada de Dados". 6. ed., Porto Alegre: Bookman, 2009, 688p.
- [HAR06] HART, S. G. "NASA-Task Load index (NASA-TLX); 20 years later". In: *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 2006, pp. 904-908.
- [HAR88] HART, S. G.; STAVELAND, L. E. "Development of NASA-TLX (Task Load

Index): Results of empirical and theoretical research". *Human mental workload*, Amsterdam, vol. 1, 1988, pp. 139-183.

- [HES60] HESS, E. H.; POLT, J. M. "Pupil Size as Related to Interest Value of Visual Stimuli". *Science*, vol. 132, 1960, pp. 349-350.
- [HES64] HESS, E. H.; POLT, J. M. "Pupil Size in Relation to Mental Activity During Simple Problem-Solving". *Science*, vol. 143-3611, 1964, pp. 1190-1192.
- [HIX93] HIX, D.; HARTSON, H. "Developing User Interfaces: ensuring usability through product & process". New Jersey: John Wiley & Sons, 1993, 416p.
- [HOD01] HODGES, L. F. *et al.* "Treating Psychological and Physical Disorders with VR". *IEEE Computer Graphics and Applications*, vol. 21-6, November 2001, pp. 25-33.
- [HOS00] HOSMER, D. W.; LEMESHOW, S. "Applied Logistic Regression". 2. ed., New York: Wiley-Interscience Publication, 2000, 392p.
- [KAR57] KARVONEN, M. J.; KENTALA, E.; MUSTALA, O. "The Effects of Training on Heart Rate: a longitudinal study". *Annales Medicinæ Experimentalis et Biologiae Fenniae*, vol. 35-3, 1957, pp. 307-315.
- [KEN93] KENNEDY, R. S. *et al.* "Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness". *International Journal of Aviation Psychology*, vol. 3-3, 1993, pp. 203-220.
- [KIM08] KIM, Y. Y. *et al.* "The Application of Biosignal Feedback for Reducing Cybersickness from Exposure to a Virtual Environment". *Presence: Teleoperators and Virtual Environments*, Massachusetts, vol. 17-1, February 2008, pp. 1-16.
- [KIR02] KIRTLEY, C. "BE513 Bioinstrumentation Projects". Courses - Clinical Gait Analysis - Universität Wien, 2002. Capturado em: <http://www.univie.ac.at/cga/courses/BE513/Projects/>, Maio 2010.
- [KOL33] KOLMOGOROV, A. N. "Sulla determinazione empirica di una legge di distribuzione". *Giornale dell' Istituto Italiano degli Attuari*, vol. 4, 1933, pp. 83-91.
- [KRI04a] KRIJN, M. *et al.* "Virtual Reality Exposure Therapy of Anxiety Disorders: a review". *Clinical Psychology Review*, vol. 24-3, July 2004, pp. 259-281.
- [KRI04b] KRIJN, M. *et al.* "Treatment of acrophobia in virtual reality: The role of immersion and presence". *Behaviour Research and Therapy*, vol. 42-2, 2004, pp. 229-239.

- [LES01] LESSITER, J. *et al.* "A cross-media presence questionnaire: The ITC-Sense of Presence Inventory". *Presence: Teleoperators and Virtual Environments*, vol. 10-3, 2001, pp. 282-297.
- [LIN04] LIN, Y.; ZHANG, W. J.; KOUBEK, R. J. "Effective Attention Allocation Behavior and its Measurement: a preliminary study". *Interacting with Computers*, vol. 16-6, 2004, pp. 1195-1210.
- [LIN05] LIN, T. *et al.* "Do physiological data relate to traditional usability indexes?" In: Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future, 2005, pp. 1-10.
- [LIN06a] LIN, T. *et al.* "An Empirical Study of Relationships Between Traditional Usability Indexes and Physiological Data". *Australasian Journal of Information Systems*, vol. 13-2, 2006, pp. 105-117.
- [LIN06b] LIN, T.; IMAMIYA, A. "Evaluating usability based on multimodal information: an empirical study". In: Proceedings of the 8th international conference on Multimodal interfaces, 2006, pp. 364-371.
- [LIN08] LIN, T.; IMAMIYA, A.; MAO, X. "Using multiple data sources to get closer insights into user cost and task performance". *Interacting with Computers*, vol. 20-3, Maio 2008, pp. 364-374.
- [LYK71] LYKKEN, D. T.; VENABLES, P. H. "Direct Measurement of Skin Conductance: a proposal for standardization". *Psychophysiology*, vol. 8-5, 1971, pp. 656-672.
- [MAD06] MADEIRO, J. P. D. V. *et al.* "Algoritmo para Análise de Variabilidade da Frequência Cardíaca no Domínio Temporal: Métodos Estatísticos e Geométricos". In: Congresso Brasileiro de Informática em Saúde, 2006, 6p.
- [MAL06] MALHOTRA, N. K. "Marketing Research: An Applied Orientation". 5. ed., [S.l.]: Pearson, 2006, 864p.
- [MAL95] MALIK, M.; CAMM, A. J. "Heart rate variability". London: Futura Publishing Company, 1995, 543p.
- [MAL96] MALIK, M. *et al.* "Heart Rate Variability: Standards of measurement, physiological interpretation, and clinical use". *European Heart Journal*, vol. 17-3, 1996, pp. 354-381.
- [McC50] MCCLEARY, R. A. "The nature of the galvanic skin response". *Psychological Bulletin*, vol. 47-2, Março 1950, pp. 97-117.
- [MEE02] MEEHAN, M. *et al.* "Physiological Measures of Presence in Stressful Virtual

Environments". In: SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 2002, pp. 645-652.

- [MEE05] MEEHAN, M. *et al.* "Review of Four Studies on the Use of Physiological Reaction as a Measure of Presence in Stressful Virtual Environments". *Applied Psychophysiology and Biofeedback*, vol. 30-3, Setembro 2005, pp. 239-258.
- [MIR10] MIRTAHERI, P. "A novel low-cost design of a digital infrared pupillometer using a web camera". In: The International Conference on Universal Technologies, 2010, pp. 82-90.
- [MON66] MONTAGU, J. D.; COLES, E. M. "Mechanism and measurement of the galvanic skin response". *Psychological Bulletin*, vol. 65-5, Maio 1966, pp. 261-279.
- [MOR05] MORETTIN, P. A.; BUSSAB, W. D. O. "Estatística Básica". São Paulo: Saraiva, vol. 5, 2005, 526p.
- [MÜH03] MÜHLBERGER, A.; WIEDEMANN, G.; PAULI, P. "Efficacy of a one-session virtual reality exposure treatment for fear of flying". *Psychotherapy Research*, vol. 13-3, 2003, pp. 323-336.
- [NIC03] NICOLAU, J. C.; POLANCZYK, C. A.; GUIMARÃES, J. I. "Diretriz de interpretação de eletrocardiograma de repouso". *Arquivo Brasileiro de Cardiologia*, vol. 80-2, 2003, pp. 1-18.
- [PAR08] PARSONS, T. D.; RIZZO, A. A. "Affective Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: a meta-analysis". *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 39-3, September 2008, pp. 250-261.
- [POW08] POWERS, M. B.; EMMELKAMP, P. M. G. "Virtual Reality Exposure Therapy for Anxiety Disorders: a meta-analysis". *Journal of Anxiety Disorders*, vol. 22-3, April 2008, pp. 561-569.
- [PRO03] THOUGH TECHNOLOGY LTD. "Procomp Infiniti Hardware Manual". Montreal: Though Technology Ltd., 2003, 48p.
- [PRO74] PROKASY, W. F.; RASKIN, D. C. "Electrodermal Activity in Psychological". New York: Academic Press Inc., 1974.
- [RES10] RESUSCITATION COUNCIL (UK). "Resuscitation Guidelines", Resuscitation Council (UK), London, 2010, 156p.
- [RIZ07] RIZZO, A. A. *et al.* "Virtual Iraq: initial case reports from a VR exposure therapy

application for combat-related post traumatic stress disorder". In: Proceedings of the IEEE 6th Virtual Rehabilitation, 2007, pp. 124-130.

- [ROS01] ROSSON, M.; CARROLL, J. "Usability Engineering: scenario-based development of human-computer interaction". San Francisco: Morgan Kaufmann Publishers, 2001, 448p.
- [ROT00] ROTHBAUM, B. O. *et al.* "A controlled study of virtual reality exposure therapy for the fear of flying". *Journal of Consulting and Clinical Psychology*, vol. 68-6, 2000, pp. 1020-1026.
- [ROT02] ROTHBAUM, B. O. *et al.* "Twelve-month follow-up of virtual reality and standard exposure therapies for the fear of flying". *Journal of Consulting and Clinical Psychology*, vol. 70-2, 2002, pp. 428-432.
- [SKI10] SKIPPING HEARTS. "Cardiac Quick Reference - Normal Heart Test Values". Arrhythmia Articles - Cardiac Tests, 2010. Capturado em: <http://www.skippinghearts.com/articles/normal-heart-test-values/>, Maio 2010.
- [SLA00] SLATER, M.; STEED, A. "A Virtual Presence Counter". *Presence: Teleoperators and Virtual Environments*, vol. 9-5, 2000, pp. 413-434.
- [SLA06] SLATER, M. *et al.* "Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment". *Presence: Teleoperators and Virtual Environments*, vol. 15-5, Outubro 2006, pp. 553-569.
- [SLA95] SLATER, M.; USOH, M.; STEED, A. "Taking steps: the influence of a walking technique on presence in virtual reality". *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 2-3, 1995, pp. 201-219.
- [SNE89] SNEDECOR, G. W.; COCHRAN, W. G. "Statistical Methods". 8. ed., Ames: Iowa State University Press, 1989, 503p.
- [STE94] STEIN, P. K. *et al.* "Heart Rate Variability: a measure of cardiac autonomic tone". *American Heart Journal*, vol. 127-5, 1994, pp. 1376-1381.
- [SWE93] SWEENEY, M.; MAGUIRE, M.; SHACKEL, B. "Evaluating User-Computer Interaction: a framework". *International Journal of Man-Machine Studies*, London, UK, vol. 38-4, April 1993, pp. 689-711.
- [TAN01] TANAKA, H.; MONAHAN, K. D.; SEALS, D. R. "Age-predicted Maximal Heart Rate Revisited". *Journal of the American College of Cardiology*, vol. 37-1, 2001, pp. 153-156.
- [THO10] THOUGHT TECHNOLOGY LTD. "CardioPro Infiniti: HRV Analysis Module for

BioGraph Infiniti", Thought Technology Ltd., Montreal, 2010, 36p.

- [TOR08] TORTORA, G. J.; DERRICKSON, B. H. "Principles of Anatomy and Physiology". 12. ed., Hoboken: Wiley, 2008, 1280p.
- [TUK77] TUKEY, J. W. "Exploratory Data Analysis". Reading: Addison-Wesley, 1977, 688p.
- [TUL08] TULLIS, T.; ALBERT, W. "Measuring the User Experience: collecting, analyzing, and presenting usability metrics". San Francisco: Morgan Kaufmann, 2008, 336p.
- [TYR39] TYRON, R. C. "Cluster Analysis". Ann Arbor: Edwards Brothers, 1939, 422p.
- [USO00] USOH, M. *et al.* "Using presence questionnaires in reality". *Presence: Teleoperators and Virtual Environments*, vol. 9-5, 2000, pp. 497-503.
- [USO99] USOH, M. *et al.* "Walking > walking-in-place > flying, in virtual environments". In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques, 1999, pp. 359-364.
- [VAN09] VANDERLEI, L. C. M. *et al.* "Noções Básicas de Variabilidade da Frequência Cardíaca e sua Aplicabilidade Clínica". *Revista Brasileira de Cirurgia Cardiovascular*, vol. 24-2, 2009, pp. 205-217.
- [VIN04] VINAYAGAMOORTHY, V. *et al.* "An Investigation of Presence Response Across Variations in Visual Realism". In: The 7th Annual International Presence Workshop, 2004, pp. 148-155.
- [WIE02a] WIEDERHOLD, B. K. "Physiological monitoring as an objective tool in virtual reality therapy". *CyberPsychology & Behavior*, vol. 5-1, Fevereiro 2002, pp. 77-82.
- [WIE02b] WIEDERHOLD, B. K. *et al.* "The Treatment of Fear of Flying: a controlled study of imaginal and virtual reality graded exposure therapy". *IEEE Transactions on Information Technology in Biomedicine*, vol. 6-3, September 2002, pp. 218-223.
- [WIL00] WILSON, G. M.; SASSE, M. A. "Do users always know what's good for them? Utilising physiological responses to assess media quality". In: Proceedings of HCI 2000 - People and Computers XIV - Usability or Else!, 2000, pp. 327-341.
- [WIL05] WILHELM, F. H. *et al.* "Mechanisms of Virtual Reality Exposure Therapy: The Role of the Behavioral Activation and Behavioral Inhibition Systems". *Applied Psychophysiology and Biofeedback*, vol. 30-3, Setembro 2005, pp. 271-284.
- [WIT98] WITMER, B. G.; SINGER, M. J. "Measuring Presence in Virtual Environments:

A Presence Questionnaire". *Presence*, vol. 7-3, Junho 1998, pp. 225-240.

- [ZIJ93] ZIJLSTRA, F. R. H. "Efficiency in work behavior: a design approach for modern tools", PhD. thesis, Delft University of Technology, 1993, 186p.

APÊNDICE A – Termo de Consentimento Livre e Esclarecido

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Primeiramente, o Grupo de Realidade Virtual da PUCRS, representado pelo doutorando **Rafael Rieder**, agradece a você pela participação no projeto “AVALIAÇÃO DA QUALIDADE DE TÉCNICAS DE INTERAÇÃO EM AMBIENTES VIRTUAIS IMERSIVOS UTILIZANDO MEDIDAS FISIOLÓGICAS”.

O objetivo deste trabalho é estudar o relacionamento entre medidas fisiológicas e índices tradicionais de usabilidade em ambientes virtuais. A partir disto, busca-se analisar se o uso de respostas fisiológicas pode ajudar a entender o comportamento do usuário durante o processo interativo, e se estas respostas podem servir como métrica para avaliação da qualidade de técnicas de interação, sem depender de dados subjetivos.

Para isto, os participantes são convidados a testar duas técnicas de interação tradicionais, em um ambiente virtual, utilizando um capacete de realidade virtual e um rastreador de posição, fixado a um mouse. Durante a interação com o programa, o monitoramento das respostas fisiológicas é realizado de forma não invasiva, por meio da colocação de pulseiras com sensores em dois dedos e nos punhos do usuário, conforme mostra a Figura 1, no verso deste documento.

Estas respostas fisiológicas, juntamente com os dados sobre a interação do usuário com os ambientes virtuais, são gravadas automaticamente pelo sistema, gerando informações importantíssimas para este experimento. Além disso, o participante responde a dois questionários: um demográfico, no início da experiência, e outro sobre a interação realizada, ao final dos testes.

Lembramos que o objetivo **não é** diagnosticar ou avaliar o participante – **mas sim** avaliar a eficiência do uso de medidas fisiológicas como métrica qualitativa para interfaces tridimensionais. O uso que se fará dos registros efetuados estará limitado à atividade de pesquisa e desenvolvimento, garantindo-se para tanto que:

1. O anonimato dos participantes está garantido em todo e qualquer documento divulgado em foros científicos (tais como conferências, periódicos, livros e assemelhados);
2. Todo participante tem livre acesso a cópias destes documentos após a publicação dos mesmos, bastando solicitar a qualquer momento, mesmo após o término do experimento;
3. Todo participante que se sentir desconfortável, constrangido ou incomodado durante os testes, pode abandonar a experiência, bem como pode colaborar com a equipe registrando por escrito as razões ou sensações que o levaram a tomar esta atitude;
4. Todo participante tem direito de expressar por escrito, na data do encontro, qualquer restrição ou condição adicional para aplicar às garantias enumeradas em (1), (2) e (3), acima. A equipe do projeto se compromete a observá-las com rigor e entende que, na ausência de tal manifestação, o participante concorda que estas diretrizes conduzam o comportamento ético da equipe de pesquisadores;
5. A equipe do projeto tem direito de utilizar os dados coletados durante os testes, mantidas as condições acima mencionadas, para quaisquer fins acadêmicos contemplados por seus membros.

() Estou de pleno acordo com os termos acima.

() Em anexo registro condições adicionais para participar dos testes.

Assinatura do participante

Assinatura do observador

Rafael Rieder

Nome completo do participante (por extenso)



Figura 1. Disposição dos sensores para monitoramento fisiológico.

APÊNDICE B – Questionário Pré-Teste (Estudo Piloto)

QUESTIONÁRIO PRÉ-TESTE

“AVALIAÇÃO DA QUALIDADE DE TÉCNICAS DE INTERAÇÃO EM AMBIENTES VIRTUAIS IMERSIVOS UTILIZANDO MEDIDAS FISIOLÓGICAS”

Idade: _____ anos

Sexo: () Masc. () Fem.

Data: ____ / ____ / 2010

Nível de Formação (concluído):

() Ensino Fundamental

() Ensino Superior

() Ensino Médio

() Pós-Graduação

Para as questões abaixo, assinale somente uma das opções. No caso das questões 1, 2, 5 e 6, considere uma escala progressiva de 1 a 7 (de um conceito inferior para um superior):

1) Em relação à utilização de computadores e respectivos sistemas, você se considera um usuário de que nível? (sendo 1 para principiante e 7 para especialista)

() 1 () 2 () 3 () 4 () 5 () 6 () 7

2) Em relação ao seu nível de conhecimento sobre Realidade Virtual, você se considera um usuário de que nível? (sendo 1 para principiante e 7 para especialista)

() 1 () 2 () 3 () 4 () 5 () 6 () 7

3) Nos últimos dois anos, quantas vezes você participou de experiências de interação em ambientes virtuais, utilizando dispositivos não convencionais, como óculos de Realidade Virtual e equipamento para rastrear movimentos?

() Nenhuma vez () Uma vez () Duas vezes () Três vezes () Quatro vezes () Cinco vezes () Mais de Cinco

4) Nos últimos dois anos, quantas vezes você já utilizou equipamentos de medição fisiológica, como relógios ou sensores de frequência cardíaca presos ao corpo, contabilizando experiências científicas e/ou exames médicos?

() Nenhuma vez () Uma vez () Duas vezes () Três vezes () Quatro vezes () Cinco vezes () Mais de Cinco

5) Quando você está aprendendo a interagir com uma nova interface gráfica, costuma sentir-se incomodado (nervoso, irritado) por não conseguir dominar as novas técnicas ou formas de interação? (sendo 1 para sem incômodo, e 7 para extremamente incomodado)

() 1 () 2 () 3 () 4 () 5 () 6 () 7

6) Quando você está interagindo com uma interface gráfica dinâmica (como jogos e ambientes virtuais), costuma sentir mal-estar (tontura, náusea, dor de cabeça) devido ao uso de dispositivos ou a interação com a própria interface? (sendo 1 para nenhum mal-estar, e 7 para mal-estar constante)

() 1 () 2 () 3 () 4 () 5 () 6 () 7

Por favor, caso deseje expressar alguma observação importante a respeito das questões acima, utilize o espaço abaixo ou o verso deste formulário:

APÊNDICE C – Instruções para o Experimento (Estudo Piloto)

Instruções para o Experimento

“Avaliação da Qualidade de Técnicas de Interação em Ambientes Virtuais Imersivos Utilizando Medidas Fisiológicas”

1. Introdução

Seja bem-vindo!

Muito obrigado por concordar em participar deste experimento, o qual pretende estudar o uso de medidas fisiológicas como métrica para avaliação da qualidade de técnicas de interação em ambientes virtuais.

Neste experimento, você irá interagir em dois ambientes tridimensionais, utilizando técnicas de interação para executar tarefas de seleção e manipulação de objetos. A descrição destas tarefas e técnicas é apresentada a seguir.

Por meio deste estudo, você estará ajudando a entender e avaliar o potencial das medidas fisiológicas como medida de usabilidade para sistemas interativos.

Para tanto, você utilizará um óculos de realidade virtual, através do qual verá o ambiente virtual; um rastreador de movimentos (fixado a um *mouse*), e dois dispositivos de medição fisiológica, que coletarão informações sobre seu estado físico e emocional.

Cabe ressaltar que, neste experimento, **você NÃO será avaliado pelo seu desempenho na execução das tarefas, nem tampouco diagnosticado pelo sistema.** O interesse é somente analisar se respostas fisiológicas podem, eficientemente, servir como métrica qualitativa para avaliar interfaces tridimensionais.

Ao final do experimento, solicita-se por gentileza que você responda a algumas questões que relatam sua impressão a respeito da experiência, contribuindo para o aperfeiçoamento do trabalho.

2. Ambiente Virtual de Treinamento

Tarefa: Manipulação de um sofá

O objetivo deste ambiente virtual é proporcionar uma prática inicial com os diferentes equipamentos, o próprio ambiente virtual e as técnicas de interação para seleção e manipulação disponíveis, a fim de familiarizar-se com estes recursos.

O ambiente simulado representa uma sala virtual, composta apenas por um sofá vermelho, posicionado no canto esquerdo da sala, com os assentos virados de frente para o observador. Nesta sala, existem também duas áreas demarcadas em verde – uma onde está o sofá e a outra vazia.

A tarefa consiste em reposicionar o sofá, girando-o em 180 graus e trocando-o de lado (mover da esquerda para a direita). Ao final, o sofá deverá estar do lado direito da sala, virado de costas, e dentro da área marcada.

Para fins de ilustração, a Figura 1 mostra o estado inicial e final do sofá dentro do ambiente virtual.

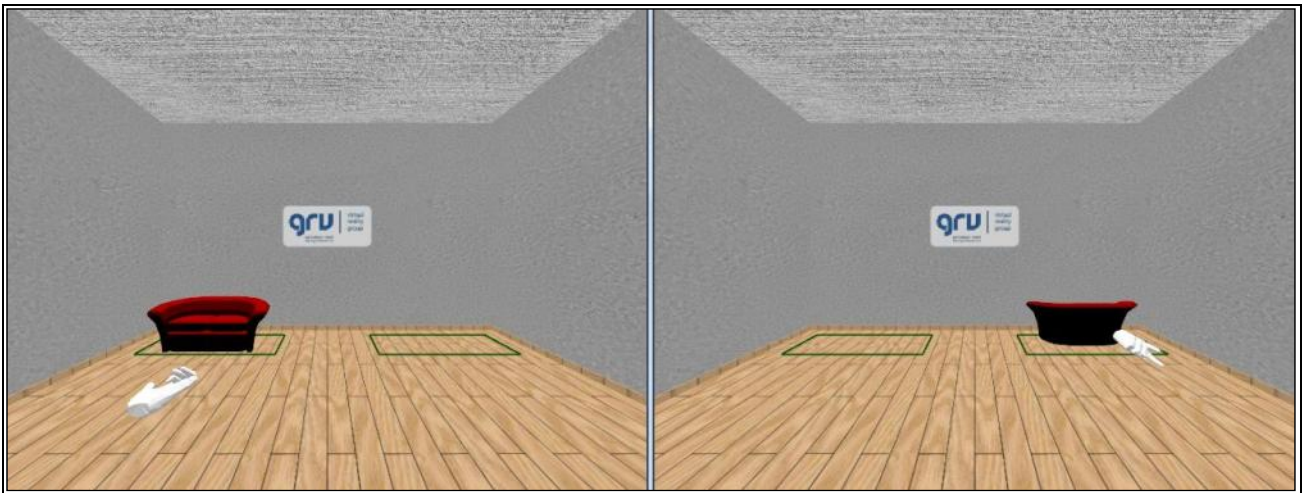


Figura 1. Posição inicial do sofá (à esquerda) e posição final desejada (à direita).

3. Ambiente Virtual de Experimento

Tarefa: Organização de livros

O objetivo deste ambiente virtual é proporcionar, de forma prática, a seleção e manipulação de diferentes objetos, tendo como base a experiência adquirida na tarefa executada anteriormente.

Neste caso, o ambiente apresentado simula uma sala virtual, na qual quatro livros numerados, de tamanhos e cores diferentes, estão espalhados pelo chão do cenário. Existem também quatro áreas demarcadas em cores transparentes no chão, similares as cores dos livros.

A tarefa do experimento consiste em pegar os livros, girá-los o quanto for necessário, e posicioná-los nas áreas demarcadas. Para realizar a tarefa corretamente, é preciso organizar os livros de forma que a capa numerada esteja visível ao observador, e a lombada do livro esteja voltada para o lado esquerdo da sala. Ao final, os livros deverão estar posicionados no chão da sala, em ordem crescente (da esquerda para a direita).

Para fins de ilustração, a Figura 2 mostra o estado inicial e final da estante e dos livros dentro do ambiente virtual.

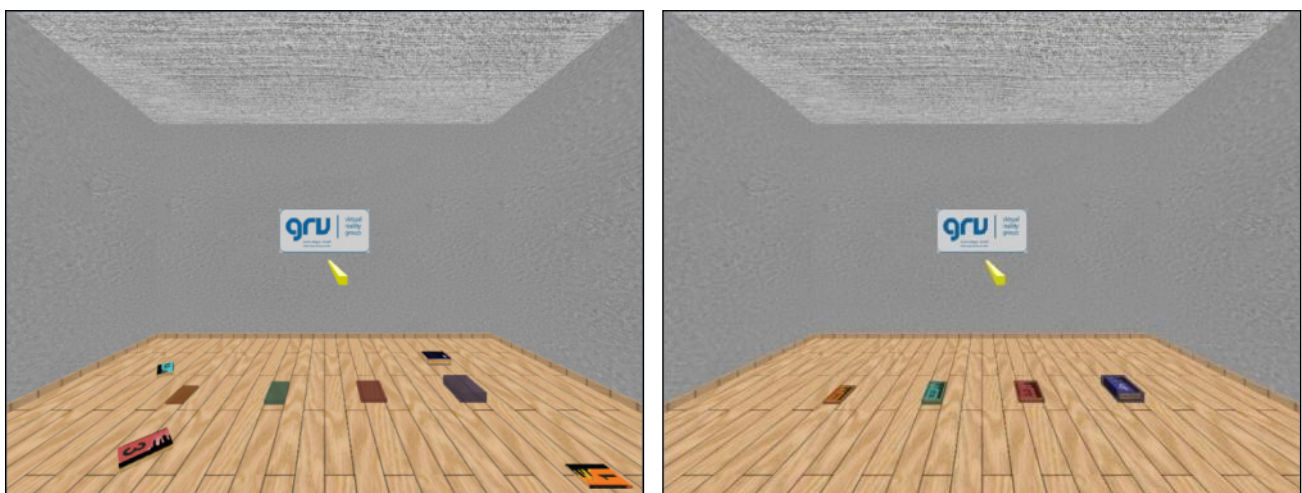


Figura 2. Posições iniciais da estante e dos livros (à esquerda) e posições finais desejadas (à direita).

4. Técnicas de Interação

4.1 Apontamento e Manuseio por Raio

Esta técnica de interação permite que o usuário indique, com um “raio-laser”, qual objeto deseja selecionar e manipular.

Nesta aplicação, o “raio-laser” é controlado pela movimentação espacial do dispositivo de rastreamento.

Para selecionar um objeto, o raio deve ser apontado na direção deste, procurando interceptá-lo. Uma vez indicado, o objeto é destacado dos demais pelo desenho de um cubo aramado, de cor vermelha, ao seu redor (Figura 3).

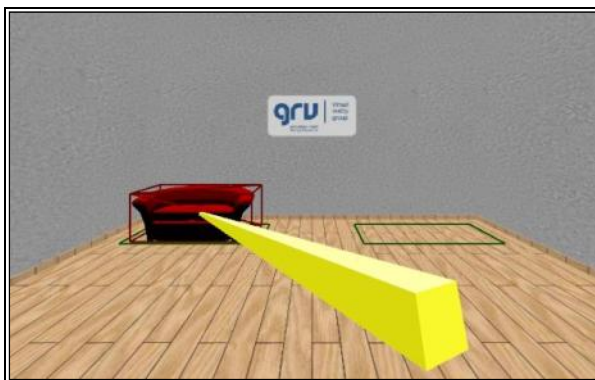


Figura 3. Exibição do cubo aramado, quando o objeto está sendo interceptado pelo raio.

Após isto, para pegá-lo e iniciar as tarefas de manipulação, basta pressionar a pequena alavanca fixada junto ao dispositivo de rastreamento. Esta ação acionará, automaticamente, o botão esquerdo de um *mouse*.

Em seguida, o objeto é anexado (grudado) ao raio. A partir deste momento, ele pode ser reposicionado para qualquer lugar do ambiente, seguindo os movimentos do raio.

Posteriormente, para soltá-lo, basta pressionar novamente a alavanca.

Importante: operações de girar o objeto requerem atenção e precisão, uma vez que o objeto acompanha os movimentos do raio. A Figura 4 ilustra este processo, mostrando como a orientação do objeto é afetada pelo giro do raio.

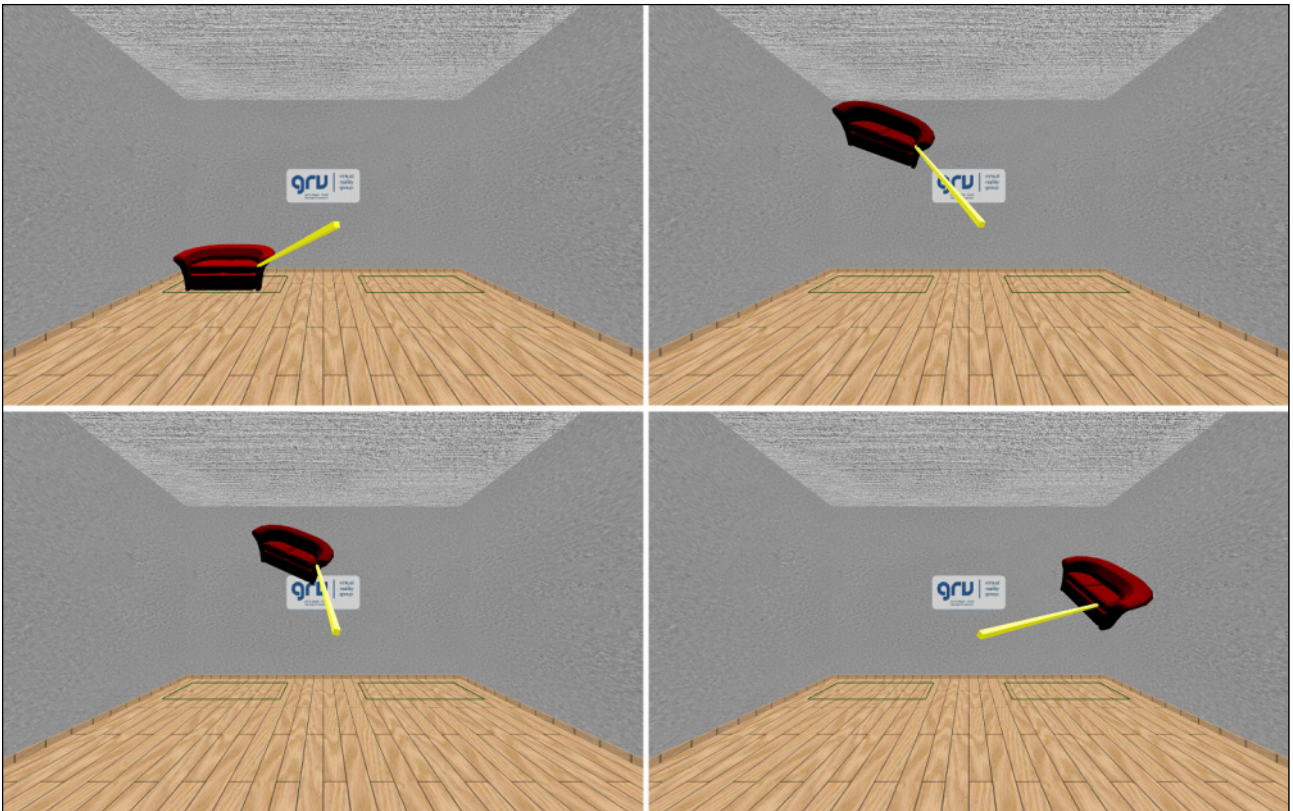


Figura 4. Exemplos de girar o raio com um objeto preso a ele, mostrando como a orientação do objeto é afetada por estes movimentos.

4.2 Apontamento por Raio e Manuseio com a Mão (HOMER)

Na técnica HOMER (*Hand-centered Object Manipulation Extending Ray-casting*), o usuário seleciona um objeto utilizando a técnica de apontamento por raio e, ao invés do objeto ser preso ao raio, ele é preso à mão virtual do usuário.

Desta forma, é possível manipular o objeto com movimentos centrados na mão virtual – ou seja, pode-se posicionar e girar o objeto como se ele estivesse na palma da mão do usuário.

O processo de seleção de objetos é o mesmo da técnica anteriormente apresentada, bem como o processo de pegar e soltar os objetos (pressionamento da alavanca).

Nesta aplicação, o “raio-laser” também é controlado pela movimentação espacial do dispositivo de rastreamento. Porém, em estado de manipulação, o raio é ocultado, e apenas o objeto selecionado segue os movimentos do rastreador. Neste caso, uma mão virtual é exibida, indicando a posição e a orientação do objeto, com base nos movimentos efetuados (Figura 5).

Após soltar o objeto, a mão virtual é ocultada, permitindo uma nova seleção a partir da movimentação do raio, que torna a ser visível.

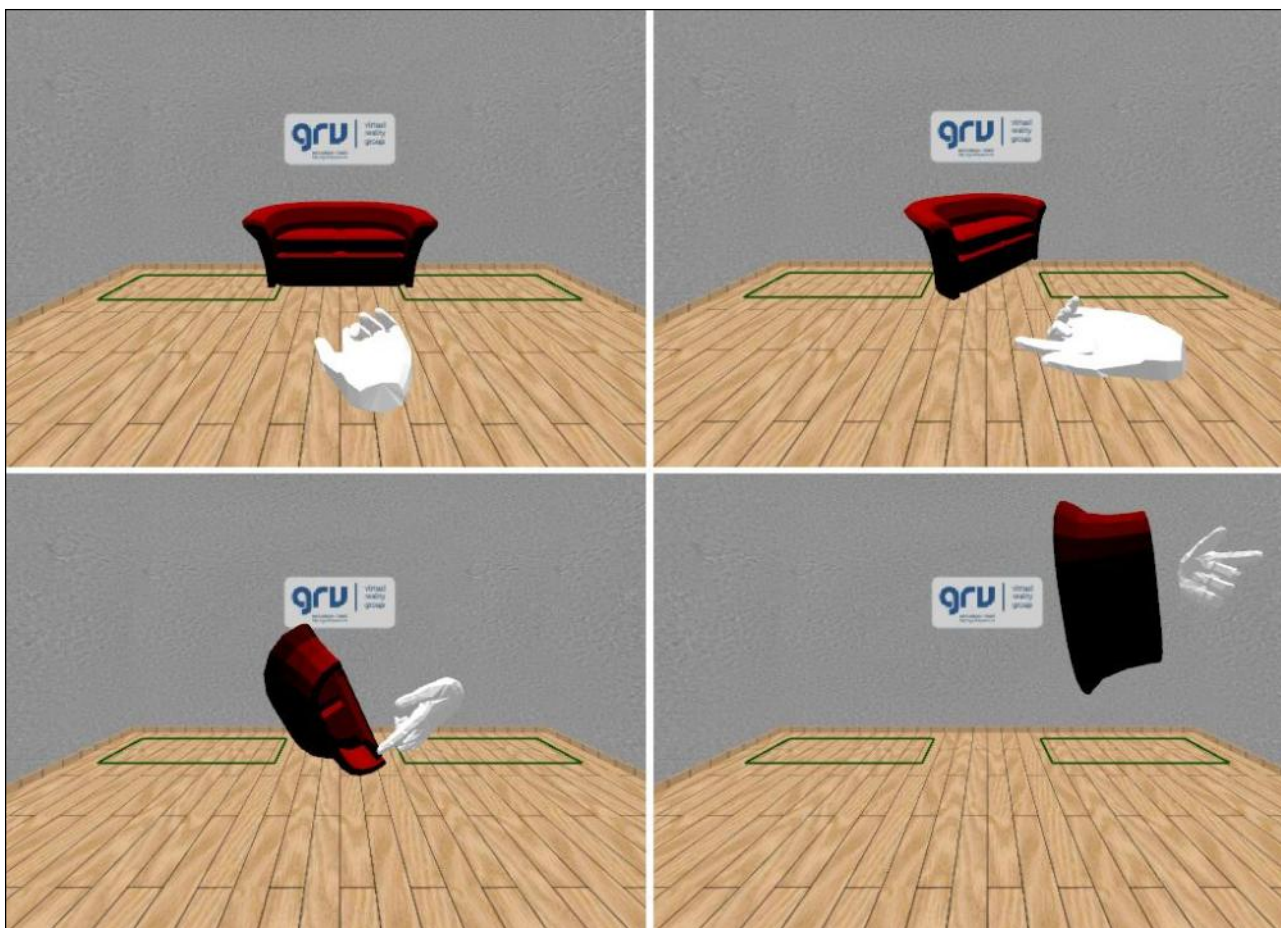


Figura 5. Exemplos de exibição da mão virtual durante tarefa de manipulação, utilizando a técnica HOMER.

Caso ainda persista alguma dúvida sobre os ambientes, os dispositivos e as técnicas de interação, por favor, consulte o examinador antes de começar o experimento.

Mais uma vez, muito obrigado por participar. Bom experimento!

APÊNDICE D – Questionário Pós-Teste (Estudo Piloto)

QUESTIONÁRIO PÓS-TESTE

“AVALIAÇÃO DA QUALIDADE DE TÉCNICAS DE INTERAÇÃO EM AMBIENTES VIRTUAIS IMERSIVOS UTILIZANDO MEDIDAS FISIOLÓGICAS”

Para as questões abaixo, assinale somente uma das opções e, quando necessário, julgue a escala de 1 a 7 como progressiva (de um conceito inferior para um superior).

- 1) Em relação ao uso de equipamentos com **FIOS** presos ao corpo, ou limitados por estes, quão confortável foi sua interação no ambiente virtual? (sendo 1 para totalmente incomodado e 7 para totalmente confortável)

1 2 3 4 5 6 7
- 2) Em relação ao uso de diferentes **DISPOSITIVOS**, quão rápida foi sua adaptação aos recursos oferecidos por estes equipamentos? (sendo 1 para não adaptado e 7 para rapidamente adaptado)

1 2 3 4 5 6 7
- 3) Considerando a comodidade durante a experiência, quão confortável foi vestir e/ou usar diferentes **DISPOSITIVOS**? (sendo 1 para extremamente desconfortável e 7 para extremamente confortável)

1 2 3 4 5 6 7
- 4) Você sentiu algum desconforto físico (náusea, tontura, dor de cabeça, irritação nos olhos, coceira, etc.) durante a experiência?

Insuportável Intenso Forte Moderado Suportável Leve Ausente
- 5) O descanso entre as etapas deste experimento foi suficiente para manter o nível de desempenho nas tarefas seguintes? (sendo 1 para insuficiente e 7 para totalmente suficiente)

1 2 3 4 5 6 7
- 6) Em relação à qualidade visual da interface, você julga satisfatórios os recursos utilizados no ambiente virtual para a execução da tarefa sugerida? (sendo 1 para insatisfatório e 7 para totalmente satisfatório)

1 2 3 4 5 6 7
- 7) Como você considera o nível de dificuldade da tarefa sugerida no ambiente? (sendo 1 para extremamente difícil e 7 para extremamente fácil)

1 2 3 4 5 6 7
- 8) As instruções apresentadas antes dos testes ajudaram no aprendizado das técnicas de interação usadas para executar tarefas no ambiente virtual? (sendo 1 para totalmente confusas e 7 para totalmente esclarecedoras)

1 2 3 4 5 6 7

9) Com relação à técnica de apontamento e manuseio por raio, você julga satisfatórios os recursos disponíveis para apontar, selecionar, manipular e soltar os objetos durante o processo de interação? (sendo 1 para muito insatisfatório e 7 para totalmente satisfatório)

1 2 3 4 5 6 7

10) Com relação à técnica HOMER (apontamento por raio e manuseio com a mão), você julga satisfatórios os recursos disponíveis para apontar, selecionar, manipular e soltar os objetos durante o processo de interação? (sendo 1 para muito insatisfatório e 7 para totalmente satisfatório)

1 2 3 4 5 6 7

11) Considerando a interação no ambiente virtual, você está convicto de que fez correta ou incorretamente as tarefas de girar e posicionar o objeto “sofá” no local proposto?

← ... convicto de que fez *incorretamente* | ... convicto de que fez *corretamente* →

Muito Bastante Pouco Não sei Pouco Bastante Muito

12) Quanto influenciou no seu desempenho o fato de que, a cada colisão no cenário virtual, era necessário pegar novamente o objeto desejado? (sendo 1 para não influenciou e 7 para influenciou muito)

1 2 3 4 5 6 7

13) O fato de que, a cada colisão no cenário virtual, era necessário pegar novamente o objeto desejado, causou-lhe irritação? (sendo 1 para não irritou e 7 para irritou muito)

1 2 3 4 5 6 7

14) O quanto você se sentiu pressionado por saber que existia um tempo-limite para alcançar o objetivo de cada ambiente virtual? (sendo 1 para não pressionou e 7 para pressionou muito)

1 2 3 4 5 6 7

15) Considerando as etapas de execução desta experiência (ler instruções, responder questionário, vestir equipamentos, executar tarefa, descansar, executar tarefa, retirar equipamentos, responder questionário), você considera satisfatória a metodologia utilizada? (sendo 1 para insatisfatório e 7 para totalmente satisfatório)

1 2 3 4 5 6 7

I) Se possível, descreva/ aponte os momentos que lhe causaram maior irritação, desconforto e/ou dificuldade para realizar a tarefa no ambiente virtual. Para tanto, utilize o espaço abaixo ou o verso deste formulário:

II) Por favor, caso deseje expressar alguma observação importante a respeito do experimento ou das questões acima, críticas, elogios ou sugestões para trabalhos futuros, utilize o espaço abaixo ou o verso deste formulário:

APÊNDICE E – Questionário Pré-Teste (Estudo de Caso)

QUESTIONÁRIO PRÉ-TESTE

“AVALIAÇÃO DA QUALIDADE DE TÉCNICAS DE INTERAÇÃO EM AMBIENTES VIRTUAIS IMERSIVOS UTILIZANDO MEDIDAS FISIOLÓGICAS”

Idade: _____ anos

Sexo: () Masc. () Fem.

Data: ____/____/2010

Nível de Formação (concluído):

() Ensino Fundamental

() Ensino Superior

() Ensino Médio

() Pós-Graduação

Para as questões abaixo, assinale somente uma das opções e, no caso das questões 1, 4 e 5, considerando uma escala progressiva de 1 a 7 (de um conceito inferior para um superior):

- 1) Em relação ao seu nível de conhecimento sobre Realidade Virtual, você se considera um usuário de que nível? (sendo 1 para principiante e 7 para especialista)
 1 2 3 4 5 6 7
- 2) Nos últimos dois anos, quantas vezes você participou de experiências de interação em ambientes virtuais utilizando dispositivos não convencionais, como óculos de Realidade Virtual e equipamento para rastrear movimentos?
 Nenhuma vez Uma vez Duas vezes Três vezes Quatro vezes Cinco vezes Mais de Cinco
- 3) Nos últimos dois anos, quantas vezes você já utilizou equipamentos de medição fisiológica, como relógios ou sensores de frequência cardíaca presos ao corpo, contabilizando experiências científicas e/ou exames médicos?
 Nenhuma vez Uma vez Duas vezes Três vezes Quatro vezes Cinco vezes Mais de Cinco
- 4) Quando você está aprendendo a interagir com uma nova interface gráfica, costuma sentir-se incomodado (nervoso, irritado) por não conseguir dominar as novas técnicas ou formas de interação? (sendo 1 para não incomodado, e 7 para muito incomodado)
 1 2 3 4 5 6 7
- 5) Quando você está interagindo com uma interface gráfica dinâmica (como jogos e ambientes virtuais), costuma sentir mal-estar (tontura, náusea, dor de cabeça) devido ao uso de dispositivos ou a interação com a própria interface? (sendo 1 para nenhum mal-estar, e 7 para mal-estar constante)
 1 2 3 4 5 6 7

Por favor, caso deseje expressar alguma observação importante a respeito das questões acima, utilize o espaço abaixo ou o verso deste formulário:

APÊNDICE F – Instruções para o Experimento

(Estudo de Caso)

Instruções para o Experimento

“Avaliação da Qualidade de Técnicas de Interação em Ambientes Virtuais Imersivos Utilizando Medidas Fisiológicas”

1. Introdução

Seja bem-vindo!

Muito obrigado por concordar em participar deste experimento, o qual pretende estudar o uso de medidas fisiológicas como métrica para avaliação da qualidade de técnicas de interação em ambientes virtuais.

Neste experimento, você irá interagir em um ambiente tridimensional, em dois momentos distintos, utilizando duas técnicas de interação para executar uma simples tarefa de seleção e manipulação de objetos. A descrição da tarefa e das técnicas é apresentada a seguir.

Antes de cada teste, será destinado um tempo para familiarização e aprendizado dos recursos de cada técnica.

Por meio deste estudo, você estará ajudando a entender e avaliar o potencial das medidas fisiológicas como medida de usabilidade para sistemas interativos.

Para realizar os testes, você utilizará um óculos de realidade virtual, através do qual verá o ambiente virtual; um rastreador de movimentos (fixado a um *mouse*), e dois dispositivos de medição fisiológica, que coletarão informações sobre seu estado físico e emocional.

Cabe ressaltar que, neste experimento, você NÃO será avaliado pelo seu desempenho na execução das tarefas, tampouco diagnosticado pelo sistema. O interesse é somente analisar se respostas fisiológicas podem, eficientemente, servir como métrica qualitativa para avaliar interfaces tridimensionais.

Ao final do experimento, solicita-se por gentileza que você responda a algumas questões que relatam sua impressão a respeito da experiência, contribuindo para o aperfeiçoamento do trabalho.

2. Ambiente Virtual

Tarefa: Organização de livros

O objetivo deste ambiente virtual é proporcionar uma prática com os diferentes equipamentos, o próprio ambiente virtual e as técnicas de interação para seleção e manipulação disponíveis, a fim de familiarizar-se com estes recursos.

O ambiente apresentado simula uma sala virtual, na qual quatro livros numerados, de tamanhos e cores diferentes, estão espalhados pelo chão do cenário. Existem também quatro áreas demarcadas em cores transparentes no chão, similares as cores dos livros.

A tarefa do experimento consiste em pegar os livros, girá-los o quanto for necessário, e posicioná-los nas áreas demarcadas. Para realizar a tarefa corretamente, é preciso organizar os livros de forma que a capa numerada esteja visível ao observador, e a lombada do livro esteja voltada para o lado esquerdo da sala. Ao final, os livros deverão estar posicionados no chão da sala, em ordem crescente (da esquerda para a direita).

Para fins de ilustração, a Figura 1 mostra os estados inicial e final dos livros dentro do ambiente virtual.

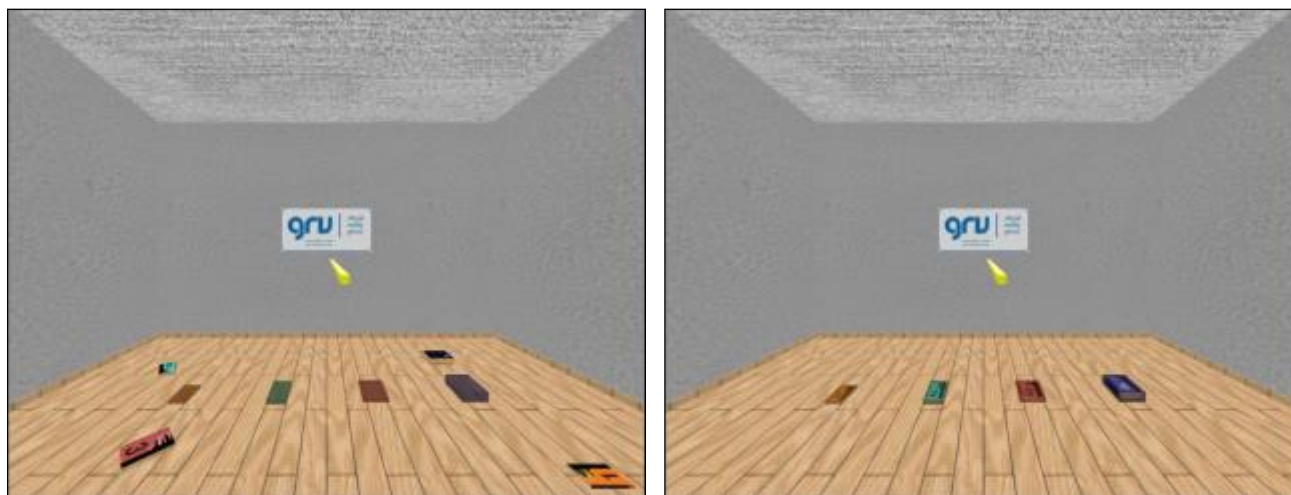


Figura 1. Posições iniciais dos livros (à esquerda) e posições finais desejadas (à direita).

3. Técnicas de Interação

3.1 Apontamento e Manuseio por Raio

Esta técnica de interação permite que o usuário indique, com um “raio-laser”, qual objeto deseja selecionar e manipular.

Nesta aplicação, o “raio-laser” é controlado pela movimentação espacial do dispositivo de rastreamento que o usuário segura com a mão.

Para selecionar um objeto, o raio deve ser apontado na direção deste, procurando interceptá-lo. Uma vez indicado, o objeto é destacado dos demais pela exibição de dois recursos visuais: um cubo aramado desenhado ao redor dele, na cor vermelha; e a mudança da textura do quadro localizado no fundo do cenário. A Figura 2 ilustra estas alterações.

Após isto, para pegá-lo e iniciar as tarefas de manipulação, basta pressionar a pequena alavanca presa ao dispositivo de rastreamento.

Em seguida, o objeto é preso ao raio e pode ser reposicionado para qualquer lugar do ambiente, seguindo os movimentos do raio.

Posteriormente, para soltá-lo, basta pressionar novamente a alavanca.

Importante: operações de girar o objeto requerem atenção e precisão, uma vez que o objeto acompanha os movimentos do raio. A Figura 3 ilustra este processo, mostrando como a orientação do objeto é afetada pelo giro do raio.



Figura 2. Exibição do cubo amarrado e da textura do quadro, quando o objeto está sendo interceptado.

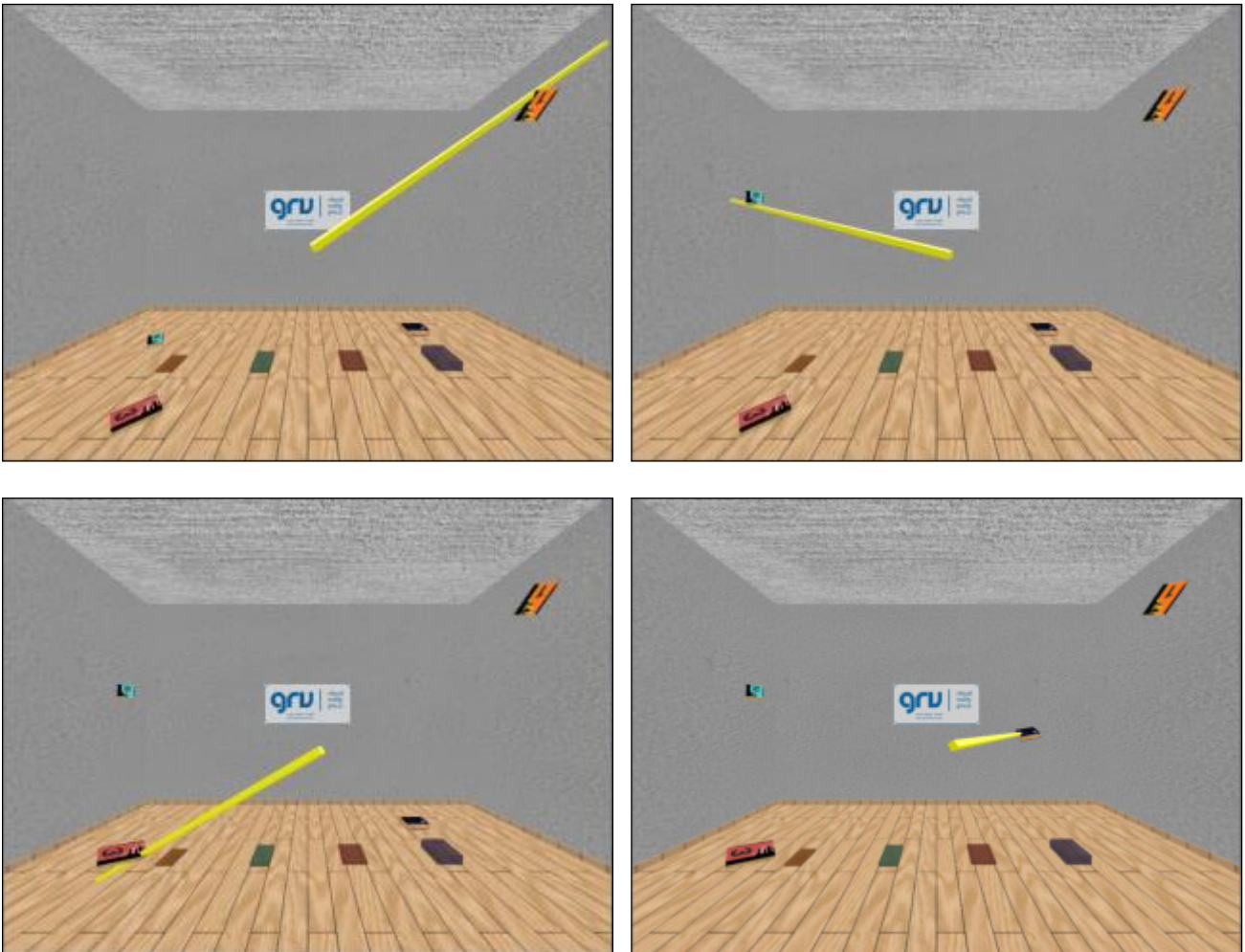


Figura 3. Exemplos de girar o raio com um objeto preso a ele, mostrando como a orientação do objeto é afetada por estes movimentos.

3.2 Apontamento por Raio e Manuseio com a Mão (HOMER)

Na técnica HOMER (*Hand-centered Object Manipulation Extending Ray-casting*), o usuário seleciona um objeto utilizando a técnica de apontamento por raio e, ao invés do

objeto ser preso ao raio, uma mão virtual é presa ao objeto, proporcionando o controle do mesmo.

Desta forma, é possível manipular o objeto com movimentos centrados nesta mão virtual – ou seja, pode-se posicionar e girar o objeto como se ele estivesse na palma da mão do usuário.

O processo de seleção de objetos é o mesmo da técnica anteriormente apresentada, bem como o processo de pegar e soltar os objetos (pressionamento da alavanca).

Nesta aplicação, o “raio-laser” também é controlado pela movimentação espacial do dispositivo de rastreamento. Porém, em estado de manipulação, o raio é ocultado, e apenas o objeto selecionado segue os movimentos do rastreador. Neste caso, uma mão virtual é exibida, indicando a posição e a orientação do objeto, com base nos movimentos efetuados (Figura 4).

Após soltar o objeto, a mão virtual é ocultada, permitindo uma nova seleção a partir da movimentação do raio, que torna a ser visível.

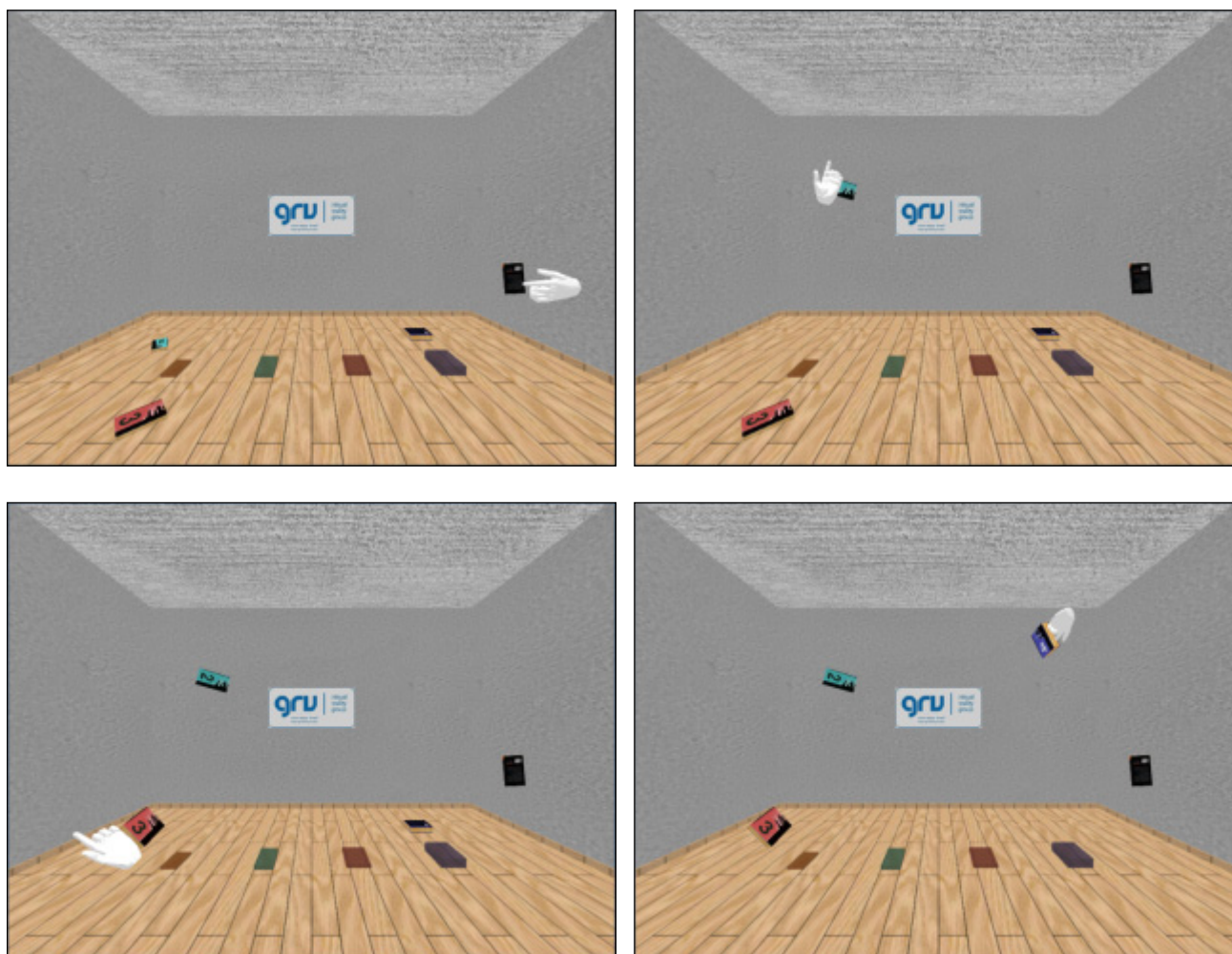


Figura 4. Exemplos de exibição da mão virtual durante tarefa de manipulação, utilizando a técnica HOMER.

Caso ainda persista alguma dúvida a respeito do ambiente, dos dispositivos e das técnicas de interação, por favor, consulte o pesquisador antes de iniciar o experimento.

Mais uma vez, muito obrigado por participar. Bom experimento!

APÊNDICE G – Questionário Pós-Teste (Estudo de Caso)

QUESTIONÁRIO PÓS-TESTE

“AVALIAÇÃO DA QUALIDADE DE TÉCNICAS DE INTERAÇÃO EM AMBIENTES VIRTUAIS IMERSIVOS UTILIZANDO MEDIDAS FISIOLÓGICAS”

Para as questões abaixo, assinale somente uma das opções e, quando necessário, julgue a escala de 1 a 7 como progressiva (de um conceito inferior para um superior).

- 1) Em relação ao uso de diferentes **DISPOSITIVOS COM FIOS** presos ao corpo, ou limitado por estes, quão confortável e natural foi sua interação no ambiente virtual? (sendo 1 para totalmente desconfortável/ incomodado e 7 para totalmente confortável/ natural)

1 2 3 4 5 6 7
- 2) O fato de que, a cada colisão no cenário virtual, era necessário pegar novamente o objeto desejado, causou-lhe irritação? (sendo 1 para não irritou e 7 para irritou muito)

1 2 3 4 5 6 7
- 3) O quanto você se sentiu pressionado por saber que existia um tempo limite para alcançar o objetivo proposto? (sendo 1 para não pressionou e 7 para pressionou muito)

1 2 3 4 5 6 7
- 4) Com relação ao uso da técnica de apontamento e manuseio por raio, você julga satisfatórios os recursos disponíveis para apontar, selecionar, manipular e soltar os objetos durante o processo de interação? (sendo 1 para totalmente insatisfatório e 7 para totalmente satisfatório)

1 2 3 4 5 6 7
- 5) Com relação ao uso da técnica HOMER (apontamento por raio e manuseio com a mão), você julga satisfatórios os recursos disponíveis para apontar, selecionar, manipular e soltar os objetos durante o processo de interação? (sendo 1 para totalmente insatisfatório e 7 para totalmente satisfatório)

1 2 3 4 5 6 7
- 6) Considerando a interação no ambiente virtual, você está convicto de que fez corretamente as tarefas de girar e posicionar os objetos nos locais propostos? (sendo 1 para totalmente incorreto e 7 para totalmente correto)

1 2 3 4 5 6 7
- 7) Em relação à qualidade visual e sonora da interface, você julga satisfatórios os recursos utilizados no ambiente virtual para a execução da tarefa sugerida? (sendo 1 para totalmente insatisfatório e 7 para totalmente satisfatório)

1 2 3 4 5 6 7

- I) Caso tenha sentido algum desconforto físico (náusea, tontura, dor de cabeça, irritação nos olhos, coceira, etc.) durante a experiência, por favor, indique abaixo o desconforto, o momento que começou a senti-lo, e a respectiva intensidade.**

- II) Se possível, descreva/ aponte abaixo os momentos que lhe causaram maior irritação e/ou dificuldade para realizar a tarefa no ambiente virtual.**

- III) Caso deseje expressar alguma observação importante a respeito do experimento ou das questões acima, utilize o espaço a seguir.**

APÊNDICE H – Glossário sobre Técnicas e Métodos Estatísticos

Alfa de Cronbach: é uma técnica estatística comumente utilizada para medir a consistência interna ou confiabilidade de um teste para um conjunto de duas ou mais variáveis. Foi desenvolvida por L. J. Cronbach, e apresenta como vantagem a possibilidade de se obter uma medida tão confiável e tão válida quanto possível, utilizando o menor número de itens. O alfa de Cronbach ou coeficiente alfa varia entre uma escala de 0 a 1, e está relacionado com a correlação entre cada par de itens e o número de itens de uma escala. Em geral, um valor de 0,7 ou acima é aceito como um indicativo de um bom instrumento.

Análise de Variância (ANOVA): também conhecido como teste F , em homenagem a R. A. Fisher, tem por objetivo verificar se existem diferenças significativas entre médias e se fatores exercem influência em alguma variável dependente. É um teste estatístico paramétrico cujo modelo de distribuição de probabilidades é o da variância para k amostras ou tratamentos independentes. As amostras podem ser de mesmo tamanho ou desiguais, no caso de k tratamentos (um critério), e devem ser iguais quando for o caso de k tratamentos e r blocos (dois critérios). Os testes ANOVA podem ser utilizados para averiguar diferenças entre diversas situações e para duas ou mais variáveis.

Análise Exploratória: é uma abordagem que define possíveis relações apenas na forma mais geral, possibilitando que outros métodos e técnicas estimem a natureza destas relações.

Análise Fatorial: é uma abordagem estatística que pode ser usada para analisar inter-relações entre um grande número de variáveis e explicar essas variáveis em termos de suas dimensões inerentes comuns. O objetivo é encontrar um meio de condensar a informação contida em um número de variáveis originais para um conjunto menor de variáveis estatísticas (fatores), com uma perda mínima de informação.

Aumento do Tamanho Amostral: é um método computacional de geração aleatória de dados, a partir de valores observados, com o objetivo de gerar novas e

maiores amostras, cujo tamanho é decidido pelo pesquisador. É possível, assim, aumentar o poder do teste, sobretudo quando as amostras iniciais são pequenas por indisponibilidade de indivíduos.

Autovalor: ou *Eigenvalue*, representa a variância total explicada por cada fator.

Correlação de Spearman: teste de associação entre duas variáveis mensuradas em nível ordinal, calculando-se a correlação pelo coeficiente de Spearman (ρ). Este coeficiente varia entre -1 a +1: quanto mais próximos desses valores, mais forte a associação entre as variáveis. Geralmente utilizado para distribuições não-normais, em substituição ao teste de correlação linear de Pearson.

Correlação linear de Pearson: teste de associação linear entre duas variáveis mensuradas a nível intervalar ou de razões, amplamente utilizado, na qual se mede o grau e a direção utilizando o coeficiente de correlação linear (r). Este coeficiente varia entre -1 a +1: quanto mais próximos desses valores, mais forte a associação das variáveis em exame (correlação negativa ou positiva). O escore zero desse coeficiente indica ausência de correlação.

Estatística Descritiva: parte da estatística que objetiva a coleta, a organização, a classificação dos dados amostrais ou das populações, as apresentações tabular e gráfica e o cálculo de determinadas medidas: média, mediana, variância, desvio padrão, coeficiente de variação, de assimetria, de curtose e outras.

Fator de Inflação da Variância: teste que indica a multicolinearidade entre variáveis – ou seja: aponta o grau em que cada variável independente é explicada pelas demais variáveis independentes. Valores superiores a 10,0 indicam a presença de multicolinearidade, o que acaba mostrando a redundância entre variáveis. Desta forma, simplifica-se o processo de análise, uma vez que este teste reduz a ocorrência de erros durante a detecção de diferenças significativas entre fatores.

Gráfico de Box-Plot: é um diagrama que sintetiza um conjunto de dados medido em uma escala intervalar. É frequentemente usado em análise de dados exploratória para mostrar a forma da distribuição, o valor central e a variabilidade dos dados. É também um gráfico especialmente útil para verificar a existência de valores *outlier* no conjunto de

dados. O resultado gráfico exibe os valores extremos de um conjunto (valores máximo e mínimo), os quartis inferior e superior, e a mediana.

Gráfico Stem-and-Leaf (“caule e folha”): é um diagrama que organiza os escores em ordem crescente, separando cada valor em dois componentes: o “caule” (*stem*), constituído por todos os valores de cada variável, exceto o último à direita, e a “folha” (*leaf*), que é o último algarismo à direita. É frequentemente utilizado em análise exploratória para ilustrar as principais características de uma distribuição de dados.

Medida de Adequação de Kaiser, Meyer e Olkin (KMO): índice usado para avaliar a adequação de uma análise fatorial. Os valores obtidos variam de 0 a 1, onde valores iguais ou acima de 0,8 indicam um ótimo ajuste para o instrumento em avaliação.

Método das Metades (*split-half*): procedimento utilizado para medir a consistência interna de um teste, no qual os escores de duas subdivisões do instrumento são comparados para determinar sua confiabilidade.

Regressão Linear: procedimento que tem por objetivo descobrir um relacionamento linear entre a variável resposta e uma possível variável preditora, por intermédio do método dos mínimos quadrados.

Regressão Múltipla: procedimento que tem por objetivo descobrir um relacionamento linear entre a variável resposta e diferentes variáveis preditoras possíveis. Para tanto, uma medida ou coeficiente de determinação (r^2) é apresentado para cada variável preditora. Esta medida apresenta valores dentro de uma escala de 0 a 1, sendo que um valor próximo de 0 indica um modelo de regressão fraco. Uma ANOVA também é realizada para cada comparação de variáveis, o qual testa se a regressão realizada é estatisticamente significativa.

Regressão Passo a Passo (*stepwise*): procedimento de seleção de variáveis para formar um modelo de regressão. Variáveis preditoras vão sendo adicionadas a um modelo, uma a uma, de acordo com sua correlação com a variável resposta.

Teste de Kolmogorov-Smirnov: para uma simples amostra de dados, este teste verifica a conformidade dos dados em relação a uma distribuição estatística teórica.

Quando existem duas amostras de dados, é utilizado para testar se estes conjuntos seguem ou não uma mesma distribuição de probabilidade – ou seja, se as características das duas amostras são compatíveis e recolhidas de uma população com mesma distribuição, desconhecida. Em suma, é um teste que verifica a normalidade dos dados.

Testes t de Student: são testes paramétricos de hipóteses para médias, baseados no modelo de distribuição de Student (W. S. Gosset). Eles podem ser aplicados sempre que as variâncias são desconhecidas e a distribuição de dados é normal. Na comparação entre dois grupos, por exemplo, a hipótese nula indica que não existe diferença entre as amostras. Os tipos de testes t são:

- Uma amostra: verifica se a média da população assume um determinado valor, ou se uma dada amostra provém de um universo com uma dada média;
- Duas amostras pareadas: verifica se existe diferença significativa entre os valores médios de uma mesma medida, feita sob duas condições diferentes, para o mesmo grupo de sujeitos. As amostras devem ser de tamanhos iguais. Um exemplo de aplicação deste tipo de teste é a comparação entre uma medida coletada antes e depois de um experimento;
- Duas amostras independentes: verifica se existe diferença significativa entre os valores médios de uma mesma medida, para grupos diferentes de sujeitos. Pode ser aplicado para amostras de tamanhos desiguais, e presumindo variâncias equivalentes (homocedástico) ou variâncias diferentes (heterocedástico).

APÊNDICE I – Artigo aceito para publicação no INTERACT 2011

As próximas páginas apresentam, na íntegra, o artigo aceito para publicação no *13th IFIP TC13 Conference on Human-Computer Interaction* (INTERACT 2011), o qual divulga os resultados obtidos por este trabalho na comparação entre medidas fisiológicas e métricas convencionais de usabilidade, como medidas de desempenho do usuário na tarefa interativa e medidas subjetivas para avaliação de interfaces 3D.

Identifying Relationships between Physiological Measures and Evaluation Metrics for 3D Interaction Techniques

Rafael Rieder^{1,2}, Christian Haag Kristensen¹, Márcio Sarroglia Pinho¹

¹ Pontifical Catholic University of Rio Grande do Sul, Av. Ipiranga, 6681, Partenon
CEP 90619-900 Porto Alegre, RS, Brazil

² University of Passo Fundo, BR 285, Km 171, São José
CEP 99052-900 Passo Fundo, RS, Brazil

rafaelrieder@gmail.com, {christian.kristensen, marcio.pinho}@pucls.br

Abstract. This project aims to present a methodology to study the relationships between physiological measures and evaluation metrics for 3D interaction techniques using methods for multivariate data analysis. Physiological responses, such as heart rate and skin conductance, offer objective data about the user stress during interaction. This could be useful, for instance, to evaluate qualitative aspects of interaction techniques without relying on solely subjective data. Moreover, these data could contribute to improve task performance analysis by measuring different responses to 3D interaction techniques. With this in mind, we propose a methodology that defines a testing protocol, a normalization procedure and statistical techniques, considering the use of physiological measures during the evaluation process. A case study comparison between two 3D interaction techniques (ray-casting and HOMER) shows promising results, pointing to heart rate variability, as measured by the NN50 parameter, as a potential index of task performance. Further studies are needed in order to establish guidelines for evaluation processes based on well-defined associations between human behaviors and human actions realized in 3D user interfaces.

Keywords: usability metrics, physiological measures, interaction techniques.

1 Introduction

In order to evaluate the characteristics of three-dimensional user interfaces (3DUI), like presence and immersion, methods and tools commonly used to evaluate two-dimensional user interfaces can be applied, such as prototypes, questionnaires and formative and summative tests. These instruments are able to get relevant usability metrics also in 3DUIs, like variables to measure system performance, user task performance and user preferences. The first two measures result in objective data for assessing, respectively, the computer or graphics system performance, and the quality of performance of specific tasks in the 3D application. The third measure results in subjective data for assessing the user satisfaction while using an interface. The

evaluation of these measures is important because they allow that different 3DUI elements, such as spatial perception, multimodal interaction and sensory stimulus are considered.

However, the adaptation of these tools to evaluate 3DUIs can lead to incomplete assessment of the particular characteristics of these applications, such as the use of non-conventional devices and 3D interaction techniques (ITs) [6]. These characteristics tend to influence the user performance and the user satisfaction, which requires a process to evaluate its various resources based on the user experience level.

A recent alternative used to evaluate interfaces it is the use of the physiological measures. According to Malik *et al* [22], the physiological monitoring provides information about the user's physiological balance, and its measures are associated with stress. Researches in the Virtual Reality (VR) area have been using this type of measurement to assess the user's physical and mental effort on the 2D games [17][18][19][20] and to evaluate presence and user comfort in immersive virtual environments (VEs) [7][8][16][23][24]. However, there are no studies about the relationship between physiological measures and metrics focused on the evaluation of the quality of ITs.

The use of physiological measures can still address other two classical problems in the evaluation of 3DUIs. The first concerns the collect of objective measures, which in some cases require modifications in the source code. In these situations, it is not always possible or desirable to alter an application, due to the complexity of the system [2] or the limited availability of development time [28]. The second problem concerns the reliability of subjective metrics, which may have influenced their results by external factors to the interaction process, such as user's physical and mental efforts, or user's cognitive mediation, such as omission or summarization of information.

Physiological measures, therefore, offer objective responses that are not controlled by the person, they are associated with factors such as fatigue and irritation, and provide data related to the organism's behavior. These are measures that can indicate, for example, the adaptation periods to a new device or new IT, because the user's stress level can be viewed along the timeline. Besides, they can aid in the comprehension of the performance results and answers of questionnaires. Doing so, physiological responses may complement the current methods of assessment [5][14][25], allowing the understanding of the interaction process as a whole, and contributing to increasing the quality of the VR applications.

In order to evaluate whether physiological measures may be or not a substitute for objective and subjective usability metrics, statistical methods of multivariate data analysis can be used. According to Hair *et al* [13], these methods allow analyzing simultaneously the influence of multiple measures on each subject or variable under investigation, regardless of the complexity or the context in which these variations occur. So, it is possible to verify if results obtained with physiological measures are able to indicate the same problems identified by traditional usability metrics.

Therefore, this work describes a methodology for assessing the quality of ITs in immersive VEs, comparing physiological measures and evaluation metrics for 3DUIs using multivariate data analysis. Our methodology contemplates the use of a testing protocol, data normalization and exclusion processes, and statistical methods for exploratory data analysis and regression analysis, in order to discover relationships

between variables that contribute to the evaluation process, complementing or assisting in the interpretation of results. In the same way, our methodology also determines the physiological measures able to indicate the same results expected by traditional usability measures, and these may eventually replace usual measures in projects in which the simplification of the testing stage is desirable. So, it is possible to reduce the dependency on subjective data, and to avoid changes to collect performance data in complex software.

This paper is organized as following: Section 2 presents the related work, whereas Section 3 describes the developed methodology. In Section 4, a case study is presented to evaluate the use of this approach. Section 5 shows a discussion about the results obtained using this methodology. Finally, Section 6 concludes the paper highlighting the potential of our approach.

2 Related Work

The use of physiological measures is a recent alternative in the evaluation of graphical interfaces. Latest researches apply this resource to measure presence and cybersickness in immersive VEs, and user's physical and cognitive effort in videogames.

Meehan *et al* [23][24] used physiological monitoring to measure presence in a stressful VE. These researches used heart rate (HR), skin conductance (SC) and skin temperature measures in four different experiments to compare participants' physiological reactions to a non-threatening virtual room and their reactions to a stressful virtual height situation. According to the authors, HR satisfied all the requirements for a reliable, valid, sensitive and objective measure of presence in a stressful VE. In addition, HR showed correlation with the well-established presence questionnaire. SC had some of the properties desired to measure presence, and skin temperature did not.

Slater *et al* [26] reported the difficulty to measure presence subjectively, and highlighted the need to evaluate alternatives that provided objective data to overcome, or, at least, supplement the use of questionnaires. With this in mind, the authors conducted an experiment to explore the relationship between physiological responses, breaks in presence and utterances by virtual characters towards the participants using a virtual bar scenario. The results showed that changes in HR and SC point to occurrence of breaks in presence during the interaction process. Changes in HR also indicated the moments when an avatar speaks to the subjects, whereas heart rate variability (HRV) parameters pointed to differ between participants with different social anxiety scores, classified by questionnaire.

Brogni *et al* [7][8] followed a similar approach to previous work. They studied the use of physiological responses to determine the impact of visual realism and the user stress level during the interaction in an urban VE. In this scenario, the sense of presence was subjected to the texture quality and the appearance of avatars. The results showed that HRV parameters indicated the reducing of level stress as time goes by. The authors also reported that these parameters were associated with the level of visual realism based on the texture strategy used in their experiments.

Kim *et al* [16] presented a study about the use of physiological measures to reduce cybersickness in immersive VEs. Their approach proposed a system based on 11 physiological signals, which detected the cybersickness and automatically reduced the user's field of view and slowed the travel velocity in the VE. The results indicated a significantly lower frequency of cybersickness when users use this system, compared to a situation in which it is not used. The authors also highlighted the gastric tachyarrhythmia as a physiological measure of a better outcome.

On the other hand, Lin *et al* [17][18] studied the use of physiological measures as a metric to evaluate the usability of video games. The experiments considered the use of HR, SC and blood pressure (BVP) measures to assess the user's performance and satisfaction while tasks were realized in three different difficulty game levels. The authors reported the SC measure as relevant in comparisons between game levels and performance user groups, indicating that a greater number of errors and difficulty consequently increases the SC. This work also studied the relationship between physiological measures and frustration events, which showed a range of more than 5% in SC in 70% of the frustration cases analyzed. Other measures did not showed significant results.

Complementary studies of Lin *et al* [19][20] added measures like pupillary response, eye tracking and HRV parameters to assess the user cognitive efforts in videogames. The results also showed the relationship between HRV parameters and game levels. Pupillary response and eye tracking presented promise results, but it was not consistent as SC and HRV measures.

From the literature review, it can be noticed that there are no studies exploring the relationship between physiological measures and usability metrics used to evaluate the quality of three-dimensional ITs in immersive VEs. The following section presents a methodology designed for this purpose.

3 Methodology

Firstly, the following sections present the platform for testing, the physiological measures, the task performance measures and the questionnaires used by our approach. Secondly, we present the steps to apply our methodology, which include the use of a test protocol, normalization and exclusion methods for physiological measures, and statistical analysis methods.

3.1 Platform for testing

In order to illustrate the use of our methodology, we built a virtual room with four numbered books, distributed on the floor inside the user's field of vision, as presented in Figure 1. Two well-known ITs also were implemented to select and manipulate objects: ray-casting [3] and HOMER [4]. These techniques were chosen because they are commonly used as parameters to evaluate new ITs. With ray-casting, the user selects and manipulates objects using a virtual light ray, with the ray's direction specified by the user's hand. With HOMER, the user selects an object using the

ray-casting technique, and manipulates it using a virtual hand, which instantly moves to the current object's position and attaches to it.

The user's task is to get the books, turning them as necessary, and place them in transparent areas marked on the floor. The books must be organized with their front cover visible to the user, and their spine toward to the left side of the virtual room. At the end of task, the books will be placed in ascending order, from left to right. The VE also provides two visual feedbacks to indicate when an object is ready for selection, and an aural feedback to inform collisions occurred into the virtual room.

This application was built using C++, OpenGL, GLUT, irrKlang¹, and the SmallVR² toolkit, which simplifies the development of VR applications by abstracting many implementation aspects such as device control and scene graph management, while maintaining the GLUT structure for the program.

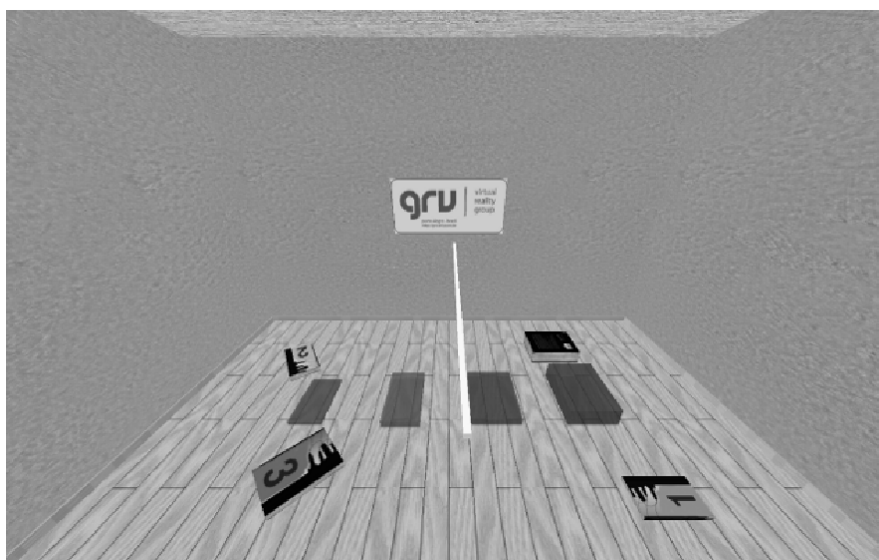


Fig. 1. The virtual room application built to our experiment.

In order to explore the VE and use the ITs resources, we used i-Glasses Head Mounted Display (HMD) and Polhemus FastTrak motion tracker with two tracking points enabled for interaction. The first tracking point was used to track the user's head movements, whereas second was used in the user's dominant hand to select and manipulate objects. Grab and release user's actions were confirmed using a push-button attached on the second tracking-point.

The physiological monitoring used an electrocardiogram (ECG) sensor and a SC sensor, on a non-invasive way. These equipments were connected to the Procomp Infiniti encoder, which captured and sent the physiological responses to the Biograph Infiniti software for data processing. Three electrodes connected to the ECG sensor

¹ <http://www.ambiera.com/irrklang>

² <http://grv.inf.pucrs.br/projects>

were fixed on the user's wrists with rubber wrist straps, whereas two electrodes connected to the SC sensor were strapped to two fingers of user's non-dominant hand using finger bands. All these solutions are manufactured by Thought Technology³.

Figure 2 presents the device's default configuration, highlighting the push-button attached to one of the tracking-points, and positions of the ECG and SC sensors.



Fig. 2. Subject wearing physiological and VR devices during the experiment.

3.2 Physiological Measures

For this work we used the HR and SC physiological measures, collected by ECG and SC sensors, respectively. Our approach also includes the use of seven different HRV measures, generated by time domain and frequency domain methods from short-term 5-minutes recordings, according to Malik *et al* [22].

This way, the following physiological measures were selected:

- Mean SC, expressed in micro-Siemens (μS);
- Mean HR, expressed in beats per minute (bpm);
- Standard deviation of the NN⁴ interval (SDNN), expressed in milliseconds (ms);

³ <http://www.thoughttechnology.com>

⁴ On an ECG recording, NN are the normal-to-normal intervals between adjacent QRS complexes.

- Number of interval differences of successive NN intervals greater than 50 ms (NN50);
- Proportion derived by dividing NN50 by the total number of NN intervals (pNN50);
- Mean total power in very low frequency range (VLF), expressed in milliseconds squared (ms^2);
- Mean total power in low frequency range (LF), expressed in ms^2 ;
- Mean total power in high frequency range (HF), expressed in ms^2 ;
- Ratio between the LF and HF measures (LF/ HF).

3.3 Task Performance Measures

In order to get the task performance during the experiment with ITs, the following measures were defined. These data were collected in real time by the own VE application:

- *Total time*: entire period of the interaction process, measured in seconds (s);
- *Accuracy*: distance between the center of an object and the center of its target position⁵, measured in meters (m);
- *Collisions*: amount of collisions occurred with an object;
- *Grabs*: number of attempts to grab an object.

3.4 Questionnaires

Two questionnaires were created for this work. The pre-test questionnaire obtained demographic information and user's past experiences with VEs, whereas the post-test questionnaire evaluated the user preferences in relation to the interface, tasks, and ITs. A progressive scale of 1 to 7 was set to evaluate each question in both instruments, from a lower to a higher concept.

Pre-test questionnaire asked about the age, gender and education of each participant, and was composed of five questions. These questions were about the VR level of knowledge, the number of times there was presence in VE experiments, the number of times there were physiological measures, the sense of discomfort when interacting with new interfaces, and the sense of discomfort when interacting with graphical interfaces like games and VEs.

Post-test questionnaire was composed of seven questions. The aim of these questions were to evaluate the influence of wired devices in the user's performance, the level of irritation generated by the fact of requiring to the user pick up again the objects after each collision, the level of pressure generated by the time limit to complete the task, the level of satisfaction in using the ray-casting and the HOMER techniques, the level of confidence in performing the task correctly, and the level of satisfaction with the visual and aural feedbacks presented by the VE. Moreover, this questionnaire also provided areas to describe the occurrence of discomfort situations, and the moments of more irritation or more difficulty during the interaction process.

⁵ The target position means the desired position of the object in the VE.

3.5 Test Protocol

In order to test our methodology, we designed a protocol based on related work presented in Section 2. Our approach supports steps to the simultaneous collection of task performance, subjective evaluation and physiological data during the experience.

The test protocol was established with nine stages, which can be executed in approximately 45 minutes, in the following order:

- Apply pre-test questionnaire;
- Prepare devices;
- Collect baseline data;
- Start training – first IT;
- Start experiment – first IT;
- Start training – second IT;
- Start experiment – second IT;
- Release devices;
- Apply post-test questionnaire.

“Apply pre-test questionnaire” stage contemplates the trainer and experiment presentations, the distribution of informed consent form to read, and the filling of pre-test questionnaire. After this it is provided to read the VE instructions, which explains how to execute tasks using the two ITs. The estimate duration to finish this stage is eight minutes. According to Kim *et al* [16], this period is also important to stabilize the physiological responses before next stages.

“Prepare devices” stage considers the arrangement of physiological sensors and VR equipments in the user’s body. Firstly, the user is invited to turn off electronic devices, to remove watch and bracelets and to accommodate in a chair, sitting in a comfortably way. After this the trainer cleans the user’s wrists with alcohol gel, and applies a conductive gel into the ECG sensors to reduce noise caused by electrical resistance of the skin. The physiological sensors are fixed as mentioned in Section 3.1 and, finally, the user wears the HMD. For this stage, we estimate approximately three minutes to complete it.

“Collect baseline data” is started after the user feels comfortable with the devices. The trainer requests the user to put his/her arms on his/her legs in a rest position. In this step, VR devices remain turned off, and it is asked for the user to keep his/her eyes open. Based on Slater *et al* [26] and Brogni *et al* [8] collecting physiologic data is done in two minutes and it will be the comparison base for the interaction process.

In the next stage, “Start training – first IT”, the subject begins to interact with the VE to learn to use the devices and the first IT. Selection and manipulation tasks can be performed during three minutes.

Again, the subject interacts with the same VE and IT in the “Start experiment – first IT” stage, and must perform all tasks in seven minutes. After this period, the subject has three minutes to rest with all devices off.

The same procedures are applied to the next two stages for training and experiment of the “second IT”.

It is important to highlight for an unbiased analysis, the use of two techniques must be balanced according the number of participants in an experiment. For this reason, our protocol divides the experiment between “first IT” and “second IT”.

The above times, defined to complete the training and experiment stages, are based on Lin *et al* [17][18][19][20] and Brogni *et al* [7][8] works, whereas the interval to rest between stages is based on Kim *et al* [16].

During the “Release devices” stage, VR and physiological equipments are removed from the user. The trainer applies procedures to clean the user’s wrist and devices, using dry wipes and dusters. The estimate time is four minutes to do it.

At last, “Apply post-test questionnaire” stage contemplates the subjective evaluation of test, which the user is invited to answer the post-test questionnaire. A brief period also is addressed for comments and thanks. For this stage, we estimate approximately five minutes, based on Slater *et al* [26] and Lin *et al* [20] works.

3.6 Normalization and Exclusion Methods for Physiological Measures

In order to statistically compare physiological responses and different usability variables, it is necessary to define data normalization and exclusion methods to the HR and SC measures.

For the SC data normalization, we adopted a scale of 0 to 1, which it attributed the minimum value of 0 to a lowest SC value, and the maximum value of 1 to a highest SC value. This procedure was applied to each user’s SC signal, generated a new and normalized Mean SC measure. So, SC values became uniform and preserved the individual characteristics of each subject.

By contrast, for the HR measures we needed to apply a procedure to exclude some data, because the adopted way to collect this physiological response is susceptible to generation of noises in the HR signal. The procedure eliminated participants who presented HR values outside the normal range for a human. The exclusion criterion was executed in the following order:

- HR rest: based on the baseline data, subjects were excluded from the dataset when their mean HR was below 60 bpm or above 100 bpm. According to Guyton and Hall [12], typical healthy resting HR in adults is 60–100 bpm;
- HR max: subjects were excluded from the dataset when their mean HR during the experiment was above to the maximum HR, which it was estimated from the Tanaka formula [27], presented by the Equation 1;
- HR target: subjects were excluded from the dataset when their mean HR during the experiment was above to the target HR, which it was estimated from the Karvonen method [15], presented by the Equation 2. In order to use this method, we determined an intensity level of 50% to the interaction task, since the physical effort during the interaction process can be considered within a moderate activity zone [1], as a result of the subjects being seated and performed spatial movements using their arms and head during the test.
- HR min: subjects were excluded from the dataset when their mean HR during the experiment was below 60 bpm.

$$HR_{\max} = 208 - (0,7 \times \text{Age}) \quad (1)$$

$$HR_{\text{target}} = [(HR_{\max} - HR_{\text{rest}}) \times (\text{Intensity level } \%)] + HR_{\text{rest}} \quad (2)$$

3.7 Statistical Analysis Methods

In order to verify the relationships between different measures, we chose to use multivariate data analysis methods. According to Hair *et al* [13], these methods are able to investigate, simultaneously, multiple measures about each subject or object under study.

This work adopted the following analytical steps:

- Apply methods for exploratory data analysis to summarize, test the normality, detect outliers of the data, and use techniques to verify correlations between variables. This approach allows to identify the consistency and distribution of the data, and avoid the redundant variables;
- Apply multiple regression techniques to generate prediction models, considering methods to select relevant variables and its coefficients of determination. This approach allows discovering what measures are associated to the task performance and subjective responses.

For the exploratory data analysis, we defined the following tests:

- Summarize data: Descriptive statistics;
- Normalization: Kolmogorov-Smirnov test;
- Outliers' detection: stem and leaf and box-plots;
- Correlations: Pearson's coefficient to linear relationships, and Spearman's coefficient to nonlinear relationships.

The stepwise regression was chosen to create regression models and selects the predictor variables. In this process, each regression model may have one or more predictor variables and their coefficient of determination (r^2). These coefficients inform the power of these measures have to explain the variability of results. In this project, these results indicate whether a physiological measure can substitute or not a traditional usability measure. An analysis of variance (ANOVA) is also applied to test the significance of the regression model.

In order to generate regression models, our analysis selected only physiological measures with results statistically significant in the correlation tests ($p < 0,05$).

4 Case Study

In order to evaluate the effectiveness of physiological measures as indexes of quality to 3DUIs, this section aims to present a case study using our methodology, according to the definitions shown in Section 3.

Our evaluation included 54 healthy participants, 28 men and 26 women aged between 17 and 57 years old. Their tests were scheduled during a period of two weeks.

The subjects were also distributed into two equal groups (14 men and 13 women), in order to balance the use of ITs. The group "A" used as first IT the ray-casting technique, whereas the group "B" used as first IT the HOMER technique.

4.1 Relationships between Physiological and Task Performance Measures

According to the Section 3.7, it is necessary to apply a set of multivariate data analysis methods to assess the relationship between physiological and task performance measures, as presented in Sections 3.2 and 3.3. Thus, it is possible to identify which physiological responses are able to indicate task performance, or whether they can at least assist the interpretation of the results.

First of all, we used the testing protocol to collect the physiological, task performance and user preferences measures. After this we applied the normalization and exclusion procedures, in order to adjust our physiological datasets for the statistical comparisons. During this last stage, we detected some abnormal HR measures in 22 subjects, which needed to be discarded. Because of this situation, the original dataset had to be subdivided into two new groups: a dataset for SC measures, which included all the experiment participants (54 subjects), and another dataset for HR and HRV measures, which included only 32 subjects.

In the next stage, we applied the statistical methods for exploratory data analysis and multiple regression, looking for physiological measures able to indicate task performance.

Since our methodology was applied, two physiological measures (NN50 e HF) had a statistically significant relationship with two task performance measures (“Total time” and “Accuracy”). However, only one of these relationships indicated, on the regression model, strongly statistically significant results by both techniques (“Total time” x NN50, $p < 0,01$), as presented in the Table 1.

According to the results of the Table 1, the “accuracy” task performance measure only has statistically significance with the NN50 and HF physiological measures, for experiments using ray-casting technique.

Table 1. Regression models for physiological and task performance measures with strong correlation.

Interaction Techniques	Task Performance Measures	Physiological Measures	Regression	ANOVA	
			r^2 (%)	F-test	p(value)
HOMER	Total Time	NN50	28,83%	12,15	0,00**
	Accuracy	NN50	7,26%	2,35	0,13
	Accuracy	HF	2,35%	0,72	0,59
Ray-Casting	Total Time	NN50	61,98%	48,91	0,00**
	Accuracy	NN50	43,28%	22,89	0,00**
	Accuracy	HF	31,33%	13,69	0,00**

On the other hand, NN50 physiological measure may be considered as the variable with the most associated with the “Total time” measure, because results were strongly significant for both experiments, independently of two techniques ($p < 0,01$). Based on Table 1, the NN50 physiological measure is able to indicate the user task performance, for the “total time” measure, with a statistical power (r^2) of 61,98% to the experiments using ray-casting technique, and 28,83% to the experiments using HOMER technique.

We also generated a regression model using the NN50 and “Total time” means, in order to join the statistical power of the selected physiological response in a single

model, independently of two techniques. The result showed a coefficient of determination of 45,16% (ANOVA, $p < 0,01$, $F = 24,70$).

However, our results presented intermediary statistical power values. The coefficients of determination showed values far from an index close to perfect correlation ($r^2 = 100\%$), showing that the variance of NN50 measure cannot explain, alone and exactly, the variance of “Total time” measure. In other words, we can say that the NN50 physiological measure still cannot be used to replace the “Total time” measure during a task performance evaluation.

4.2 Relationships between Physiological and User Preferences Measures

Using the same approach presented in the Section 3.7, we also applied a multivariate data analysis to verify the relationships between physiological and user preferences measures, already presented in the Sections 3.2 and 3.4. In this study, we also used the data normalization and exclusion procedures, defined in the Section 3.6, which subdivided our dataset in two new sets (54 subjects for the SC measure, and 32 subjects for the HR and HRV measures).

In order to compare the questionnaire answers and physiological measures, we generated new physiological measure means from the values of experiences using the two ITs. Results can be visualized in the Table 2.

Table 2. Regression models for physiological and user preferences measures with strong correlation.

Questionnaires	User Preferences Measures	Physiological Measures	Regression	ANOVA	
			r^2 (%)	F-test	p(value)
Pre-test	Question 1	SC	11,44%	6,71	0,01*
	Question 1	NN50	15,26%	5,40	0,03*
	Question 2	NN50	13,82%	4,81	0,03*
	Question 2	LF	10,65%	3,57	0,07
	Question 2	HF	7,50%	2,43	0,13
Post-test	Question 4	SC	10,49%	6,09	0,02*
	Question 4	HR	12,83%	4,42	0,04*
	Question 4	NN50	17,02%	6,15	0,02*
	Question 4	HF	23,39%	9,16	0,01*
	Question 7	LF/HF	13,72%	4,77%	0,03*

Firstly, tests were applied to verify the relationships between physiological measures and pre-test questionnaire answers. In this study, NN50 and SC physiological measures presented statistically significant relationships with questions addressed the level of knowledge about VR (Question 1), the experience with non-conventional devices in VEs (Question 2) and the tendency to feel discomfort or irritation when interact with new interfaces (Question 4).

Based on these analysis, we may note that the regression models presented in the Table 2 showed significant results ($p < 0,05$) for SC and NN50 measures as predictors of assessment, but with low statistical power for the Question 1 (SC, $r^2 = 11,44\%$; NN50, $r^2 = 15,26\%$), Question 2 (NN50, $r^2 = 13,82\%$) and Question 4

(SC, $r^2 = 10,49\%$). In this way, we can say that these physiological measures – when solely used – still cannot be employed to indicate the user level of knowledge about VR, the user level of experience in VEs, and the user level of irritation during learning in new graphical interfaces.

Secondly, we applied the same tests to verify the relationships between physiological measures and post-test questionnaire answers. In this study, we did not compare the physiological measure means and the answers related to the Questions 4 and 5, because these questions aimed to evaluate the ITs, separately. In this case, the Question 4 responses were compared with physiological measures collected during the user experiences with the ray-casting technique, and Question 5 responses were compared with physiological data of the experiences using HOMER technique.

This study presented only one physiological measure (LF/HF) with statistically significant relationship ($p < 0,05$) for the evaluation about the quality of visual and aural feedbacks displayed during the interaction process (Question 7). However, only 13,72% of the variance of LF/HF can explain the variance of the Question 7 responses. So, we can say that the LF/HF is not able to indicate this item evaluation.

Comparisons between physiological measures and Questions 4 and 5, which aimed to evaluate the level of satisfaction in using the ITs, presented statistically significant relationship ($p < 0,05$) only between HR, NN50 and HF measures and the Question 4 answers (ray-casting technique evaluation), as shown in the Table 2. The generated regression models also showed determination coefficients of low explanatory power (HR, $r^2 = 12,83\%$; NN50, $r^2 = 17,02\%$; HF, $r^2 = 23,39\%$), impossible to indicate the level of satisfaction with the use of ray-casting technique through physiological measures.

The post-test questionnaire also evaluated aspects about physical discomfort, sense of irritation and difficulties to perform tasks during the interaction process. The subjects' responses showed that these sensations did not affect the user's performance during the test, pointing no significant results.

5 Discussion

Based on the previous section, physiological measures still cannot be considered as indexes of user task performance and user preferences measures for VEs. From the results, the physiological responses only indicate user behavior tendencies during the interaction process, which can help as an additional resource to understand usability metrics during the evaluation process.

Comparisons between physiological and task performance measures only highlight the NN50 measure, which presented statistically significant results for both evaluated techniques and statistical power near an acceptable level to the “Total time” measure. According to the Section 3.2, the NN50 attests the amount of interval differences of successive NN intervals greater than 50 ms, which indicates the level of stabilization of the heart rhythm. Figure 3 shows a correlation plot between “Total time” and “NN50” measures, considering both experiences using the two ITs. This figure also indicates the trend of the NN50 increases as the “Total time” increases too.

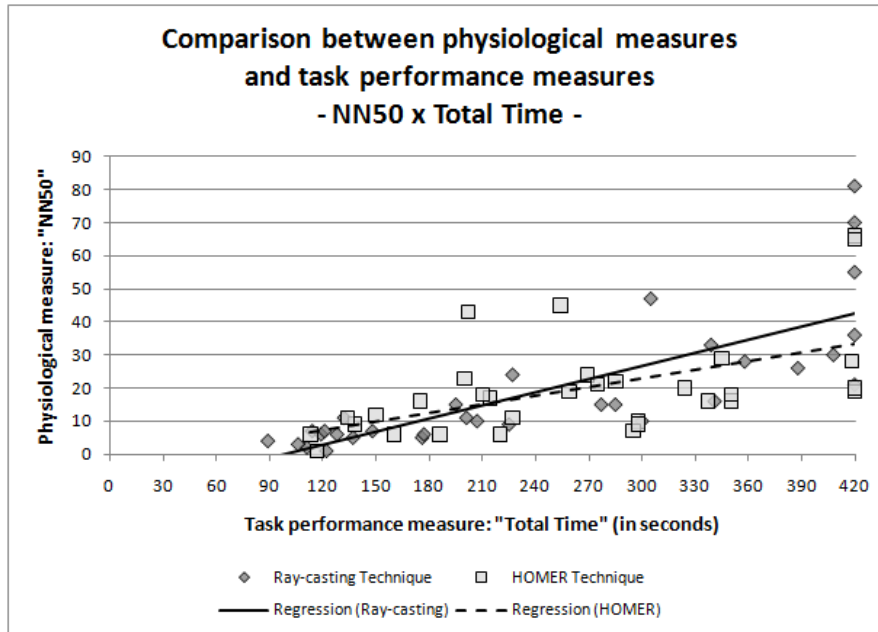


Fig. 3. Correlations and trends between “Total time” and “NN50” using the two ITs.

This result can be interpreted on two different points of views. Firstly, experiments completed in less time show more concentrated subjects, which also use more physical effort to perform the tasks, compared with those who spend more time. On the other hand, we can say that the subjects spend more time learning to use the 3DUI, which ends up leaving them more relaxed and with their HRs in levels around a baseline measurement. Anyway, it is a promising measure to be explored and evaluated again in future research.

Comparisons between physiological user preferences measures presented some relevant measures, such as SC, NN50 and LF/HF – but none of them showed significant statistical power. Probably, these low relationships can be associated with the first application of the pre- and post-test questionnaires, and their progressive scales.

With this in mind, we applied some tests to verify the questionnaire’s reliability. Firstly, Cronbach’s alpha [10] and split-half tests [11] were used, which presented the results shown in Table 3. According to Hair *et al* [13], the reliability coefficients indicate a good assessment tool when their index results in values above 0,7, which were not observed in this analysis. A probably reason for this result is the sample size used (54 subjects), which was below the literature recommendations, considering the scale dimension used by the instruments. As our scale ranged from 1 to 7, it would be necessary to apply our questionnaires to 70 subjects, at least, which guarantees obtaining more reliable results.

We also applied a factor analysis to the questionnaires to verify their relevance. According to Malhotra [21], it is recommendable to have the largest possible number

of factors to this analysis, which are generated based on the number of questions, satisfying at least 60% of the total variance. According to the Table 4, for the pre-test questionnaire, 63,54% of the total variance can be explained using two factors, considering eigenvalues greater than 1,00. For the post-test questionnaire, Table 5 presents that 67,60% of the total variance can be explained using three factors, but only two of them have eigenvalues greater than 1,00. These results are also corroborated by low values of the Kaiser-Meyer-Olkin (KMO) tests, which examine the appropriateness of factor analysis.

These results recommend the review of the progressive scale used in our questionnaires, before new evaluation. Certainly, these changes will result in a more refined analysis to discover the relationships between physiological measures and subjective data.

Table 3. Reliability tests for the pre- and post-test questionnaires.

Questionnaires	Cronbach's Alpha	Guttman Split-Half Coefficient	Items
Pre-test	0,50	0,10	5
Post-test	0,35	0,48	7

Table 4. Factor analysis of the Pre-test questionnaire (KMO measure = 0,51).

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	1,92	38,30	38,30
2	1,26	25,24	63,54
3	0,92	18,48	82,02
4	0,60	12,05	94,07
5	0,30	5,93	100,00

Table 5. Factor analysis of the Post-test questionnaire (KMO measure = 0,60).

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2,24	32,04	32,04
2	1,50	21,36	53,40
3	0,99	14,20	67,60
4	0,73	10,46	78,06
5	0,70	10,05	88,11
6	0,49	6,93	95,04
7	0,35	4,96	100,00

Moreover, our evaluation was partially hampered because some data were discarded during the normalization and exclusion procedures. Comparisons involving HR and HRV measures had a loss of almost 60% of data.

According to Combatalade [9], despite the precaution taken in relation to skin preparation, conductive gel application, electrode placement and user instructions, it is very difficult to save HR data absolutely clean and no noise. It forces the use of a

normalization process to the HR signal, especially to detect two types of artifacts: missed beats and extra beats.

Missed beats can occur when the signal is so distorted that the software is unable to identify the beat pattern and only picks up on the next good beat, which results in long inter beat interval value. On other hand, extra beats can occur when the program confuses a distortion in the signal for a beat and detect two or more beats, when there should be only one, resulting in short inter beat interval value at the end of analysis.

In order to reduce artifacts, a software solution to process HR signals and analyze HRV measures can be adopted. In this case, it is important to be assisted by a medical professional in order to ensure that the data cleaning does not interfere in future results.

It is important to mention that there are also real natural physiological events, similar to these artifacts, like premature trial or ventricular contractions. As a result of this, it is recommended to follow a specific protocol if there is presence of subjects with heart diseases and/or under medication.

Another suggestion to minimize the occurrence of noise in HR signal, it is the use of self-adjustable wrist straps, which prevents the electrodes become tightened or loosened on the user's wrists. It is possible to use a non-invasive device fixed on the user's chest, closer to the heart, which reports the HR to the ECG sensor without using wires.

At last, it is also recommended a collaborative effort between Computer Science and Medicine experts, in order to define guidelines allowing a better understanding about the user's behavior during the interaction process using physiological measures to do it.

6 Conclusions

This work presents a methodology for assessing the quality of ITs in 3DUIs, based on the use of physiological measures during the interaction process. A case study comparing two ITs was executed, in order to apply this new method through a well-defined testing protocol, two procedures to normalize and exclude some samples from physiological dataset, and the application of multivariate data analysis methods, which allow to highlight the relationships between physiological measures and usability metrics.

The use of our methodology pointed to promising results and interesting tendencies. We expect our methodology to evolve and to be consolidate through new comparisons with other ITs, not only to evaluate selection, manipulation or navigation tasks in VEs, but also to consider another features of 3DUIs, like computer stereo vision, different degrees of freedom and multiple sensory stimuli.

Physiological measures still cannot be considered as substitutes of user task performance or user preferences. However, these measures can be used as a complementary resource for the interpretation of usability metrics commonly observed during the evaluation processes, highlighting the relationship between NN50 psychological measure and "Total time" task performance measure. We recommend a

more detailed study of this measure for future work, in order to contribute and qualify the ITs evaluation process.

Our research also presents a mode to validate the proposed questionnaires, as a means of establishing better relationships between physiological measures and subjective user preferences. This analysis pointed to the importance of scale adjustment for subjective evaluation before testing, based on the available sample size of subjects. This fact may have contributed to measures with significant differences, such as SC and NN50, have shown low statistical power in this work.

Regarding the comparisons between physiological measures and usability metrics, we highlighted the need to use multivariate data analysis techniques during the evaluation process. These statistical methods allowed the understanding and the detailed case study of relationships between different measures, encouraging the formulation of more refined conclusions and the indication of trends for future studies in this promising area.

Finally, we emphasize that our project is a first step to define a specific methodology to the 3DUI evaluation process, considering the use of physiological measures and appropriated statistic techniques to compare multiple datasets. The continuity of our research involves new case studies and a multidisciplinary effort between Computer Science and Medicine experts to create guidelines for clear and wide association of the user's behavior and their actions performed in 3DUIs using different interaction resources.

References

1. American College of Sports Medicine: ACSM's Advanced Exercise Physiology. Lippincott Williams & Wilkins, Philadelphia (2005)
2. Bisbal, J., Lawless, D., Wu, B., Grimson, J.: Legacy Information Systems: Issues and directions. *IEEE Software* 16-5, 103-111 (2002)
3. Bolt, R. A.: "Put-That-There": Voice and Gesture at the Graphics Interface. In: *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 262-270. ACM, New York (1980)
4. Bowman, D. A., Hodges, L. F.: An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In: *Proceedings of the 1997 Symposium on Interactive 3D graphics*, pp. 35-38. ACM, New York (1997)
5. Bowman, D. A., Gabbard, J. L., Hix, D.: A Survey of Usability Evaluation in Virtual Environments: classification and comparison of methods. *Presence: Teleoperators and Virtual Environments* 11-4, 404-424 (2002)
6. Bowman, D. A., Kruijff, E., LaViola, J.J., Poupyrev, I.: *3D User Interfaces: theory and practice*. Addison-Wesley, Boston (2004)
7. Brogni, A., Vinayagamoorthy, V., Steed, A., Slater, M.: Variations in Physiological Responses of Participants during Different Stages of an Immersive Virtual Environment Experiment. In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pp. 376-382. ACM, New York (2006)
8. Brogni, A., Vinayagamoorthy, V., Steed, A., Slater, M.: Responses of Participants during an Immersive Virtual Environment Experience. *The International Journal of Virtual Reality* 6-2, 1-10 (2007)

9. Combatalade, D.: Basics of Heart Rate Variability Applied to Psychophysiology. Technical report MAR953-00, Thought Technology Ltd. (2010)
10. Cronbach, L. J.: Coefficient Alpha and the Internal Structure of Tests. *Psychometrika* 16–3, 297–334 (1951)
11. Guttman, L.: A Basis for Analyzing Test-Retest Reliability. *Psychometrika* 10–4, 255–282 (1945)
12. Guyton, A. C., Hall, J. E.: *Textbook of Medical Physiology*. Saunders, Philadelphia (2005)
13. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E.: *Multivariate Data Analysis*. Prentice Hall, New Jersey (2005)
14. Hix, D., Hartson, H.: *Developing User Interfaces: ensuring usability through product & process*. John Wiley & Sons, New Jersey (1993)
15. Karvonen, M. J., Kentala, E., Mustala, O.: The Effects of Training on Heart Rate: a longitudinal study. *Annales Medicinæ Experimentalis et Biologiae Fenniae* 35–3, 307–315 (1957)
16. Kim, Y. Y., Kim, E. N., Park, M. J., Park, K. S., Ko, H. D., Kim, H. T.: The Application of Biosignal Feedback for Reducing Cybersickness from Exposure to a Virtual Environment. *Presence: Teleoperators and Virtual Environments* 17–1, 1–16 (2008)
17. Lin, T., Omata, M., Hu, W., Imamiya, A.: Do physiological data relate to traditional usability indexes? In: *Proceedings of the 17th Australia conference on Computer-Human Interaction*, pp. 1–10. Computer-Human Interaction Special Interest Group of Australia, Narrabundah (2005)
18. Lin, T., Imamiya, A., Omata, M., Hu, W.: An Empirical Study of Relationships Between Traditional Usability Indexes and Physiological Data. *Australasian Journal of Information Systems* 13–2, 105–117 (2006)
19. Lin, T., Imamiya, A.: Evaluating usability based on multimodal information: an empirical study. In: *Proceedings of the 8th International Conference on Multimodal Interfaces*, pp. 364–371. ACM, New York (2006)
20. Lin, T., Imamiya, A., Mao, X.: Using Multiple Data Sources to get Closer Insights into User Cost and Task Performance. *Interacting with Computers* 20–3, 364–374 (2008)
21. Malhotra, N. K.: *Marketing Research: an applied orientation*. Prentice Hall, New Jersey (2006)
22. Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., Schwartz, P. J.: Heart Rate Variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal* 17–3, 354–381 (1996)
23. Meehan, M.F., Insko, B., Whitton, M., Brooks Jr, F. P.: Physiological Measures of Presence in Stressful Virtual Environments. In: *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 645–652. ACM, New York (2002)
24. Meehan, M.F., Razzaque, S., Insko, B., Whitton, M., Brooks Jr, F. P.: Review for Four Studies on the Use of Physiological Reaction as a Measure of Presence in Stressful Virtual Environments. *Applied Psychophysiology and Biofeedback* 30–3, 239–258 (2005)
25. Rosson, M., Carroll, J.: *Usability Engineering: scenario-based development of human-computer interaction*. Morgan Kaufmann, San Francisco (2001)
26. Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., Garau, M., Brogni, A., Friedman, D.: Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment. *Presence: Teleoperators and Virtual Environments* 15–5, 553–569 (2006)
27. Tanaka, H., Monahan, K. D., Seals, D. R.: Age-predicted Maximal Heart Rate Revisited. *Journal of the American College of Cardiology* 37–1, 153–156 (2001)
28. Tullis, T., Albert, W.: *Measuring the User Experience: collecting, analyzing, and presenting usability metrics*. Morgan Kaufmann, San Francisco (2008)