

John Lennon Oliveira Couto

**Avaliação do desempenho de técnicas de interação com as
mãos em Realidade Aumentada imersiva**

São Paulo

2020

John Lennon Oliveira Couto

Avaliação do desempenho de técnicas de interação com as mãos em Realidade Aumentada imersiva

Versão Corrigida

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo como parte das exigências do Programa de Pós-Graduação em Engenharia Elétrica para obtenção do título de Mestre em Ciências.

Área de Concentração:
Sistemas Eletrônicos

Orientadora: Prof.^a Dra. Roseli de Deus Lopes

São Paulo

2020

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 11 de dezembro de 2020

Assinatura do autor: _____



Assinatura do orientador: _____



Catálogo-na-publicação

Couto, John Lennon Oliveira

Avaliação do desempenho de técnicas de interação com as mãos em Realidade Aumentada imersiva / J. L. O. Couto -- versão corr. -- São Paulo, 2020.

165 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos.

1.Interface Homem-Computador 2.Realidade Virtual 3.Teste e Avaliação de Software I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos II.t.

Aos meus pais, à minha avó materna, aos meus irmãos e em homenagem aos meus falecidos avós, todos eles pilares fundamentais que me nutrem com afeto e me sustentam neste mundo.

AGRADECIMENTOS

Apreendi com este trabalho que realizar uma pesquisa científica exige, além de habilidade para conciliar os diversos desejos e interesses da mente humana, capacidade para encontrar, processar e organizar uma elevada quantidade de informações. Eu sabia que desenvolver este tipo de trabalho não seria trivial, mas em nenhum momento eu havia imaginado que o processo seria tão longo e tão exaustivo – cognitiva e emocionalmente – como foi. De fato, a conclusão deste trabalho só foi possível graças a uma rede de apoios de diferentes naturezas que recebi ao longo deste percurso, e a ela deixo meus sinceros agradecimentos.

Agradeço à minha orientadora, Prof.^a Dra. Roseli de Deus Lopes, pelas horas de orientação, pela paciência nos momentos em que eu estive impaciente e pelo entendimento, consolo e estímulo quando minhas energias faltaram.

Agradeço ao Dr. Marcelo Archanjo José, pelas horas dedicadas em discussões produtivas nas diversas etapas deste trabalho, tais como aprovação do Comitê de Ética em Pesquisa, delineamentos experimentais, análise de dados e organização geral das informações.

Agradeço ao Centro Interdisciplinar de Pesquisas Interativas (CITI-USP), pelo acesso ao seu espaço, equipamentos e *expertise* profissional, que foram essenciais nas fases de exploração tecnológica e condução dos experimentos descritos neste trabalho.

Agradeço aos 12 participantes voluntários que cederam seu tempo, paciência e energia, para colaborarem com os experimentos realizados nesta investigação científica.

Agradeço à toda equipe do LSI-TEC, pelas dicas sobre o como trilhar o percurso acadêmico, pelo tempo dedicado e considerações feitas na prévia da minha apresentação de qualificação, e pelos cafés quando eu precisava espairecer.

Agradeço ao Professor Doutor Scott Mackenzie (Universidade de York), por gentilmente ter fornecido o código fonte do programa GoFitts, utilizado como referência neste trabalho.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão de bolsa de mestrado (processo número 167205/2017-0).

Agradeço ao meu amigo Andy, por ter aplicado seus conhecimentos especializados, empenhado tempo e esforços para me ajudar a produzir uma figura maravilhosa que ilustrou muito bem este estudo científico.

Por fim, agradeço àqueles que, apesar de não contribuírem diretamente para o conteúdo deste trabalho, indiretamente criaram as condições afetivas, emocionais, materiais e ambientais para que eu pudesse escrever um manuscrito de artigo original e esta dissertação: meus pais, minha avó, meus irmãos e meus amigos.

“ A Lei de Fitts esconde alguns interessantes paradoxos. Primeiramente, apesar da lei se referir a um paradigma experimental extremamente restrito – movimentos elementares e uni-dimensionais – seu escopo é verdadeiramente geral: a Lei de Fitts captura aspectos do movimento orientado a um objetivo – um traço geral do comportamento dos seres vivos, e obviamente uma preocupação central para Interação Humano-Computador. Em segundo lugar, a Lei de Fitts é curiosamente simples e ainda assim, muito complexa. Apesar da lei ser expressa por uma relação matemática rudimentar, ela abriga um número de problemas difíceis e, até o momento, não solucionados [...].”

(GUIARD; Beaudouin-Lafon, 2004, p. 748, tradução do autor)

RESUMO

COUTO, J. L. O. **Avaliação do desempenho de técnicas de interação com as mãos em Realidade Aumentada imersiva**. 2020. 159 p. Dissertação (Mestrado em Engenharia Elétrica) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2020.

A busca por uma experiência imersiva em Realidade Aumentada (RA) tem levado a um rápido desenvolvimento das tecnologias para óculos de RA. Entretanto, falta a esse novo sistema computacional uma técnica de interação natural que seja rápida, precisa e robusta como a que o mouse permite em monitores 2D. Uma opção é a interação direta com as mãos, mas as tecnologias atuais de rastreamento das mãos apresentam limitações que dificultam o desenvolvimento até mesmo de aplicativos básicos. A avaliação formal do desempenho da interação pode ajudar a aprimorá-las. A Lei de Fitts e os métodos descritos na ISO 9241-411 são amplamente aplicados para avaliar interação 2D. Após pesquisa não sistemática da literatura, este trabalho propôs um método formal e reproduzível para avaliar o desempenho de interações diretas com as mãos em óculos de RA. O software 3DFitts foi desenvolvido para aplicar o método proposto, e quatro experimentos foram conduzidos para estudar como ele se comporta na avaliação de interações 2D e 3D. Os experimentos 1 e 2 investigaram interação em 2D sem e com participantes humanos, objetivando determinar se os resultados do 3DFitts eram estatisticamente equivalentes aos de um software de referência da literatura. O Experimento 3 foi um estudo de usuário para avaliar o desempenho de interação em um ambiente semelhante ao Experimento 2, mas visualizado em 3D com um óculos de RA denominado Meta 2. O Experimento 4 também foi um estudo de usuário, mas focado em discriminar o desempenho dos participantes ao usar seis diferentes técnicas de interação, as quais incluíam três dispositivos (um mouse, o Leap Motion Controller e o sistema de interação embutido no Meta 2), dois métodos de seleção (clique com o botão esquerdo do mouse e tempo de permanência) e duas orientações espaciais do plano de testes (frontal e horizontal). Essas técnicas foram avaliadas por métricas objetivas (computadas pelo 3DFitts) e subjetivas (obtidas a partir de questionários aplicados aos participantes). Os resultados dos experimentos 1 e 2 indicaram que o 3DFitts se comporta de maneira semelhante ao software de referência. A comparação entre os resultados dos Experimentos 2 e 3 revelou que a adição do óculos de RA não teve impacto estatisticamente significativo no desempenho dos participantes. Os dados do Experimento 4 forneceram fortes evidências quantitativas de que o desempenho ao usar o mouse foi significativamente maior do que ao usar o Leap Motion Controller, que por sua vez foi significativamente maior do que ao usar o sistema de interação do Meta 2. As respostas aos questionários corroboraram esses resultados, indicando que o método de avaliação foi capaz de prever a percepção humana geral sobre o desempenho de uma determinada técnica de interação.

Palavras-chave: Interação Humano-Computador. Realidade Aumentada. Interação com as mãos. Lei de Fitts.

ABSTRACT

COUTO, J. L. O. **Performance evaluation of hand interaction techniques in immersive Augmented Reality**. 2020. 159 p. Dissertation (Master's degree in Electrical Engineering) – School of Engineering, Sao Paulo University, Sao Paulo, 2020.

The search for an immersive experience in Augmented Reality (AR) has led to a fast development of optical see-through head mounted displays. However, this new computational system lacks a natural interaction technique that can provide fast, accurate and robust interaction like the mouse delivers for 2D monitors. Direct hand interaction is an option, but current hand tracking technologies present limitations that hinders the development of even basic applications. Formal evaluation of interaction performance can help improving them. Fitts' Law and the methods described in ISO 9241-411 are widely applied to evaluate 2D interaction. After a non-systematic literature survey, this work proposed a formal and reproducible method to evaluate performance of direct hand interaction in an AR headset. The 3DFitts software was developed to apply the proposed method, and four experiments were conducted to study how it behaves when evaluating 2D and 3D interactions. Experiments 1 and 2 investigated 2D interaction without and with human participants and aimed at determining if 3DFitts results were statistically equivalent to a reference software from the literature. Experiment 3 was a user study to evaluate interaction performance in a similar setting from Experiment 2, but visualized in 3D through an AR headset named Meta 2. Experiment 4 was also a user study, but focused on discriminating the participants' performance when using six different interaction techniques, which included three interaction devices (a mouse, the Leap Motion Controller and the interaction system embedded on the Meta 2), two selection methods (mouses' left button click and dwell time) and two test plane orientations (frontal and horizontal). These techniques were evaluated by objective metrics (computed by 3DFitts) and subjective ones (gathered from questionnaires applied to participants). Results for Experiments 1 and 2 indicated that the 3DFitts behaves in a similar fashion as the reference software. Comparison between results from Experiments 2 and 3 revealed that the addition of the AR headset had no statistically significant impact on participants' performance. Data from Experiment 4 provided strong quantitative evidence that performance when using the mouse was significantly higher than when using the Leap Motion Controller, which in turn was significantly higher than when using the Meta 2 interaction system. Answers to the questionnaires corroborated these results, indicating that the evaluation method was able to predict the overall human perception about interaction performance.

Keywords: Human-Computer Interaction. Augmented Reality. Hand interaction. Fitt's Law.

LISTA DE ABREVIATURAS E SIGLAS

<i>A</i>	amplitude de movimento
A_e	amplitude efetiva do movimento
<i>D</i>	distância entre alvos
<i>DT</i>	<i>Distance Travelled</i>
<i>ED</i>	<i>Error Distance</i>
<i>ER</i>	taxa de erros
<i>ID</i>	Índice de Dificuldade
ID_e	Índice de Dificuldade Efetivo
<i>IR</i>	<i>Inefficiency Ratio</i>
<i>MCL</i>	<i>Missed Clicks</i>
<i>MT</i>	tempo médio de movimento
<i>TP</i>	<i>throughput</i>
<i>TRE</i>	<i>Target Re-Entry</i>
<i>W</i>	tamanho dos alvos
W_e	tamanho efetivo dos alvos
1D	uma dimensão
2D	duas dimensões
3D	três dimensões
CAVE	<i>Cave Automatic Virtual Environment</i>
CEP	Comitê de Ética em Pesquisa
IHC	interação humano-computador
IMU	Unidade de Medidas Inerciais
LMPF	Controlador Leap Motion avaliado com tempo de permanência no plano frontal

LMPH	Controlador Leap Motion avaliado com tempo de permanência no plano horizontal
MCF	<i>mouse</i> avaliado com clique no plano frontal
MetaTracker	Sistema de Interação com as Mãos do Meta 2
MPF	<i>mouse</i> avaliado com tempo de permanência no plano frontal
MTPF	MetaTracker avaliado com tempo de permanência no plano frontal
MTPH	MetaTracker avaliado com tempo de permanência no plano horizontal
NASA-TLX	<i>NASA Task Load Index</i>
OST-HMD	<i>Optical See-Through Head-Mounted Display</i>
RA	Realidade Aumentada
RV	Realidade Virtual
SDK	<i>Kit de Desenvolvimento de Software</i>
ST	Sequência de Teste
SUS	<i>System Usability Scale</i>

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Contextualização	19
1.2	Motivação	20
1.3	Interação em 3D	21
1.4	Avaliação formal de técnicas de interação	22
1.5	Objeto de estudo	24
1.6	Perguntas norteadoras	25
1.7	Hipóteses	25
1.8	Objetivos	25
1.9	Organização do conteúdo	26
2	REVISÃO BIBLIOGRÁFICA	29
2.1	A Lei de Fitts	29
2.2	Avaliação de dispositivos para interação 2D segundo o padrão ISO 9241-411	30
2.3	Extensão da Lei de Fitts para movimentos em 3D	35
2.4	Avaliação de técnicas de interação com as mãos em 3D	37
2.4.1	Técnicas de interação e <i>feedback</i> visual	38
2.4.2	Forma e disposição dos alvos	42
2.4.3	Procedimentos adotados nos testes	43
2.4.4	Formulações da Lei de Fitts	44
2.4.5	Métricas reportadas	44
3	MÉTODOS DA PESQUISA	47
3.1	Proposta de um método para avaliar técnicas de interação com as mãos em RA	48
3.1.1	Forma e disposição dos alvos	49
3.1.2	Forma do cursor e mecânica de interação	49
3.1.3	Mecânica de teste	50
3.1.4	Formulação da Lei de Fitts	51
3.1.5	Procedimentos gerais	51
3.1.5.1	Planos de teste	51
3.1.5.2	Validação externa por meio da avaliação de um <i>mouse</i>	53
3.1.6	Métricas a serem reportadas	53
3.2	3DFitts: <i>software</i> para aplicar o método de avaliação proposto	54
3.2.1	Arquitetura do <i>software</i>	54
3.2.2	Interface do usuário	56
3.2.3	Configurações possíveis	56

3.2.4	Dispositivos suportados	58
3.2.4.1	Dispositivos para visualização	58
3.2.4.2	Dispositivos para interação	59
3.3	Experimentos conduzidos	59
3.3.1	Autorização do Comitê de Ética em Pesquisa	59
4	EXPERIMENTOS CONDUZIDOS	61
4.1	Experimento 1 - Comparativo entre o 3DFitts e o GoFitts em um cenário sem humanos	61
4.1.1	Método	61
4.1.1.1	Aparato	61
4.1.1.2	Delineamento experimental	61
4.1.1.3	Procedimentos	62
4.1.2	Resultados e discussão	62
4.1.3	Conclusão (Experimento 1)	64
4.2	Experimento 2 - Comparativo entre o 3DFitts e o GoFitts em um cenário com humanos	64
4.2.1	Métodos	64
4.2.1.1	Participantes	64
4.2.1.2	Aparato	65
4.2.1.3	Delineamento experimental	65
4.2.1.4	Procedimentos	66
4.2.2	Resultados e discussão	67
4.2.2.1	Efeitos de aprendizagem	67
4.2.2.2	Efeitos devido ao grupo do participante	68
4.2.2.3	Modelos preditivos pela Lei de Fitts	69
4.2.2.4	Variáveis dependentes	69
4.2.2.5	Características de distribuição estatística	71
4.2.3	Conclusão (Experimento 2)	72
4.3	Experimento 3 - Uso do 3DFitts para transpor o teste multi-direcional 2D para RA	72
4.3.1	Métodos	73
4.3.1.1	Participantes	73
4.3.1.2	Aparato	73
4.3.1.3	Delineamento experimental	74
4.3.1.4	Procedimentos	75
4.3.2	Resultados e discussão	76
4.3.2.1	Efeitos de aprendizagem	76
4.3.2.2	Modelos preditivos pela Lei de Fitts	77
4.3.2.3	Variáveis dependentes	78

4.3.2.4	Características de distribuição estatística	80
4.3.3	Conclusão (Experimento 3)	80
4.4	Experimento 4 - Uso do 3DFitts para avaliar duas técnicas de interação com as mãos em óculos de RA	81
4.4.1	Métodos	81
4.4.1.1	Participantes	82
4.4.1.2	Aparato	82
4.4.1.3	Delineamento experimental	83
4.4.1.4	Procedimentos	86
4.4.2	Resultados e discussão	88
4.4.2.1	Efeitos de aprendizagem	89
4.4.2.2	Efeitos devido ao grupo do participante	89
4.4.2.3	Modelos preditivos pela Lei de Fitts	89
4.4.2.4	Variáveis dependentes	91
4.4.2.5	Características de distribuição estatística	94
4.4.2.6	Questionários de avaliação de conforto	94
4.4.3	Conclusão (Experimento 4)	96
4.5	Discussão geral	98
5	CONCLUSÃO	101
5.1	Retomando as perguntas norteadoras	102
5.2	Principais contribuições	104
5.3	Limitações	104
5.4	Trabalhos futuros	105
	REFERÊNCIAS	107
	APÊNDICE A – QUESTIONÁRIO PARA AVALIAÇÃO DE CONFORTO	113
	ANEXO A – QUESTIONÁRIO PARA AVALIAÇÃO DE CONFORTO DO PADRÃO ISO/TS 9241-411	115
	ANEXO B – MANUSCRITO DE ARTIGO ORIGINAL	117

1 INTRODUÇÃO

1.1 Contextualização

Sutherland (1965) imaginou, em seu famoso texto intitulado *The Ultimate Display* (“o dispositivo de exibição perfeito”, em tradução do autor), um mundo em que o domínio científico e tecnológico sobre a realidade fosse tão elevado que nos permitiria programar a própria realidade. Neste mundo, o dispositivo de exibição perfeito seria, segundo o autor, uma sala dentro da qual um computador teria controle da realidade e na qual objetos virtuais e objetos reais se confundiriam. Em um cenário como esse, seria possível, por exemplo, sentir cheiros e sabores, sentar em uma cadeira virtual e até mesmo ser atingido por uma bala de revólver programados pelo computador.

Tal cenário ainda está mais próximo da ficção científica do que da realidade. Contudo, avanços tecnológicos nos últimos 60 anos permitiram o surgimento de tecnologias com potencial imersivo que nos aproximam da visão de Sutherland, como por exemplo a Realidade Virtual (RV) e Realidade Aumentada (RA).

A RA pode ser definida como sistemas computacionais que permitem suplementar a realidade por meio do registro, nas três dimensões (3D), de objetos virtuais tridimensionais com os quais se possa interagir em tempo real (AZUMA, 1997). As tecnologias de RA poderão levar a aplicações disruptivas nas mais diversas áreas das atividades e interesses humanos, tais como medicina, educação, arquitetura, marketing, entretenimento, indústria, comércio e defesa militar, além do uso pessoal para utilidades diversas como navegação, comunicação e turismo (KREVELEN; POELMAN, 2010; BILLINGHURST; CLARK; LEE, 2015). Apesar de muitos dos problemas fundamentais da RA terem sido resolvidos nas últimas décadas, muitas limitações persistem e precisam ser superadas para que seja possível alcançar seu verdadeiro potencial (BILLINGHURST; CLARK; LEE, 2015).

Atualmente, as experiências mais populares de RA são, provavelmente, aquelas oferecidas por aplicativos para *smartphones* e *tablets* tais como *Pokémon Go*, *Instagram* e *Snapchat*. Elas permitem ao usuário filmar o ambiente e visualizar, pela tela do dispositivo, elementos virtuais sobrepostos à imagem capturada pela câmera, como se pertencessem e interagissem com os elementos do mundo real.

Outra forma de experienciar a RA é por meio do uso de óculos de realidade aumentada, também conhecidos pelo termo em inglês *Optical See-Through Head-Mounted Display* (OST-HMD). Tais dispositivos permitem que o usuário visualize estereoscopicamente informações virtuais tridimensionais projetadas uma lente transparente, aumentando a sensação de imersão. Barreiras tecnológicas mantiveram o interesse a esses óculos restrito a pesquisas acadêmicas até meados de 2010. Nos últimos anos, porém, alguns óculos de RA começaram a ser comercializados

para uso geral, como por exemplo o Microsoft HoloLens (Microsoft, 2020), o Magic Leap One (Magic Leap, 2020) e o Meta 2 (MATNEY, 2016). Não obstante, tais dispositivos ainda são comercializados apenas em versão para desenvolvedores e apresentam limitações como imprecisões no registro espacial dos objetos virtuais, pequeno campo de visualização ou baixa resolução, além de baixa qualidade geral de interação em 3D.

1.2 Motivação

Há cerca de 40 anos, com o aparecimento e popularização dos computadores de mesa, iniciou-se uma profunda revolução social. O surgimento do *mouse*, do teclado e de monitores permitiu exibir e manipular intuitivamente informação digital em duas dimensões (2D), o que fomentou novas possibilidades de criação e consumo de conhecimento que alteraram de modo profundo como o ser humano se organiza, trabalha e se entretém. De forma semelhante, o desenvolvimento e futura popularização dos óculos de RA, junto com outros sistemas computacionais como os de RV, tem o potencial de desencadear uma nova revolução social. Tais tecnologias expandem as possibilidades de interagir com informações virtuais para as três dimensões (3D). O cérebro humano está biologicamente adaptado a compreender e manipular objetos nestas dimensões e, portanto, estas novas tecnologias permitirão que as interações com os sistemas computacionais ocorram de modo cada vez mais imersivo e natural (II; GABBARD, 2005).

A popularização dos computadores de mesa nessas últimas décadas só foi possível porque os seres humanos não-especialistas conseguiram utilizá-los para executar suas tarefas cotidianas. Para que ela ocorresse, portanto, primeiro foi necessário melhorar a qualidade da interação humano-computador (IHC), facilitando a troca de informações — ou seja, a comunicação — entre o usuário e o computador. Já em 1983, Card, Moran e Newell comentavam, no livro “A psicologia da interação humano-computador”, que

“A sociedade está no meio de um processo de auto-transformação para usar o poder dos computadores em todo o seu tecido — onde quer que a informação seja usada — e essa transformação depende criticamente da qualidade da interação humano-computador” (CARD; MORAN; NEWELL, 1983, p. 3, tradução do autor).

De fato, foi apenas a partir do surgimento do teclado e do *mouse*, além dos monitores 2D, que os computadores puderam ser facilmente utilizados por pessoas não familiarizadas com a tecnologia. E foi apenas a partir do uso intuitivo destes sistemas computacionais que as pessoas em geral puderam entender e apreciar os benefícios potenciais que aquelas novas tecnologias poderiam trazer às suas vidas — fator essencial para que o uso destas tecnologias pudesse se difundir por toda a sociedade. Da mesma maneira que a popularização dos computadores de mesa e seus monitores 2D dependeu do surgimento e aperfeiçoamento do teclado e *mouse*,

a popularização dos óculos de RA também dependerá do surgimento e desenvolvimento de formas de interação que facilitem a comunicação entre usuário e o sistema computacional nesse novo meio digital.

1.3 Interação em 3D

Para se tornarem mais intuitivos, ambientes tridimensionais imersivos necessitam de dispositivos que ofereçam interação com seis graus de liberdade, ou seja, que permitam controlar tanto a posição quanto a orientação tridimensional dos objetos virtuais simultaneamente (TEATHER, 2013). Algumas das modalidades de interação que tem sido exploradas incluem controles 3D (ZHAI, 1998; KOUROUPETROGLOU et al., 2012), rastreamento do olhar e movimentos da cabeça (ESTEVES; SHIN; OAKLEY, 2020), e interação gestual (LAVIOLA, 2013). Como cada uma delas terão suas vantagens e desvantagens, é comum combiná-las em soluções multimodais (KREVELEN; POELMAN, 2010). Há vinte anos atrás, Zhai (1998) afirmou que nenhuma das soluções existentes até o momento para prover interação tridimensional poderia ser declarada como o padrão *de facto* equivalente ao que o *mouse* desempenhou para interação 2D. Esta constatação parece soar verdadeira ainda hoje.

Segundo Billinghurst, Clark e Lee (2015), as diferentes interfaces utilizadas para promover a interação humano-computador em ambientes de RA podem ser classificadas em:

- a) navegadores de informação — são muito simples e permitem ao usuário apenas visualizar informações virtuais conforme um contexto real (por exemplo, ver elementos virtuais associados a um objeto que a câmera de um celular está apontando);
- b) interfaces de usuário tridimensionais — em geral envolvem o uso de sensores nas mãos ou de controles manuais, o que dificulta a interação com objetos reais (que é uma das características principais de aplicações RA);
- c) interfaces de usuário tangíveis — envolvem o uso de objetos do mundo real para transmitir e receber informações do sistema de RA, apresentando uma forma de interação muito intuitiva para o cérebro humano;
- d) interfaces de usuário naturais — se beneficiam dos avanços recentes na área de visão computacional para entender os movimentos do próprio corpo do usuário e sua forma natural de se expressar e usá-los como meio de interação com o computador, apresentando as vantagens de reduzir a curva de aprendizagem e de não exigir que o usuário vista sensores;
- e) interfaces multimodais — buscam unir diferentes modos de interação, como por exemplo reconhecimento de fala e de gestos, para fornecer uma experiência mais rica de interação.

As interfaces de usuário naturais, por não exigirem o uso de controles e sensores

vestíveis, se adequam bem à proposta tecnológica dos óculos de RA portáteis e móveis: tornar a RA uma tecnologia tão ubíqua quanto os computadores e *smartphones*. Um usuário, por exemplo, poderia caminhar pela rua utilizando o seu óculos e receber informações virtuais conforme necessário no contexto do mundo real. Para interagir com o sistema, ele poderia apenas levantar as mãos — sem precisar vestir nada nem carregar nenhum controle — e usar gestos que comumente usaria para se comunicar com uma outra pessoa, como por exemplo apontar o dedo indicador, traçar trajeteto no ar, apontar o polegar para cima ou para baixo ou se despedir balançando a mão. A complementação com reconhecimento de voz pode tornar tal interação ainda mais simples e flexível. Interfaces naturais também são intuitivas para o usuário e podem permitir a transposição direta de conhecimento de como interagir com objetos em 3D no mundo real para o mundo virtual.

Talvez em parte por esses motivos, um dos tópicos de pesquisa que tem recebido maior atenção na última década é o de reconhecimento de gestos manuais (Al-Shamayleh et al., 2018). Adicionalmente, todos os óculos de RA mencionados anteriormente possuem um sistema de interação por gestos que é capaz de rastrear as mãos e reconhecer gestos simples. Também é possível encontrar outros dispositivos de interação gestual independentes, como o Controlador Leap Motion (Ultraleap, 2020). Apesar desses dispositivos de interação gestual já estarem sendo comercializados, o rastreamento das mãos e o reconhecimento dos gestos ainda é falho e a qualidade de interação é baixa. Estudos evidenciaram que a interação por rastreamento das mãos desempenham significativamente pior que interação com *mouse* e *touchscreen* em tarefas simples de apontar e selecionar alvos em 2D (TEATHER; STUERZLINGER, 2008; SAMBROOKS; WILKINSON, 2013; BURNO et al., 2015).

Mesmo considerando este baixo desempenho e a clara desvantagem de provocar maior cansaço muscular, a busca por melhor tecnologia para permitir interação com as mãos talvez se mantenha por ter um potencial de prover interação direta com seis graus de liberdade e co-localizada com o espaço motor do usuário. Além de ser mais intuitiva e reduzir a curva de aprendizagem, a co-localização aumenta o sentimento de imersão uma vez que os objetos virtuais são apresentados no mesmo local espacial em que o usuário realiza seus movimentos de interação (TEATHER, 2013).

1.4 Avaliação formal de técnicas de interação

Nesta dissertação, o termo “técnicas de interação” foi escolhido para referenciar todo o sistema que determina diretamente como a interação com o sistema computacional se dará. Ou seja, ele se refere ao conjunto de elementos que afetam a forma com que o usuário interage com o sistema, tais como: o dispositivo responsável para traduzir ações humanas em informações digitais (por exemplo, um *mouse*), o método de seleção ou de entrada de comandos (por exemplo, um clique com o botão esquerdo do *mouse*) e os algoritmos implementados para interpretar

tais informações de interação (por exemplo, uma configuração no sistema operacional que determina a velocidade do cursor do *mouse* na tela).

A avaliação formal do desempenho que uma determinada técnica de interação permite alcançar na realização de uma tarefa é um importante passo para auxiliar no seu processo de desenvolvimento e aprimoramento. Adicionalmente, avaliações formais de diferentes técnicas de interação possibilitam realizar comparativos para descobrir qual delas oferece melhor desempenho para um determinado tipo de aplicação. As métricas de desempenho usualmente estudadas são o tempo médio gasto para completar uma tarefa e a taxa de erros cometidos durante a execução. A primeira informa a velocidade de resposta que a modalidade de interação permite, enquanto a segunda permite revelar sua precisão média.

Estudos formais avaliando desempenho de interação em 2D são abundantes e ocorrem desde a década de 1970, quando CARD, ENGLISH e BURR (1978) publicaram um estudo avaliando o desempenho de diferentes dispositivos (*mouse*, *joystick* e teclado) em tarefas de seleção de texto. Por meio de um método de avaliação formal, este estudo identificou, por métricas objetivas, a superioridade do *mouse* sobre os demais dispositivos de interação tanto em termos de velocidade quanto de precisão. Além disso, por meio da aplicação da Lei de Fitts – uma formulação empírica que permite prever o tempo gasto para realizar determinados movimentos humanos – eles puderam concluir que a movimentação do *mouse* ocorreu com uma taxa próxima da máxima predita para movimentos humanos.

Por três décadas a falta de um método padronizado dificultou a comparação confiável entre avaliações formais realizadas por diferentes pesquisadores (SOUKOREFF; MACKENZIE, 2004). Este problema foi resolvido com o surgimento, no ano 2000, do padrão ISO 9241-9 – “*Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices*” (International Organization for Standardization, 2000) – que estabeleceu uma padronização do método e das formulações a serem utilizadas na avaliação de dispositivos para interação 2D (SOUKOREFF; MACKENZIE, 2004). Em 2012, o documento ISO/TS 9241-411:2012 – “*Ergonomics of human-system interaction - Part 411: Evaluation methods for the design of physical input devices*” (International Organization for Standardization, 2012) – atualizou e substituiu o documento ISO 9241-9:2000.

MacKenzie (2017), ao defender o uso deste método padronizado para avaliação da IHC em 2D, argumentou que a padronização é importante porque fortalece comparações entre diferentes estudos. Ele fornece três razões para defender esse argumento:

- a) o método de testes passa a ser consistente entre estudos;
- b) o padrão sugere um dispositivo representativo para estabelecer uma “linha-base” comparativa entre os estudos, permitindo comparar quão bem o mesmo dispositivo se sai entre diferentes estudos e estabelecendo níveis de confiança para as avaliações dos demais dispositivos;
- c) a métrica de comparação utilizada é a mesma – o *throughput* – e inclui tanto a

velocidade quanto a acurácia do desempenho do usuário, permitindo normalizar diferenças comportamentais entre diferentes usuários e diferentes estudos.

As interfaces e técnicas para interação em 3D tem sido pesquisadas e desenvolvidas desde a década de 1960. Apesar disso, métodos de avaliação não receberam muita atenção até recentemente. Nas palavras de Bowman et al. (2005),

“Avaliação tem sido, com frequência, o componente ausente das pesquisas em interação em 3D. Por muitos anos, os campos de Ambientes Virtuais e interfaces de usuário em 3D foram tão novos e as possibilidades tão ilimitadas que muitos pesquisadores simplesmente focaram no desenvolvimento de novos dispositivos, técnicas de interação e metáforas de interface de usuário — explorando o espaço de desenho — sem gastar muito tempo para avaliar quão bons estes novos projetos eram. À medida em que os campos amadurecem, entretanto, estamos observando mais de perto a usabilidade. Devemos analisar criticamente, avaliar e comparar dispositivos, técnicas de interação, interfaces de usuário e aplicações se as interfaces de usuário em 3D se destinam a ser utilizadas no mundo real”(BOWMAN et al., 2005, p. 349, tradução do autor).

Apesar da importância da padronização de métodos de avaliação, ainda não há um padrão estabelecido para a avaliação de dispositivos em interação 3D (RAYNAL; DUBOIS; SCHMITT, 2013; NIEUWENHUIZEN, 2015). Estabelecer meios apropriados para medir e comparar diferentes técnicas para interações em 3D é um passo necessário para se determinar qual dispositivo de interação será eficiente e bem aceito para contribuir para o avanço do desenvolvimento das novas técnicas avançadas de interação tridimensional (BACHMANN; WEICHERT; RINKENAUER, 2018).

1.5 Objeto de estudo

A presente pesquisa pretende contribuir para o desenvolvimento e aprimoramento de dispositivos e formas de interação para óculos de RA investigando formas de se avaliar formalmente tais interações. Especificamente, o foco da pesquisa será na avaliação da interação direta com as mãos, considerando a relevância que tal modalidade tem tido como forma de interagir com tais sistemas computacionais.

Faz-se necessário esclarecer que, neste trabalho, ao se mencionar ambientes tridimensionais imersivos ou RA imersiva (inclusive em seu título), está-se referindo especificamente à experiência estereoscópica, com foco principalmente na providenciada por óculos de RA ou RV. O leitor deve ser alertado, entretanto, que imersão não é um conceito binário, mas sim uma escala – à medida que uma experiência digital se aproxima cada vez mais da realidade, mais imersiva ela é.

1.6 Perguntas norteadoras

As seguintes perguntas guiarão o desenvolvimento deste trabalho:

- a) Quais métodos tem sido utilizados para avaliar o desempenho de técnicas de interação em 2D e em 3D?
- b) Quais métodos tem sido utilizados para avaliar o desempenho de técnicas de interação direta com as mãos em ambientes tridimensionais imersivos?
- c) Como avaliar o desempenho da interação direta com as mãos em um óculos de RA?

1.7 Hipóteses

Um dos motivadores pessoais do autor para escolher este tema de pesquisa durante seu curso de mestrado foi a experiência pessoal de, ao buscar desenvolver aplicações para um óculos de RA (Meta 2), se sentir frustrado com o baixo desempenho da interação com as mãos fornecido pelo dispositivo. A experiência de usuário foi tão sofrível que impediu a realização até de tarefas simples como capturar e soltar um objeto virtual. Na tentativa de solucionar este problema, o autor adaptou os óculos e a aplicação que estava sendo desenvolvida para permitir interação não pelo sistema de interação dos óculos mas por um dispositivo externo – o Controlador Leap Motion. A interação com as mãos permitida por tal dispositivo ainda provocou frustração, mas na percepção subjetiva do autor esta segunda técnica parecia oferecer uma experiência mais fluida que a primeira. A partir desta impressão, o autor decidiu testar as seguintes hipóteses, que acabou por direcionar todo o trabalho desenvolvido:

- H1. O desempenho da interação com as mãos no Meta 2, em tarefas simples de seleção de objetos virtuais, é *objetivamente melhor – segundo métricas quantitativas* – quando executadas com o Controlador Leap Motion do que quando executadas com o sistema de interação do próprio Meta 2.
- H2. O desempenho da interação com as mãos no Meta 2, em tarefas simples de seleção de objetos virtuais, é *percebido subjetivamente como sendo melhor* quando executadas com o Controlador Leap Motion do que quando executadas com o sistema de interação do próprio Meta 2.

1.8 Objetivos

Esta pesquisa tem como objetivo geral avaliar o desempenho de técnicas de interação com as mãos em um óculos de RA, quando utilizados para interagir com objetos virtuais que

estejam ao alcance das mãos.

Para atingir esse objetivo geral, será necessário observar os seguintes objetivos específicos:

- a) propor um método de avaliação para técnicas de interação com as mãos em óculos de RA baseado em princípios científicos e na literatura científica existente;
- b) desenvolver um software para óculos de RA que implemente o método de avaliação proposto, sendo capaz de executar os testes com diferentes técnicas de interação e coletar as métricas necessárias para medir o desempenho da interação;
- c) realizar um estudo de caso com participantes humanos, utilizando o software desenvolvido, para verificar se o método tem poder discriminatório e pode ser utilizado como ferramenta de avaliação formal do desempenho de técnicas de interação com as mãos para óculos de RA. Para isso, é importante que:
 - seja obtida uma autorização de um Comitê de Ética em Pesquisa para realizar os estudos com seres humanos;
 - os testes sejam executados também com um dispositivo de interação que seja considerado "padrão-ouro" na literatura, permitindo comparar os resultados desta pesquisa com a literatura;
 - se avalie a opinião dos participantes sobre cada técnica de interação, permitindo analisar se a percepção subjetiva geral corrobora os resultados obtidos quantitativamente pelo *software*.
- d) verificar se as hipóteses que originaram o trabalho são verdadeiras.

1.9 Organização do conteúdo

O conteúdo desta dissertação está organizado em cinco capítulos. Neste primeiro capítulo, contextualizou-se o assunto e justificou-se a importância de se avaliar técnicas de interação direta com as mãos para auxiliar no avanço das tecnologias que dão suporte à RA.

No Capítulo 2, a revisão bibliográfica que embasou as decisões posteriores será apresentada. Discutir-se-á a Lei de Fitts, sua aplicação no padrão ISO 9241-411 e tentativas de estendê-la para movimentos em 3D. Também serão apresentados os resultados de uma pesquisa bibliográfica não sistemática de trabalhos que promoveram avaliação de interação 3D em configurações similares às requeridas em um óculos de RA.

A partir dos resultados da revisão, decidiu-se propor um método de avaliação da interação direta com as mãos em óculos de RA e utilizá-lo em experimentos formais para verificar sua capacidade de discriminar objetivamente dois sistemas de interação por rastreamento das mãos. Os detalhes deste método de avaliação e demais procedimentos para atingir esse objetivo estão descritos no Capítulo 3.

O Capítulo 4 apresenta os métodos, resultados e discussão para cada experimento conduzido durante esta pesquisa, sendo encerrado com uma discussão geral que resume os resultados obtidos em todos eles.

Por fim, o Capítulo 5 conclui esta dissertação resumizando o trabalho desenvolvido, elencando suas principais contribuições e sugerindo trabalhos futuros.

Este trabalho de pesquisa deu origem a um artigo original que foi submetido para apreciação ao corpo editorial do *International Journal of Human-Computer Interaction* (ISSN 1071-5819). Até o momento de depósito desta dissertação o processo de análise ainda não havia sido concluído. Como este documento reproduz material primeiramente produzido para o artigo, referências serão feitas ao manuscrito quando for pertinente. O manuscrito submetido para análise foi incluído no Anexo B.

2 REVISÃO BIBLIOGRÁFICA

2.1 A Lei de Fitts

O Teorema 17 da Teoria da Comunicação de Shannon (1948, p. 43) expressa que (tradução do autor) “a capacidade [de transmissão de informação C] de um canal [analógico] de banda B perturbado por um ruído branco aditivo Gaussiano de potência N quando a potência média do sinal transmitido é limitada por S é dada por

$$C = B \log_2 \frac{S + N}{N} \text{ ”} \quad (2.1)$$

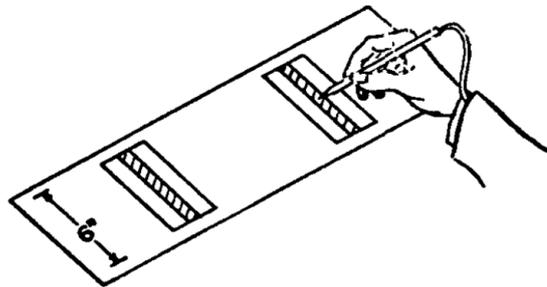
Fitts (1954) se inspirou neste teorema para modelar a capacidade informacional do sistema motor humano. O racional por trás da proposta de Fitts é que o sistema motor humano tem uma capacidade fixa para transmissão de informação motora, enquanto os diferentes movimentos requerem diferentes quantidades de informação motora para serem realizados conforme desejado. Movimentos que requerem maior amplitude e/ou maior precisão, portanto, necessitam de maior de informação de movimento para serem realizados com sucesso. Como a capacidade de transmissão de um determinado ser humano é constante, movimentos que requerem mais informação necessitarão de mais tempo para serem concluídos.

Fitts realizou experimentos em que solicitou a participantes humanos que tocassem alvos repetidamente e na maior velocidade possível, realizando apenas movimentos laterais, conforme ilustrado na Figura 1. Segundo ele, ao requerer movimentos repetitivos e rápidos, reduz-se a influência de outros sistemas humanos – como por exemplo os sistemas sensores e neurais – na resposta final, maximizando a influência do sistema motor, que era seu objeto de estudo ¹. Para modelar a precisão requerida por um movimento, Fitts propôs controlar o tamanho dos alvos. A amplitude do movimento foi controlada variando-se a distância entre os alvos.

Os resultados empíricos obtidos nestes experimentos indicaram que a quantidade de informação necessária para concretizar um movimento direcionado podia ser expressa com o logaritmo da razão entre o tamanho dos alvos e a distância entre eles. Esta relação foi chamada por ele de Índice de Dificuldade (ID). Adicionalmente, os dados obtidos por Fitts indicaram uma relação linear entre a duração de um movimento e o ID associado. Eventualmente, esta

¹ Fitts deixa claro que, ao se referir ao sistema motor, também está considerando os sistemas visual e proprioceptor que compõe o laço de *feedback* que permite a um ser humano monitorar seu próprio movimento

Figura 1 – Representação de um dos experimentos conduzidos por Fitts em seu artigo original que estabeleceu as bases para a relação que hoje é conhecida como Lei de Fitts.



Fonte: Fitts (1954).

relação passou a ser conhecida como Lei de Fitts, e pode ser expressa como:

$$MT = a + b \times ID \quad (2.2)$$

onde:

MT é o tempo médio do movimento, em segundos;

a é o coeficiente linear obtido por regressão linear, expresso em s ;

b é o coeficiente angular obtido por regressão linear, expresso em $s/bits$.

Apesar de originalmente proposta para movimentos uni-direcionais, a Lei de Fitts tem sido extensivamente utilizada por pesquisadores de IHC como ferramenta para avaliar técnicas de interação tanto para movimentos em uma dimensão (1D) quanto em 2D (SOUKOREFF; MACKENZIE, 2004). O trabalho original de Fitts propôs uma expressão para ID ligeiramente diferente da forma logarítmica da Equação 2.1. Desde então, revisões posteriores por diversos autores sugeriram diferentes variações para aprimorar o modelo inicial proposto por Fitts. Segundo Soukoreff e MacKenzie (2004, p. 753, tradução do autor), isso resultou em uma “aparente multiplicidade sem fim de variações da lei, tornando grande parte da literatura inconsistente e incomparável”. O surgimento e a difusão do padrão ISO 9241-9 (International Organization for Standardization, 2000) sanou este problema.

2.2 Avaliação de dispositivos para interação 2D segundo o padrão ISO 9241-411

Conforme mencionado no Capítulo 1, o documento ISO 9241-411 (International Organization for Standardization, 2012) atualizou e substituiu o documento da ISO 9241-9 (International Organization for Standardization, 2000), que originalmente estabeleceu métodos padronizados

para avaliar dispositivos de entrada para interação em 2D. O documento atualizado manteve as recomendações originais para a seção de métodos de avaliação e, portanto, neste texto estes documentos serão mencionados de forma intercambiável.

A ISO 9241-411 estabelece métodos de teste que permitem avaliar o desempenho de um dispositivo de interação na execução de tarefas primitivas de interação, tais como apontar, selecionar, arrastar e traçar. Interações mais complexas em geral são construídas a partir da combinação destas tarefas primitivas. Um dispositivo ou técnica de interação podem ser avaliados em uma ou mais destas tarefas, conforme for necessário para melhor caracterizá-los, considerando o uso a que se destinam.

Um dos testes propostos pela norma e muito utilizado por pesquisadores é o de toque (*tapping test*), que doravante será chamado de “apontar-e-selecionar” por ser um termo que indica mais claramente as ações interativas envolvidas. Neste tipo de teste, o usuário deve executar tarefas de apontar (mover o cursor até um alvo) e selecionar (realizar uma ação para indicar que se deseja selecionar um alvo, como por exemplo clicar em um botão). O padrão descreve dois tipos de testes de apontar-e-selecionar: uni e multi-direcional.

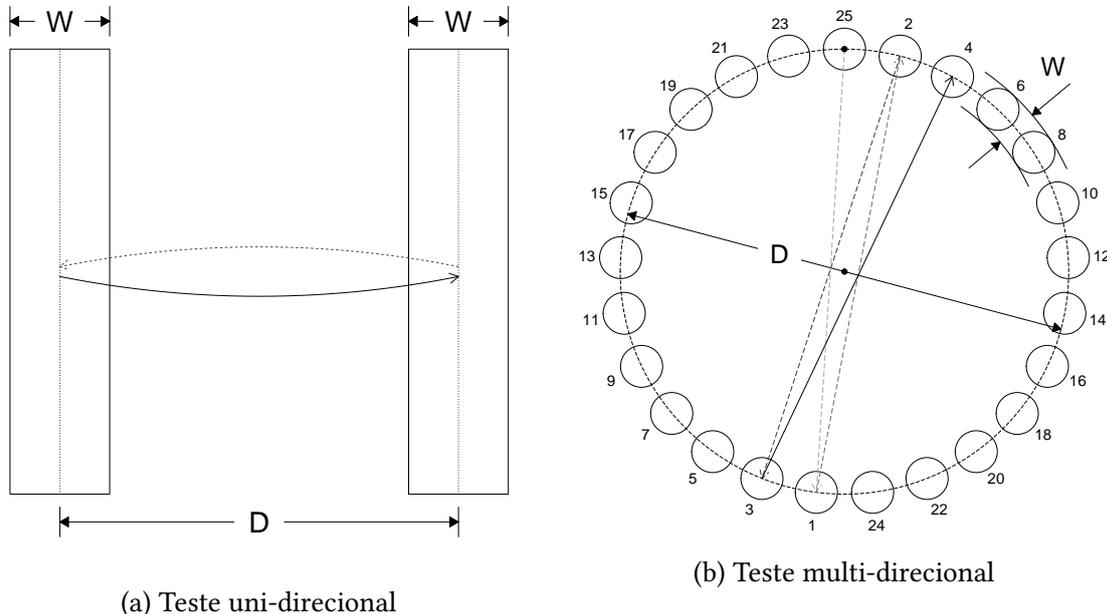
O teste uni-direcional é utilizado quando se deseja avaliar o desempenho do dispositivo em um eixo de interação (horizontal ou vertical). Conforme ilustrado na Figura 2a, o participante é orientado a selecionar dois alvos alternadamente, na maior velocidade possível, mas evitando cometer erros. Um erro é computado se um evento de seleção ocorrer quando o cursor de interação estiver fora do alvo. Se um erro ocorrer, o teste deve prosseguir normalmente. O padrão ISO sugere que uma sessão de teste envolva um total de 25 seleções de alvos.

O teste multi-direcional, ilustrado na Figura 2b, é recomendado quando se está interessado em avaliar o desempenho geral de um dispositivo ou técnica de interação em diferentes direções bidimensionais. Ele segue a mesma mecânica do teste unidirecional, com a diferença de que agora o usuário deverá selecionar alvos executando movimentos em diversas direções. Os alvos são dispostos em uma circunferência, espaçados de forma homogênea. A sequência de seleção se inicia por um dos alvos (círculo 25 na Figura 2b). O próximo alvo sempre será aquele do lado oposto ao último alvo selecionado, de modo que todos os alvos sejam selecionados apenas uma vez à medida que se escolhe alvos sempre no sentido horário (conforme setas ilustrativas apresentadas na Figura 2b). O padrão ISO sugere, de modo não explícito, que uma sessão de teste consista na seleção de 25 alvos.

Na literatura, estes tipos de teste que requerem movimentos rápidos e repetitivos que se iniciam a partir do final do último movimento e prosseguem na direção oposta são chamados de testes de movimentos recíprocos (ou seriais). Também existem os testes de movimentos discretos, no qual apenas um movimento é realizado por vez, geralmente iniciado a partir de um mesmo ponto de partida e somente após um estímulo sonoro e/ou auditivo.

Em qualquer um destes testes, a ação de “selecionar” um alvo dependerá da técnica de interação em teste, podendo variar tanto em termos de *hardware* quanto de *software*. Por

Figura 2 – Representação dos testes de apontar-e-selecionar uni-direcional e multi-direcional propostos pela ISO 9241-411. W denota o diâmetro dos alvos e D , o diâmetro da circunferência em que os alvos são posicionados.



(a) Teste uni-direcional

(b) Teste multi-direcional

Fontes: (a) adaptado de Soukoreff e MacKenzie (2004); (b) Couto, José e Lopes (2020), adaptado de Soukoreff e MacKenzie (2004).

exemplo, para um *mouse* essa ação poderia ser um clique em um botão, mas também poderia ser uma seleção automática, na qual o alvo é selecionado assim que o cursor o alcança.

Um outro teste estabelecido pela ISO 9241-411 é o de “arrastar” (*dragging test*). Ele é recomendado quando se deseja avaliar o desempenho de um dispositivo para executar tarefas de clicar-arrastar-soltar, que são a base de interações como, por exemplo, arrastar um ícone para a lixeira ou redimensionar uma imagem. Para este tipo de tarefa, o padrão recomenda adaptar os testes de apontar-e-selecionar unidirecional ou multi-direcional para que, ao invés de simplesmente selecionar um alvo, o usuário tenha que arrastá-lo até o próximo alvo.

O objetivo final destes testes é obter dados para computar uma métrica denominada *throughput* (TP), proposta para quantificar o desempenho geral da técnica de interação na realização do teste em estudo. Para os testes de apontar-e-selecionar e arrastar, o cálculo do TP se baseia na formulação de Shannon para a Lei de Fitts proposta por Soukoreff e MacKenzie (2004), na qual o ID é expresso por:

$$ID = \log_2 \left(\frac{A}{W} + 1 \right) \quad (2.3)$$

onde:

ID é o Índice de Dificuldade, em *bits*;

A é a amplitude de movimento requerida entre dois alvos do teste;

W é a largura dos alvos, medida no eixo de aproximação do movimento.

Em um teste uni-direcional, o valor da amplitude de movimento (A) será estritamente igual à distância entre alvos (D) expresso na Figura 2a. Para o teste multi-direcional, contudo, ele será ligeiramente diferente do diâmetro da circunferência ao longo da qual os alvos são dispostos. Ainda assim, é comum desprezar essa diferença e utilizar A igual a D também para o teste multi-direcional, conforme é possível inferir do trabalho de Soukoreff e MacKenzie (2004).

O ID pode ser entendido como uma medida que indica quão difícil é para um ser humano concluir um movimento rápido e repetitivo. Movimentos com baixo ID possuem amplitude pequena ou que não exigem grande precisão (por exemplo, encostar alternadamente em dois pratos posicionados um do lado do outro), enquanto movimentos com altos valores de ID apresentam amplitudes maiores ou exigem maior precisão (por exemplo, encostar alternadamente em duas moedas de dez centavos distantes 1 metro entre si). Naturalmente, quanto maior for o ID , maior será o tempo necessário para completar o movimento com sucesso.

Em um teste típico de avaliação, tais como os da Figura 2, o participante deve realizar repetidas tarefas que envolvem mover o cursor até diferentes alvos e interagir com eles. Isso deve ser feito para diferentes valores de ID , ou seja, para diferentes pares de D e tamanho dos alvos (W). O padrão ISO recomenda que o pesquisador escolha valores de D e W suficientes para que seja possível obter um número de ID s suficiente para representar as condições de uso do dispositivo (valores que variem entre 2 a 8 *bits* é recomendado).

Para quantificar o desempenho final do participante em um teste, deve-se coletar, a cada interação realizada, o tempo e as coordenadas do cursor no momento em que uma interação é finalizada. Para o teste de apontar-e-selecionar, isto acontece quando o usuário realiza uma ação de seleção (mesmo quando o cursor estiver fora do alvo), disparando as seguintes ações:

- a) a posição atual do cursor deve ser salva como as coordenadas da interação;
- b) o tempo transcorrido desde o início da interação deve ser salvo como tempo da interação;
- c) o próximo alvo deve ser ativado;
- d) uma nova interação deve ser iniciada.

De modo similar, para o teste de arrastar o fim da interação acontecerá quando o usuário soltar o objeto virtual que estava arrastando (mesmo que não tenha alcançado o alvo final). Em decorrência disso, as seguintes ações acontecem:

- a) a posição atual do cursor deve ser salva como as coordenadas da interação;
- b) o tempo transcorrido desde o início da interação deve ser salvo como tempo da interação;
- c) o alvo que estava sendo arrastado deve reaparecer na sua posição original;

- d) o próximo alvo a ser arrastado deve ser ativado;
- e) o próximo alvo a ser o destino final da interação de arrastar deve ser ativado;
- f) uma nova interação deve ser iniciada.

O tempo médio de movimento (MT) de um participante em um teste é computado por meio da média dos tempos de cada interação realizada, incluindo os casos em que erros foram cometidos (MACKENZIE, 2015). A variabilidade do movimento pode ser analisada a partir das coordenadas de seleção de cada interação. Para isso, o Índice de Dificuldade Efetivo (ID_e) do teste deve ser calculado utilizando a equação:

$$ID_e = \log_2 \left(\frac{A_e}{W_e} + 1 \right) \quad (2.4)$$

onde:

ID_e é o Índice de Dificuldade Efetivo, em *bits*;

A_e é a amplitude efetiva do movimento projetada no eixo de movimento; W_e é o tamanho efetivo dos alvos.

Repare que a ISO 9241-411 sugere o uso de A ao invés da amplitude efetiva do movimento (A_e). O uso de A_e geralmente não tem impacto relevante no resultado final. Mesmo assim, seu uso é recomendado para prevenir casos em que as coordenadas de interação não estejam distribuídas homogeneamente ao redor do centro do alvo (MACKENZIE, 2015).

O ID_e é uma correção do ID realizada para refletir a dificuldade real da tarefa ao ser executada por um usuário, normalizando a taxa de erros cometidos entre participantes. Por exemplo, um usuário que prefira adotar movimentos mais rápidos tenderá a ser menos preciso, o que pode indicar que a tarefa real executada por ele teve um W maior do que o esperado inicialmente. O tamanho efetivo dos alvos (W_e) é calculado de modo a garantir que 96% de todas as amostras coletadas no teste estejam dentro de um alvo cuja largura seja W_e (SOUKOREFF; MACKENZIE, 2004). Para computar este valor, utiliza-se a equação

$$W_e = 4,133 \times SD_x \quad (2.5)$$

onde:

SD_x é o desvio padrão das coordenadas de interação das amostras coletadas no teste, na direção do eixo de movimento².

Um vez computados os valores de ID_e e de posse do MT , é possível calcular o valor de TP , em *bits/s*, que indicará o desempenho geral dos participantes ao utilizar a técnica de

² Veja MacKenzie (2015) para maiores detalhes em como calcular tanto A_e quanto SD_x .

interação estudada:

$$TP = \sum_{j=1}^{N_p} \left(\sum_{i=1}^{N_s} \frac{ID_{e_{ij}}}{MT_{ij}} \right) \quad (2.6)$$

onde:

N_s é o número de sequências de testes realizadas por um participante; N_p é o número de participantes no estudo.

O valor geral de TP nada mais é que a média geral dos valores de TP obtidos para cada participante individualmente. Esta métrica só pode ser considerada confiável caso o movimento dos participantes tenha obedecido a Lei de Fitts. Por isso, é importante aplicar a Equação 2.2, realizando uma regressão linear utilizando os pares de pontos (MT , ID_e) obtidos no experimento. Se a variabilidade observada nos dados não for bem explicada por uma relação linear, é possível que o movimento de um ou mais participantes não tenha obedecido aos requisitos previstos na formulação da lei: serem rápidos, direcionados e sem pausas. Uma análise dos dados por participante pode revelar aqueles cujos movimentos não conformaram com a Lei de Fitts, que poderão ser removidos do cálculo final do TP .

O desempenho dos movimentos humanos é caracterizado por uma relação de dependência mútua entre a velocidade e precisão: quanto mais rápido um participante tentar selecionar os alvos, menos preciso ele será. Como o TP da técnica de interação é computado por meio de média geral dos participantes em diferentes condições de testes, ele pode ser visto como uma medida estatística que é obtida a partir da compressão de várias medidas de tempo de movimento e que combina tanto as características de precisão quanto de velocidade de uma técnica de interação (SOUKOREFF; MACKENZIE, 2004). Por causa disso, o TP é uma métrica útil para comparar resultados entre diferentes estudos.

Além do TP , é recomendado que pesquisadores também reportem os valores médios de MT e da taxa de erros (ER) cometidos ao longo das interações, obtida pela relação entre o número de interações que não foram bem sucedidas (isto é, alvos que não foram corretamente selecionados) e o total de interações realizadas (SOUKOREFF; MACKENZIE, 2004).

2.3 Extensão da Lei de Fitts para movimentos em 3D

A Lei de Fitts originalmente foi proposta para modelar movimentos em 1D, e posteriormente foi usada com sucesso para modelar também movimentos em 2D. Porém, quando aplicada a movimentos genéricos em 3D, ela geralmente resulta em uma pior correlação entre MT e ID_e . Por causa disso, pesquisadores tem proposto estender a versão original modificando ou adicionando novos termos tanto à formulação do ID quanto ao modelo linear.

Atsuo Murata e Hirokazu Iwase (2001) conduziram um experimento em que participantes realizaram interações discretas com círculos físicos dispostos ao redor de uma circunferência, a qual foi disposta em um plano frontal e centralizada no ombro da mão dominante. Os autores observaram que o MT varia de maneira senoidal com o ângulo de azimute do alvo, isto é, quanto mais alto a posição de um alvo, mais tempo se levou para interagir com ele. Assim, eles propuseram adicionar um novo termo – o seno do ângulo de azimute – à formulação do ID , modelando essa variação da dificuldade com a direção do movimento.

Grossman e Balakrishnan (2004) conduziram um experimento em uma tela volumétrica, na qual dois paralelepípedos com variadas dimensões foram apresentados em um plano horizontal. Os participantes deveriam realizar interações recíprocas entre os dois alvos utilizando um controle com seis graus de liberdade. Cinco modelos para ID foram discutidos e testados. As modificações incluíam dimensões dos alvos e o ângulo de movimento.

Cha e Myung (2013) conduziram um experimento no qual os participantes usaram suas mãos para interagir de modo discreto (não-serial) com placas metálicas quadradas que foram dispostas à sua frente em uma pirâmide com oito lados e quatro níveis de profundidade. O aparato permitia que os pesquisadores gravassem o tempo de interação conforme o participante tocava nas placas. Os autores observaram que os ângulos de inclinação e azimute tiveram efeitos estatisticamente significativos no MT e confirmaram o padrão senoidal reportado anteriormente por Atsuo Murata e Hirokazu Iwase (2001). Porém, ao invés de proporem uma modificação na formulação do ID , eles sugeriram um modelo linear com dois novos termos para representar os ângulos de inclinação e o azimute.

Machuca e Stuerzlinger (2019) estudaram o efeito de telas estereoscópicas no desempenho dos usuários ao interagir com sistemas de RV e propuseram uma versão estendida da Lei de Fitts para este tipo de aplicação. Neste estudo, os participantes executaram movimentos recíprocos selecionando duas esferas posicionadas à altura dos olhos e dispostas nos eixos lateral e frontal. Esta tarefa foi executada com óculos de RV e também em uma versão com alvos físicos. Eles observaram que o desempenho dos participantes no sistema de RV foi inferior, justificando que isso se deve provavelmente à uma pior percepção de profundidade devido às deficiências da tela estereoscópica. Para considerar estes efeitos em avaliações formais de interação em sistemas estereoscópicos, eles propuseram adicionar um novo termo (*Change of Depth*) ao modelo linear da Lei de Fitts.

Clark, Bhagat e Riggs (2020) também propuseram recentemente uma versão estendida da Lei de Fitts para sistemas de RV. Contudo, diferentemente de Machuca e Stuerzlinger (2019), o experimento que eles conduziram exigiu movimentos em todas as 3D. Alvos cúbicos foram dispostos em um volume cúbico e os participantes executaram movimentos discretos para selecioná-los em ordem aleatória, utilizando para isso um controle manual com seis graus de liberdade. Os resultados indicaram que o ângulo de inclinação e a interação entre o tamanho de um alvo e o ângulo de inclinação tiveram efeitos estatisticamente significativos no MT . Os autores argumentaram que isto provavelmente ocorre porque o tamanho visual de um

alvo depende tanto do ângulo de inclinação quanto do tamanho real do alvo – um alvo menor, mas mais próximo do olho, por exemplo, poderá ser percebido como sendo visualmente maior do que um alvo distante, mas maior que este primeiro. Contrariando os estudos anteriores (Atsuo Murata; Hirokazu Iwase, 2001; CHA; MYUNG, 2013), os resultados não indicaram efeito significativo do azimute no *MT*. A versão final proposta para estender a Lei de Fitts para interação em sistemas de RV incluiu no modelo linear tanto o ângulo de inclinação quanto o produto entre o tamanho e o ângulo de inclinação dos alvos.

Estes cinco estudos ilustram a complexidade de modelar movimentos em 3D: em cada um deles, os movimentos são restritos a certas regiões espaciais; um deles utilizou interação direta com as mãos enquanto outros usaram controles manuais ou tecnologias para rastrear o dedo dos participantes; alguns usaram alvos físicos no mundo real enquanto outros estudaram interação especialmente em ambientes virtuais com tela estereoscópica; alguns utilizaram movimentos discretos enquanto outros usaram movimentos recíprocos. Cada uma destas variáveis pode afetar o resultado final e assim, limitar a aplicabilidade dos modelos propostos.

Adicionalmente, apesar de todos os modelos sugeridos terem efetivamente culminado em uma melhor correlação entre a variável dependente e as variáveis independentes, Teather e Stuerzlinger (2013) argumenta que a adição de novos parâmetros livres a qualquer análise de regressão *sempre* melhorará a correlação final, fato que dificulta determinar qual foi a real contribuição dos novos parâmetros para o poder preditivo do modelo. Estas consideração possivelmente explicam porque até o momento nenhum modelo estendido da Lei de Fitts se tornou amplamente aceito e utilizado para avaliar interações em 3D.

2.4 Avaliação de técnicas de interação com as mãos em 3D

A área de IHC como um todo possui muitos métodos de avaliação de usabilidade estabelecidos para interfaces em 2D, tais como métodos de inspeção (avaliação heurística, revisão de *guidelines*, inspeção de consistência e percurso cognitivo) e testes de usabilidade com usuários e coleta de medidas de desempenho (ROCHA; BARANAUSKAS, 2000). Parte destes métodos naturalmente foram utilizados para avaliação no campo da interação em 3D, especialmente para avaliação de aplicações, interfaces gráficas e técnicas de interação, tais como avaliação heurística, percurso cognitivo, avaliações formativa e somativa, questionários e entrevistas (BOWMAN et al., 2005).

Bowman et al. (2005) apontaram uma ausência de trabalhos que usavam modelos preditivos, como a Lei de Fitts, para avaliar técnicas de interação em 3D, e sugeriram esta como uma área de interesse para pesquisas futuras. Desde então, talvez pela popularização da ISO 9241-9 como instrumento de avaliação de interfaces para 2D, alguns autores buscaram aplicar métodos de teste semelhantes para o caso tridimensional. É possível encontrar uma multitude de trabalhos que utilizam fundamentos da Lei de Fitts e a ISO 9241-411 para avaliar

interação 3D tanto em telas bidimensionais quanto para ambientes virtuais tridimensionais como os da RV (WARE; LOWTHER, 1997; TEATHER et al., 2009; TEATHER; NATAPOV; JENKIN, 2010; TEATHER; STUERZLINGER, 2011; KOUROUPETROGLOU et al., 2012; TEATHER; STUERZLINGER, 2013; PINO et al., 2013; SAMBROOKS; WILKINSON, 2013; RAYNAL; DUBOIS; SCHMITT, 2013; BRUDER; STEINICKE; STÜRZLINGER, 2013; BACHMANN; WEICHERT; RINKENAUER, 2014; LUBOS; BRUDER; STEINICKE, 2014; LUBOS; BRUDER; STEINICKE, 2015; BURNO et al., 2015; DELONG; MACKENZIE, 2018; YOUNG; TEATHER; MACKENZIE, 2017; HANSEN et al., 2018; ESTEVES; SHIN; OAKLEY, 2020). Porém, a ausência de um método padronizado dificulta a comparação entre resultados de diferentes estudos. Além disso, nenhum deles estudou interação dentro de um ambiente imersivo de RA.

Com o objetivo de entender melhor quais métodos tem sido utilizados para realizar as avaliações formais de técnicas de interação em 3D, foi feito um levantamento da literatura especializada. O interesse principal foi selecionar trabalhos que apresentavam avaliação de técnicas de interação que se aproximavam das mecânicas da interação com as mãos utilizadas nos óculos de RA, na qual os alvos estão co-localizados com o espaço motor das mãos do usuário.

Os resultados deste levantamento estão sintetizados na Tabela 2 e serão comentados a seguir. Apesar de alguns dos trabalhos listados descreverem testes com diferentes técnicas de interação e *feedbacks* visuais, a tabela lista apenas aquela que mais se aproximou da interação direta em 3D, co-localizada com as mãos e com *feedback* visual estereoscópico. Os acrônimos não mencionados até então e que aparecem na tabela significam: Unidade de Medidas Inerciais (IMU); *Cave Automatic Virtual Environment* (CAVE); *Missed Clicks* (MCL), significando “seleções incorretas”; *Target Re-Entry* (TRE), significando “re-entradas nos alvos”; *Distance Travelled* (DT), significando “distância percorrida”; *Error Distance* (ED), significando “erro de distância”; e *Inefficiency Ratio* (IR), significando “razão de ineficiência”.

2.4.1 Técnicas de interação e *feedback* visual

Conforme pode ser observado na Tabela 2, as técnicas de interação que se aproximam da interação direta com as mãos usualmente envolvem o uso de câmeras e outros dispositivos para realizar o rastreamento espacial tridimensional do dedo indicador ou de controles que o usuário segura com sua mão. Dentre as técnicas reportadas, interação direta, co-localizada com a mão do usuário, pôde ser identificada nos trabalhos de Teather, Natapov e Jenkin (2010), Teather e Stuerzlinger (2011), Bruder, Steinicke e Stürzlinger (2013), Lubos, Bruder e Steinicke (2015) e Özacar et al. (2017). Estes são os estudos que mais se aproximam do que o autor deseja realizar neste trabalho.

Com relação ao *feedback* visual, a literatura pesquisada utilizou uma variedade de métodos, incluindo ambientes virtuais tridimensionais não-estéreos (KOUROUPETROGLOU

Tabela 2 – Síntese das diferentes técnicas de interação com as mãos em 3D avaliadas pela literatura e do método de avaliação aplicado.

Referência	Técnica de interação	Feedback visual	Forma e disposição dos alvos	Procedimento	Lei de Fitts	Métricas reportadas
(BÉRARD et al., 2009)	Controle manual (<i>Wii Remote</i> rastreado)	RV com projetores	Cubos posicionados aleatoriamente num volume cúbico	Mover um cubo para envolver o alvo	Não reportado	MT, ER
(TEATHER; NATAPOV; JENKIN, 2010)	Controle manual (caneta rastreada)	CAVE	Esféras no teste ISO multi-direcional, plano frontal	Teste multi-direcional recíproco da ISO	ISO 9241-9 (seção 2.2)	TP, MT, ER
(TEATHER; STUERZLINGER, 2011)	Controle manual (caneta rastreada)	RV em monitor 2D	Esféras sobre pilares em teste ISO multi-direcional, plano inclinado	Teste multi-direcional recíproco da ISO	ISO 9241-9 (seção 2.2)	TP, MT, ER
(KOUROUPETROGLOU et al., 2012; PINO et al., 2013)	Controle manual (<i>Wii Remote</i>) e mãos livres (<i>Kinect</i>)	Monitor TFT (2D)	8 esferas posicionadas nos vértices de um cubo	Seleções seriais de alvos opostos diagonalmente	ISO 9241-9 (seção 2.2)	TP, MT, DT, MCL, TRE
(BRUDER; STEINICKE; STÜRZLINGER, 2013)	Dedo indicador rastreado (<i>WorldViz PPT X4</i>)	RV em mesa interativa	Esféras no teste ISO multi-direcional, plano horizontal em alturas variadas	Teste multi-direcional recíproco da ISO	ISO 9241-9 (seção 2.2)	TP, MT, ER, ED
(RAYNAL; DUBOIS; SCHMITT, 2013)	Controle manual (<i>Polhemus Patriot</i>)	Monitor LCD (2D)	Esféras homogeneamente distribuídas em uma casca esférica	Seleções discretas de alvos opostos diagonalmente	ISO 9241-9, adaptada	TP, MT, ER, IR
(LUBOS; BRUDER; STEINICKE, 2014)	Dedo indicador rastreado (<i>Optitrack V120:Trio</i>)	Óculos de RV	Esféras em teste ISO multi-direcional, 9 planos em variadas orientações	Seleções seriais a partir de uma esfera fixa	Não reportado	TP, MT, ER, ED
(LUBOS; BRUDER; STEINICKE, 2015)	Dedo indicador rastreado (<i>WorldViz PPT X4</i>)	Óculos de RV	Esféras em teste ISO multi-direcional, 3 planos em variadas orientações	Teste multi-direcional recíproco da ISO	ISO 9241-9 (seção 2.2)	TP, MT, ER
(YOUNG; TEATHER; MACKENZIE, 2017)	IMU vestível	RV em monitor 2D	Esféras sobre pilares em teste ISO multi-direcional, planos frontais em profundidades variadas	Teste multi-direcional recíproco da ISO	ISO 9241-9 (seção 2.2)	TP, MT, ER
(ÖZACAR et al., 2017)	Dedo indicador rastreado (câmeras infravermelhas)	Óculos de RA	Esféras distribuídas aleatoriamente em um espaço de interação cúbico	Seleção serial dos alvos	Não foi aplicada	MT, ER

Fonte: adaptado de Couto, José e Lopes (2020).

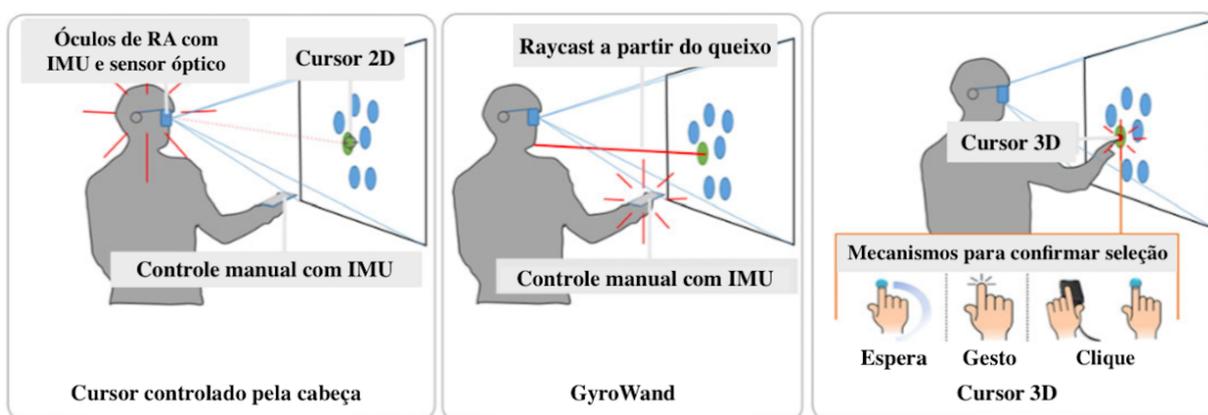
et al., 2012; PINO et al., 2013; RAYNAL; DUBOIS; SCHMITT, 2013), RV imersiva com uso de diferentes tecnologias (BÉRARD et al., 2009; TEATHER; STUERZLINGER, 2011; BRUDER; STEINICKE; STÜRZLINGER, 2013; LUBOS; BRUDER; STEINICKE, 2015; YOUNG; TEATHER; MACKENZIE, 2017), sistema CAVE (TEATHER; NATAPOV; JENKIN, 2010) e óculos de RA (ÖZACAR et al., 2017).

O objetivo central da pesquisa da literatura foi encontrar trabalhos que avaliaram interação direta com as mãos em óculos de RA. Contudo, dentre os trabalhos identificados, somente o de Özacar et al. (2017) reportou este tipo de estudo. Conforme ilustrado na Figura 3, os autores investigaram como diferentes técnicas de seleção de conteúdo em um ambiente imersivo de RA afetavam o desempenho dos participantes. Entretanto, este trabalho não aplicou a Lei de Fitts e utilizou um método pouco estruturado, o que dificulta a comparação de resultados com outros estudos.

Outros pesquisadores estudaram técnicas de interações que se aproximam da experiência oferecida por óculos de RA. Por exemplo, Teather, Natapov e Jenkin (2010) estudaram interação direta em um teste ISO multi-direcional visualizado estereoscopicamente em uma CAVE, conforme ilustrado na Figura 4.

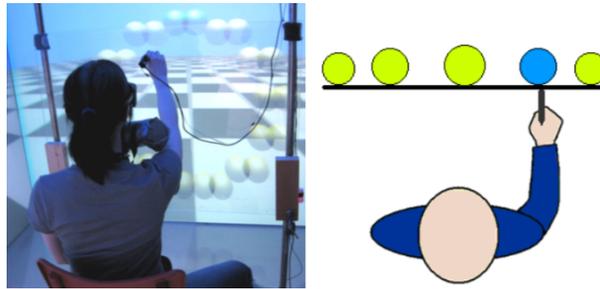
Os trabalhos de Teather e Stuerzlinger (2011) e Bruder, Steinicke e Stürzlinger (2013) também investigaram técnicas de interação muito semelhantes às presentes em óculos de RA, porém utilizando tecnologia de RV por meio de monitores 2D habilitados para reprodução estereoscópica. Conforme ilustrado na Figura 5 e na Figura 6, nestes experimentos os alvos foram exibidos em planos horizontais de modo estereoscópico e apresentados em alturas visualmente acima da tela, permitindo ao usuário realizar interação direta ou com uma caneta ou com o dedo indicador.

Figura 3 – Avaliação de diferentes técnicas de interação em óculos de RA conduzida por Özacar et al.. Além da interação direta (cursor 3D), o autor também avaliou interação usando duas técnicas de *raycasting*, em que o raio é lançado a partir da referência da cabeça ou do queixo do participante.



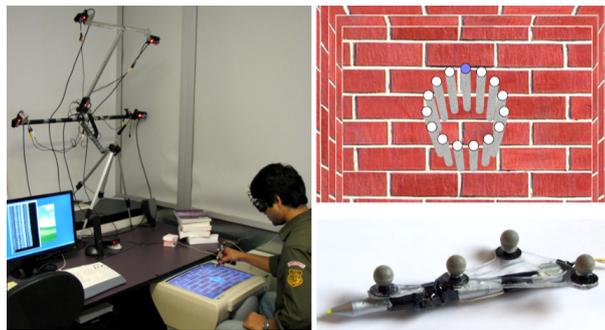
Fonte: adaptado de Özacar et al. (2017).

Figura 4 – Avaliação de interação com as mãos em um ambiente CAVE usando uma caneta rastreada espacialmente, conforme trabalho de Teather, Natapov e Jenkin (2010).



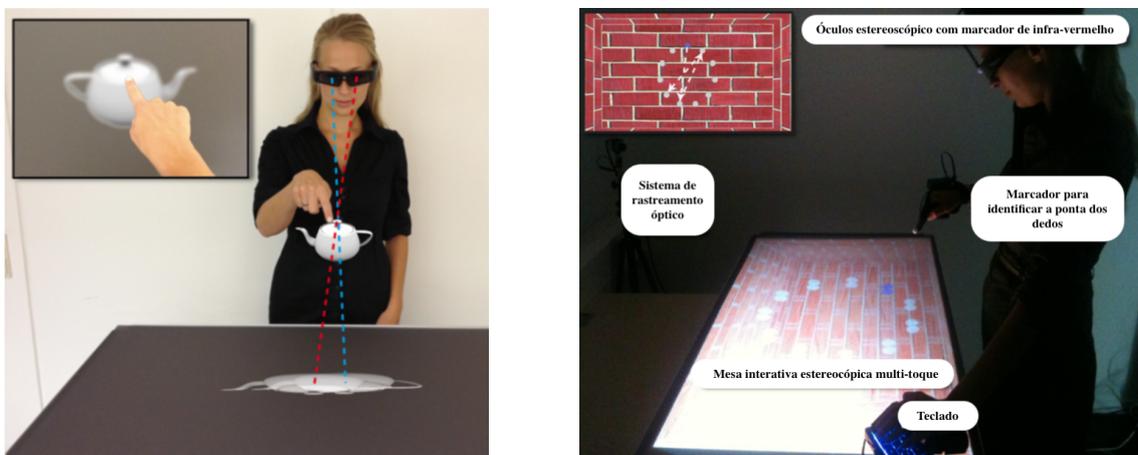
Fonte: Teather, Natapov e Jenkin (2010).

Figura 5 – Aparato do experimento de Teather e Stuerzlinger (2011). Uma das técnicas de interação estudadas foi interagir com os alvos usando uma caneta rastreada espacialmente. Os alvos, apresentados estereoscopicamente, também foram apresentados como estando visualmente “acima” da tela do monitor 2D.



Fonte: Teather e Stuerzlinger (2011).

Figura 6 – Avaliação da interação direta com as mãos em um ambiente estereoscópico de RV realizada por Bruder, Steinicke e Stürzlinger (2013). Os alvos foram apresentados tanto na superfície da tela da mesa interativa quanto visualmente “acima” dela, em uma experiência final semelhante à de óculos de RA



(a) Ilustração da técnica de interação direta

(b) Aparato utilizado no experimento

Fonte: adaptado de Bruder, Steinicke e Stürzlinger (2013).

2.4.2 Forma e disposição dos alvos

A maioria dos trabalhos utilizaram esferas como alvos nos testes. A única exceção foi o trabalho de Bérard et al. (2009), que utilizou cubos. A escolha de esferas é uma extensão natural para os círculos, que é o formato usualmente utilizado nos experimentos em 2D. A justificativa mais pertinente é que o comprimento de alvos circulares e esféricos independe da direção do movimento e, portanto, tornam o resultado do teste mais uniforme. Esferas também foram utilizadas como alvos pelos trabalhos de Machuca e Stuerzlinger (2019) e Clark, Bhagat e Riggs (2020), ambos mencionados na seção 2.3. A forma do cursor de interação sempre foi a mesma da escolhida para os alvos. As cores mais comuns utilizadas nos estudos foram branco, cinza, azul, vermelho e verde. Alvos inativos foram frequentemente representados pelas cores branca ou cinza, enquanto a cor dos alvos ativos e do cursor variaram entre branco, azul, vermelho e verde.

Alguns trabalhos mencionaram o uso de alvos ou cursor semi-transparentes. Na maioria dos trabalhos, entretanto, os autores não deixam explícito se esta decisão foi tomada. O uso de alvos opacos com cursor semi-transparente é evidente nos trabalhos de Bérard et al. (2009) e Teather e Stuerzlinger (2011). Esta decisão é baseada no trabalho de Zhai, Buxton e Milgram (1994), que demonstrou que um cursor volumétrico semi-transparente permitiu interação com maior desempenho em tarefas de seleção de objetos 3D do que usar cursor equivalente representado apenas como *wireframe* (somente com arestas e vértices).

Por outro lado, o uso de alvos opacos possui a desvantagem de provocar oclusão do cursor quando este é movido para determinadas posições visualmente atrás dos alvos. A oclusão do cursor durante a realização de um teste de avaliação de desempenho pode confundir o usuário, potencialmente impactando o resultado final. Por esta razão, o uso de alvos transparentes costuma ser uma escolha em interações 3D (STUERZLINGER; TEATHER, 2014). Dentre os trabalhos analisados, esta escolha ficou evidente apenas no trabalho de Raynal, Dubois e Schmitt (2013).

Com relação à disposição dos alvos, o arranjo preferido foi o definido pelo teste multi-direcional da ISO 9241-411 (Figura 2b), tendo sido aplicado em (TEATHER; NATAPOV; JENKIN, 2010; TEATHER; STUERZLINGER, 2011; BRUDER; STEINICKE; STÜRZLINGER, 2013; LUBOS; BRUDER; STEINICKE, 2014; LUBOS; BRUDER; STEINICKE, 2015; YOUNG; TEATHER; MACKENZIE, 2017). A escolha frequente para adaptá-lo para 3D foi dispor os alvos esféricos ao longo de uma circunferência, efetivamente restringindo o teste a um plano. Apesar dessa decisão reduzir as possibilidades de interação a 2D e, portanto, não permitir avaliar interações genéricas no espaço tridimensional, ela possui a vantagem de manter os estudos em 3D comparáveis ao corpo literário de interação em 2D (TEATHER; STUERZLINGER, 2011; YOUNG; TEATHER; MACKENZIE, 2017).

Outros arranjos utilizados para os alvos foram o posicionamento aleatório em um

espaço cúbico (BÉRARD et al., 2009; ÖZACAR et al., 2017), posicioná-los nos vértices de um cubo (KOUROUPETROGLOU et al., 2012; PINO et al., 2013), e distribuí-los homogeneamente em uma casca esférica (RAYNAL; DUBOIS; SCHMITT, 2013). Estes trabalhos, entretanto, adotam métodos raramente utilizados na literatura, sendo pouco comparáveis com trabalhos anteriores.

Quando explícito, os trabalhos pesquisados reportaram que um algoritmo de colisão volumétrica foi utilizado para identificar quando o cursor poderia interagir com um alvo. Todos os trabalhos que seguiram o arranjo do teste multi-direcional da ISO também destacaram o alvo para indicar visualmente quando a interação era possível. O destaque foi feito mudando a cor do alvo. Teather e Stuerzlinger (2014) recomendaram que pesquisadores usem esta técnica após experimentos indicarem que destacar o alvo tem um efeito relevante no desempenho final do participante.

2.4.3 Procedimentos adotados nos testes

Todos os trabalhos analisados estudaram apenas a tarefa de apontar-e-selecionar. Três tipos de movimentos foram utilizados: discretos, seriais e recíprocos. As tarefas com movimentos discretos foram utilizadas em (BÉRARD et al., 2009; RAYNAL; DUBOIS; SCHMITT, 2013) e consistem em iniciar o movimento a partir de um ponto inicial e finalizá-la assim que o alvo é selecionado. Outros trabalhos utilizaram tarefas com movimentos seriais (KOUROUPETROGLOU et al., 2012; PINO et al., 2013; LUBOS; BRUDER; STEINICKE, 2014; ÖZACAR et al., 2017), que são compostas por uma sequência de movimentos discretos em que o início de uma interação começa a partir do fim da última interação, até que todos os alvos desejados tenham sido selecionados. Os demais trabalhos utilizaram tarefas com movimentos recíprocos baseadas na mecânica do teste multi-direcional da ISO 9241-411.

O mesmo procedimento dos testes de apontar-e-selecionar utilizado na literatura 2D foi aplicado nos trabalhos analisados. Inicialmente os participantes são instruídos a realizar as seleções na maior velocidade possível, buscando manter uma baixa taxa de erros, mas continuando o teste caso um erro seja cometido. O teste formal é iniciado após seleção do alvo inicial, que é apresentado em uma cor diferenciada. Uma vez selecionado, o próximo alvo é destacado com outra cor diferenciada e, a partir de então, o participante segue realizando as seleções conforme requerido pela tarefa: para seleções discretas, o teste se encerra e o próximo alvo inicial é destacado, permitindo ao participante pausas momentâneas; para seleções seriais e recíprocas, o próximo alvo é imediatamente destacado e o participante deve continuar as seleções sem pausar. O teste é finalizado quando o participante disparar o evento de seleção correspondente ao último alvo.

2.4.4 Formulações da Lei de Fitts

Os estudos que reportaram explicitamente o uso da Lei de Fitts utilizaram a formulação expressa pelo padrão ISO 9241-411 (conforme descrito na seção 2.2). Nota-se, contanto, que Raynal, Dubois e Schmitt (2013) adaptaram a formulação do *ID* para considerar um cursor com dimensões não infinitesimais. O objetivo desta modificação foi garantir que o índice de dificuldade refletirá mais fielmente a dificuldade das tarefas, dado que, para uma mesma configuração dos alvos, as seleções serão mais fáceis caso o cursor tenha uma dimensão maior. Apesar desta ser uma consideração válida, nenhum dos demais trabalhos que se basearam na ISO 9241-411 utilizou uma formulação diferenciada para analisar e reportar seus resultados.

Nenhuma versão estendida da Lei de Fitts para 3D foi utilizada nos trabalhos analisados. Particularmente, Teather e Stuerzlinger (2011) argumentaram que os dados que obtiveram foram bem modelados pela formulação tradicional da Lei de Fitts e, portanto, não pareceu haver necessidade de avaliar os dados com outras formulações. Dentre as diversas técnicas de interação que eles avaliaram neste trabalho, a pior correlação apresentou um coeficiente de determinação $R^2 = 0.66$. Este valor não é incomum entre estudos avaliando interações 2D com a Lei de Fitts.

2.4.5 Métricas reportadas

Zhai (1998) propôs que modalidades de interação com seis graus de liberdade fossem avaliadas com relação a seis aspectos: velocidade, precisão, coordenação, facilidade de aprendizado, fadiga, e persistência e aquisição do dispositivo. Os três primeiros aspectos estão mais associados às características tecnológicas (*hardware* e *software*) que permitem a interação, e os últimos três estão principalmente relacionados à questões de usabilidade e desenho conceitual.

Os aspectos de velocidade e precisão medem a eficiência com que tarefas podem ser concluídas com a modalidade de interação, sendo os mais estudados em avaliações formais. Eles são medidos e reportados pelas métricas *MT*, *ER* e *TP*, conforme descrito na seção 2.2.

Coordenação é um aspecto menos estudado e se refere à capacidade de se controlar o cursor para alcançar a trajetória mais eficiente para completar uma tarefa. Zhai e Milgram (1998) propuseram a métrica *IR* para medir esse aspecto:

$$IR = \frac{\text{distância percorrida} - \text{menor distância possível}}{\text{menor distância possível}} \quad (2.7)$$

Desta forma, *IR* é capaz de medir o quanto a trajetória final realizada pelo usuário desviou da trajetória ideal. Raynal, Dubois e Schmitt (2013) propuseram incluir essa métrica em sua proposta de método para avaliar interação em ambientes virtuais não-imersivos. Kouroupetoglou et al. (2012), por outro lado, reportaram a distância percorrida (*DT*), métrica que

permite computar *IR* indiretamente.

Aspectos como facilidade de aprendizado e fadiga tem como objetivo medir o esforço cognitivo e físico que os usuários precisam executar para conseguir interagir com a modalidade de entrada. Uma das formas utilizadas para avaliar a facilidade de aprendizado é analisar quanto tempo um participante leva para atingir um desempenho máximo realizando uma mesma tarefa. Nenhum dos trabalhos estudados realizou uma avaliação formalizada com este objetivo. Raynal, Dubois e Schmitt (2013) argumentou que este aspecto pode ser avaliado por meio do questionário *System Usability Scale* (SUS) (BROOKE, 1996).

A ISO 9241-411 sugere aplicar o questionário para avaliação do conforto (reproduzido no Anexo A) e a Escala Borg (BORG, 1998) para avaliar aspectos de usabilidade da técnica de interação. Estes questionários são úteis para avaliar fadiga e foram incluídos por Raynal, Dubois e Schmitt (2013) em seu método de avaliação. Özacar et al. (2017) aplicaram a Escala Borg para avaliar fadiga física e o *NASA Task Load Index* (NASA-TLX) (HART; STAVELAND, 1988) para avaliar a demanda mental da técnica de interação.

Por fim, o aspecto de persistência e aquisição do dispositivo de entrada se referem a características de usabilidade e devem medir a capacidade do cursor de se manter no mesmo local após o usuário soltar o dispositivo (persistência) e a facilidade de se readquirir o dispositivo e recomeçar a controlar o cursor (aquisição). Nenhum dos trabalhos pesquisados avaliaram este aspecto, mas Raynal, Dubois e Schmitt (2013) sugeriram que o questionário SUS e entrevistas com os participantes poderiam ser utilizados como instrumentos para atingir este objetivo.

3 MÉTODOS DA PESQUISA

A revisão bibliográfica apresentada no Capítulo 2 indicou que há poucos trabalhos avaliando interação direta com as mãos em ambientes imersivos de RA. De fato, este tipo de estudo só foi reportado por Özacar et al. (2017). Ainda assim, os autores não utilizaram nenhum modelo preditivo, como a Lei de Fitts, para analisar os resultados. As únicas métricas reportadas foram tempo de interação e taxa de erros, que sozinhas não apresentam a mesma robustez que o *throughput* para comparar resultados de diferentes estudos.

Ainda que alguns pesquisadores tenham preferido sugerir novos métodos de avaliação para avaliar interação em 3D, os trabalhos analisados indicam uma preferência em seguir mecânicas próximas do teste multi-direcional da ISO 9241-411. Foi comum observar a adoção de pequenas adaptações à versão original deste teste, como utilizar esferas dispostas em uma circunferência, mas escolhendo manter a interação restrita a um plano espacial (TEATHER; NATAPOV; JENKIN, 2010; TEATHER; STUERZLINGER, 2011; BRUDER; STEINICKE; STÜRZLINGER, 2013; LUBOS; BRUDER; STEINICKE, 2015; YOUNG; TEATHER; MACKENZIE, 2017). Teather (2013) explorou este tipo de adaptação em diversos experimentos para avaliar tanto interação em 2D quanto em 3D, e recomendou que fosse utilizada. Uma das principais justificativas é que o método já é amplamente utilizado para avaliar interação em 2D e, portanto, utilizá-lo como ponto de partida para avaliar interação em 3D garante que os resultados se mantenham em patamares mais comparáveis com toda a literatura de interação 2D.

É importante notar que o método proposto por Teather e Stuerzlinger (2011) foi replicado nos trabalhos de Bruder, Steinicke e Stürzlinger (2013) e Young, Teather e MacKenzie (2017), tornando-os comparáveis e indicando um esforço em direção a uma padronização da forma de avaliar interação em ambientes 3D. Entretanto, uma clara fraqueza deste método é que ele estuda a interação apenas em um plano. Isto significa que nem todas as direções do espaço tridimensional são avaliadas, e portanto, pode-se argumentar que os resultados não representam o desempenho em todo o espaço tridimensional de interação.

Por outro lado, um método que permita avaliar as diversas direções de movimento permitidas pelas 3D ou deverá se afastar da mecânica já utilizada nos testes 2D – e, portanto, se tornar menos comparável com a literatura já existente – ou será custoso demais para ser realizado – por exemplo, sendo necessário testar vários planos com diferentes orientações espaciais até garantir que boa parte do espaço de interação foi analisada.

Baseando-se principalmente nestas constatações, o presente trabalho buscou apresentar um método de avaliação que, partindo das propostas mais utilizadas pela literatura, possa ser aplicado de forma mais padronizada e que garanta que mais regiões do espaço tridimensional sejam avaliadas sem, contudo, onerar demais o custo da pesquisa.

Neste sentido, os seguintes procedimentos foram estabelecidos para prosseguir com o

desenvolvimento desta pesquisa:

- a) propor um método para avaliar técnicas de interação com as mãos em um óculos de RA;
- b) desenvolver um *software* para permitir aplicar o método em avaliações formais com seres humanos;
- c) solicitar autorização ao Comitê de Ética em Pesquisa para realizar testes com humanos;
- d) aplicar o método em experimentos com o objetivo de:
 - analisar a resposta do *software* desenvolvido com relação a outro utilizado na literatura para verificar se os algoritmos para coleta de dados e cálculo das métricas foram corretamente implementados;
 - verificar se o método proposto é capaz de discriminar o desempenho de diferentes técnicas de interação;
 - demonstrar como o método proposto e o *software* desenvolvido pode ser aplicado como ferramenta de avaliação;
 - testar as hipóteses $H1$ e $H2$ que deram origem a esta pesquisa.

Neste capítulo, serão apresentados a proposta de método de avaliação, o *software* desenvolvido para permitir sua aplicação e um resumo dos experimentos que foram conduzidos. Por uma questão de clareza, decidiu-se descrever cada um experimentos em um capítulo à parte (Capítulo 4), no qual serão apresentados os métodos utilizados, os resultados obtidos e a discussão destes resultados.

3.1 Proposta de um método para avaliar técnicas de interação com as mãos em RA

A revisão da literatura apresentada no Capítulo 2 sugere que as adaptações propostas por Teather e Stuerzlinger (2011) tem sido replicadas por mais trabalhos e, portanto, tem sido aceita como ferramenta de avaliação por mais de um pesquisador. Porém, na forma como os trabalhos reportaram o método, não há indicação de que planos devem ser testados para garantir que a técnica de interação foi avaliada em diferentes direções do espaço tridimensional. Além disso, sem essa formalização, diferentes pesquisadores tenderão a testar orientações de planos diversas, reduzindo a comparabilidade entre seus estudos. Por essa razão, o autor decidiu propor um método mais estruturado, com procedimentos melhor estabelecidos, baseando-se no levantamento bibliográfico realizado. As recomendações para aplicar este método serão descritas nas próximas subseções.

É importante ressaltar que este método tem como objetivo avaliar apenas duas tarefas primitivas de interação: apontamento e seleção de objetos. Por meio delas, é possível construir um amplo conjunto de interações mais complexas e, assim, avaliá-las permite obter uma visão geral da capacidade de uma técnica de interação para realizar interações diversas no ambiente de RA.

O método aqui proposto tem como objetivo avaliar os aspectos ergonômicos da interação em um dado sistema de RA, uma vez que ele busca identificar o custo cognitivo e físico demandado de um usuário para executar uma tarefa específica. Esta avaliação é feita de modo indireto, por meio da coleta de métricas quantitativas como o tempo de execução, a taxa de erros e a precisão da interação, e de modo mais direto por meio da aplicação de questionários para capturar a impressão subjetiva do usuário sobre sua experiência ao executar as tarefas propostas.

3.1.1 Forma e disposição dos alvos

Os alvos devem ser representados por esferas semi-transparentes e seus centros devem ser dispostos em uma mesma circunferência, seguindo o arranjo do teste multi-direcional da ISO 92141-411 (Figura 2b). O tamanho dos alvos (W) é determinado pelo diâmetro da esfera e controla a precisão requerida para a tarefa. A distância entre os alvos (D) é determinado pelo diâmetro da circunferência e que estão dispostos. Variando-se estes parâmetros, é possível obter tarefas com diferentes ID s.

A escolha de alvos semi-transparentes é comum pois evita a oclusão do cursor (STUERZLINGER; TEATHER, 2014). Além disso, ela ajuda a aumentar a percepção visual de profundidade, uma vez que a cor do cursor variará conforme ele entra dentro de um alvo ou se mantém por trás dele (RAYNAL; DUBOIS; SCHMITT, 2013). Este fenômeno é similar ao proposto por Zhai, Buxton e Milgram (1994) com seu “cursor de seda”.

Para garantir uma maior homogeneidade entre estudos, sugere-se que os pesquisadores utilizem cor branca ou cinza claro para indicar que um alvo não está ativo e cor azul para indicar o próximo alvo. Para tornar o teste mais intuitivo, sugere-se também que os pesquisadores utilizem uma cor diferenciada para indicar o primeiro alvo (aquele que inicia o teste).

3.1.2 Forma do cursor e mecânica de interação

O cursor também deve ser esférico, garantindo que a distância de interação com qualquer alvo será independente da direção de aproximação. O tamanho do cursor pode ser escolhido conforme o objetivo do estudo. É importante, entretanto, reportar o valor utilizado, já que ele impactará o resultado final.

Não houve, na literatura pesquisada, nenhuma recomendação para utilizar um cursor opaco ou semi-transparente. Considerando que o cursor é o elemento central do teste de avaliação – todas as interações dependem dele ser movido até diferentes alvos – e levando em conta que ele será o elemento com maior dinamismo, é importante que sua identificação visual seja fácil a qualquer momento do teste, sob risco de afetar negativamente o desempenho do participante. Neste sentido, recomenda-se utilizar um cursor opaco. Quanto à cor do cursor, para uniformizar condições de teste entre diferentes estudos, sugere-se utilizar uma cor verde.

A localização do cursor dependerá da técnica de interação a ser estudada. Para avaliar um *mouse* com o presente método, por exemplo, basta mapear as coordenadas 2D do dispositivo para as coordenadas espaciais do plano de teste e posicionar o cursor virtual nessa posição. Caso se deseje estudar interação com as mãos, o pesquisador poderá escolher em que região da mão posicionar o cursor. Para tarefas básicas de apontar-e-selecionar, recomenda-se posicioná-lo ou na ponta do dedo indicador ou no centro da mão dominante do usuário.

Para identificar quando o cursor pode interagir com um alvo, deve-se utilizar o algoritmo de colisão volumétrica. Como tanto os alvos quanto o cursor são esferas, isso pode ser feito facilmente, bastando verificar se a distância entre o centro do cursor e o centro do alvo é menor ou igual à soma dos seus raios.

Por fim, baseado no trabalho de Teather e Stuerzlinger (2014), destaque deve ser dado ao alvo com que o cursor estiver interagindo em um dado momento. Recomenda-se, para isso, modificar a cor do alvo para a cor vermelha enquanto o cursor estiver em colisão com um ele.

3.1.3 Mecânica de teste

As demais recomendações do padrão ISO 9241-411 para a execução de testes multidirecionais continuam válidas. Dessa forma, deve-se testar diferentes condições de D e W , explorando um intervalo de ID s que atenda à técnica de interação sendo avaliada. Os movimentos devem ser recíprocos, conforme expresso na Figura 2b. O próximo alvo deve ser destacado com uma cor específica, enquanto os demais alvos se mantêm em uma cor neutra. O número de alvos em um teste também pode variar, visto que o padrão não determina explicitamente quantos alvos devem ser utilizados.

Sugere-se o uso de *feedback* sonoro para indicar ao usuário quando um alvo for corretamente selecionado (usar um som que estimule a sensação de acerto) ou não (usar um som que estimule a sensação de erro). Este retorno sonoro é importante pois auxilia o participante a identificar quando um evento de seleção ocorreu e a controlar sua taxa de erro, incentivando-o a reduzir a velocidade caso muitos erros estejam sendo cometidos.

Ao fim de um teste, recomenda-se apresentar os resultados do teste ao participante, para mantê-lo motivado e interessado em se desempenhar de forma melhor à medida que prossegue a realização dos demais testes.

3.1.4 Formulação da Lei de Fitts

Os dados a serem coletados nos testes permanecem os mesmos exigidos pelo padrão ISO 9241-411: tempo de movimento e posição final do cursor em que a interação foi encerrada. Com estes dados será possível aplicar a formulação da Lei de Fitts, conforme seção 2.2, para obter o valor de TP para a técnica de interação estudada.

3.1.5 Procedimentos gerais

3.1.5.1 Planos de teste

As recomendações apresentadas até o momento seguem na mesma direção dos métodos adotados pelos trabalhos analisados no Capítulo 2. Nesta subseção será descrito o principal procedimento que diferencia este método de avaliação dos demais: uma recomendação expressa de que pelo menos duas orientações de planos sejam testadas.

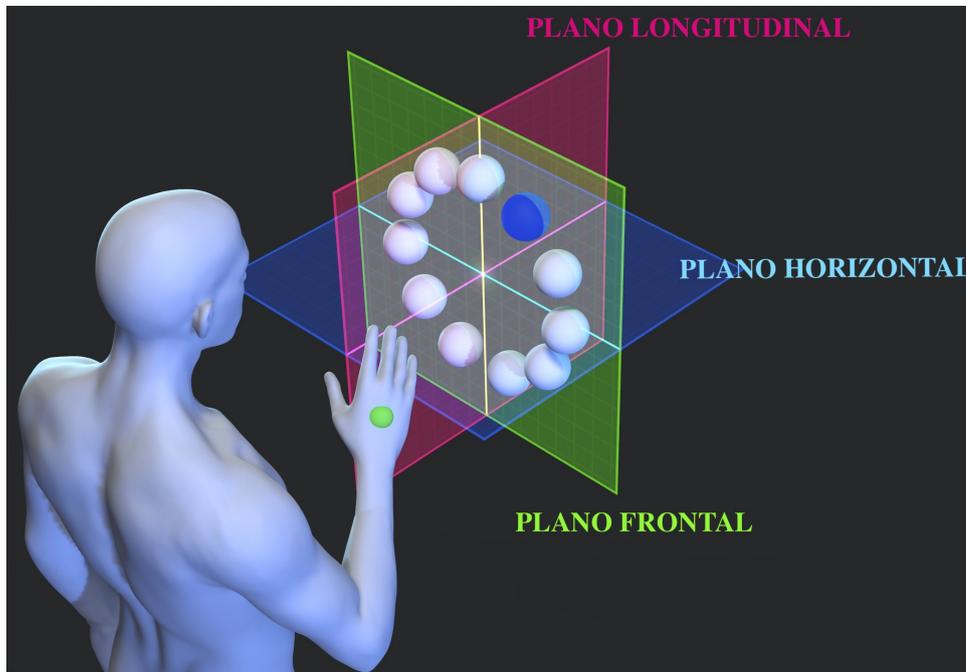
A principal justificativa já foi mencionada anteriormente: permitir avaliar a qualidade de interação em todas as 3D do espaço. Além disso, técnicas de interação com as mãos em geral dependem de sensores de infravermelho para identificar a profundidade espacial da mão do usuário. Uma das variáveis que afetam a qualidade de rastreamento quando estes sensores são utilizados é justamente a sua distância da mão do usuário. Assim, a qualidade da interação tende a ser ótima quando a mão se encontra a meio alcance do sensor, e tende a se reduzir quando está muito próxima ou muito distante do dispositivo. É importante, portanto, que um método de avaliação de técnicas de interação com as mãos busquem avaliar todo um volume espacial de interação a que se destinam, pois uma determinada técnica de interação pode não ter bom desempenho em alguma região específica.

Os candidatos naturais para se tornarem planos de teste em avaliações de interação com as mãos são os planos frontal, horizontal e longitudinal. A Figura 7 ilustra a orientação de cada um destes planos com relação a um participante do teste de avaliação. Como estes planos são ortogonais entre si, movimentos realizados neles podem ser combinados para representar qualquer movimento em 3D, permitindo avaliar, ainda que de modo indireto, movimentos tridimensionais genéricos.

Os seres humanos frequentemente interagem com planos frontais e horizontais, tais como superfícies de mesas, do chão, de telas de televisores ou monitores, de paredes e lousas. Por esse motivo, eles são uma escolha natural e adequada para avaliar interação direta com as mãos.

Por outro lado, avaliar interação no plano longitudinal é uma tarefa mais desafiadora, pois os alvos mais próximos do participante confundirão visualmente aqueles mais distantes,

Figura 7 – Representação dos planos frontal, horizontal e longitudinal. A imagem também ilustra a visão que um participante tem ao realizar o teste multi-direcional para avaliar interação com as mãos em um óculos de RA. Apesar da imagem não deixar evidente, os alvos devem ser apresentados com semi-transparência suficiente para permitir visualizar o cursor através deles.



Fonte: adaptado de Couto, José e Lopes (2020).

impactando o desempenho final do participante. Uma possível solução para este problema seria alterar a mecânica do teste – por exemplo, exibindo somente o próximo alvo a ser selecionado, ao invés de sempre mostrar os demais. Porém, ao mudar a mecânica, reduz-se a comparabilidade dos testes com a literatura de avaliação 2D, além de possivelmente implicar em novos desafios. Sabe-se, por exemplo, que a qualidade de interação é afetada pela capacidade de identificar a profundidade dos alvos (HOFFMAN et al., 2008; MACHUCA; STUERZLINGER, 2019; CLARK; BHAGAT; RIGGS, 2020) e, portanto, é importante manter dicas visuais para auxiliar no processo de percepção de profundidade. Ao se esconder todos os alvos exceto o próximo a ser selecionado, reduz-se as dicas visuais, possivelmente impactando no desempenho final do usuário.

Outra consideração é que a adição de cada novo plano de teste aumenta a duração total de uma avaliação, demandando mais tempo tanto dos participantes quanto do pesquisador. Felizmente, é possível estudar movimentos que abarcam todas as 3D do espaço somente com os planos frontal e horizontal.

Deste modo, este método recomenda que *ao menos* os planos frontal e horizontal sejam testados ao se avaliar técnicas de interação com as mãos em 3D. Outros planos, incluindo o longitudinal, podem ser incluídos no processo de avaliação conforme necessário para o objetivo do estudo.

Esta recomendação deixa o presente método flexível o suficiente para permitir ava-

liações em outras orientações de planos. Se um pesquisador desejar ter uma avaliação mais precisa espacialmente, por exemplo, ele poderá testar planos inclinados de 45° com relação aos planos principais de teste (frontal e horizontal). Como os dois planos principais também terão sido testados, isto significa que um estudo mais preciso ainda assim poderá ser comparado com os outros que se limitaram a estudar apenas os dois planos principais.

Outro benefício desta recomendação é que toda avaliação de técnica de interação em 3D que seguir este método poderá ser comparada também com a maioria da literatura de interação em 2D, uma vez que ela usualmente testa interação em um plano frontal (o monitor de um computador).

Se o teste multi-direcional for executado com um intervalo suficiente de *IDs*, para cada um dos planos propostos, será possível inspecionar, durante os testes, a região central do volume esférico de interação à frente do usuário. A avaliação, portanto, tem o potencial de discriminar a capacidade de diferentes técnicas de interação para desempenhar interações com objetos virtuais localizados em diferentes regiões espaciais. Também será possível quantificar a taxa de erro em diferentes posições espaciais, o que pode contribuir para caracterizar a qualidade de interação 3D fornecida por uma técnica de interação.

3.1.5.2 Validação externa por meio da avaliação de um *mouse*

Os trabalhos de Teather e Stuerzlinger (2011), Stuerzlinger e Teather (2014) e Young, Teather e MacKenzie (2017) recomendam fortemente que todo estudo avaliando interação em 3D também inclua avaliação de interação utilizando um *mouse*. Como este dispositivo é considerado o padrão-ouro da interação 2D, sua inclusão no processo de avaliação aumentará a validade externa do estudo. Deste modo, o *mouse* se torna a técnica de interação ponte para permitir ancorar os valores de *TP* obtidos no estudo de interação 3D com os demais valores de *TP* reportados na literatura 2D.

Neste sentido, o presente método recomenda que, além das técnicas de interação a serem estudadas, seja também avaliada interação com um *mouse*. Neste caso, basta avaliar o plano frontal, que é o plano natural de uso deste dispositivo e que é comumente testado na literatura. Para esta avaliação, a posição do cursor deve ser controlada pelo *mouse* e ficar restrita ao plano frontal, tornando a interação semelhante àquela experimentada em um monitor 2D.

3.1.6 Métricas a serem reportadas

Seguindo a maioria dos trabalhos da literatura, recomenda-se reportar ao menos os valores médios e desvios padrão para as métricas *MT*, *ER* e *TP*, para cada técnica de interação e para cada plano avaliados. Sugere-se ainda que seja reportado também a taxa de ineficiência

(*IR*) para quantificar o aspecto de coordenação.

Por fim, é importante que o método de avaliação preveja também uma forma de medir a percepção subjetiva do participante sobre cada dispositivo utilizado nos testes propostos. Uma avaliação subjetiva, além de ajudar a validar se a avaliação quantitativa condiz com a opinião do usuário, também pode revelar informações importantes que não seriam obtidas somente pela avaliação quantitativa, como por exemplo quão cansativo é o uso de cada dispositivo. Sugere-se aplicar o questionário para avaliação de conforto indicado no Anexo C do documento da ISO 9241-411 para realizar essa avaliação subjetiva (este documento está reproduzido no Anexo A).

3.2 3DFitts: *software* para aplicar o método de avaliação proposto

Para permitir aplicar o método da seção 3.1, foi necessário desenvolver um *software* para executar as mecânicas propostas em um óculos de RA. Durante o desenvolvimento, o autor utilizou como referência o programa GoFitts (MACKENZIE, 2020), que é frequentemente utilizado por pesquisadores nas avaliações de interação em 2D e que implementa as mecânicas de testes previstas no padrão ISO 9241-411.

O programa foi nomeado “3DFitts” e foi desenvolvido em linguagem C# na plataforma Unity®. A escolha da plataforma de desenvolvimento considerou que atualmente todos os grandes fabricantes de óculos de RV e RA e de dispositivos de interação em 3D dão suporte a ela, fornecendo um *Kit* de Desenvolvimento de *Software* (SDK) que facilita o processo de integração da tecnologia com esta plataforma.

3.2.1 Arquitetura do *software*

O 3DFitts implementa um algoritmo central que permite: a disposição de esferas de tamanhos variados em uma circunferência também de tamanho variado; o controle da orientação do plano da circunferência; a identificação de colisão do cursor com os alvos; a realização dos testes segundo a mecânica prevista no teste multi-direcional descrito na ISO 9241-411 (seção 2.2); e a coleta dos dados (tempo e coordenadas de seleção) requeridos para computar as métricas *MT*, *ER* e *TP*. Este algoritmo é apresentado de modo simplificado e em formato de pseudo-código no Algoritmo 1.

Para controlar o cursor esférico, este algoritmo central se comunica com módulos que representam a técnica de interação a ser estudada. Tais módulos devem ser codificados pelo pesquisador, caso as que ele deseje reproduzir ainda não estejam disponíveis. São nestes módulos que diferentes dispositivos de interação podem ser adicionados ao sistema de avaliação, como por exemplo permitir controlar o cursor com um *mouse* ou com um rastreador de mãos.

Algoritmo 1: Pseudo-código do algoritmo para condução dos testes de interação do 3DFitts

Dado de entrada: *lista_de_comprimentos* contendo valores de W a testar.

Dado de entrada: *lista_de_distancias* contendo valores de D a testar.

Dado de entrada: N contendo o número de alvos em cada teste.

Dado de entrada: *plano_de_teste* contendo a orientação do plano de interação.

Resultado : Dados coletados durante os testes e as métricas objetivas calculadas são escritas em arquivo com identificador único do teste realizado.

INÍCIO

PARA CADA W em *lista_de_comprimentos* **FAÇA**

PARA CADA D em *lista_de_distancias* **FAÇA**

Disponha N esferas cinzas com diâmetro W em uma circunferência de diâmetro D em um plano orientado conforme *plano_de_teste*;

$num_tentativas \leftarrow 0$; $num_erros \leftarrow 0$; Destaque com cor rosa o alvo que deve ser selecionado para iniciar o teste;

ENQUANTO $num_tentativas \leq N$ **FAÇA**

Atualize a posição da esfera que representa o cursor com a posição do cursor informada pelo módulo de interação;

Destaque com cor vermelha os alvos interseccionados pelo cursor;

SE o módulo de interação indicar que houve comando de seleção **ENTÃO**

$contabilizar_selecao \leftarrow FALSO$;

SE $num_tentativas > 0$ **ENTÃO**

$contabilizar_selecao \leftarrow VERDADEIRO$;

SE cursor não intersecciona o próximo alvo do teste **ENTÃO**

| $num_erros \leftarrow num_erros + 1$

FIM

SENÃO SE o cursor intersecciona o alvo inicial **ENTÃO**

| $contabilizar_selecao \leftarrow VERDADEIRO$;

FIM

SE $contabilizar_selecao = VERDADEIRO$ **ENTÃO**

$num_tentativas < num_tentativas + 1$;

Registre o tempo atual e a posição do cursor;

Mude a cor do alvo atual para cinza;

SE $num_tentativas \leq N$ **ENTÃO**

| Destaque com cor azul o próximo alvo, conforme teste da multidirecional da ISO 9241-411;

FIM

FIM

FIM

FIM

Calcule as métricas objetivas do teste (MT , ER e TP) a partir dos resultados registrados;

Registre os dados coletados e as métricas calculadas em um arquivo;

FIM

FIM

FIM

Todo este ambiente 3D pode ser visualizado em diferentes dispositivos de visualização, bastando ao pesquisador adicionar um objeto “câmera” que o represente e configurando-o adequadamente para funcionar com os demais elementos do sistema. Deste modo, o 3DFitts pode ser utilizado em diferentes telas, sejam elas monitores 2D, óculos de RV ou óculos de RA.

Com esta arquitetura, o 3DFitts pode ser utilizado para executar avaliação tanto de interações em 3D quanto em 2D, sem que seja necessário modificar o algoritmo central que executa as mecânicas do teste e coleta e computa as métricas finais. Para isso, durante um teste em 2D, o *software* utiliza uma câmera posicionada ortogonalmente ao plano de testes, configurando-a para apresentar os alvos no modo ortogonal, ou seja, como se fossem circunferências. Assim, o teste multi-direcional tradicional da ISO 9241-411 pode ser reproduzido bastando restringir o movimento do cursor ao plano de teste e definindo seu tamanho para 1 pixel. Como tanto o teste em 2D como o em 3D são executados pelo mesmo algoritmo central, reduz-se o efeito de quaisquer fatores relacionados a possíveis diferenças nas mecânicas de teste.

3.2.2 Interface do usuário

A Figura 8 apresenta quatro capturas de telas para ilustrar a interface de usuário do 3DFitts. Na tela principal (Figura 8a), o pesquisador pode configurar o tipo de teste que deseja realizar, incluindo a técnica de interação, o dispositivo de visualização e a orientação do plano que deseja testar. Nesta tela também é possível acionar procedimentos para calibrar diferentes dispositivos conforme for necessário para um teste. As demais telas representam a execução de um teste no modo 2D, ilustrando o início (Figura 8b), a execução (Figura 8c) e a tela de resultados de um teste (Figura 8d).

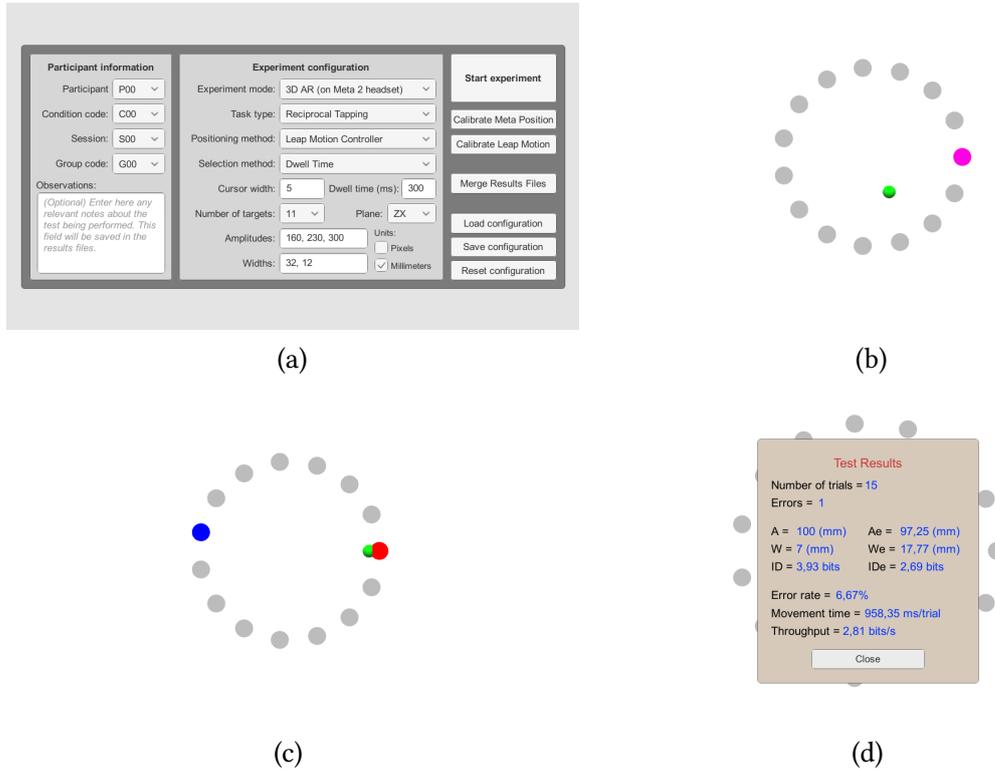
A Figura 9 ilustra a visão que um participante tem ao avaliar interação com as mãos no modo 3D utilizando um óculos de RA. Neste caso, as esferas do teste suplementam a realidade e o cursor é posicionado em uma região da mão do usuário.

3.2.3 Configurações possíveis

O pesquisador pode controlar a execução dos testes configurando diversos parâmetros na tela inicial:

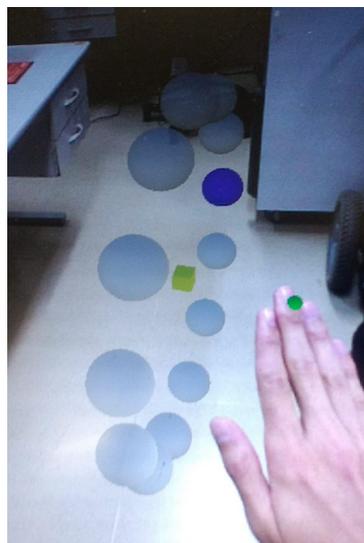
- a) modo do experimento – 2D ou 3D;
- b) tipo da tarefa – além da tarefa de apontar-e-selecionar recíproca, também está implementado (mas ainda em fase de testes) a tarefa de arrastar, prevista no padrão ISO 9241-411;
- c) método de posicionamento do cursor – determina diferentes formas de se controlar

Figura 8 – Capturas de tela da interface do 3DFitts. (a) Tela inicial do programa. (b) Visão do início de um teste multi-direcional da norma ISO para avaliar interação em 2D. O teste se inicia após o participante selecionar o alvo rosa. A esfera verde representa o cursor. (c) Demonstração da execução de um teste. Quando interação com um alvo é possível, ele é destacado em vermelho. O próximo alvo a ser selecionado é indicado pela cor azul. (d) Tela de resultados exibida ao fim de um teste.



Fonte: Couto, José e Lopes (2020).

Figura 9 – Visão que um participante tem ao executar o teste multi-direcional em 3D usando o 3DFitts com um óculos de RA.



Fonte: autoria própria.

- a posição do cursor, como por exemplo, utilizando o *mouse* ou o Controlador Leap Motion;
- d) método de seleção – determina a forma como o participante pode acionar a seleção de um alvo, sendo que atualmente isto é possível utilizando o clique do botão esquerdo do *mouse*, a barra de espaço do teclado, ou a técnica de tempo de permanência¹;
 - e) tamanho do cursor – determina o diâmetro do cursor;
 - f) tempo de permanência – determina quanto tempo o sistema deve esperar antes de considerar uma seleção quando usando a técnica de tempo de permanência como método de seleção;
 - g) número de alvos – determina quantas esferas devem ser apresentadas em um teste;
 - h) plano – determina a orientação do plano de teste;
 - i) amplitudes – determina o parâmetro D para os testes, isto é, quais distâncias entre alvos devem ser testadas;
 - j) comprimentos – determina o parâmetro W para os testes, isto é, quais diâmetros de alvos devem ser testados;
 - k) unidades – o pesquisador pode escolher entrar com informações de tamanho em *pixels* ou milímetros.

3.2.4 Dispositivos suportados

A seguir, detalha-se os dispositivos atualmente suportados pelo 3DFitts. Como o *software* será disponibilizado como código aberto, outros pesquisadores poderão adicionar novos dispositivos ou modos de interação, permitindo realizar testes em uma ampla variedade de sistemas interativos.

3.2.4.1 Dispositivos para visualização

É possível visualizar o conteúdo do teste tanto por meio de um monitor 2D (a própria tela do computador em que o 3DFitts estiver executando) quanto por meio dos óculos de RA Meta 2, que foram utilizados como base para desenvolver este trabalho. O Meta 2 foi integrado ao 3DFitts utilizando a versão 2.0.7.38 do *MetaSDK*.

¹ *Dwell Time*, em inglês. Método em que o evento de seleção só acontece após a interação entre o cursor e o alvo se mantiver por um tempo pré-determinado.

3.2.4.2 Dispositivos para interação

O 3DFitts possui módulos que permite controlar o cursor com o sistema de rastreamento das mãos do Meta 2, com um Controlador Leap Motion e com um *mouse*. O Controlador Leap Motion foi integrado ao 3DFitts utilizando o SDK *Leap Motion Core*, versão 4.4.0.

Um módulo para dar suporte ao controle HTC VIVE também foi desenvolvido, mas seu uso atualmente é complicado, pois demanda realizar configurações extras no sistema operacional.

3.3 Experimentos conduzidos

Quatro experimentos foram conduzidos durante esta pesquisa e serão detalhados no Capítulo 4. O primeiro e o segundo experimentos foram concebidos para verificar se o algoritmo implementado no 3DFitts retorna resultados similares ao *software* de referência (GoFitts) quando utilizados sem e com participantes humanos, respectivamente. Nestes experimentos, ambos os programas foram utilizados para avaliar interação em 2D com um *mouse*, utilizando para isso o teste multi-direcional tradicional da ISO 9241-411. O terceiro experimento teve como objetivo principal utilizar o 3DFitts para estudar o impacto que a adição dos óculos de RA causam no desempenho dos participantes. Para isso, eles foram requisitados a realizarem uma tarefa semelhante à do experimento dois, porém visualizando os alvos em 3D com o Meta 2. Por fim, o experimento quatro teve como objetivo avaliar duas técnicas de interação com as mãos em um óculos de RA, conforme o método estabelecido na seção 3.1.

3.3.1 Autorização do Comitê de Ética em Pesquisa

Os experimento 2, 3 e 4 são estudos com usuários e, portanto, necessitaram de aprovação do Comitê de Ética em Pesquisa (CEP). Um projeto detalhando o desenho destes experimentos foi submetido para aprovação ao CEP do Hospital Universitário da Universidade de São Paulo, tendo sido aprovado no parecer de número 3.461.869.

4 EXPERIMENTOS CONDUZIDOS

4.1 Experimento 1 - Comparativo entre o 3DFitts e o GoFitts em um cenário sem humanos

Um experimento foi conduzido para verificar se o 3DFitts e o GoFitts produzem resultados equivalentes quando uma mesma sequência de interações é utilizada como entrada para realizar o mesmo teste de interação em 2D. Para isso, utilizou-se um *software* de auto-clique – que é capaz de gravar e reproduzir tanto a posição quanto os eventos de clique do *mouse* ao longo do tempo – para capturar uma interação inicialmente realizada pelo autor e posteriormente reproduzida em ambos os programas.

4.1.1 Método

4.1.1.1 Aparato

Um *notebook* Dell G7 (tela de 15,6”, placa de vídeo NVIDIA GeForce GTX 1060 Max-Q, processador Intel Core i7-8750H, memória RAM de 16GB) executando Windows 10 foi utilizado neste experimento. Tanto o GoFitts quanto o 3DFitts foram utilizados para executar o teste multi-direcional da ISO 9241-411. O *software* *AutoClicker*¹ foi utilizado para gravar e reproduzir interações com o *mouse*.

O 3DFitts foi configurado para executar o teste no modo 2D e com um tamanho de cursor de 1 *pixel*, de modo que a interação se assemelhasse com aquela do GoFitts.

4.1.1.2 Delineamento experimental

Este experimento utilizou um delineamento de medidas repetidas, sem participantes, com duas variáveis independentes:

- *Software*: { GoFitts, 3DFitts }
- *Sessão*: { 1, 2, 3, 4, 5 }

O *software* é a principal variável independente a ser analisada. A variável de sessão foi adicionada para coletar medidas repetidas e, assim, verificar a variabilidade das respostas

¹ <<https://sourceforge.net/projects/orphamielautoclicker/>>

devido ao comportamento do sistema computacional.

A distância entre alvos foi configurada para 800 *pixels* e tamanho dos alvos foi 15 *pixels*. O teste multi-direcional foi configurado com 15 alvos e, portanto, 15 tentativas de seleção foram coletadas em cada sessão. Começando com o GoFitts, cada programa foi utilizado alternadamente até que todas as 5 sessões fossem realizadas com cada um deles.

As variáveis dependentes foram *MT*, *ER* e *TP*.

Um total de 150 tentativas foram coletadas (15 tentativas * 5 sessões * 2 programas).

4.1.1.3 Procedimentos

As janelas dos programas GoFitts e 3DFitts foram maximizadas para ocupar todo o espaço acima da barra de tarefas do Windows. Foi observado que o centro dos alvos não estavam localizados exatamente na mesma posição da tela, o que implicaria em diferenças nas respostas. Um procedimento de calibração foi adotado para reduzir este problema. Primeiramente, o 3DFitts foi adaptado para permitir ao autor mover os alvos meio pixel por vez pressionando as teclas direcionais do teclado. Em seguida, o autor utilizou o programa de auto-clique para gravar a realização de um teste no GoFitts, durante a qual buscou-se clicar o mais próximo possível das bordas mais externas dos alvos. Essa interação foi então reproduzida repetidamente – pela aplicação de auto-clique – em teste equivalente no 3DFitts, e a cada reprodução o autor moveu os alvos na direção mais conveniente para verificar se os resultados dos programas se aproximavam ou não.

O experimento foi realizado após este procedimento de calibração. Para isso, o autor executou uma única vez um teste multi-direcional no GoFitts, gravando as interações com o programa de auto-clique. O *trackpad* do notebook foi utilizado para realizar este teste e nenhum alvo foi selecionado incorretamente.

Em seguida, o programa de auto-clique foi utilizado para reproduzir a interação gravada cinco vezes em cada um dos programas, de modo alternado. Os dados foram coletados durante esta etapa.

4.1.2 Resultados e discussão

A Tabela 3 apresenta a média geral para *MT*, *ER* e *TP* obtidos para cada um dos programas, considerando todas as sessões. Ambos apresentaram taxas de erro nulas, o que era esperado já que a interação original foi realizada sem nenhum erro. Uma análise de variância aplicando ANOVA revelou que a variável *software* não teve efeito estatisticamente significativo nem no *TP* ($F_{1,8} = 0,30, p >, 05$) nem no *MT* ($F_{1,8} = 0,95, p >, 05$).

A Figura 10 apresenta todas as tentativas coletadas em cada sessão e para cada *software*

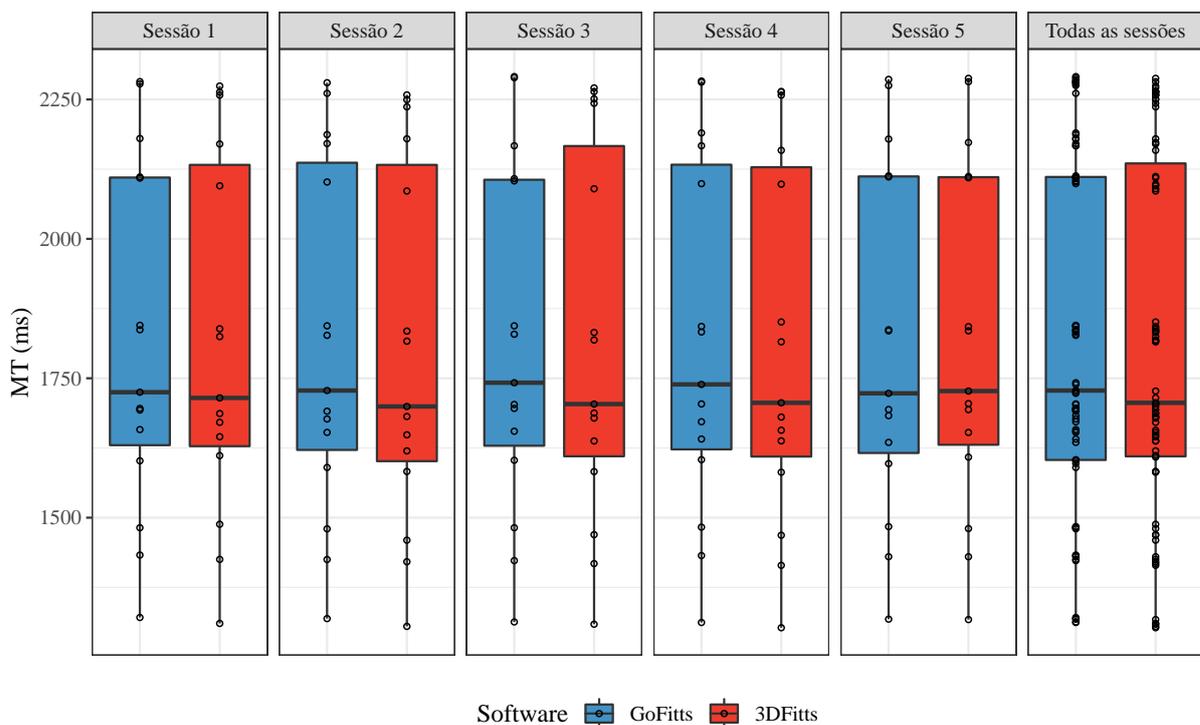
no formato de diagramas de caixa (*boxplots*). Cada círculo apresentado na figura representa uma tentativa de seleção. É possível notar que a sua distribuição ao longo do eixo vertical não varia significativamente, mantendo um formato semelhante ao longo de diferentes combinações de *software* e sessão. Considerando este fator e dado o resultado do teste estatístico, é razoável afirmar que o algoritmo dos dois programas apresentam comportamento semelhante quando estimulados por uma mesma entrada.

Tabela 3 – Valores médios e desvios padrão das variáveis dependentes computados para cada programa testado, considerando todas as sessões.

<i>Software</i>	<i>ER (%)</i>	<i>MT (ms)</i>	<i>TP (bits/s)</i>
GoFitts	0 (SD = 0)	1816 (SD = 2,0)	3,254 (SD = 0,010)
3DFitts	0 (SD = 0)	1814 (SD = 5,6)	3,257 (SD = 0,004)

Fonte: autoria própria.

Figura 10 – Tempo de movimento para que o programa de auto-clique realizasse cada seleção no GoFitts e no 3DFitts, em cada uma das sessões.



Fonte: autoria própria.

4.1.3 Conclusão (Experimento 1)

Os resultados deste experimento trazem evidências de que o algoritmo do teste multi-direcional em 2D do 3DFitts se comporta de forma semelhante ao do GoFitts, indicando que o programa desenvolvido é robusto para avaliar interação em 2D. Como o mesmo algoritmo é utilizado para executar o teste em 3D, este resultado sugere que interações em 3D serão avaliadas pelo 3DFitts com um algoritmo semelhante ao usualmente utilizado na literatura de interação 2D.

4.2 Experimento 2 - Comparativo entre o 3DFitts e o GoFitts em um cenário com humanos

A partir deste experimento o autor passou a conduzir estudos com usuários para avaliar o desempenho de participantes em diferentes condições de interação com sistemas computacionais.

Neste experimento, os participantes avaliaram um *mouse* em um estudo formal de interação 2D utilizando tanto o 3DFitts quanto o GoFitts na execução do teste multi-direcional do padrão ISO 9241-411.

O objetivo principal foi obter a resposta de ambos os programas em uma avaliação que incluísse o fator humano. O objetivo secundário foi medir o desempenho dos participantes com o *mouse* em um teste que pode ser diretamente comparado com boa parte da literatura de interação 2D.

4.2.1 Métodos

4.2.1.1 Participantes

Doze voluntários participaram deste experimento (sete homens, quatro mulheres cis e uma mulher trans). Todos os participantes eram frequentadores do *campus* Butantã da Universidade de São Paulo e incluíam estudantes de graduação, pós-graduação e pesquisadores. A idade do grupo variou entre 21 e 45 anos, apresentando uma média de 28,1 anos ($SD = 6, 2$). Todos eles declararam serem destros e usuários frequentes de computadores, com uma média de uso semanal de 42 horas ($SD = 18$). Quatro declararam ter miopia, um declarou ter hipermetropia e outro, presbiopia. Todos estes participantes foram requisitados a usarem seus óculos durante os testes.

4.2.1.2 Aparato

O mesmo sistema computacional utilizado no Experimento 1 foi utilizado neste experimento. Os programas 3DFitts e GoFitts foram utilizados para executar o teste multi-direcional da ISO 9241-411, conforme seção 2.2. O tamanho do cursor foi configurado para 1 *pixel* no 3DFitts.

Os testes foram realizados com o *Mouse Óptico Básico v2.0* (Modelo 1113) da Microsoft®, que é um dispositivo com fio. Todos os participantes executaram o teste nas mesmas condições ambientais, utilizando a mesma mesa e cadeira. As mesmas aplicações estavam rodando no sistema operacional durante todo o período de testes.

4.2.1.3 Delineamento experimental

O experimento foi delineado principalmente como intraparticipantes com medidas repetidas, incluindo quatro variáveis independentes:

- *Software*: { GoFitts, 3DFitts }
- *D*: { 400, 575, 750 } *pixels*
- *W*: { 30, 80 } *pixels*
- Sessão: {1, 2, 3, 4, 5}

A variável *software* permitirá comparar como cada um dos programas avalia um mesmo tipo de técnica de interação. Os valores de *D* e *W* definem as sequências de teste que devem ser executadas e foram escolhidos para permitir testar a interação em um intervalo razoável de *IDs* (2,58 a 4,70 *bits/s*). A variável sessão foi adicionada para coletar dados repetidos para cada participante, aumentando a capacidade de análise da variabilidade das respostas e permitindo estudar possíveis efeitos de aprendizado.

O método de seleção utilizado em ambos os programas foi o clique com o botão esquerdo do *mouse*. As sequências de teste foram apresentadas a cada participante sempre na mesma ordem: da mais fácil (menor *ID*) para a mais difícil (maior *ID*). Esta decisão foi tomada para uniformizar a resposta entre os participantes, garantindo que possíveis variáveis de confusão fossem reduzidas e maximizando a possibilidade de identificar se ambos os programas retornam resultados equivalentes sob as mesmas condições de teste.

Para reduzir a influência de possíveis efeitos de aprendizado devido aos programas, eles foram testados em ordem alternada. Para isso, os participantes foram aleatoriamente atribuídos a um de dois grupos. O grupo determinava com qual dos programas o teste deveria ser iniciado,

conforme apresentado na Tabela 4. Em termos de delineamento experimental, grupo foi uma variável independente entre participantes.

As variáveis dependentes foram *MT*, *ER* e *TP*.

Cada um dos testes foi executado com 15 alvos, resultando em 15 tentativas de seleção por sequência. O total de tentativas coletadas, portanto, foi de 10800 (15 alvos * 3 *D* * 2 *W* * 5 sessões * 2 programas * 12 participantes).

4.2.1.4 Procedimentos

Os testes foram executados com um participante por vez. Antes de iniciar os testes, os participantes receberam uma explicação oral sobre a pesquisa e sobre os três experimentos a que estavam sendo convidados a participar (Experimentos 2, 3 e 4), assinando o Termo de Consentimento Esclarecido aprovado pelo CEP.

Em seguida, os programas 3DFitts e GoFitts foram apresentados e os participantes foram instruídos sobre as mecânicas do teste multi-direcional. Foi solicitado que realizassem uma sessão de prática para se familiarizarem com o procedimento de teste e esclarecerem possíveis dúvidas. Todos foram instruídos a selecionar os alvos o mais rápido e precisamente possível, isto é, buscando selecionar os alvos em seu centro. Também foram instruídos a

Tabela 4 – Ordem de teste dos programas utilizada pelo participante, conforme seu grupo.

Ordem de teste	Sessão	Grupo	
		A	B
1	1	GoFitts	3DFitts
2		3DFitts	GoFitts
3	2	GoFitts	3DFitts
4		3DFitts	GoFitts
5	3	GoFitts	3DFitts
6		3DFitts	GoFitts
7	4	GoFitts	3DFitts
8		3DFitts	GoFitts
9	5	GoFitts	3DFitts
10		3DFitts	GoFitts

Fonte: autoria própria.

continuarem o teste caso um erro fosse cometido e a evitarem o cometimento de muitos erros.

Finalizada esta etapa, iniciou-se o teste formal. Seguindo a ordem da Tabela 4, cada participante executou as cinco sessões de teste alternando o programa utilizado no teste.

Pausas eram permitidas entre cada sessão caso o participante desejasse, mas durante uma sessão os participantes foram requisitados a manterem o foco. Todo o experimento durou cerca de 30 minutos por participante.

4.2.2 Resultados e discussão

Um total de 184 tentativas de seleção (1,7%) apresentaram MT e A_e que desviaram mais de 3 desvios padrão da sua respectiva média. Elas foram consideradas pontos fora da normalidade (*outliers*) e removidos das análises posteriores. Os valores de média e desvio padrão desta análise foram computados para todas as tentativas de um participante em uma determinada sequência de teste, envolvendo as tentativas de todas as sessões. Deste modo, os pontos eliminados representam respostas incomuns para um mesmo participante em uma determinada sequência.

Após a remoção destas tentativas, as remanescentes foram utilizadas para calcular o MT , ER e TP de cada sequência avaliada, resultando em 720 amostras de dados (6 sequências * 5 sessões * 2 programas * 12 participantes) que foram utilizadas nas análises subsequentes.

4.2.2.1 Efeitos de aprendizagem

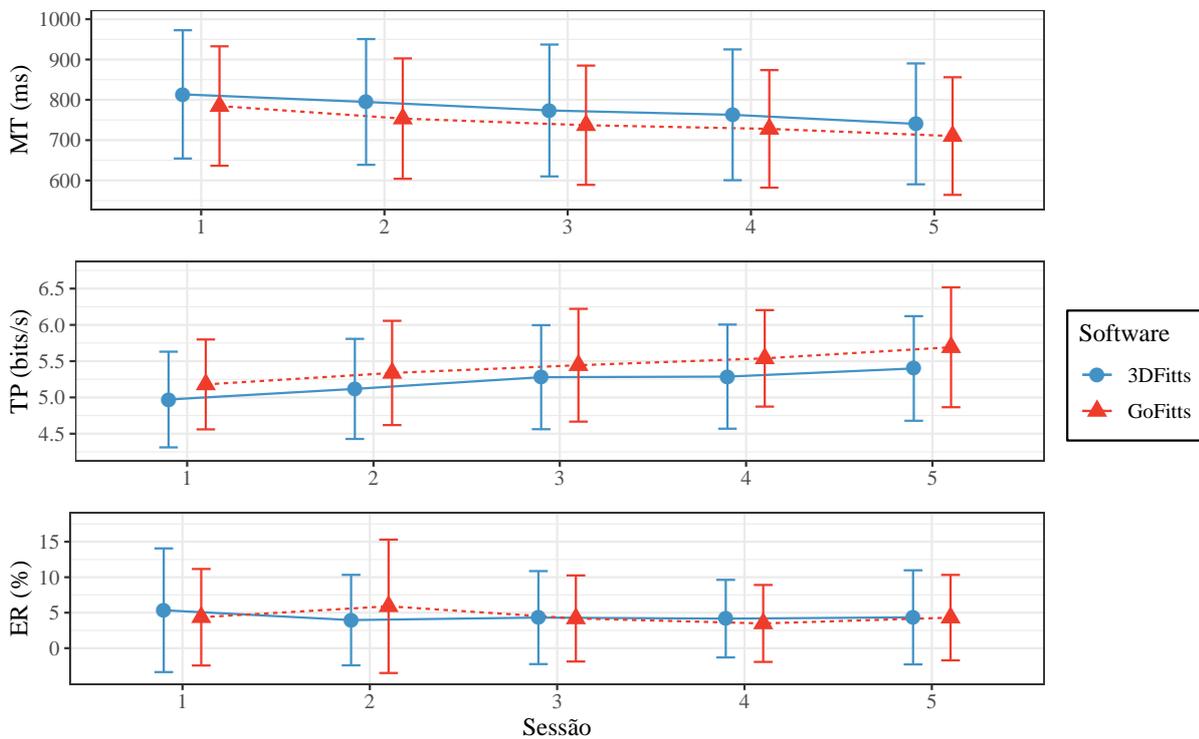
A Figura 11 apresenta as médias gerais, considerando todos os participantes, das variáveis dependentes para cada sessão e programa.

Em uma análise geral, comparando os valores obtidos para a última sessão com a média dos valores das outras quatro sessões, observou-se uma queda de 5,6% no MT , um aumento de 5,2% no TP e uma queda de 3,1% na ER . Uma análise de variância indicou que a sessão apresentou um efeito estatisticamente significativo tanto para o MT ($F_{4,44} = 23,16, p <, 001$) quanto para o TP ($F_{4,44} = 19,27, p <, 001$). Nenhum efeito significativo foi observado para ER ($F_{4,44} = 0,62, ns$).

O teste de Tukey foi executado para as variáveis MT e TP em uma análise *post hoc* e revelou um efeito estatisticamente significativo entre todos os pares de sessão não-consecutivos. Para os pares consecutivos, todos resultaram em efeitos não significativos, com exceção do par formado pela primeira e segunda sessões quando analisado para MT .

Contrastes de Helmert com a correção *bonferroni* do valor p indicaram que todas as sessões foram estatisticamente significativas quando comparadas com a média das sessões anteriores a ela, tanto para MT quanto para TP .

Figura 11 – Média geral de todos os participantes para cada sessão e *software* para as variáveis dependentes *MT*, *ER* e *TP*. Os intervalos de confiança indicam um desvio padrão.



Fonte: autoria própria.

Estes resultados indicam que a sessão utilizada como prática e o número de sessões escolhidas para o experimento não foram suficientes para remover todos os efeitos de aprendizagem. Eles também sugerem que a taxa de aprendizado não foi elevada, visto que não houve efeito significativo para as métricas entre duas sessões consecutivas. Devido a estas considerações, decidiu-se manter todos os dados para as análises subsequentes.

4.2.2.2 Efeitos devido ao grupo do participante

A Tabela 5 apresenta os valores médios para cada variável dependente computadas para cada grupo do experimento. Eles foram calculados primeiramente para cada participante, considerando todas as amostras coletadas em todas as sequências (todas as sessões e *software*), e então para todos os participantes de cada grupo.

Quando analisado como efeito principal em uma análise de variância intraparticipantes com medidas repetidas, a variável independente grupo não apresentou efeito significativo em nenhuma das variáveis dependentes. Entretanto, houve efeito significativo em *MT* na interação entre grupo e sessão e entre grupo e *software*, assim como houve efeito significativo em *ER* na interação grupo-*software*-sessão.

Tabela 5 – Valores médios e desvios padrão para ER , MT e TP , para cada grupo.

Grupo	ER (%)	MT (ms)	TP (bits/s)
A	6,3 (SD = 4,0)	710 (SD = 62)	5,49 (SD = 0,51)
B	2,5 (SD = 2,1)	810 (SD = 119)	5,16 (SD = 0,54)

Fonte: autoria própria.

Em teoria, tais resultados não eram esperados: como a diferença entre grupos foi somente o programa de início dos testes, nenhuma diferença de desempenho deveria ser observada. Porém, este efeito pode ocorrer não devido à mecânica atribuída ao grupo, mas aos participantes que foram alocados a eles. De fato, uma análise mais detalhada dos dados revelou que, por pura chance, os dois participantes que desempenharam melhor nos testes estavam no grupo A, enquanto os dois mais lentos estavam no grupo B.

4.2.2.3 Modelos preditivos pela Lei de Fitts

O desempenho de cada participante com cada programa foi modelado utilizando a formulação da Lei de Fitts proposta pela ISO 9241-411. Para cada sequência e *software*, computou-se os pares (MT, ID_e) por meio do valor médio dos valores de MT , A_e e W_e obtidos para as 60 sequências testadas pelos 12 participantes nas 5 sessões. O cálculo de ID_e seguiu a Equação 2.4. Os pontos resultantes deste processo estão tabulados na Tabela 6.

A Figura 12 apresenta os modelos lineares obtidos. Como pode ser observado, ambos os modelos apresentaram um coeficiente de determinação elevado, indicando que mais de 98% da variabilidade nos dados pode ser explicada pelo modelo linear. Este resultado indica que os movimentos dos participantes conformaram à Lei de Fitts. Além disso, a figura demonstra que ambos os programas testados apresentaram resultados que retornaram modelos preditivos semelhantes, indicando que o algoritmo do 3DFitts consegue avaliar interação em 2D de modo similar ao do GoFitts. Entretanto, o desempenho geral medido pelo 3DFitts foi consistentemente pior, mesmo que por uma margem pequena, aos retornados pelo GoFitts.

4.2.2.4 Variáveis dependentes

As médias gerais e desvios padrão obtidos para as variáveis dependentes são apresentadas na Tabela 7. Estes valores foram calculados a partir dos valores médios observados para cada participante, considerando todas as sequências de teste e sessões.

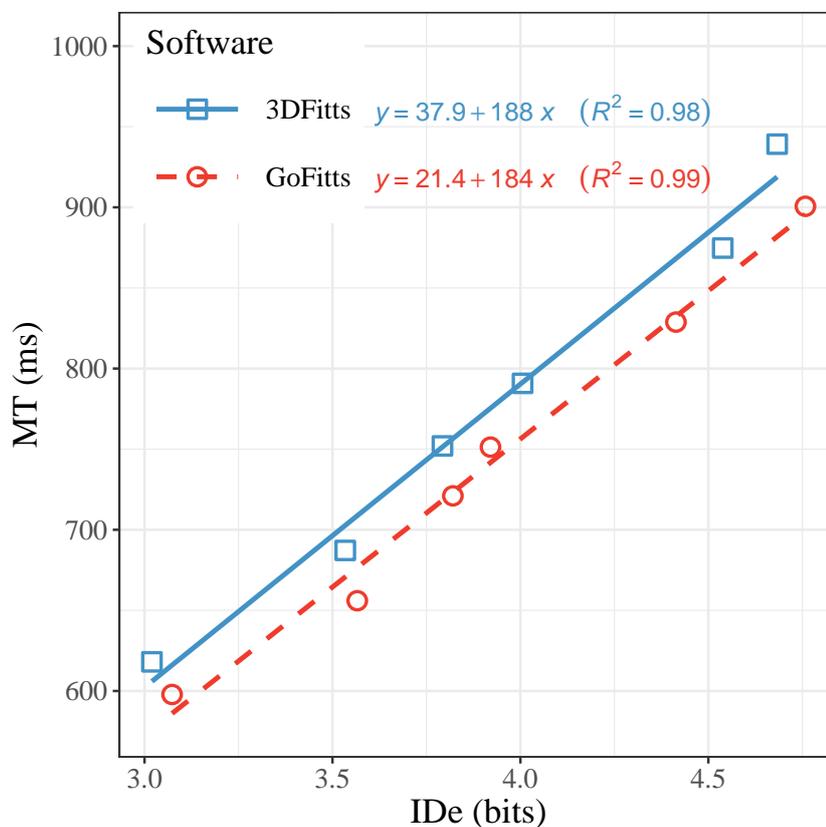
Ambos os programas apresentaram taxas de erros semelhantes e a diferença não foi estatisticamente significativa ($F_{1,11} = 0,009, ns$). Interação com o GoFitts apresentou MT

Tabela 6 – Pares dos índices de dificuldade efetivo e tempo médio de movimento para cada programa, em cada sequência de testes, computados para todos os participantes e sessões.

D (pixels)	W (pixels)	ID (bits)	GoFitts		3DFitts	
			ID_e (bits)	MT (ms)	ID_e (bits)	MT (ms)
400	30	2,59	3,09	598	3,04	618
575	30	3,03	3,60	656	3,58	687
750	30	3,38	3,84	721	3,83	752
400	80	3,84	3,93	751	4,02	791
575	80	4,33	4,43	829	4,55	875
750	80	4,70	4,77	901	4,69	939

Fonte: autoria própria.

Figura 12 – Regressões lineares resultantes ao aplicar a Lei de Fitts para modelar o desempenho dos participantes ao realizar o teste multi-direcional com o 3DFitts e GoFitts.



Fonte: adaptado de Couto, José e Lopes (2020).

Tabela 7 – Médias gerais desvios padrão das variáveis dependentes ao se utilizar o 3DFitts e o GoFitts para realizar o teste multi-direcional da ISO 9241-411 na avaliação de interação com um *mouse*.

<i>Software</i>	<i>ER (%)</i>	<i>MT (ms)</i>	<i>TP (bits/s)</i>
GoFitts	4,5 (SD = 3,6)	743 (SD = 101)	5,44 (SD = 0,55)
3DFitts	4,4 (SD = 3,7)	777 (SD = 109)	5,21 (SD = 0,52)

Fonte: adaptado de Couto, José e Lopes (2020).

4,4% menor do que o reportado pelo 3DFitts, sendo que esta diferença foi estatisticamente significativa ($F_{1,11} = 52,52, p <, 001$). Da mesma forma, *TP* para o GoFitts foi 4,4% maior que o do 3DFitts, diferença também estatisticamente significativa ($F_{1,11} = 108,6, p <, 001$). Os tamanhos de efeitos, segundo número de Cohen, foram $d = 0,22$ para o *MT* e $d = 0,31$ para o *TP*. Ambos são considerados pequenos e sugerem que, apesar de haver uma diferença significativa, ela não apresenta uma magnitude relevante.

Ainda que o efeito seja pequeno, os resultados indicam que a realização de testes com o 3DFitts apresenta um desempenho menor que com o GoFitts. Porém, conforme observado no Experimento 1, quando o fator humano não está presente, esta diferença não aparece. Portanto, é plausível assumir que as diferenças significativas observadas neste experimento se devem ao comportamento humano e não a possíveis diferenças no algoritmo de ambos os programas. Existe, de fato, diferenças entre os *feedbacks* visuais e sonoros providenciados por cada programa. Além dos sons no 3DFitts serem mais pronunciados que no GoFitts, os alvos são apresentados em alto contraste no GoFitts – são apenas circunferências pretas num fundo branco – enquanto que no 3DFitts eles são exibidos como círculos cinzas e sem borda (conforme Figura 8b e Figura 8c). Estas diferenças podem ter provocado diferenças nas respostas dos sistemas visual, sonoro e motor, levando ao menor desempenho no 3DFitts. Este é um tópico interessante para pesquisas futuras.

O *TP* geral obtido para o *mouse* com ambos os programas foi 5,32 bits/s. Este valor é maior do que o intervalo geralmente observado na literatura, de 3,7 a 4,9 *bits/s* (SOUKOREFF; MACKENZIE, 2004). O valor mais elevado possivelmente pode ser explicado tanto por uma melhora tecnológica envolvendo interações com o *mouse* quanto por uma maior experiência por parte dos participantes deste estudo no uso do dispositivo em seu dia-a-dia.

4.2.2.5 Características de distribuição estatística

O tempo de movimento e as coordenadas de seleção das amostras coletadas – desconsiderando os pontos fora da normalidade que foram eliminados – foram analisados quanto à normalidade de sua distribuição estatística para garantir que as suposições requeridas para

aplicar as análises de variância e o teste de Tukey foram atendidas.

Os dados coletados de 720 sequências (12 participantes * 5 sessões * 2 programas * 3 distâncias de alvos * 2 tamanhos de alvos) foram analisados. Cada sequência continha entre 13 a 15 tentativas, que foram submetidas à análise de assimetria, curtose e ao teste de normalidade Shapiro-Wilk (com $\alpha = 0,05$).

Para o tempo de movimento, a média dos valores de assimetria para as 720 sequências foi 0,50 ($SD = 0,61$), indicando que os dados apresentaram curvas normais próximas da simetria, mas levemente enviesados à esquerda. A curtose média foi 2,99 ($SD = 1,15$), indicando dados que se conformam à distribuição normal. O teste de Shapiro-Wilk não pôde rejeitar a hipótese nula para normalidade dos dados para 81% das sequências.

As coordenadas de seleção, por sua vez, apresentaram assimetria média igual a -0,001 ($SD = 0,60$) e curtose média igual a 2,80 ($SD = 0,96$), ambos os valores sugerindo uma boa conformidade dos dados à distribuição normal. O teste de Shapiro-Wilk não rejeitou a hipótese nula de normalidade de 93% das sequências.

Todos estes resultados sugerem que boa parte dos dados coletados podem ser modelados por uma distribuição normal.

4.2.3 Conclusão (Experimento 2)

O principal objetivo deste experimento foi verificar se a avaliação da interação com o *mouse* em 2D utilizando o 3DFitts reportaria métricas equivalentes às informadas pelo *software* de referência. Os resultados indicam que as duas aplicações retornaram resultados similares. Além disso, como objetivo secundário desejava-se obter um valor de TP para comparar este trabalho com literatura 2D. O valor obtido foi de 5,32 *bits/s*.

Deste modo, os Experimentos 1 e 2 fornecem evidências de que o algoritmo central do 3DFitts foi corretamente implementado e é capaz de executar as mecânicas do teste multidirecional, coletar os dados e computar as métricas de desempenho conforme a literatura de referência.

4.3 Experimento 3 - Uso do 3DFitts para transpor o teste multidirecional 2D para RA

Os dois experimentos anteriores produziram evidências que indicam que o algoritmo do 3DFitts se comporta de maneira similar ao do GoFitts. Como o 3DFitts também utiliza o mesmo algoritmo para avaliar interação em 3D, pode-se concluir os resultados que ele reporta são condizentes com a literatura de avaliação de interação em 2D.

Antes de prosseguir e utilizar o 3DFitts em uma avaliação formal de técnicas de interação com as mãos em óculos de RA, porém, é importante entender qual o impacto que a adição destes óculos ao teste pode causar no desempenho dos participantes. O novo *hardware* pode, por exemplo, adicionar um atraso substancial no sistema de interação – afetando em especial o retorno visual – o que poderá provocar um impacto significativo no desempenho. O objetivo deste experimento foi mensurar este impacto.

4.3.1 Métodos

Para medir o impacto que a adição dos óculos de RA provoca no desempenho dos participantes, é necessário comparar duas condições de teste similares: uma com o *hardware* e outra sem ele, com tudo o mais semelhante. O Experimento 2 já permitiu medir o *TP* do *mouse* com o 3DFitts no modo de interação 2D. Assim, ao conduzir um teste em condições semelhantes a este último experimento em um ambiente imersivo de RA, pode-se atribuir eventuais divergências nas respostas ao novo *hardware*.

Uma possível variável de confusão que pode afetar este experimento é a diferença de percepção visual que o participante tem em uma tela 2D e no ambiente de visualização estereoscópica. Por exemplo, os participantes verão alvos tridimensionais com os óculos de RA, e não bidimensionais, e estes alvos estarão suplementando o mundo virtual, e portanto serão mais difíceis de discernir. Estas diferenças possivelmente afetarão o desempenho, refletindo em menor *TP*. Para minimizar estes efeitos, decidiu-se realizar os testes posicionando os alvos na frente de um monitor 2D desligado. Deste modo, o contraste dos alvos ficará mais nítido e a condição de teste ficará mais próxima daquelas utilizadas no Experimento 2.

4.3.1.1 Participantes

Os mesmos 12 voluntários do Experimento 2 participaram deste estudo. Sete deles já haviam experimentado óculos de RV no passado, sendo que um deles declarou ser usuário frequente deste tipo de tecnologia, utilizando-a cerca de 8 horas por semana. Nove dos participantes nunca havia utilizado óculos de RA. Os outros três já haviam experimentado a tecnologia em atividades demonstrativas.

4.3.1.2 Aparato

O mesmo sistema computacional dos experimentos anteriores foi utilizado neste. Escolheu-se realizar os testes com os óculos de RA Meta 2, da Metavision, que possui uma resolução de 2550x1440, um campo de visão de 90° e uma taxa de atualização de 60 Hz. Uma

versão *standalone* do 3DFitts foi utilizado para executar a versão 3D do teste multi-direcional, porém apenas no plano frontal. Conforme ilustrado na Figura 13, um monitor 2D com tela de 19" foi utilizado, desligado, como superfície de referência onde o plano frontal foi posicionado.

O Meta 2 apresenta uma funcionalidade de oclusão que, quando ativada, permite ao sistema identificar quando elementos do mundo estão à frente dos elementos virtuais e redesenhe os objetos virtuais de forma a escondê-los. Esta funcionalidade aprimora a percepção de profundidade, mas em testes prévios seu comportamento errático foi considerado prejudicial à realização dos testes. Por exemplo, ao utiliza-lo em testes com as mãos, o cursor por vezes ficava escondido dentro da mão do usuário e a qualidade da oclusão variava constantemente, acabando por confundir a percepção de profundidade ao invés de aprimorá-la. Por estes motivos, esta funcionalidade foi desativada. Por causa desta decisão, os objetos virtuais (alvos e cursor) estarão sempre visíveis por cima dos objetos do mundo real.

Os testes foram executados com o mesmo *mouse* e técnica de seleção (clique com o botão esquerdo do *mouse*) do Experimento 2. A posição espacial do cursor foi limitada ao plano em que o centro dos alvos estavam. Neste modo 3D, o cursor foi representado por uma esfera verde com 5 mm de diâmetro.

4.3.1.3 Delineamento experimental

O experimento foi delineado como intraparticipantes com medidas repetidas e incluiu três variáveis independentes:

- $D: \{ 72, 103,5, 135 \}$ mm

Figura 13 – Aparato utilizado no Experimento 3. Por meio dos óculos de RA, os participantes podiam visualizar os alvos esféricos posicionados no plano frontal sobreposto ao monitor de tela plana.



Fonte: autoria própria.

- $W: \{ 5,4, 14,4 \}$ mm
- Sessão: $\{1, 2, 3, 4, 5\}$

Os valores de D e W são expressos em milímetros pois esta unidade é mais adequada para representar os objetos virtuais no mundo tridimensional virtual. O SDK MetaSDK é responsável por reproduzir este tamanho nos óculos, tornando-o semelhante à medida do mundo real. Para obter os valores destas variáveis, primeiro obteve-se a taxa de milímetros/*pixel* do *notebook* Dell G7 – medindo-se a tela com uma régua e dividindo-se o valor em milímetros pela resolução da tela em *pixels* – e em seguida multiplicou-se esse fator pelos tamanhos em *pixels* utilizados no Experimento 2. A taxa foi determinada como sendo 0,18 milímetros/*pixel*. Após configurar o 3DFitts com estes valores, o autor verificou o tamanho visual aparente das esferas com os alvos do Experimento 2 (sobrepondo-os visualmente). Foi constatado que os tamanhos são semelhantes. Como todos os valores são proporcionais com os do Experimento 2, o mesmo intervalo de valores de ID s foi testado neste experimento (2,58 to 4,70 *bits*).

A variável sessão foi adicionada para coletar medidas repetidas, melhorando a capacidade de análise estatística dos resultados e permitindo analisar possíveis efeitos de aprendizado ao longo das repetições.

O método de seleção utilizado foi o mesmo do Experimento 2, ou seja, cliques com o botão esquerdo do *mouse*.

As variáveis dependentes foram MT , ER e TP .

Cada sequência foi executada com 15 alvos, resultando em 15 tentativas de seleção por sequência. Assim, um total de 5400 tentativas foram coletadas neste experimento (15 alvos * 3 distâncias entre alvos * 2 tamanhos de alvos * 5 sessões * 12 participantes).

4.3.1.4 Procedimentos

Antes do experimento, a posição do mundo virtual do Meta 2 foi manualmente ajustada pelo autor para manter o centro do plano de testes no centro do monitor. O Meta 2 possui um procedimento para realizar uma calibração ótica que adequa o dispositivo às condições visuais de diferentes usuários. Entretanto, este procedimento não foi realizado e todos os participantes utilizaram os óculos com a calibração inicialmente feita pelo autor.

Os testes foram executados com um participante por vez. Os participantes foram solicitados a utilizarem seus óculos cotidianos para correção de visão durante os testes. O autor explicou os objetivos do experimento, apresentou os óculos Meta 2 e solicitou que o participante os vestisse. Uma sessão prática foi conduzida para que os participantes se ambientassem ao ambiente tridimensional e às novas condições do teste.

Os testes formais foram conduzidos logo após a sessão prática. Cada participante

completou as cinco sessões de testes, sendo permitido pausas entre as sessões. Cada sessão envolvia testar todas as seis sequências de teste, o que foi feito sempre na mesma ordem, seguindo a mesma mecânica do Experimento 2: da mais fácil (menor ID) à mais difícil (maior ID). Utilizar esta ordem tem a vantagem de reduzir o esforço cognitivo do participante ao longo da realização dos testes. Porém, ela também pode trazer efeitos de confusão. Por exemplo, os participantes sempre estarão mais cansados – tanto física quanto psicologicamente – ao realizarem as últimas sequências (que também serão as mais difíceis). Contudo, ao mesmo tempo, quanto mais praticam, mais hábeis os participantes ficam e, assim, um efeito pode balancear o outro.

A duração média do experimento foi de 10 minutos para cada participante.

4.3.2 Resultados e discussão

Um total de 90 tentativas de seleção (1,7%) apresentaram MT ou A_e que desviaram mais de 3 vezes o desvio padrão da média (computada por sequência, considerando todas as tentativas de um mesmo participante). Elas foram consideradas pontos fora da normalidade e removidas das demais análises. As tentativas remanescentes foram utilizadas para calcular o MT , a ER e o TP de cada sequência, para cada sessão e participante, resultando num banco de dados com 360 amostras (3 distâncias entre alvos * 2 tamanhos de alvos * 5 sessões * 12 participantes) que foi analisado para obter os resultados apresentados a seguir.

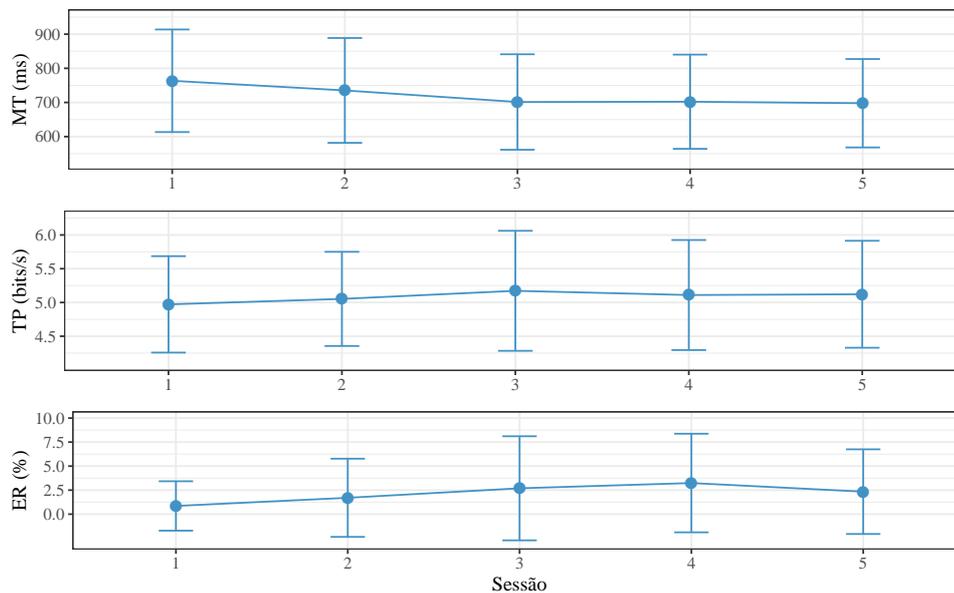
4.3.2.1 Efeitos de aprendizagem

A Figura 14 apresenta os valores médios obtidos para cada variável dependente considerando todas as interações de todos os participantes em cada sessão.

Comparando-se a última sessão com a primeira, o valor médio do MT caiu 8,6%. Uma análise de variância indicou que a sessão teve um efeito estatisticamente significativo nesta variável dependente ($F_{4,44} = 11,53, p < ,001$). Em uma análise *post-hoc* com o teste Tukey, observou-se que o efeito significativo ocorreu nos pares que incluem as duas primeiras sessões com as últimas sessões, o que sugere que os participantes se tornaram mais rápidos na execução do teste nas duas primeiras sessões, mas atingiram um limite de desempenho a partir de então.

Do mesmo modo, comparando-se a última com a primeira sessão, observa-se um aumento de 3% no TP . Entretanto, neste caso a análise de variância não revelou nenhum efeito significativo de sessão no TP ($F_{4,44} = 1,60, ns$). Como o TP é uma métrica que representa melhor o desempenho geral da interação do que MT , este resultado sugere que qualquer efeito de aprendizagem foi irrelevante para a capacidade de realizar as tarefas do teste ao longo das cinco sessões.

Figura 14 – Valores médios e desvio padrão para as variáveis dependentes ao longo de todas as sessões do Experimento 3.



Fonte: autoria própria.

Apesar do *MT* ter decrescido de maneira significativa, é possível que o aumento na velocidade de interação tenha comprometido a precisão geral da seleção, o que explicaria um efeito não significativo no *TP*. De fato, *ER* aumentou 274% comparando-se a última sessão com a primeira. Também é possível observar na Figura 14 que o aumento da velocidade ao longo das sessões em geral foi acompanhado por um aumento na taxa de erros cometidos, o que indicaria que os participantes não se tornaram mais eficientes na tarefa, apenas trocaram precisão por velocidade.

Portanto, considerando que nenhum efeito de aprendizado foi identificado no desempenho geral dos participantes, dados de todas as sessões serão utilizados nas análises posteriores.

4.3.2.2 Modelos preditivos pela Lei de Fitts

Para modelar os dados com a Lei de Fitts, calculou-se a média dos valores de *MT*, A_e e W_e (obtidos para cada sequência de testes) primeiramente para cada participante e então, entre todos os participantes. Uma análise por participante revelou que o movimento desempenhado por dois dos 12 voluntários não conformou com a Lei de Fitts, apresentando um coeficiente de determinação $R^2 < 0,01$ enquanto todos os demais apresentaram valores acima de 0,60. Portanto, decidiu-se apresentar dois modelos: o Modelo A inclui todos os 12 participantes e o Modelo B, somente os 10 cujos movimentos puderam ser modelados com alguma precisão pela Lei de Fitts quando analisados separadamente. Os dados resultantes estão tabulados na

Tabela 8.

Os dois participantes que apresentaram movimentos que não conformaram com a Lei de Fitts chegaram a mencionar que estavam vendo dois círculos sobrepostos ao invés de uma esfera. Portanto, uma possível explicação para seu comportamento atípico foi a falta de calibração óptica dos óculos de RA para suas características visuais.

Os modelos lineares obtidos são apresentados na Figura 15. Tanto o Modelo A quanto o Modelo B conseguiram explicar mais de 80% da variabilidade observada nos dados, o que sugere que a Lei de Fitts pode ser utilizada para modelar interações em um teste multi-direcional em ambientes imersivos de RA.

4.3.2.3 Variáveis dependentes

O impacto que a adição dos óculos de RA teve no desempenho dos participantes será analisado comparando os dados deste experimento com os obtidos no Experimento 2. Para isso, serão considerados apenas os dados coletados com o 3DFitts e somente os dados referentes aos 10 participantes cujo movimento conformaram com a Lei de Fitts. Deste modo, a análise resulta em um delineamento experimental intraparticipantes em que a variável “experimento” pode ser tratada como variável independente. As médias gerais obtidas das variáveis dependentes para estes 10 participantes são apresentadas para cada experimento na Tabela 9. As médias foram calculadas primeiro para cada participante, considerando todas as sequências de teste e sessão, e então para todos os 10 participantes.

O MT para o Experimento 3 foi 7,9% menor que o observado no Experimento 2 e essa diferença foi estatisticamente significativa ($F_{1,9} = 21,99, p <, 01$), porém com um tamanho de efeito pequeno ($d = 0,45$). Da mesma forma, ER caiu 54% no Experiment 3 e a diferença foi estatisticamente significativa ($F_{1,9} = 10,41, p <, 05$), com um tamanho de efeito pequeno ($d = 0,45$). Apesar destes resultados sugerirem que o desempenho dos participantes melhorou no Experimento 3, a análise do TP não corrobora esta conclusão: essa métrica foi 1,9% maior no Experimento 2, uma diferença não estatisticamente significativa ($F_{1,9} = 2,57, ns$).

O aumento na velocidade média dos participantes e a melhora na taxa de erros observados neste experimento possivelmente podem ser atribuídos à prática e ao aprendizado que todos os participantes tiveram ao realizarem o Experimento 2.

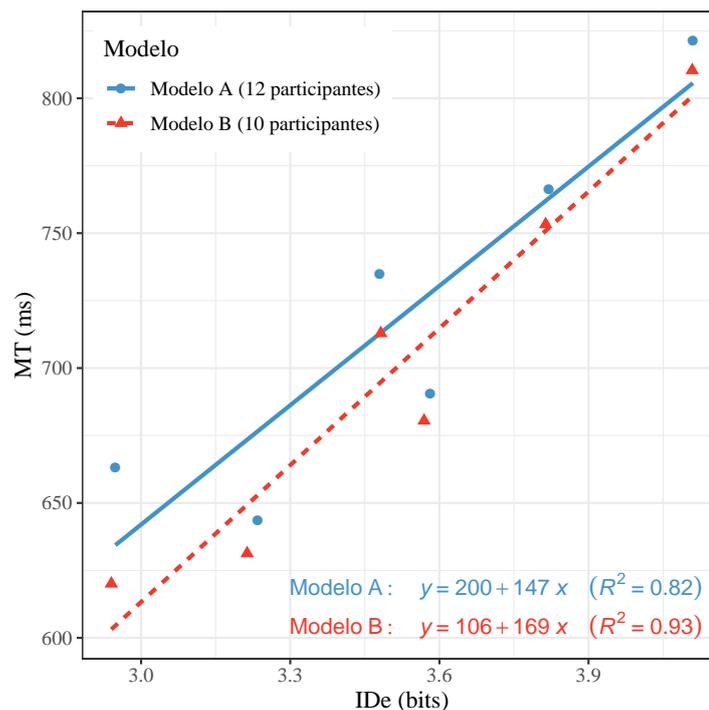
Considerando os resultados obtidos para o TP , pode-se concluir que a adição do Meta 2 não impactou de maneira significativa o desempenho dos participantes quando utilizado para uma interação em condições semelhantes às do Experimento 2.

Tabela 8 – Pares dos índices de dificuldade efetivo e tempo médio de movimento utilizados para modelar o experimento com a Lei de Fitts. O Modelo A considera dados de todos os 12 participantes. O Modelo B considera apenas os 10 participantes cujo movimento conformaram com a Lei de Fitts.

<i>D</i> (mm)	<i>W</i> (mm)	<i>ID</i> (bits)	Modelo A (12 participantes)		Modelo B (10 participantes)	
			<i>ID_e</i> (bits)	<i>MT</i> (ms)	<i>ID_e</i> (bits)	<i>MT</i> (ms)
72	14,4	2,58	2,95	663	2,94	620
103,5	14,4	3,03	3,23	644	3,21	631
135	14,4	3,38	3,58	691	3,57	680
72	5,4	3,84	3,48	735	3,48	713
103,5	5,4	4,33	3,82	766	3,82	753
135	5,4	4,70	4,11	821	4,11	810

Fonte: autoria própria.

Figura 15 – Modelos de regressão lineares obtidos aplicando a Lei de Fitts nos dados coletados. O Modelo A considerou todos os 12 participantes na análise e o Modelo B, somente aqueles cujos movimentos conformaram com a Lei de Fitts.



Fonte: autoria própria.

Tabela 9 – Médias gerais e desvios padrão para MT , ER e TP para o Experimento 2 e 3, considerando apenas os 10 participantes do Modelo B do Experimento 3.

Experimento	ER (%)	MT (ms)	TP (bits/s)
2	5,1 (SD = 3,7)	761 (SD = 95)	5,27 (SD = 0,43)
3	2,4 (SD = 1,7)	701 (SD = 84)	5,17 (SD = 0,42)

Fonte: autoria própria.

4.3.2.4 Características de distribuição estatística

A mesma análise de distribuição estatística realizada no Experimento 2 foi executada para os dados coletados neste experimento após remover os dois participantes cujos movimentos não obedeceram à Lei de Fitts e desconsiderando os *outliers*. Um total de 10 participantes * 5 sessões * 3 distâncias entre alvos * 2 tamanhos dos alvos = 300 sequências de testes foram analisadas, cada uma delas contendo entre 12 e 15 tentativas de seleção.

A assimetria e curtose médias para o MT foram, respectivamente, 0,56 ($SD = 0,62$) e 3,02 ($SD = 1,29$). O teste de normalidade Shapiro-Wilk não pôde rejeitar a hipótese nula de normalidade para 78% das sequências de testes.

As coordenadas de seleção apresentaram uma assimetria média igual a 0,06 ($SD = 0,60$) e uma curtose média igual a 2,78 ($SD = 0,92$). O teste de normalidade não rejeitou a hipótese nula de normalidade para 91% das sequências.

Estes resultados são semelhantes aos obtidos para os dados no Experimento 2, sugerindo que a variabilidade observada no comportamento dos participantes se manteve similar após a adição dos óculos de RA, podendo ser modelados por uma distribuição normal para a maioria das sequências testadas.

4.3.3 Conclusão (Experimento 3)

O objetivo deste experimento foi medir o impacto que a adição do Meta 2 causaria no desempenho dos participantes quando submetidos a uma avaliação com condições semelhantes às do Experimento 2. Os resultados indicaram que, considerando a métrica TP , nenhuma diferença significativa foi observada no desempenho dos participantes.

Este experimento também indicou que o movimento da maioria dos participantes em uma interação com *mouse* em um ambiente imersivo de RA pôde ser modelado pela Lei de Fitts. Porém, para dois participantes isto não foi verdade. A principal hipótese para explicar este comportamento atípico é que a percepção de profundidade destes participantes foi prejudicada devido a uma decisão de não se calibrar opticamente os óculos de RA para cada participante.

Recomenda-se, portanto, que trabalhos futuros envolvendo interação em ambientes estereoscópicos sempre incluam uma etapa de calibração para adequar o dispositivo de visualização ao participante.

4.4 Experimento 4 - Uso do 3DFitts para avaliar duas técnicas de interação com as mãos em óculos de RA

Os experimentos anteriores permitiram confirmar que o algoritmo do 3DFitts apresenta resultados de avaliação equivalentes ao da literatura 2D e indicaram que o ambiente imersivo de RA, a princípio, não impacta o desempenho dos participantes de maneira significativa.

Este experimento foi desenhado como um estudo com usuários para demonstrar o uso e avaliar a capacidade do 3DFitts de atuar como uma ferramenta para avaliar técnicas de interação com as mãos em um sistema com óculos de RA.

O principal objetivo deste experimento é aplicar o método proposto na seção 3.1 para avaliar duas técnicas de interação com as mãos no Meta 2 e verificar se os resultados quantitativos obtidos utilizando o 3DFitts e o método de avaliação podem ser utilizados para discriminar diferenças de desempenho que também são percebidas subjetivamente pelos participantes. A hipótese é que as métricas objetivas computadas pelo 3DFitts permitirão prever qual técnica de interação é preferida pelos participantes.

Um objetivo secundário é verificar se o desempenho das técnicas de interação com as mãos varia com a orientação espacial do plano de teste. Este é um aspecto interessante a ser estudado, uma vez que, a princípio, a qualidade de rastreamento das mãos cai à medida que as mãos se afastam do sensor do dispositivo rastreador.

Um terceiro objetivo é verificar se interações 3D com as mãos num ambiente de RA imersivo podem ser modeladas pela Lei de Fitts, conforme formulação da norma ISO 9241-411.

Um último objetivo é verificar se as hipóteses $H1$ e $H2$ que deram origem a este trabalho são verdadeiras.

4.4.1 Métodos

Para atingir os objetivos propostos, conduzir-se-á o experimento seguindo o método proposto na seção 3.1 para obter os valores de TP para duas técnicas de interação com as mãos, sendo uma delas providenciada pelo Sistema de Interação com as Mãos do Meta 2 (MetaTracker) e outra pelo Controlador Leap Motion. A avaliação será realizada nos planos frontal e horizontal. Adicionalmente, conforme recomendado na seção 3.1, um *mouse* também será avaliado.

A percepção subjetiva dos participantes será mensurada por meio do questionário de

avaliação do conforto sugerido pela ISO 9241-411 (reproduzido no Anexo A). Este questionário é composto por doze afirmações que devem ser julgadas conforme uma escala likert de 7 pontos. As afirmações avaliam aspectos como satisfação, conforto e esforço com a técnica de interação ao executar as tarefas do teste. A versão em português apresentada no Apêndice A foi utilizada nos testes.

4.4.1.1 Participantes

Os mesmos participantes dos experimentos 2 e 3 foram convidados para contribuir com este estudo. Um deles não pôde participar e um outro não conseguiu concluir os testes por ter sido afetado por náusea digital. Portanto, os dados foram coletados para 10 participantes. Destes, quatro declararam ter miopia, um declarou ter hipermetropia e um declarou ter presbiopia. Como mencionado anteriormente, todos são destros.

Os dois participantes que foram excluídos da análise de resultados no experimento 3 por terem apresentado comportamento que não pôde ser modelado pela Lei de Fitts também participaram deste experimento.

4.4.1.2 Aparato

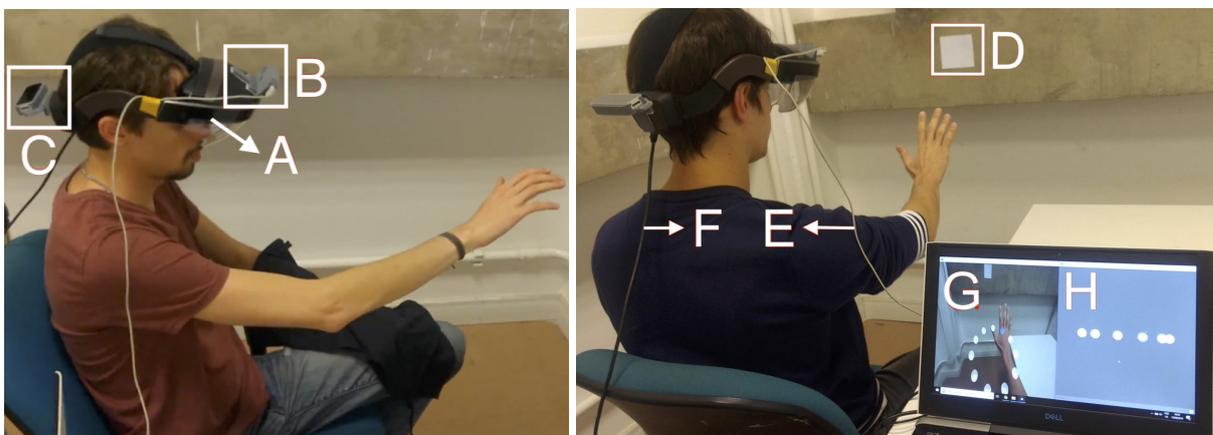
O mesmo sistema computacional utilizado nos experimentos anteriores e os mesmos óculos Meta 2 utilizado no Experimento 3 foram utilizados neste estudo. A aplicação 3DFitts em versão *standalone* foi utilizada para executar a versão 3D do teste multi-direcional, conforme seção 3.1. A funcionalidade de oclusão do Meta 2, que permite a ele ocultar objetos virtuais quando a sua posição espacial é detectada como estando atrás de objetos reais, foi mantida desativada para este experimento.

Foram utilizados três dispositivos para controlar a posição do cursor: o mesmo *mouse* dos Experimentos 2 e 3; o MetaTracker; e o Controlador Leap Motion. A escolha destes últimos dois dispositivos permitirá testar as hipóteses deste trabalho.

O cursor foi configurado para ter 5 mm de diâmetro e, nos casos de interação com as mãos, para ficar posicionado no centro da mão rastreada.

A Figura 16 ilustra o ambiente em que os testes foram executados e destaca os principais elementos do aparato. Para garantir que o campo de visão do Controlador Leap Motion fosse semelhante àquele do MetaTracker, um suporte foi customizado e impresso em uma impressora 3D e colado no topo dos óculos Meta 2 com fita dupla face. O Controlador Leap Motion foi anexado à este suporte. A adição deste conjunto ao Meta 2 alterou seu centro de gravidade, tornando-o muito desconfortável para uso prolongado. Para reduzir este desbalanceamento, um segundo suporte com outro Controlador Leap Motion foi anexado na parte anterior dos

Figura 16 – Aparato envolvido nos testes do Experimento 4. Os óculos Meta 2 (A) foram adaptados para incluir dois Controladores Leap Motion: o primeiro (B) é responsável por rastrear as mãos do usuário, sendo utilizado para controlar a interação durante os testes; o segundo (C) foi adicionado para atuar como um contra-peso à (B). Um quadrado de papel (D) foi colado em uma posição fixa na parede e foi utilizado como referência durante o procedimento de mapeamento do espaço pelo Meta 2, que é responsável por ancorar o mundo virtual no mundo real. O Controlador Leap Motion (B) foi conectado ao *notebook* por meio de um cabo USB (E) e o Meta 2, por meio de cabos USB e HDMI (F). A visão aumentada da realidade que o usuário tem ao utilizar os óculos podia ser visualizada no *Meta 2 Viewer* (G). A aplicação 3DFitts (H) controlou a execução dos testes. A mesa posicionada à direita do participante foi utilizada como suporte durante os testes utilizando o *mouse*.



Fonte: Couto, José e Lopes (2020).

óculos, de modo a atuar como um contra-peso ao primeiro.

4.4.1.3 Delineamento experimental

Seis métodos de interação foram testados e modelados em uma variável independente denominada “condição de teste”. A nomenclatura utilizada para os seis possíveis valores desta variável estão listados na Tabela 10, relacionando-as com os respectivos dispositivos de interação, método de seleção e plano de teste.

Dois métodos de seleção foram utilizados: o clique com o botão esquerdo do *mouse* e seleção por tempo de permanência igual a 300 ms (método conhecido como *dwell time*). O tempo de permanência foi escolhido como método de seleção para avaliar a interação com as mãos para reduzir fontes de ruídos que poderiam afetar as variáveis de resposta (ao clicar um botão, por exemplo, há uma tendência de se mover a mão no espaço, o que pode impactar a precisão final da interação).

Conforme o método proposto na seção 3.1, cada um dos dispositivos que permitem interação com as mãos serão testados nos planos frontal e horizontal, o que determina quatro condições de teste: Controlador Leap Motion avaliado com tempo de permanência no plano

Tabela 10 – Nomenclatura e definições das condições de teste do Experimento 4.

Condição de teste	Descrição	Dispositivo para interação	Método de seleção	Plano de teste
MCF	<i>Mouse</i> avaliado com Clique no plano Frontal	<i>Mouse</i>	Clique com botão esquerdo	Frontal
MPF	<i>Mouse</i> avaliado com tempo de Permanência no plano Frontal	<i>Mouse</i>	Tempo de permanência (300 ms)	Frontal
LMPF	Controlador Leap Motion avaliado com tempo de Permanência no plano Frontal	Controlador Leap Motion	Tempo de permanência (300 ms)	Frontal
LMPH	Controlador Leap Motion avaliado com tempo de Permanência no plano Horizontal	Controlador Leap Motion	Tempo de permanência (300 ms)	Horizontal
MTPF	MetaTracker avaliado com tempo de Permanência no plano Frontal	MetaTracker	Tempo de permanência (300 ms)	Frontal
MTPH	MetaTracker avaliado com tempo de Permanência no plano Horizontal	MetaTracker	Tempo de permanência (300 ms)	Horizontal

Fonte: autoria própria.

frontal (LMPF), Controlador Leap Motion avaliado com tempo de permanência no plano horizontal (LMPH), MetaTracker avaliado com tempo de permanência no plano frontal (MTPF) e MetaTracker avaliado com tempo de permanência no plano horizontal (MTPH).

A avaliação de um *mouse* também é recomendada pelo método da seção 3.1 e deve ser realizada no plano frontal. Decidiu-se avaliar este dispositivo com dois métodos de seleção diferentes: o clique com o botão esquerdo e o tempo de permanência de 300 ms. O primeiro permitirá comparar os resultados deste experimento diretamente com o Experimento 3, que também avaliou o mesmo dispositivo no plano frontal. O segundo método permitirá tanto medir o impacto que a técnica de tempo de permanência provoca no desempenho dos participantes quanto atuar como ponte para permitir comparar as técnicas de interação com as mãos com outros estudos que aplicarem este mesmo método de avaliação. Define-se, assim, outras duas condições de teste: *mouse* avaliado com clique no plano frontal (MCF) e *mouse* avaliado com tempo de permanência no plano frontal (MPF).

Deste modo, este experimento adota um delineamento intraparticipantes com medidas repetidas e quatro variáveis independentes:

- Condições de teste: { MCF, MPF, LMPF, LMPH, MTPF, MTPH } (conforme Tabela 10)
- D : { 160, 230, 300 } mm
- W : { 12, 32 } mm
- Sessão: {1, 2}

Tabela 11 – Configuração de cada Sequência de Teste (ST) a ser avaliada no Experimento 4.

	D	W	ID
ST	(mm)	(mm)	($bits$)
S1	160	32	2,58
S2	230	32	3,03
S3	300	32	3,38
S4	160	12	3,84
S5	230	12	4,33
S6	300	12	4,70

Fonte: autoria própria.

Para determinar os valores escolhidos para as variáveis D e W , alguns testes preliminares foram realizados pelo autor. Valores de $D > 300mm$ começam a ficar difíceis demais, podendo exigir movimentação do corpo quando testando o plano horizontal, o que pode acabar interferindo nos resultados do teste. Por outro lado, valores de $W < 10mm$ dificultaram a seleção dos alvos devido à precisão geral do MetaTracker e do Controlador Leap Motion. Os valores finais foram escolhidos levando em consideração estas primeiras percepções e buscando manter os valores de ID s das sequências de testes equivalentes aos valores testados nos Experimentos 2 e 3. A Tabela 11 sumariza as configurações das Sequência de Teste (ST) avaliadas no experimento.

A variável sessão foi adicionada para coletar medidas repetidas para permitir uma melhor análise estatística dos dados. Um número maior de sessões era desejado, porém testes preliminares indicaram que a duração de uma única sessão seria de cerca de 50 minutos, causando grande cansaço físico e mental e desmotivando o participante com o passar do tempo. Por causa disto, somente duas sessões foram coletadas.

A ordem dos testes seguiu um delineamento de quadrado latino balanceado. Para isso, os participantes foram atribuídos aleatoriamente a um de seis possíveis grupos. A ordem de execução dos testes para cada um dos grupos é apresentada na Tabela 12. Como dois participantes não puderam participar de todos os testes, os grupos 1 e 4 ficaram com apenas 1 participante cada.

Cada ST foi executada com 11 alvos, resultando em 11 tentativas de seleção por sequência.

As variáveis dependentes foram MT , ER e TP .

Um total de 7920 tentativas de seleção (11 alvos * 3 distâncias entre alvos * 2 tamanhos de alvos * 2 sessões * 6 condições de teste * 10 participantes) foram coletadas neste experimento.

Tabela 12 – Grupos para balancear os possíveis efeitos de aprendizagem devido à ordem dos testes. As condições de teste foram distribuídas para criar um quadrado latino balanceado.

Grupo	Participantes	Ordem de teste					
		1°	2°	3°	4°	5°	6°
1	1	MCF	MPF	MTPH	LMPF	MTPF	LMPH
2	2	MPF	LMPF	MCF	LMPH	MTPH	MTPF
3	2	LMPF	LMPH	MPF	MTPF	MCF	MTPH
4	1	LMPH	MTPF	LMPF	MTPH	MPF	MCF
5	2	MTPF	MTPH	LMPH	MCF	LMPF	MPF
6	2	MTPH	MCF	MTPF	MPF	LMPH	LMPF

Fonte: autoria própria.

4.4.1.4 Procedimentos

Os testes foram executados com um participante por vez. Os participantes inicialmente foram apresentados aos objetivos do experimento e aos procedimentos gerais que seriam realizados. Eles puderam inspecionar um questionário de conforto e esclarecer quaisquer dúvidas que tivessem. Foi requisitado que utilizassem, durante os testes, seus óculos cotidianos para correção da visão.

Em seguida, motivado pelos resultados do Experimento 3, os participantes foram submetidos a uma etapa de calibração óptica do Meta 2. O próprio procedimento do Meta 2 foi utilizado com esse fim. Ao fim deste processo, os óculos estavam configurados para atender às características visuais de cada participante. A aplicação do 3DFitts foi inicializada em seguida.

Um dos maiores desafios ao se avaliar cientificamente a interação em um sistema de RA é garantir que o mapeamento espacial que ele realiza do mundo físico será similar para as diversas condições de testes e participantes. Isto é importante porque é este processo que determina como ancorar o mundo virtual ao mundo real, afetando a forma como os objetos virtuais são posicionados no mundo real e, portanto, o fundo visual que todos os participantes terão durante os testes. Diferentes fundos podem provocar diferentes respostas do sistema visual, afetando o desempenho final. O mapeamento espacial também pode afetar o desempenho geral do sistema computacional.

O Meta 2 realiza este mapeamento espacial sempre que é acionado. Assim, para reduzir erros devido a diferenças neste processo, os participantes foram requisitados a fixarem seu olhar em um ponto fixo em uma parede à sua frente (elemento D na Figura 16) antes de ativar os óculos.

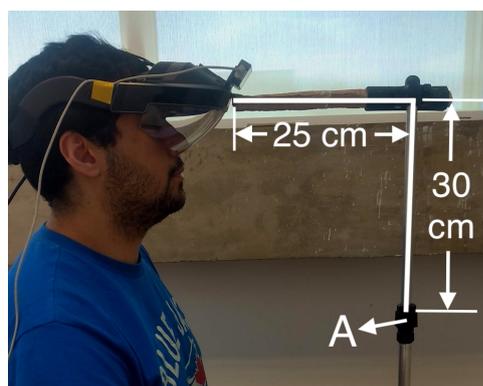
Finalizada esta etapa, um procedimento presente no 3DFitts foi utilizado para calibrar o Controlador Leap Motion, utilizando-o para rastrear a mão dominante do participante enquanto o óculos Meta 2 apresentava um modelo de uma mão virtual sobreposto à mão real do participante. Por meio de comandos dados pelo teclado, controlou-se a orientação espacial da mão virtual, de modo a melhor ajustá-la à mão do participante. Desta forma, garantiu-se um melhor posicionamento visual do cursor à mão do usuário.

Outro desafio envolvido neste tipo de estudo é que o desempenho da técnica de interação com as mãos depende não só do fundo visual, mas também da distância dos alvos ao participante e ao dispositivo de rastreamento. É necessário, portanto, uniformizar a localização espacial do plano de testes entre condições de teste e participantes. Para isso, decidiu-se posicionar o centro do plano de testes a uma distância fixa dos dispositivos de rastreamento das mãos. A Figura 17 ilustra o procedimento utilizado para calibrar esta posição: utilizando um aparato físico, foi possível determinar o ponto A no espaço, localizado a 25 cm de distância frontal e 30 cm abaixo do óculos de RA. O participante foi solicitado a colocar sua mão dominante sobre o ponto A e um procedimento de calibração do 3DFitts foi utilizado para posicionar o centro do plano de testes neste ponto. Este procedimento foi realizado antes de iniciar cada uma das condições de teste.

Após estas etapas de calibração dos sistemas, os participantes foram apresentados ao Controlador Leap Motion e ao MetaTracker. Para se familiarizarem com ambos os dispositivos e com os dois planos de orientação, foi solicitado que executassem uma sessão prática com as condições de teste MTPF e LMPH.

Os testes formais começaram após a sessão prática. As condições de testes foram avaliadas na ordem determinada pelo grupo a que o participante foi inicialmente atribuído (conforme Tabela 12). Antes de iniciar uma nova condição de teste, os participantes foram informados sobre qual dispositivo, método de seleção e orientação do plano eles iriam testar.

Figura 17 – Aparato utilizado para uniformizar a posição espacial dos alvos para as diversas condições de testes e participantes. O centro da circunferência dos alvos de teste foi posicionado no ponto A.



Fonte: Couto, José e Lopes (2020).

Cada participante completou a primeira sessão para todas as condições de teste previstas. Em seguida, repetiu todos os testes na segunda sessão. Para cada condição de teste, testou-se todas as seis ST apresentadas na Tabela 11. Do mesmo modo que nos Experimentos 2 e 3, as ST foram apresentadas sempre na mesma ordem: da mais fácil para a mais difícil (S1 a S6). Foi permitido pausas de curta duração entre cada sequência, mas durante a execução de uma sequência eles foram solicitados a manter o foco e completar a tarefa o mais rápido e precisamente possível.

Testes preliminares realizados pelo autor indicaram que o uso prolongado do Meta 2 poderia causar cansaço na testa e na região anterior da cabeça, provocando um desconforto crescente com o passar do tempo. Para reduzir este efeito, pausas mandatórias de dois minutos foram incluídas entre cada condição de teste. Uma pausa maior, de cinco minutos, foi obrigatória entre as duas sessões. O Meta 2 foi removido da cabeça dos participantes durante estes períodos e os participantes foram estimulados a estendê-los caso estivessem se sentindo cansados.

Ao fim de cada condição de testes, exceto para a condição MCF, os participantes responderam ao questionário para avaliação do conforto (Apêndice A) referente à técnica de interação testada. A condição MCF não foi avaliada por questionários porque o objetivo de testá-la foi apenas para permitir comparar este experimento com o Experimento 3. No total, cada participante respondeu a 10 questionários (5 condições de teste * 2 sessões).

Os participantes preencheram o segundo questionário consultando as respostas dadas na primeira sessão. Isto foi feito para garantir que as respostas na segunda sessão representassem mais fielmente a percepção subjetiva do participante, já que os primeiros questionários foram preenchidos antes que ele tivesse experimentado todas as condições de teste, e portanto ainda não tinha uma referência clara de qual técnica seria melhor em cada aspecto.

Todos os participantes eram destros e utilizaram sua mão direita para controlar as interações em todos os testes. Durante os testes envolvendo o Controlador Leap Motion e MetaTracker a mão esquerda dos participantes foi bloqueada da visão dos dispositivos de interação utilizando uma sacola de pano opaca. Isto foi feito pois testes preliminares permitiram identificar que, por vezes, estes sistemas de rastreamento das mãos poderiam deixar de rastrear a mão dominante e passar a rastrear somente a mão não-dominante caso fossem identificadas por acaso.

A duração média de todo o experimento foi de 108 minutos por participante.

4.4.2 Resultados e discussão

Um total de 91 tentativas de seleção (1,2%) apresentaram MT ou A_e que desviaram mais do que três desvios padrão da respectiva média (calculada considerando os pontos das duas sessões, para cada sequência, condição de teste e participante). Estas tentativas foram consideradas fora do padrão de normalidade e removidas das análises subsequentes. As tentativas

remanescentes foram utilizadas para calcular o MT , a ER e o TP de cada ST, resultando num conjunto de dados com 720 amostras (6 ST * 2 sessões * 6 condições de teste * 10 participantes), que serão analisadas a seguir.

4.4.2.1 Efeitos de aprendizagem

Este experimento só coletou dados para duas sessões, um número insuficiente para realizar uma análise de efeitos de aprendizagem capaz de apresentar algum significado relevante.

4.4.2.2 Efeitos devido ao grupo do participante

O Grupo 1 conclui as interações com o menor tempo médio ($MT = 1354$ ms), um desempenho 19% mais rápido que o Grupo 3, que foi o mais lento ($MT = 1672$ ms). Apesar disso, uma análise de variância considerando o grupo como fator entre participantes revelou que ele não teve um impacto estatisticamente significativo em MT ($F_{5,4} = 0,48, ns$).

O Grupo 5 apresentou o melhor desempenho geral com um $TP = 3,11$ bits/s, valor 21% maior que o grupo com pior desempenho (Grupo 3, com $TP = 2,57$ bits/s). Contudo, uma análise de variância também não revelou impacto estatisticamente significativo do fator grupo em TP ($F_{5,4} = 0,75, ns$).

Estes resultados sugerem que a ordem dos testes não impactou significativamente o resultado dos participantes.

4.4.2.3 Modelos preditivos pela Lei de Fitts

O desempenho geral dos participantes para cada uma das condições de teste estudadas foram modelados com a Lei de Fitts usando o mesmo procedimento dos Experimentos 2 e 3. Assim, a média geral de MT , A_e e W_e para cada condição foi computada primeiramente para cada participante e, em seguida, para todos os participantes. Os valores de A_e e W_e foram então utilizados para computar ID_e . A Tabela 13 apresenta todos os pares de pontos obtidos, sendo que cada um deles representa a resposta média de 20 tentativas de seleção (2 sessões * 10 participantes).

As regressões lineares obtidas com estes pares de dados estão representadas na Figura 18 e sugerem que o dispositivo de interação impactou o desempenho dos participantes, com o *mouse* apresentando melhor resposta e o MetaTracker, a pior. A orientação do plano de teste não parece ter tido um impacto relevante na resposta final do modelo preditivo.

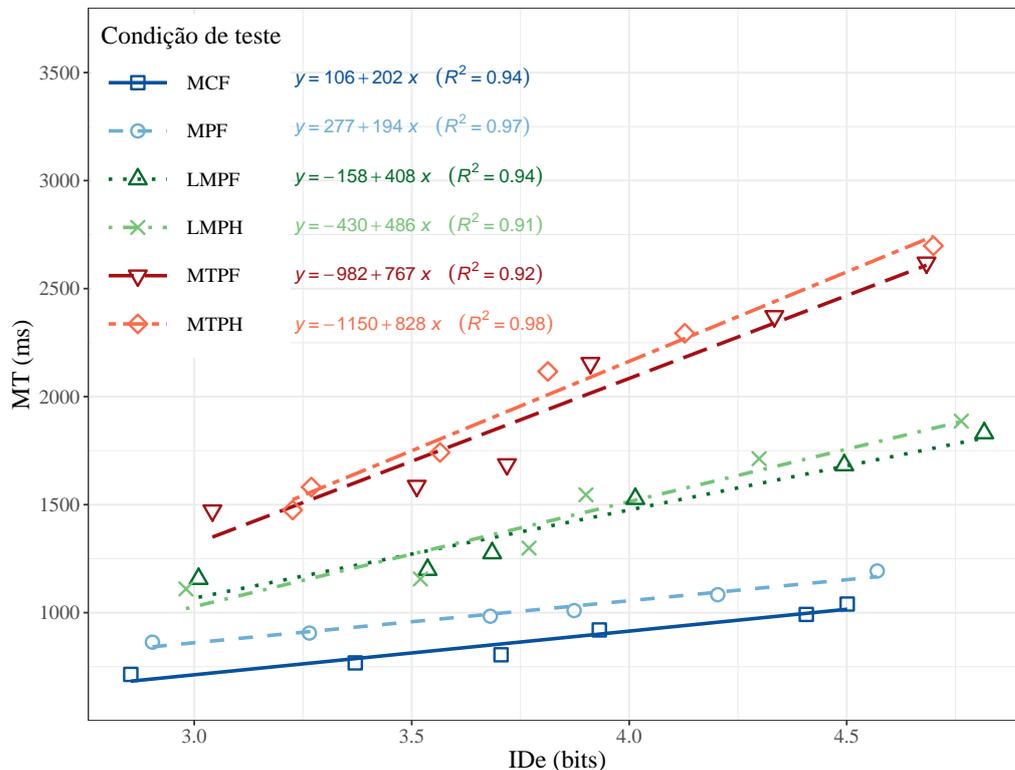
Os coeficientes de determinação obtidos para todas as seis regressões foi maior que

Tabela 13 – Pares de valores MT e ID_e utilizados para modelar cada uma das condições com teste do Experimento 4 com a Lei de Fitts. As condições de teste foram definidas na Tabela 10. Os valores de ST foram definidos na Tabela 11.

ST	Condição de teste											
	MCF		MPF		LMPF		LMPH		MTPF		MTPH	
	ID_e (bits)	MT (ms)	ID_e (bits)	MT (ms)	ID_e (bits)	MT (ms)	ID_e (bits)	MT (ms)	ID_e (bits)	MT (ms)	ID_e (bits)	MT (ms)
S1	2,85	714	2,90	864	3,05	1157	3,00	1110	3,01	1473	3,18	1475
S2	3,40	768	3,26	906	3,56	1199	3,55	1156	3,51	1587	3,26	1581
S3	3,73	805	3,70	984	3,71	1276	3,80	1299	3,72	1687	3,57	1742
S4	3,95	920	3,89	1010	4,04	1527	3,92	1546	3,93	2155	3,82	2116
S5	4,43	992	4,22	1083	4,51	1683	4,32	1713	4,35	2372	4,14	2293
S6	4,52	1040	4,59	1193	4,84	1831	4,78	1887	4,70	2620	4,71	2698

Fonte: adaptado de Couto, José e Lopes (2020).

Figura 18 – Regressões lineares obtidos com o modelo preditivo da Lei de Fitts para cada uma das condições de teste avaliadas.



Fonte: adaptado de Couto, José e Lopes (2020).

0,90, o que indica que o modelo foi capaz de explicar mais de 90% da variabilidade dos dados. Este resultado sugere, portanto, que a Lei de Fitts ainda é válida para modelar interação com as mãos em um ambiente de RA imersivo no caso em que os movimentos estejam restritos a um plano frontal ou horizontal.

Os coeficientes lineares observados para as condições de teste de interação com as mãos foram negativos. Valores negativos são comuns na literatura (GUIARD; OLAFSDOTTIR, 2011), mas o seu significado ainda é objeto de debate. Baseando-se em trabalhos de interação em 2D, Soukoreff e MacKenzie (2004) sugeriram que valores menores do que -200 ms ou maiores que 400 ms poderiam indicar falhas nos métodos do estudo ou na execução ou coleta de dados dos testes. Guiard e Olafsdottir (2011) rebateram este ponto de vista, defendendo que o coeficiente linear não carrega nenhum significado por si só e que seu valor simplesmente reflete as condições gerais do procedimento de regressão linear. De fato, valores negativos aparecerão sempre que uma técnica de interação apresentar uma taxa de degradação de desempenho elevada o suficiente para um determinado intervalo de *IDs* testados, aumentando a inclinação da reta e “empurrando” o coeficiente linear para valores cada vez mais baixos. No caso específico do MetaTracker, este efeito é evidente: a inclinação da reta é cerca de quatro vezes maior que aquela observada para o *mouse*, o que resultou num coeficiente linear abaixo de -1000 ms.

4.4.2.4 Variáveis dependentes

As médias gerais e desvios padrão para as variáveis dependentes foram computadas para cada condição de teste e estão tabuladas na Tabela 14 e representadas em gráfico de barras na Figura 19. As médias foram obtidas primeiramente para cada participante, considerando todas as 12 sequências testadas (6 ST * 2 sessões) por cada um, e depois a média geral foi calculada para todos os 10 participantes.

É impossível cometer erros quando o método de seleção por tempo de permanência é utilizado, uma vez que ele só dispara um evento de seleção caso o cursor seja mantido em colisão com o alvo pelo tempo determinado (no caso deste experimento, 300 ms). Portanto, somente a condição MCF apresentou uma taxa de erros maior que zero.

Efeito do método de seleção

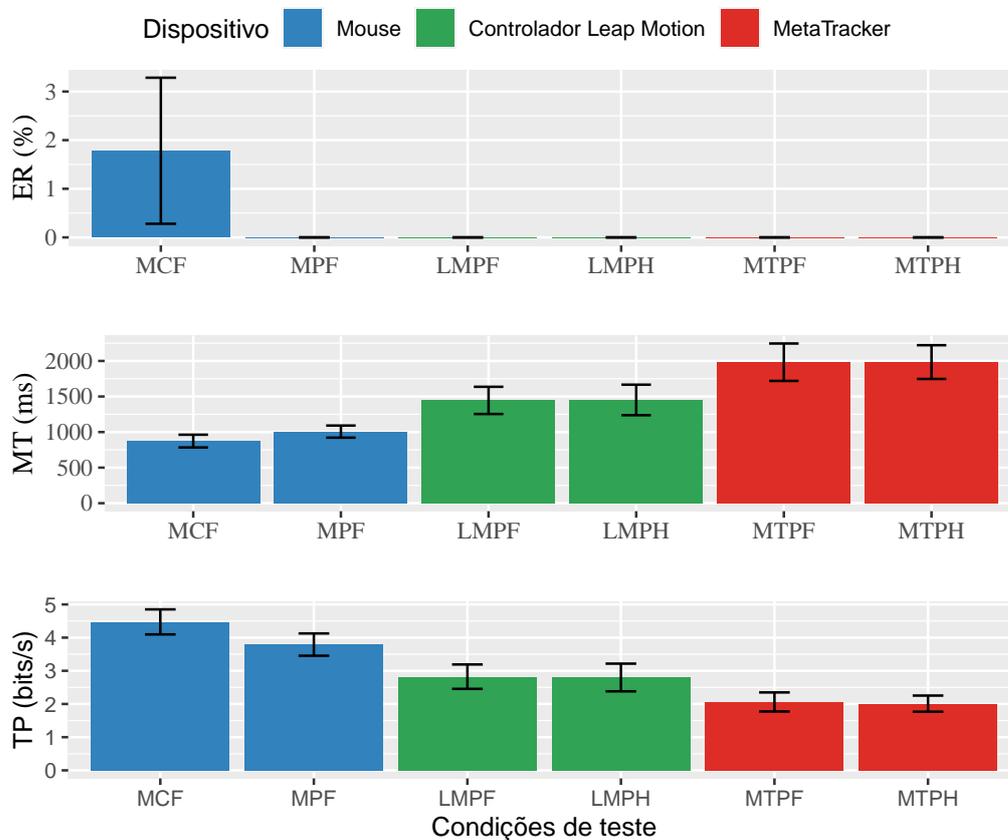
A comparação entre as condições MCF e MPF permite verificar o impacto que o método de seleção por tempo de permanência teve no desempenho dos participantes. Os valores da Tabela 14 sugerem este método reduziu o desempenho geral. De fato, a execução dos testes na condição MCF foi 13% mais rápida que na MPF, uma diferença estatisticamente significativa ($F_{1,9} = 24,6, p <, 001$). Análise da métrica de *TP* aponta para a mesma conclusão: ela foi 18% maior para a condição MCF e essa diferença também foi estatisticamente significativa ($F_{1,9} = 52,5, p <, 001$). Conclui-se, portanto, que a seleção por tempo de permanência afetou

Tabela 14 – Médias gerais para o *MT*, a *ER* e o *TP* de cada condição avaliada no Experimento 4. Condições com seleção por tempo de permanência sempre resultam em *ER* nula.

Condição de teste	<i>ER</i> (%)	<i>MT</i> (ms)	<i>TP</i> (bits/s)
MCF	1,8 (SD = 1,5)	873 (SD = 90)	4,47 (SD = 0,38)
MPF	0 (SD = 0)	1007 (SD = 85)	3,79 (SD = 0,34)
LMPF	0 (SD = 0)	1445 (SD = 191)	2,82 (SD = 0,37)
LMPH	0 (SD = 0)	1452 (SD = 215)	2,80 (SD = 0,42)
MTPF	0 (SD = 0)	1982 (SD = 263)	2,06 (SD = 0,29)
MTPH	0 (SD = 0)	1984 (SD = 237)	2,01 (SD = 0,24)

Fonte: autoria própria.

Figura 19 – Representação em barras das médias gerais para o *MT*, a *ER* e o *TP* de cada condição avaliada no Experimento 4. Os intervalos de confiança representam um desvio padrão. Condições com seleção por tempo de permanência sempre resultam em *ER* nula.



Fonte: adaptado de Couto, José e Lopes (2020).

negativamente o desempenho dos participantes.

Efeito do dispositivo de interação

Para estudar o efeito que o dispositivo de interação teve no desempenho final dos participantes, pode-se comparar as variáveis dependentes para as condições MPF, LMPF e MTPF. Elas usaram o mesmo método de seleção e plano de teste, variando apenas o dispositivo que controlou a posição do cursor.

O *MT* para a condição MPF foi 1007 ms, que foi 30% menor que o valor observado para LMPF (1445 ms) e 49% menor que o observado para MTPF (1982 ms). Uma análise de variância indicou que há ao menos uma diferença estatisticamente significativa entre estas médias ($F_{2,18} = 97,5, p < ,001$). Uma análise *post hoc* com um teste de Tukey confirmou que a diferença entre todos os três pares de condições foram estatisticamente significativas.

De modo semelhante, o *TP* para a condição MPF (3,79 *bits/s*) foi 34% maior que aquele da condição LMPF (2,82 *bits/s*) e 84% maior que a da condição MTPF (2,06 *bits/s*). Uma análise de variância indicou a presença de efeito estatisticamente significativo ($F_{2,18} = 96,9, p < ,001$) e o teste de Tukey confirmou que as diferenças entre todos os três pares são estatisticamente significativas.

Adicionalmente, é possível comparar as condições LMPH e MTPH para estudar o efeito do Controlador Leap Motion e do MetaTracker no desempenho dos participantes durante interação no plano horizontal. Também neste caso, os resultados indicam que a interação com o Controlador Leap Motion é melhor do que com o MetaTracker: o *MT* foi 27% menor para o Controlador Leap Motion, diferença estatisticamente significativa ($F_{1,9} = 112,7, p < ,001$), e o *TP* foi 39% maior, com diferença também estatisticamente significativa ($F_{1,9} = 60,68, p < ,001$).

Estes resultados indicam que interações com o Controlador Leap Motion podem ser realizadas com um desempenho superior àquelas com o MetaTracker. Ainda assim, interação utilizando o *mouse* se sobressaiu mesmo no ambiente de RA, o que sugere que as tecnologias de interação com as mãos estudadas ainda são mais lentas, imprecisas e suscetíveis a mais erros que o *mouse*. Estas limitações, se não solucionadas, pode impedir sua difusão como técnica de interação com as mãos no futuro.

Efeito da orientação espacial do plano

As condições LMPF, LMPH, MTPF e MTPH foram analisadas em conjunto para verificar o impacto que a orientação espacial do plano de testes teve no desempenho dos participantes. A média de *MT* para o plano frontal foi 1714 ms e para o plano horizontal, foi 1718 ms. A diferença entre estas médias não foi estatisticamente significativa ($F_{1,9} = 0,01, ns$). A média do *TP* para o plano frontal foi 2,44 *bits/s* e para o plano horizontal, foi 2,41 *bits/s*. A diferença observada também não foi estatisticamente significativa ($F_{1,9} = 0,53, ns$). Consequentemente, conclui-se que a orientação espacial do plano de testes não afetou significativamente o desempenho dos

participantes ao interagir com alvos no ambiente de RA.

4.4.2.5 Características de distribuição estatística

A mesma análise da distribuição estatística executada nos Experimentos 2 e 3 foi replicada para os dados deste experimento. Foram analisadas um total de 720 sequências (10 participantes * 2 sessões * 6 ST * 6 condições de teste), cada uma contendo entre 9 e 11 tentativas de seleção.

A média da assimetria e curtose para as distribuições do *MT* foram, respectivamente, 0,56 ($SD = 0,65$) e 2,84 ($SD = 1,15$). O teste de normalidade Shapiro-Wilk não rejeitou a hipótese nula de normalidade para 80% das sequências.

Já as distribuições para as seleções de coordenadas apresentaram uma média de assimetria igual a 0,07 ($SD = 0,56$) e uma média de curtose igual a 2,44 ($SD = 0,82$). A hipótese nula não foi rejeitada para 95% das sequências.

Todos estes resultados são semelhantes aos obtidos nos Experimentos 2 e 3 e sugerem que o comportamento dos participantes neste experimento ainda pode ser modelado por uma distribuição normal para a maioria das sequências coletadas.

4.4.2.6 Questionários de avaliação de conforto

Somente os questionários aplicados na segunda sessão serão analisados, uma vez que eles apresentam as respostas que melhor representam a percepção geral dos participantes sobre cada uma das técnicas de interação testadas. A Figura 20 apresenta os resultados em forma de pilhas de barras. Menores pontuações estão representadas com a cor vermelha e indicam uma percepção negativa sobre o aspecto analisado. Pontuações maiores, ao contrário, denotam um sentimento positivo e estão representadas em verde.

Uma análise visual da Figura 20 deixa claro que há uma tendência de queda nas pontuações quando se analisa o *mouse*, o Controlador Leap Motion e o MetaTracker, com o primeiro adquirindo um perfil verde-escuro, o segundo um perfil verde-amarelado e o terceiro apresentando um perfil mais amarelo-avermelhado.

Os dados coletados estão em forma categorizada ordinal e, portanto, é imprudente analisá-los interpretando-os como dados intervalares. Não se pode garantir, por exemplo, que a distância entre a pontuação 1 e 2 equivale a uma mesma medida de “sentimento de satisfação” que uma distância entre 6 e 7. Por este motivo, as pontuações foram analisadas com testes estatísticos não-paramétricos. Entretanto, é importante lembrar que testes não-paramétricos possuem um poder estatístico menor que os testes paramétricos.

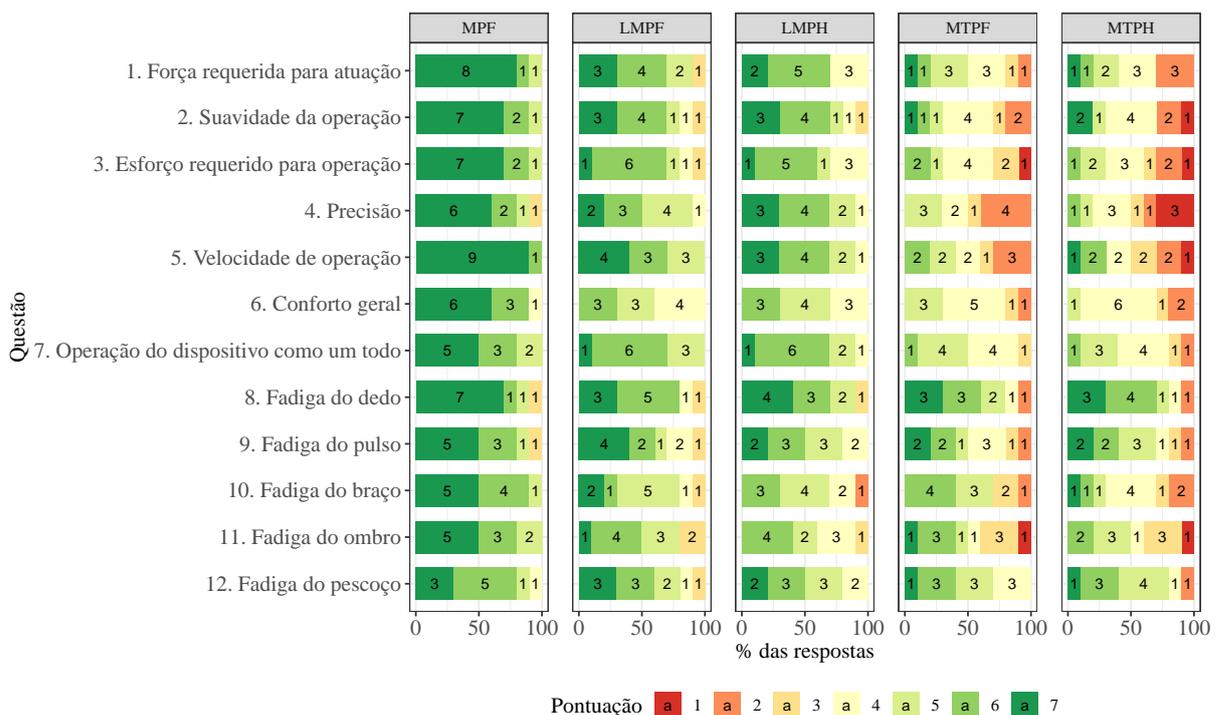
O teste de Friedman foi aplicado para cada uma das questões com o objetivo de avaliar

o impacto que a condição de teste teve nas pontuações dadas pelos participantes. Para todas as questões, com exceção da 12, houve efeito estatisticamente significativo ($p < 0,05$). Uma análise *post hoc* com o teste de Conover e correção do valor p com *bonferroni* revelou diferenças estatisticamente significativas conforme indicado na Tabela 15.

Para grande maioria dos aspectos analisados, as pontuações do *mouse* foram significativamente maiores do que para as condições envolvendo o MetaTracker, indicando uma percepção clara de que o *mouse* permite melhor qualidade de interação. Apesar do Controlador Leap Motion ter recebido pontuações consistentemente menores que as do *mouse*, o teste só identificou uma diferença significativa para o par (MPF, LMPH) com relação ao aspecto de fadiga no braço (questão 10). A mesma análise se aplica às comparações entre as condições envolvendo o Controlador Leap Motion e o MetaTracker: existe uma tendência de menor pontuação para o MetaTracker em todas as questões, mas somente diferenças entre as pontuações para os aspectos “precisão” (questão 4) e “operação do dispositivo como um todo” (questão 7) foram estatisticamente significativas.

Os resultados estatísticos permitem afirmar que o aspecto de fadiga, em especial a fadiga no braço e no ombro (questões 10 e 11), foi percebido como maior para os dispositivos de interação gestual do que para o *mouse*. Além disso, a precisão e a operação geral (questões 4 e 7)

Figura 20 – Respostas dos participantes para os 12 aspectos avaliados pelo questionário de avaliação de conforto da norma ISO 9241-411. Apenas as pontuações dadas durante a segunda sessão são apresentadas. Os questionários não foram aplicados para a condição MCF. Os números apresentados dentro das barras indicam quantos participantes deram a mesma pontuação para cada aspecto e condição de teste.



Fonte: adaptado de Couto, José e Lopes (2020).

do MetaTracker foi percebida como a pior dentre os três dispositivos. A velocidade de operação (questão 5) do MetaTracker também foi considerada como pior que a do *mouse*. Estes resultados estão alinhados com comentários informais feitos por alguns participantes durante os testes, em que reclamaram que a interação com o MetaTracker apresentava atraso e falhava com certa frequência. Estas impressões sugerem que o *hardware* do MetaTracker pode ser menos potente que o do Controlador Leap Motion, exigindo mais do processador do computador principal e, por consequência, causando maior arrasto de processamento.

O *TP* obtido para cada uma das condições de teste é uma métrica objetiva capaz de medir o desempenho quando se considera tanto a velocidade de operação quanto a precisão de interação. Assim, considerando os resultados da subseção 4.4.2.4, que indicam uma diferença significativa entre os valores de *TP* para cada um dos três dispositivos de interação, pode-se afirmar que os resultados estatísticos dos questionários suportam os resultados quantitativos (se não de modo forte, ao menos com uma evidência de força moderada).

Com relação ao plano de orientação, a análise estatística resultou em $p = 1,00$ tanto para as diferenças de pontuação dadas ao par (LMPF, LMPH) quanto ao par (MTPF, MTPH), fornecendo uma evidência forte de que a percepção dos participantes sobre a qualidade da interação não foi alterada pelo plano e corroborando os resultados obtidos a partir das métricas objetivas na subseção 4.4.2.4.

4.4.3 Conclusão (Experimento 4)

Este experimento demonstrou como o método proposto na seção 3.1 e o 3DFitts podem ser utilizados como ferramenta para avaliar interação direta com as mãos em um ambiente de RA imersiva experienciado com um OST-HMD.

Tabela 15 – Questões em que foram observados níveis de significância com um teste de Friedman, por par de condições de teste. O hífen entre números denota intervalo – por exemplo, 1-5 inclui questões de 1 a 5. Um hífen sozinho foi usado para evitar espelhar o conteúdo já apresentado em outras células. A diferença de pontuação entre alguns pares não apresentaram efeito estatisticamente significativo para nenhuma questão e foram tabulados com *ns* (não significativo).

	MPF	LMPF	LMPH	MTPF
MTPH	1-7, 10-11	7	4, 7	<i>ns</i>
MTPF	1-11	<i>ns</i>	4	-
LMPH	10	<i>ns</i>	-	-
LMPF	<i>ns</i>	-	-	-

Fonte: autoria própria.

O objetivo principal foi verificar se as métricas quantitativas computadas pelo 3DFitts são capazes de distinguir diferenças em desempenho que também fossem observadas por meio de uma avaliação subjetiva dos participantes. Para isso, uma série de testes foi realizada para avaliar três dispositivos de interação (o *mouse*, o Controlador Leap Motion e o MetaTracker) em interações com alvos dispostos em planos com duas orientações espaciais (frontal e horizontal). Para mensurar a percepção subjetiva dos participantes, eles responderam ao questionário de avaliação do conforto recomendado pela ISO 9241-411 para cada condição testada. O *mouse* foi avaliado para permitir comparações dos resultados deste estudo com a literatura de interação em 2D e com futuros trabalhos que apliquem este mesmo método para avaliar outras técnicas de interação em 3D.

Os resultados obtidos nas seções 4.4.2.4 e 4.4.2.6 forneceram evidências para afirmar que o programa desenvolvido foi capaz de discriminar qual método de interação foi percebido subjetivamente como pior, mediano e melhor. Este resultado sugere que o *TP* é uma métrica objetiva com potencial de mensurar aspectos ergonômicos que são relevantes para provocar mudanças na percepção de conforto dos participantes.

Uma segunda questão guiou este experimento: a orientação espacial do plano de testes afeta o desempenho dos participantes ao interagirem diretamente com as mãos em um ambiente tridimensional? O autor esperava observar algum impacto, uma vez que a qualidade de rastreamento das mãos dos dispositivos testados tende a cair à medida que a mão se afasta do sensor, o que poderia prejudicar interações no plano horizontal. Entretanto, tanto os resultados quantitativos quanto os subjetivos indicaram que não houve diferença entre os planos. Esta conformidade das respostas objetivas com a percepção dos participantes é outra evidência que indica que os resultados retornados pelo 3DFitts efetivamente avaliam aspectos em conformidade com as percepções humanas.

Os dados deste experimento apresentaram alta correlação com o modelo da Lei de Fitts ($R^2 > 0.9$), indicando que ela é um instrumento válido para se modelar a interação direta com as mãos em um óculos de RA, ao menos para interações de apontar-e-selecionar restritas aos planos frontal e horizontal.

Por fim, com este experimento foi possível verificar as hipóteses que deram origem a este trabalho, conforme enunciadas na seção 1.7. A hipótese *H1* afirma que o Controlador Leap Motion apresenta um desempenho objetivamente melhor que o MetaTracker. Os resultados da subseção 4.4.2.4 suportam fortemente a hipótese *H1*. A hipótese *H2*, por outro lado, afirma que o Controlador Leap Motion é percebido subjetivamente como sendo melhor para interações no Meta 2 do que o MetaTracker. Uma inspeção visual das pontuações dadas pelos participantes a cada um destes dispositivos, conforme Figura 20, indica uma tendência clara a se pontuar o Controlador Leap Motion com notas maiores do que o MetaTracker. Contudo, as análises estatísticas apresentadas na subseção 4.4.2.6 não tiveram poder suficiente para denunciar uma diferença estatisticamente significativa para a maioria dos aspectos subjetivos avaliados. Portanto, pode-se afirmar que os resultados obtidos suportam moderadamente a hipótese *H2*.

4.5 Discussão geral

As respostas das aplicações 3DFitts e GoFitts foram analisadas no Experimento 1 e os resultados indicaram que ambas reportaram ER , MT e TP equivalentes estatisticamente quando utilizadas em um experimento uniformizado e sem fatores humanos. O Experimento 2 apresentou novas evidências desta equivalência de resultados, porém desta vez utilizando ambas aplicações com participantes humanos em um teste de avaliação de interação com um *mouse* em um monitor 2D. Estes resultados iniciais encorajaram o prosseguimento dos testes para o ambiente de RA em mais dois experimentos.

O Experimento 3 deu um passo nesta direção ao tentar manter todas as condições de teste semelhantes às do Experimento 2, porém adicionando um óculos de RA como dispositivo de visualização. Já o Experimento 4 buscou avaliar interações diretas com as mãos no óculos de RA, utilizando para isso dois dispositivos de rastreamento das mãos.

A Figura 21 sumariza todos as condições de teste e os respectivos resultados em termos de TP para todos os experimentos envolvendo seres humanos (Experimentos 2, 3 e 4). Ela deixa evidente que cada nova condição de teste (representadas pelos identificadores A a E) diferem da anterior por uma única variável independente. Por causa disto, é possível inferir quanto cada variável afetou o desempenho geral dos participantes. Entretanto, deve-se ter cautela com possíveis conclusões, uma vez que cada experimento empregou métodos distintos, o que pode ter afetado de algum modo as respostas.

As condições identificadas por A, B e C apresentaram resultados de TP similares, indicando que as variáveis *software* e modo de visualização não tiveram efeito significativo no desempenho dos participantes.

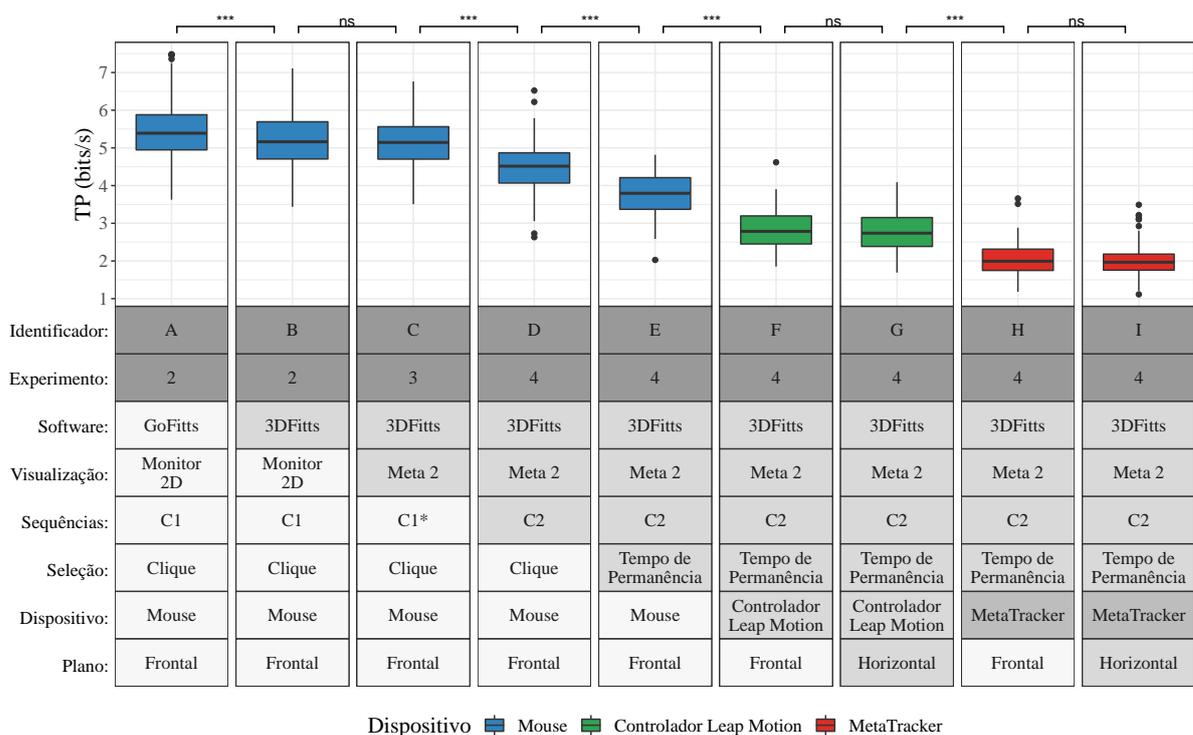
Os dados coletados para a condição MCF no Experimento 4 (teste identificado por D) podem ser diretamente comparados com os dados do teste C para estudar o impacto que a mudança nos valores de D e W teve no desempenho dos participantes. Para isso, removeu-se os dados dos participantes do Experimento 3 que não conformaram com a Lei de Fitts e os dois participantes que não concluíram os testes do Experimento 4. Deste modo, somente dados de 8 participantes foram analisados. O valor médio de TP obtido para a condição D foi 4,40 *bits/s*, um valor 14% menor que o observado para a condição C (5,14 *bits/s*). Esta diferença foi estatisticamente significativa ($F_{1,7} = 58,99, p < ,001$). O tamanho de efeito foi grande ($d = 1.32$). Pode-se concluir, portanto, que as diferentes condições de teste entre o Experimento 3 e 4 tiveram um impacto relevante no desempenho dos participantes.

Estes efeitos provavelmente não foram causados apenas pela mudança nas configurações das sequências e podem ter sido influenciados por outros fatores, como por exemplo pela remoção do monitor 2D desligado, que foi utilizado somente no Experimento 3 e que permitiu visualizar os objetos virtuais com maior contraste do que no modo do Experimento 4. Um outro fator que pode ter prejudicado o desempenho na condição D é o delineamento do Experimento

4, que exigiu a execução de um número maior de testes e com interações mais cansativas. Estes possíveis efeitos são tópicos interessantes para serem avaliados em estudos futuros.

Os testes D a I são todos do Experimento 4 e foram analisados na subseção 4.4.2.4. As mesmas conclusões gerais obtidas nessa seção podem ser extraídas da Figura 21: o impacto significativo que o tempo de permanência teve com relação à seleção por clique pode ser visto comparando as condições D e E; o impacto significativo dos diferentes dispositivos de interação (*mouse*, Controlador Leap Motion e MetaTracker) pode ser inferido comparando as condições E, F e H; e a ausência de impacto relevante devido à mudança na orientação do plano de teste pode ser observada comparando-se as condições F com G e H com I.

Figura 21 – Valores de TP para cada condição de teste dos Experimentos 2, 3 e 4. Os valores das variáveis independentes são apresentados em cores diferentes para evidenciar mudanças entre condições de teste consecutivas. Por exemplo, somente o *software* mudou da condição A para B. Os níveis de significância observados entre as condições consecutivas são apresentados no topo da figura. Os valores para a variável “seqüências” se referem aos pares (D, W) testados, sendo que C1 se refere aos valores utilizados no Experimento 2 e 3 e C2, aos utilizados no Experimento 4. Os valores apresentados para o Experimento 3 não incluem os dois participantes cujos movimentos não conformaram com a Lei de Fitts. (*) Apesar do Experimento 3 utilizar valores (D, W) equivalentes aos do Experimento 2, eles foram transformados de *pixels* para milímetros, medida utilizada pelo óculos de RA.



Fonte: adaptado de Couto, José e Lopes (2020).

5 CONCLUSÃO

A velocidade de surgimento de novas tecnologias para RA nesta última década sugere que este sistema computacional pode se tornar um novo meio para produção e consumo de informações digitais. Assim como aconteceu com os monitores 2D e os computadores de mesa, que revolucionaram a forma como a sociedade se organiza e produz conhecimento, a RA também tem um potencial de impactar profundamente as estruturas e atividades sociais. Um dos fatores limitantes para que esse potencial se realize é a inexistência de técnicas de interação que permitam, nos ambientes de RA imersivos, interações humano-computador tão eficazes, eficientes e robustas como aquelas que o *mouse* e as telas de toque proporcionam em ambientes bidimensionais.

A interação natural por gestos é uma das técnicas de interação estudadas para possibilitar comunicação humano-computador nos ambientes imersivos de RA. Nela, o usuário utiliza as próprias mãos para comunicar ao sistema computacional os comandos que deseja executar. Não obstante, as tecnologias existentes para promover este tipo de interação ainda possuem muitas limitações que afetam o desempenho final do usuário.

Uma das etapas do processo de desenvolvimento e aprimoramento de novas técnicas de interação é sua avaliação formal. A Lei de Fitts tem sido amplamente utilizada na literatura de IHC desde a década de 1970 para mensurar objetivamente o desempenho de interação permitido por uma determinada técnica de interação. A norma ISO 9241-9 trouxe, a partir do ano 2000, uma padronização de métodos que consolidou o uso da Lei de Fitts como instrumento de avaliação. Contudo, tal norma se restringe a interação em 2D, enquanto que os ambientes de RA requerem interação em 3D.

Este trabalho estudou os métodos que a literatura tem utilizado para avaliar interação em 3D e buscou contribuir com esta área de pesquisa ao propor um método de avaliação para mensurar objetivamente o desempenho de técnicas de interação direta com as mãos em óculos de RA. A proposta apresenta um pequeno incremento em métodos já utilizados na literatura, partindo de uma adaptação para 3D dos testes de apontar-e-selecionar propostos na norma ISO 9241-411.

Para verificar a capacidade do método proposto de atuar como uma ferramenta de avaliação, o autor se propôs a realizar um estudo com usuários para avaliar duas técnicas de interação direta com as mãos em um óculos de RA. Para isso, foi necessário desenvolver um *software* que permitisse executar as mecânicas do método de avaliação no ambiente de RA. Ele foi denominado 3DFitts e foi implementado usando como referência um programa denominado GoFitts (MACKENZIE, 2020), comumente utilizado na literatura para avaliar interação em 2D.

Como o 3DFitts foi implementado como um novo algoritmo, em uma nova plataforma e linguagem de programação, ele poderia conter erros ou diferenças de comportamento que

comprometeria os resultados informados. Por isso, decidiu-se executar dois experimentos para comparar seus resultados com os do GoFitts, quando ambos são submetidos às mesmas condições de teste. O primeiro experimento utilizou um programa de auto-clique para reproduzir os mesmos eventos de interação em ambas aplicações e, portanto, não envolveu fatores humanos. No segundo experimento, avaliou-se a interação de participantes usando um mesmo *mouse* em ambos os programas. Os resultados indicaram que o 3DFitts retorna resultados estatisticamente equivalentes àqueles reportados pelo GoFitts.

Um terceiro experimento foi conduzido para avaliar o impacto que a adição dos óculos de RA pode provocar no desempenho dos participantes. Os resultados deste experimento, quando comparados com uma condição de teste equivalente no segundo experimento, sugeriram que não houve impacto estatisticamente significativo.

Por fim, o quarto e último experimento foi delineado para permitir avaliar duas técnicas de interação em um óculos de RA. Os participantes foram submetidos ao método de avaliação proposto pelo autor e, utilizando o 3DFitts, avaliaram seis modos de interação que incluíram: três dispositivos de interação, dois métodos de seleção e duas orientações espaciais de planos no espaço 3D. Além das métricas objetivas retornadas pelo 3DFitts, aplicou-se questionários para avaliar o grau de conforto percebido pelos participantes para cada condição de teste. A análise dos resultados indicaram que as métricas objetivas calculadas pelo 3DFitts corroboraram os resultados obtidos por meio dos questionários. Os testes realizados no 3DFitts apresentaram, portanto, potencial para indicar qual a técnica de interação será mais confortável para os participantes realizarem as tarefas de apontar-e-selecionar no espaço 3D.

O método proposto para avaliar a interação em 3D e o *software* 3DFitts podem auxiliar *designers* e engenheiros a desenharem melhores interfaces de usuário e dispositivos de interação. Como o método propõe avaliar ao menos duas orientações espaciais de planos (frontal e horizontal), ele permite obter a resposta espacial de um determinado sistema de interação e pode, desta forma, auxiliar um *designer* de interfaces a entender em qual volume espacial os usuários terão melhor desempenho interativo. Com esta informação, o profissional poderá tomar decisões sobre posicionamento de botões e outros elementos interativos no campo de visão do usuário, além de criar estratégias que auxiliem os usuários a terem uma melhor experiência interativa em determinada aplicação. Do mesmo modo, engenheiros podem se beneficiar ao delinear e executar testes padronizados que permitam identificar se diferentes propostas de *hardware* e *software* do sistema de interação provocam ou não melhorias efetivas na capacidade de interação dos usuários.

5.1 Retomando as perguntas norteadoras

Este trabalho foi desenvolvido tendo como base três perguntas norteadoras, que serão revisitadas agora.

Pergunta 1. Quais métodos tem sido utilizados para avaliar o desempenho de técnicas de interação em 2D e em 3D?

A partir do levantamento bibliográfico realizado, observou-se que a Lei de Fitts tem sido frequentemente utilizada no âmbito de avaliações objetivas do desempenho de IHC. Inicialmente aplicada em interações em uma dimensão, ela foi posteriormente aplicada com sucesso para avaliar também interações bidimensionais e seu uso em 3D vem sendo investigado por vários autores. Com relação aos métodos de avaliação, em geral, os pesquisadores aplicam aqueles propostos pela norma ISO 9241-411 (originalmente estabelecido na norma ISO 9241-9). Também foi possível encontrar trabalhos utilizando métodos customizados, porém estes métodos não costumam ser replicados por outros pesquisadores.

Pergunta 2. Quais métodos tem sido utilizados para avaliar o desempenho de técnicas de interação direta com as mãos em ambientes tridimensionais imersivos?

Para interação em ambientes tridimensionais imersivos – aqueles nos quais o mundo virtual é experienciado visualmente com estereoscopia – foi possível identificar tanto propostas de novos métodos de avaliação quanto adaptações dos métodos propostos na ISO 9241-411. A partir do levantamento bibliográfico realizado observou-se que certas adaptações da ISO 9241-411 começaram a ser aceitas e aplicadas por diferentes pesquisadores em estudos de diferentes técnicas de interação. Estas adaptações foram incorporadas ao método proposto neste trabalho e incrementadas com a orientação de se testar a interação em pelo menos dois planos: o plano frontal e o horizontal.

Pergunta 3. Como avaliar o desempenho da interação direta com as mãos em um óculos de RA?

O levantamento bibliográfico realizado encontrou apenas um trabalho que buscou avaliar desempenho de interação direta com as mãos em óculos de RA. Neste trabalho, Özacar et al. (2017) adotaram um método próprio para avaliar a interação, solicitando que os participantes selecionassem esferas dispostas aleatoriamente em um espaço cúbico à sua frente e medindo o tempo médio de movimento até a seleção.

Também foi possível identificar trabalhos que aplicaram adaptações da ISO 9241-411 para avaliar interação em óculos de RV, como por exemplo os trabalhos de Lubos, Bruder e Steinicke (2014) e Machuca e Stuerzlinger (2019). Clark, Bhagat e Riggs (2020), por outro lado, propuseram um método de avaliação que se distancia do padrão ISO 9241-411 para avaliar interação em 3D com um óculos de RV.

Todos estes trabalhos foram estudados para embasar as escolhas do método de avaliação proposto nesta dissertação, priorizando-se aquelas decisões que foram incorporadas em um maior número de trabalhos de diferentes pesquisadores.

5.2 Principais contribuições

Este trabalho de mestrado apresenta diferentes contribuições à comunidade científica, tais como:

- a) propor um método formalizado para avaliar aspectos ergonômicos da interação direta com as mãos em ambiente de RA;
- b) desenvolver e disponibilizar¹ um *software* para conduzir avaliações de técnicas de interação em ambientes de RA;
- c) conduzir e reportar experimentos que indicaram que os resultados computados pelo 3DFitts são equivalentes aos obtidos por um *software* comumente utilizado na literatura;
- d) conduzir e reportar um experimento para avaliar formalmente dois modos de interação com as mãos – utilizando o Controlador Leap Motion e o MetaTracker – em um óculos de RA, demonstrando como o método proposto e o 3DFitts são capazes de discriminar objetivamente diferenças de desempenho que também são percebidas subjetivamente por seres humanos;
- e) avaliar também um *mouse* em todos os estudos com usuários, permitindo comparar os resultados deste trabalho com os da literatura de interação em 2D e com possíveis futuros trabalhos que apliquem métodos semelhantes para avaliar interações em 3D;
- f) escrever e submeter um manuscrito de artigo original a um periódico de alcance internacional, relatando as principais contribuições científicas advindas deste trabalho.

5.3 Limitações

O método proposto para avaliar interações em 3D com as mãos possui algumas desvantagens que podem dificultar sua aplicação. A principal diz respeito à duração dos testes: executar todos os passos propostos pode tomar muito tempo, consumindo recursos humanos e financeiros e inclusive afetando a qualidade da avaliação uma vez que a fadiga cognitiva e física dos participantes afetará seu desempenho à medida que a duração dos testes em uma sessão aumenta. Neste sentido, dois fatores principais diferenciam o método proposto dos demais comumente realizados na literatura e contribuem para aumentar o tempo dos testes: a exigência de se testar ao menos dois planos de interação (frontal e horizontal) e a sugestão de se aplicar questionários de avaliação do conforto a cada condição de teste executada. O

¹ O código do 3DFitts é aberto e pode ser encontrado em <<https://github.com/jlcouto/3DFitts>>.

Experimento 4 descrito neste trabalho é um exemplo de como avaliar muitas condições de testes ao mesmo tempo – foram seis no total – pode incrementar rapidamente o tempo necessário para coletar todos os dados. Para mitigar este problema, sugere-se que o pesquisador evite realizar experimentos com mais do que quatro condições de teste e que, caso seja necessário, realize apenas uma sessão de coleta de dados por dia, garantindo que os participantes começarão cada sessão sempre descansados. Também é possível aplicar os questionários de conforto apenas na última sessão dos testes.

Pode-se citar o baixo número de participantes como uma possível limitação dos experimentos conduzidos neste trabalho. Os dados coletados com os dez participantes do Experimento 4 foram suficientes para encontrar diferenças estatisticamente significativas entre as métricas objetivas das diferentes técnicas de interação estudadas, mas não foram suficientes para apontar diferenças estatisticamente significativas entre os dados subjetivos coletados por meio dos questionários de conforto. Um maior número de participantes teria aumentado o poder do teste estatístico não-paramétrico, possivelmente revelando diferenças estatisticamente significativas também entre os dados subjetivos coletados para cada técnica de interação.

Ainda com relação aos participantes dos experimentos deste trabalho, pode-se destacar que constituem um grupo com características homogêneas: adultos, alta escolaridade, elevado conhecimento técnico e usuários frequentes de tecnologias computacionais. Os resultados encontrados neste trabalho não são, portanto, generalizáveis para a população em geral. Um grupo mais heterogêneo de participantes poderia revelar, por exemplo, que crianças teriam maior facilidade em interagir com as mãos do que com um *mouse*, visto que crianças atualmente crescem com maior contato com telas de toque (interação direta) do que com controles indiretos como o *mouse*.

Por fim, toda a discussão deste trabalho se limitou a abordar a avaliação de tarefas simples de interação, como o apontamento e seleção de objetos virtuais 3D. Deve-se ressaltar, porém, que a interatividade em ambientes 3D deve incluir muitas outras possibilidades de interação, tais como navegar pelo ambiente virtual, manipular a posição e a orientação espacial dos objetos virtuais e aumentar ou reduzir sua escala. Estas formas de interação são mais complexas e faz-se necessário desenvolver métodos para avaliá-las.

5.4 Trabalhos futuros

Trabalhos futuros podem utilizar o 3DFitts para testar outras técnicas de interação direta com as mãos, contribuindo para identificar quais apresentam melhores desempenhos em um ambiente de RA imersivo. O *software* também pode ser utilizado para estudar como mudanças nos algoritmos de interação ou na interface do usuário podem afetar o desempenho final do usuário. Com adaptações simples, também é possível avaliar outros dispositivos de interação em 3D, como controles rastreados por sistemas eletromagnéticos, rastreadores de

pupila e movimentos da cabeça. Como o 3DFitts foi desenvolvido na plataforma Unity, é fácil adaptá-lo para funcionar com dispositivos de realidade virtual (RV). Portanto, é possível utilizá-lo para estudar desempenho de interação em 3D também num ambiente de RV.

O presente trabalho focou em avaliar apenas dois primitivos de interação no ambiente 3D: apontamento e seleção. Uma extensão natural deste trabalho envolve avaliar tarefas compostas - ou seja, tarefas que envolvem o encadeamento das interações primitivas - e que permitam manipular objetos no ambiente 3D. Um exemplo seriam as tarefas de mover objetos virtuais de um ponto a outro do espaço, manipulando tanto sua posição quanto sua orientação espacial. Uma primeira opção de trabalho futuro neste caminho seria estender o 3DFitts para realizar também os testes de arrastar objetos previstos na norma ISO 9241-411.

O programa 3DFitts foi desenvolvido com o intuito de avaliar o impacto que diferentes técnicas de interação podem ter na qualidade da interação entre seres humanos e o computador em ambientes de RA. Porém, ele também pode ser utilizado para avaliar o impacto que outros fatores relacionados ao sistema computacional pode causar na qualidade da interação. Por exemplo, é possível utilizar o *software* em desenhos de experimentos que busquem avaliar como diferentes óculos de RA podem afetar a qualidade de interação, mantendo todo o resto igual. Um experimento interessante nesta linha seria estudar como diferentes sistemas de registro espacial de objetos virtuais em 3D afetam a qualidade da interação, ou seja, tentar mensurar o quanto a instabilidade no posicionamento espacial dos objetos virtuais pode afetar a capacidade de interação do participante.

Durante todo este trabalho, discutiu-se técnicas de interação com as mãos considerando que apenas uma mão atua por vez. Esta é uma limitação impensável para um ambiente de RA imersivo. Assim, estudos que proponham e apliquem métodos de avaliação de interação usando as duas mãos *simultaneamente* podem contribuir para avançar no desenvolvimento de novas técnicas de interação em RA, ajudando a humanidade a dar mais um passo em direção ao dispositivo de exibição perfeito imaginado há mais de 50 anos por Sutherland (1965).

REFERÊNCIAS

Al-Shamayleh, A. S. et al. A systematic literature review on vision based gesture recognition techniques. *Multimedia Tools and Applications*, v. 77, n. 21, p. 28121–28184, nov. 2018. ISSN 1573-7721. Citado na página 22.

Atsuo Murata; Hirokazu Iwase. Extending Fitts' law to a three-dimensional pointing task. *Human Movement Science*, v. 20, n. 6, p. 791–805, dez. 2001. ISSN 01679457. Citado 2 vezes nas páginas 36 e 37.

AZUMA, R. T. A Survey of Augmented Reality. *Presence: Teleoper. Virtual Environ.*, v. 6, n. 4, p. 355–385, ago. 1997. ISSN 1054-7460. Citado na página 19.

BACHMANN, D.; WEICHERT, F.; RINKENAUER, G. Evaluation of the leap motion controller as a new contact-free pointing device. *Sensors (Basel, Switzerland)*, v. 15, n. 1, p. 214–233, dez. 2014. ISSN 1424-8220. Citado na página 38.

BACHMANN, D.; WEICHERT, F.; RINKENAUER, G. Review of Three-Dimensional Human-Computer Interaction with Focus on the Leap Motion Controller. *Sensors (Basel, Switzerland)*, v. 18, n. 7, jul. 2018. ISSN 1424-8220. Citado na página 24.

BÉRARD, F. et al. Did “Minority Report” Get It Wrong? Superiority of the Mouse over 3D Input Devices in a 3D Placement Task. In: GROSS, T. et al. (Ed.). *Human-Computer Interaction – INTERACT 2009*. [S.l.]: Springer Berlin Heidelberg, 2009. (Lecture Notes in Computer Science), p. 400–414. ISBN 978-3-642-03658-3. Citado 4 vezes nas páginas 39, 40, 42 e 43.

BILLINGHURST, M.; CLARK, A.; LEE, G. A Survey of Augmented Reality. *Found. Trends Hum.-Comput. Interact.*, v. 8, n. 2-3, p. 73–272, mar. 2015. ISSN 1551-3955. Citado 2 vezes nas páginas 19 e 21.

BORG, G. *Borg's Perceived Exertion and Pain Scales*. Champaign, IL, US: Human Kinetics, 1998. viii, 104 p. (Borg's Perceived Exertion and Pain Scales). ISBN 978-0-88011-623-7. Citado na página 45.

BOWMAN, D. A. et al. Chapter 11 - Evaluation of 3D User Interfaces. In: *3D User Interfaces: Theory and Practice*. Boston: Addison-Wesley, 2005. ISBN 978-0-201-75867-2. Citado 2 vezes nas páginas 24 e 37.

BROOKE, J. SUS: A “quick and dirty” usability. *Usability evaluation in industry*, CRC press, p. 189, 1996. Citado na página 45.

BRUDER, G.; STEINICKE, F.; STÜRZLINGER, W. Effects of visual conflicts on 3D selection task performance in stereoscopic display environments. In: *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. [S.l.: s.n.], 2013. p. 115–118. Citado 6 vezes nas páginas 38, 39, 40, 41, 42 e 47.

BURNO, R. A. et al. Applying Fitts' Law to Gesture Based Computer Interactions. *Procedia Manufacturing*, v. 3, p. 4342–4349, 2015. ISSN 23519789. Citado 2 vezes nas páginas 22 e 38.

CARD, S. K.; ENGLISH, W. K.; BURR, B. J. Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT. *Ergonomics*, Taylor & Francis, v. 21, n. 8, p. 601–613, ago. 1978. ISSN 0014-0139. Citado na página 23.

- CARD, S. K.; MORAN, T. P.; NEWELL, A. *The Psychology of Human-Computer Interaction*. Hillsdale, N.J.: L. Erlbaum Associates, 1983. ISBN 978-0-89859-243-6. Citado na página 20.
- CHA, Y.; MYUNG, R. Extended Fitts' law for 3D pointing tasks using 3D target arrangements. *International Journal of Industrial Ergonomics*, v. 43, n. 4, p. 350–355, jul. 2013. ISSN 0169-8141. Citado 2 vezes nas páginas 36 e 37.
- CLARK, L. D.; BHAGAT, A. B.; RIGGS, S. L. Extending Fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies*, v. 139, p. 102413, jul. 2020. ISSN 10715819. Citado 4 vezes nas páginas 36, 42, 52 e 103.
- COUTO, J. L.; JOSÉ, M. A.; LOPES, R. d. D. A method for evaluating 3D mid-air hand interaction in immersive Augmented Reality. Manuscrito Submetido Para Análise Ao International Journal of Human-Computer Studies. 2020. Citado 12 vezes nas páginas 32, 39, 52, 57, 70, 71, 83, 87, 90, 92, 95 e 99.
- DELONG, S.; MACKENZIE, I. S. Evaluating Devices for Object Rotation in 3D. In: ANTONA, M.; STEPHANIDIS, C. (Ed.). *Universal Access in Human-Computer Interaction. Methods, Technologies, and Users*. Cham: Springer International Publishing, 2018. v. 10907, p. 160–177. ISBN 978-3-319-92048-1 978-3-319-92049-8. Citado na página 38.
- ESTEVEVES, A.; SHIN, Y.; OAKLEY, I. Comparing selection mechanisms for gaze input techniques in head-mounted displays. *International Journal of Human-Computer Studies*, v. 139, p. 102414, jul. 2020. ISSN 10715819. Citado 2 vezes nas páginas 21 e 38.
- FITTS, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, v. 47, n. 6, p. 381–391, jun. 1954. ISSN 0022-1015. Citado 2 vezes nas páginas 29 e 30.
- GROSSMAN, T.; BALAKRISHNAN, R. Pointing at trivariate targets in 3D environments. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Vienna, Austria: Association for Computing Machinery, 2004. (CHI '04), p. 447–454. ISBN 978-1-58113-702-6. Citado na página 36.
- GUIARD, Y.; Beaudouin-Lafon, M. Fitts' law 50 years later: Applications and contributions from human-computer interaction. *International Journal of Human-Computer Studies*, v. 61, n. 6, p. 747–750, dez. 2004. ISSN 10715819. Citado na página 7.
- GUIARD, Y.; OLAFSDOTTIR, H. B. On the Measurement of Movement Difficulty in the Standard Approach to Fitts' Law. *PLoS ONE*, v. 6, n. 10, p. e24389, out. 2011. ISSN 1932-6203. Citado na página 91.
- HANSEN, J. P. et al. A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In: *Proceedings of the Workshop on Communication by Gaze Interaction - COGAIN '18*. Warsaw, Poland: ACM Press, 2018. p. 1–5. ISBN 978-1-4503-5790-6. Citado na página 38.
- HART, S. G.; STAVELAND, L. E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: HANCOCK, P. A.; MESHKATI, N. (Ed.). *Advances in Psychology*. [S.l.]: North-Holland, 1988, (Human Mental Workload, v. 52). p. 139–183. Citado na página 45.

- HOFFMAN, D. M. et al. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, v. 8, n. 3, p. 33.1–3330, mar. 2008. ISSN 1534-7362. Citado na página 52.
- II, J. E. S.; GABBARD, J. L. Survey of User-Based Experimentation in augmented Reality. In: *In 1st International Conference on Virtual Reality, Las Vegas*. [S.l.: s.n.], 2005. Citado na página 20.
- International Organization for Standardization. *ISO 9241-9:2000 - Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) - Part 9: Requirements for on-Keyboard Input Devices*. Geneva, 2000. Citado 2 vezes nas páginas 23 e 30.
- International Organization for Standardization. *ISO/TS 9241-411 - Ergonomics of Human-System Interaction - Part 411: Evaluation Methods for the Design of Physical Input Devices*. Switzerland, 2012. 64 p. Citado 2 vezes nas páginas 23 e 30.
- KOUROUPETROGLOU, G. et al. Using Wiimote for 2D and 3D Pointing Tasks: Gesture Performance Evaluation. In: EFTHIMIOU, E.; KOUROUPETROGLOU, G.; FOTINEA, S.-E. (Ed.). *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*. [S.l.]: Springer Berlin Heidelberg, 2012. (Lecture Notes in Computer Science), p. 13–23. ISBN 978-3-642-34182-3. Citado 6 vezes nas páginas 21, 38, 39, 40, 43 e 44.
- KREVELEN, R. V.; POELMAN, R. A Survey of Augmented Reality Technologies, Applications and Limitations. *International Journal of Virtual Reality (ISSN 1081-1451)*, v. 9, p. 1, jun. 2010. Citado 2 vezes nas páginas 19 e 21.
- LAVIOLA, J. J. 3D Gestural Interaction: The State of the Field. *ISRN Artificial Intelligence*, v. 2013, p. 1–18, 2013. ISSN 2090-7443. Citado na página 21.
- LUBOS, P.; BRUDER, G.; STEINICKE, F. Analysis of direct selection in head-mounted display environments. In: *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. [S.l.: s.n.], 2014. p. 11–18. Citado 5 vezes nas páginas 38, 39, 42, 43 e 103.
- LUBOS, P.; BRUDER, G.; STEINICKE, F. Influence of Comfort on 3D Selection Task Performance in Immersive Desktop Setups. *Journal of Virtual Reality and Broadcasting*, v. 12, n. 2, p. 15, 2015. Citado 5 vezes nas páginas 38, 39, 40, 42 e 47.
- MACHUCA, M. D. B.; STUERZLINGER, W. The Effect of Stereo Display Deficiencies on Virtual Hand Pointing. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow, Scotland Uk: Association for Computing Machinery, 2019. (CHI '19), p. 1–14. ISBN 978-1-4503-5970-2. Citado 4 vezes nas páginas 36, 42, 52 e 103.
- MACKENZIE, I. S. Fitts' Throughput and the Remarkable Case of Touch-Based Target Selection. In: KUROSU, M. (Ed.). *Human-Computer Interaction: Interaction Technologies*. Cham: Springer International Publishing, 2015. v. 9170, p. 238–249. ISBN 978-3-319-20915-9 978-3-319-20916-6. Citado na página 34.
- MACKENZIE, I. S. *A Note on ISO Testing of Computer Pointing Devices*. 2017. Disponível em: <<http://www.yorku.ca/mack/RN-ISO.html>>. Citado na página 23.
- MACKENZIE, I. S. *GoFitts*. 2020. Disponível em: <<http://www.yorku.ca/mack/FittsLawSoftware/doc/index.html?GoFitts.html>>. Citado 2 vezes nas páginas 54 e 101.
- Magic Leap. *Magic Leap 1*. 2020. Disponível em: <<https://www.magicleap.com/magic-leap-1>>. Citado na página 20.

- MATNEY, L. *Hands-on with the \$949 Mind-Bending Meta 2 Augmented Reality Headset | TechCrunch*. 2016. Disponível em: <<https://techcrunch.com/2016/03/02/hands-on-with-the-949-mind-bending-meta-2-augmented-reality-headset/>>. Citado na página 20.
- Microsoft. *HoloLens 2—Overview, Features, and Specs | Microsoft HoloLens*. 2020. Disponível em: <<https://www.microsoft.com/en-us/hololens/hardware>>. Citado na página 20.
- NIEUWENHUIZEN, K. *Quantitative Analysis of Computer Interaction Movements*. Eindhoven: Technische Universiteit Eindhoven: [s.n.], 2015. ISBN 978-90-386-3839-3. Citado na página 24.
- ÖZACAR, K. et al. 3D Selection Techniques for Mobile Augmented Reality Head-Mounted Displays. *Interacting with Computers*, v. 29, n. 4, p. 579–591, jul. 2017. ISSN 0953-5438. Citado 7 vezes nas páginas 38, 39, 40, 43, 45, 47 e 103.
- PINO, A. et al. Using Kinect for 2D and 3D Pointing Tasks: Performance Evaluation. In: KUROSU, M. (Ed.). *Human-Computer Interaction. Interaction Modalities and Techniques*. [S.l.]: Springer Berlin Heidelberg, 2013. (Lecture Notes in Computer Science), p. 358–367. ISBN 978-3-642-39330-3. Citado 4 vezes nas páginas 38, 39, 40 e 43.
- RAYNAL, M.; DUBOIS, E.; SCHMITT, B. Towards Unification for Pointing Task Evaluation in 3D Desktop Virtual Environment. In: HOLZINGER, A. et al. (Ed.). *Human Factors in Computing and Informatics*. [S.l.]: Springer Berlin Heidelberg, 2013. (Lecture Notes in Computer Science), p. 562–580. ISBN 978-3-642-39062-3. Citado 9 vezes nas páginas 24, 38, 39, 40, 42, 43, 44, 45 e 49.
- ROCHA, H. V. D.; BARANAUSKAS, M. C. C. *Design e avaliação de interfaces humano-computador*. Campinas, São Paulo: Universidade Estadual de Campinas, 2000. Citado na página 37.
- SAMBROOKS, L.; WILKINSON, B. Comparison of Gestural, Touch, and Mouse Interaction with Fitts' Law. In: *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*. New York, NY, USA: ACM, 2013. (OzCHI '13), p. 119–122. ISBN 978-1-4503-2525-7. Citado 2 vezes nas páginas 22 e 38.
- SHANNON, C. E. *A Mathematical Theory of Communication*. p. 55, 1948. Citado na página 29.
- SOUKOREFF, R. W.; MACKENZIE, I. S. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, v. 61, n. 6, p. 751–789, dez. 2004. ISSN 1071-5819. Citado 8 vezes nas páginas 23, 30, 32, 33, 34, 35, 71 e 91.
- STUERZLINGER, W.; TEATHER, R. J. Considerations for targets in 3D pointing experiments. In: *Proceedings of HCI Korea*. Seoul, Republic of Korea: Hanbit Media, Inc., 2014. (HCIK '15), p. 162–168. ISBN 978-89-6848-752-1. Citado 3 vezes nas páginas 42, 49 e 53.
- SUTHERLAND, I. E. The Ultimate Display. In: *Proceedings of the Congress of the International Federation of Information Processing (IFIP)*. [S.l.: s.n.], 1965. volume 2, p. 506–508. Citado 2 vezes nas páginas 19 e 106.
- TEATHER, R. J. *Evaluating 3D Pointing Techniques*. Tese (Doutorado) — York University, Toronto, Ontario, maio 2013. Citado 3 vezes nas páginas 21, 22 e 47.

TEATHER, R. J.; NATAPOV, D.; JENKIN, M. Evaluating haptic feedback in virtual environments using ISO 9241-9. In: *2010 IEEE Virtual Reality Conference (VR)*. [S.l.: s.n.], 2010. p. 307-308. ISSN 2375-5334. Citado 6 vezes nas páginas 38, 39, 40, 41, 42 e 47.

TEATHER, R. J. et al. Effects of tracking technology, latency, and spatial jitter on object movement. In: *2009 IEEE Symposium on 3D User Interfaces*. [S.l.: s.n.], 2009. p. 43-50. Citado na página 38.

TEATHER, R. J.; STUERZLINGER, W. Assessing the Effects of Orientation and Device on (Constrained) 3D Movement Techniques. In: *2008 IEEE Symposium on 3D User Interfaces*. Reno, NV, USA: IEEE, 2008. p. 43-50. ISBN 978-1-4244-2047-6. Citado na página 22.

TEATHER, R. J.; STUERZLINGER, W. Pointing at 3D targets in a stereo head-tracked virtual environment. In: *2011 IEEE Symposium on 3D User Interfaces (3DUI)*. [S.l.: s.n.], 2011. p. 87-94. Citado 9 vezes nas páginas 38, 39, 40, 41, 42, 44, 47, 48 e 53.

TEATHER, R. J.; STUERZLINGER, W. Pointing at 3D Target Projections with One-eyed and Stereo Cursors. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2013. (CHI '13), p. 159-168. ISBN 978-1-4503-1899-0. Citado 2 vezes nas páginas 37 e 38.

TEATHER, R. J.; STUERZLINGER, W. Visual aids in 3D point selection experiments. In: *Proceedings of the 2nd ACM Symposium on Spatial User Interaction*. Honolulu, Hawaii, USA: Association for Computing Machinery, 2014. (SUI '14), p. 127-136. ISBN 978-1-4503-2820-3. Citado 2 vezes nas páginas 43 e 50.

Ultraleap. *Tracking | Leap Motion Controller | Ultraleap*. 2020. Disponível em: <<https://www.ultraleap.com/product/leap-motion-controller/>>. Citado na página 22.

WARE, C.; LOWTHER, K. Selection Using a One-eyed Cursor in a Fish Tank VR Environment. *ACM Trans. Comput.-Hum. Interact.*, v. 4, n. 4, p. 309-322, dez. 1997. ISSN 1073-0516. Citado na página 38.

YOUNG, T. S.; TEATHER, R. J.; MACKENZIE, I. S. An arm-mounted inertial controller for 6DOF input: Design and evaluation. In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. Los Angeles, CA, USA: IEEE, 2017. p. 26-35. ISBN 978-1-5090-6716-9. Citado 6 vezes nas páginas 38, 39, 40, 42, 47 e 53.

ZHAI, S. User Performance in Relation to 3D Input Device Design. *SIGGRAPH Comput. Graph.*, v. 32, n. 4, p. 50-54, nov. 1998. ISSN 0097-8930. Citado 2 vezes nas páginas 21 e 44.

ZHAI, S.; BUXTON, W.; MILGRAM, P. The "Silk Cursor": Investigating transparency for 3D target acquisition. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*. Boston, Massachusetts, United States: ACM Press, 1994. p. 459-464. ISBN 978-0-89791-650-9. Citado 2 vezes nas páginas 42 e 49.

ZHAI, S.; MILGRAM, P. Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Los Angeles, California, USA: ACM Press/Addison-Wesley Publishing Co., 1998. (CHI '98), p. 320-327. ISBN 978-0-201-30987-4. Citado na página 44.

APÊNDICE A – QUESTIONÁRIO PARA AVALIAÇÃO DE CONFORTO¹

Nome: _____ Idade: _____ anos

Sexo: () F () M

Tarefa: () Arrastar () Apontar-e-selecionar

Método de interação: () Mouse () Meta 2 () Controlador Leap Motion

Questionário para Avaliação de Conforto (ISO 9241-411 Anexo C)

1. Força exigida para movimentar o cursor:						
1	2	3	4	5	6	7
Muito desconfortável				Muito confortável		
2. Suavidade durante a operação:						
1	2	3	4	5	6	7
Muito duro				Muito suave		
3. Esforço requerido para a operação:						
1	2	3	4	5	6	7
Muito grande				Muito baixo		
4. Precisão:						
1	2	3	4	5	6	7
Muito impreciso				Muito preciso		
5. Velocidade de operação:						
1	2	3	4	5	6	7
Inaceitável				Aceitável		
6. Conforto geral:						
1	2	3	4	5	6	7
Muito desconfortável				Muito confortável		
7. Operação do dispositivo de entrada como um todo:						
1	2	3	4	5	6	7
Muito difícil de usar				Muito fácil de usar		
8. Fadiga do dedo:						
1	2	3	4	5	6	7
Muito alta				Não há fadiga		
9. Fadiga do pulso:						
1	2	3	4	5	6	7
Muito alta				Não há fadiga		
10. Fadiga do braço:						
1	2	3	4	5	6	7
Muito alta				Não há fadiga		
11. Fadiga do ombro:						
1	2	3	4	5	6	7
Muito alta				Não há fadiga		
12. Fadiga do pescoço:						
1	2	3	4	5	6	7
Muito alta				Não há fadiga		

¹ A tradução do questionário foi feita pelo autor a partir do original reproduzido no Anexo A

ANEXO A – QUESTIONÁRIO PARA AVALIAÇÃO DE CONFORTO DO PADRÃO ISO/TS 9241-411¹

1. Force required for actuation:						
1	2	3	4	5	6	7
Very uncomfortable				Very comfortable		
2. Smoothness during operation:						
1	2	3	4	5	6	7
Very rough				Very smooth		
3. Effort required for operation:						
1	2	3	4	5	6	7
Very high				Very low		
4. Accuracy:						
1	2	3	4	5	6	7
Very inaccurate				Very accurate		
5. Operation speed:						
1	2	3	4	5	6	7
Unacceptable				Acceptable		
6. General comfort:						
1	2	3	4	5	6	7
Very uncomfortable				Very comfortable		
7. Overall operation of input device:						
1	2	3	4	5	6	7
Very difficult (to use)				Very easy (to use)		
8. Finger fatigue:						
1	2	3	4	5	6	7
Very high				None		
9. Wrist fatigue:						
1	2	3	4	5	6	7
Very high				None		
10. Arm fatigue:						
1	2	3	4	5	6	7
Very high				None		
11. Shoulder fatigue:						
1	2	3	4	5	6	7
Very high				None		
12. Neck fatigue:						
1	2	3	4	5	6	7
Very high				None		

¹ Segundo nota do próprio documento ISO/TS 9241-411, esta tabela não está protegida pelos direitos de cópia do restante do documento e pode ser livremente reproduzida.

ANEXO B – MANUSCRITO DE ARTIGO ORIGINAL¹

¹ Manuscrito submetido ao International Journal of Human-Computer Studies em 31 de Julho de 2020

A method for evaluating 3D mid-air hand interaction in immersive Augmented Reality

John Lennon Couto^{a,*}, Marcelo Archanjo José^b, Roseli de Deus Lopes^{a,b}

^a*Department of Electrical Engineering, Polytechnic School, University of Sao Paulo, Sao Paulo, Brazil*

^b*Interdisciplinary Center in Interactive Technologies, University of Sao Paulo, Brazil*

Abstract

Hand tracking is an intuitive and stimulating interaction option for augmented and virtual realities. However, existing technologies still lack the speed, accuracy and reliability of traditional 2D interaction devices. Evaluating interaction performance can help improving such technologies, but currently there is no standard evaluation method for arm's reach hand interaction in Augmented Reality (AR) systems. Based on previous Fitts' Law studies and on the multi-directional test of ISO 9241-411, this work proposes a method to evaluate near hand interaction in an AR optical See-Through Head-Mounted Display (OST-HMD). The 3DFitts software was developed to implement this method and two user studies were conducted to assess its use as an evaluation tool for 3D hand interaction in AR. A mouse and two short-range hand tracking devices - the Leap Motion Controller and the interaction system of the Meta 2 headset - were evaluated using the software and questionnaires. Results indicate that participants' perception of each input modality corroborate the objective throughput metric computed by 3DFitts, suggesting that the proposed evaluation method was able to effectively measure 3D hand interaction performance in AR environments.

Keywords: Fitts' Law, 3D interaction, Hand interaction, Augmented Reality, Head-Mounted Display

*Corresponding author

Email addresses: john.couto@usp.br (John Lennon Couto), marcelo.archanjo@usp.br (Marcelo Archanjo José), roseli.lopes@usp.br (Roseli de Deus Lopes)

1. Introduction

The field of Augmented Reality (AR) is arguably the one that gets closer to accomplish Sutherland’s vision of the Ultimate Display (Sutherland, 1965). AR is a real-time interactive system that supplements the real world by registering
 5 3D virtual information over its three dimensions (Azuma, 1997). Many of the fundamental challenges in AR have been solved in the last three decades, but limitations remain and hinder the realization of its full potential (Billinghurst et al., 2015). Recent advances allowed mobile AR to become popular, but the golden promise of AR - the immersive experience with stereoscopic visualization
 10 and fluid interaction - is yet to be commercially achieved.

Immersive AR can be experienced using Optical See-Through Head-Mounted Displays (OST-HMDs) such as the Microsoft HoloLens (Microsoft, 2020), the Meta 2 from former Metavision company (Matney, 2016) and the Magic Leap One (Magic Leap, 2020). Although these products have been commercially
 15 available, they remain in a development phase and have many limitations. One is the poor overall quality of 6 Degrees Of Freedom (DOF) interaction, which is usually preferred in an immersive experience because it allows users to control position and orientation of 3D virtual objects simultaneously, providing a more natural and intuitive interaction (Teather, 2013).

20 Researchers have explored a range of input modalities for interacting with 3D environments, such as 3D controllers (Zhai, 1998; Kouroupetroglou et al., 2012), gaze and head movement (Esteves et al., 2020), and 3D gestural interaction (LaViola, 2013). Each of them has its own benefits and drawbacks and combining them in a multimodal interface can result in a better solution
 25 (Van Krevelen and Poelman, 2010). Zhai stated 20 years ago that none of the current 3D input devices could be established as the de facto standard for 3D interaction like the mouse was for 2D interfaces (Zhai, 1998). We believe this statement remains true today.

The research field of 3D gestural interaction has seen important advances and
 30 has grown in maturity (LaViola, 2013). Despite the obvious drawback of arm

fatigue, hand gesture input has received wide attention (Al-Shamayleh et al., 2018) and is present in the three OST-HMDs mentioned above. One possible reason is that it can supposedly supports 6 DOF interactions co-located with the user's motor space. Co-location increases immersion because the virtual content is displayed in the same user movement space (Teather, 2013). Despite its high promises, hand tracking interaction proved to perform significantly worse than touch and mouse interactions for simple 2D pointing tasks (Teather and Stuerzlinger, 2008; Sambrooks and Wilkinson, 2013; Burno et al., 2015).

Immersive AR lacks a 3D interaction system that can provide fast, intuitive and accurate interaction as the mouse and touch screen have done for personal computers and mobile devices. Formal evaluation is an important step to develop better input methods because it allows comparing different interaction solutions to decide which one performs better. Fitts' Law has been widely used as an evaluation tool for 2D interaction (Soukoreff and MacKenzie, 2004). It has also been used to evaluate 3D interaction in both 2D displays and in Virtual Reality (VR) environments (Ware and Lowther, 1997; Teather et al., 2009, 2010; Teather and Stuerzlinger, 2011; Kouroupetroglou et al., 2012; Teather and Stuerzlinger, 2013; Pino et al., 2013; Sambrooks and Wilkinson, 2013; Raynal et al., 2013; Bruder et al., 2013; Bachmann et al., 2014; Lubos et al., 2014, 2015; Burno et al., 2015; DeLong and MacKenzie, 2018; Young et al., 2017; Hansen et al., 2018; Esteves et al., 2020). Still, to the best of our knowledge, it has never been studied for interaction in an AR OST-HMD. Moreover, the lack of a standard such as the ISO 9241-411 for 2D evaluation, but that contemplates the 3D case, makes it difficult to compare these studies.

In this work, we propose a method for objectively evaluating user performance in 3D pointing tasks with targets within arms' reach in AR OST-HMDs. It was adapted from ISO 9241-411 and is based on previous works by HCI research community. We applied it to evaluate two short-range hand-tracking systems with an AR OST-HMD. A mouse has also been evaluated to provide external validity.

The content of this work is organized in eight sections. In the first one, we

discussed the importance of 3D interaction evaluation for further advances in immersive AR experience. Section 2 reviews how the ISO 9241-411 standard has been used to evaluate 2D and 3D interaction. Section 3 presents a method
 65 for evaluating within arm's reach interaction in an AR OST-HMD, which was implemented in software that is outlined in Section 4. Sections 5 and 6 describe the methods and results of two user studies we conducted using the developed software. In Section 7 we discuss results from both studies. We conclude summarizing our work and pointing out future research directions in Section 8.

70 **2. Related works**

2.1. Fitts' Law and ISO 9241-411

The duration of a rapid aimed arm's movement is linearly related to the logarithm of the ratio of the amplitude to the variability of the movement (Fitts, 1954). This relation is known as Fitts' Law and has been extensively used as a
 75 scientific tool in HCI research (Guiard and Beaudouin-Lafon, 2004). Although originally proposed for unidirectional movements, it has been successfully applied to evaluate 2D pointing tasks. Many refinements have been proposed in an effort to strengthen the underlying theory and improve its predictive power (Soukoreff and MacKenzie, 2004), but the multiplicity of models hampered between-study comparisons. The publication of the ISO 9241-9 standard
 80 (International Organization for Standardization, 2000) - later updated and replaced by ISO 9241-411 (International Organization for Standardization, 2012) - established common models and test procedures for evaluating computer input devices for 1D and 2D interactions and mitigated this problem.

85 The standard describe methods to evaluate interaction primitives such as pointing, selecting, dragging, and tracing. The multi-directional tapping test (Figure 1), in particular, is widely used to evaluate 2D interaction. In such a test, participants point and select targets as fast and accurate as possible in tasks with varying difficulties. Time and coordinates of targets' selection are

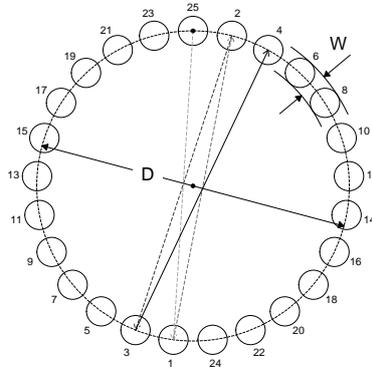


Figure 1: The multi-directional tapping test described in ISO 9241-411. W denotes target's width and D , distance between targets. Interaction starts and ends at target 25. Figure was adapted from (Soukoreff and MacKenzie, 2004).

90 collected to measure movement duration and variability, and to compute the throughput (TP) metric.

The difficulty of a task is controlled by changing targets' width (W) and distance (D), which affects the movement amplitude (A). It is modeled by the Index of Difficulty (ID), whose unit is given in *bits*:

$$ID = \log_2 \left(\frac{A}{W} + 1 \right) \quad (1)$$

95 Movement duration is represented by the mean movement time (MT), which is computed by averaging the selection times. Movement variability is expressed by the Effective Amplitude (A_e) and Effective Width (W_e), which can be computed from selection coordinates. A_e is the actual movement distance projected on the task axis. W_e is given by $4.133 * SD_x$, where SD_x is the standard deviation of the projection of the selection coordinates on the task axis. With these
100 values, one can compute the Effective Index of Difficulty (ID_e) using Equation 2¹. These four metrics are computed for each task and participant (MacKenzie,

¹ISO 9241-411 uses A instead of A_e . Using A_e has little impact in the final result but is recommended to prevent cases when selection coordinates are not distributed around the target's center (MacKenzie, 2015).

2015), and therefore ID_e represents the overall participant's accuracy in a task.

$$ID_e = \log_2 \left(\frac{A_e}{W_e} + 1 \right) \quad (2)$$

TP may then be computed by averaging the ratio of ID_e to MT across
 105 the total number of sequences of trials (N_s) by participant and then across all
 participants (N_p), as shown in Equation 3 (Soukoreff and MacKenzie, 2004).

$$TP = \frac{1}{N_p} \sum_{j=1}^{N_p} \left(\frac{1}{N_s} \sum_{i=1}^{N_s} \frac{ID_{e_{ij}}}{MT_{ij}} \right) \quad (3)$$

Human performance is subject to the speed-accuracy tradeoff: the faster a
 participant tries to finish a pointing task, the lower the final accuracy of the
 selection coordinates. Because throughput combines both participants' speed
 110 and accuracy, it is a metric that quantify human performance better than move-
 ment duration or variability alone and may be used to compare different test
 conditions over different studies (Soukoreff and MacKenzie, 2004).

HCI studies usually report TP , MT and error rate (ER) metrics. Some au-
 thors also report the resulting Fitts' Law model, which can be obtained by linear
 115 regression using Equation 4. Regressions with poor fit may indicate that par-
 ticipants' movements did not obey Fitts' Law or that there was a problem with
 the methodology or the execution of the experiment (Soukoreff and MacKenzie,
 2004).

$$MT = a + b * ID_e \quad (4)$$

2.2. Fitts' Law extensions to 3D movements

120 Equations 1, 2 and 4 are commonly used to model 1D and 2D human inter-
 action but when applied to general 3D movements, it usually results in worse
 correlation between MT and ID . This has led some researchers to propose
 extensions to the model either by modifying ID computation or by adding new
 terms to the linear model (Atsuo Murata and Hirokazu Iwase, 2001; Grossman

125 and Balakrishnan, 2004; Cha and Myung, 2013; Barrera Machuca and Stuerzlinger, 2019; Clark et al., 2020). Extended models usually improves the correlation between the variables of the model. Nevertheless, Teather and Stuerzlinger (2013) pointed out that adding any new free parameter to a regression analysis will always improve correlation, and, therefore, it is not clear what was the real
130 contribution of the new parameters to the predictability of the model. This may partly explain why no 3D model of Fitts' Law has yet become widely adopted in formal 3D interaction evaluations.

2.3. Evaluation of mid-air hand interaction

We conducted a non-systematic survey of the literature to understand which
135 methods researchers have been using to evaluate mid-air interaction in 3D environments. Results are summarized in Table 1 and will be discussed in the following subsections. Because we want to evaluate hand interaction in an AR headset, we were mainly interested in stereoscopic experience with direct hand interaction. Thus, we report only the main mid-air 3D task of each study and
140 ignore other input methods or 2D evaluation that may have been conducted.

2.3.1. Input method and display

For each work, we reported the input methods and display that provided the closest experience to direct 3D interaction in an stereoscopic environment.

The main input method for hand interaction was the use of mid and long-
145 range external trackers to track handheld controllers or participant's index finger. Direct interaction was present in (Teather et al., 2010; Teather and Stuerzlinger, 2011; Bruder et al., 2013; Lubos et al., 2014, 2015; Özacar et al., 2017). We consider these to be the works that are closest to ours.

The surveyed works employed a myriad of methods for visual feedback: non-
150 stereo 3D virtual environments (Kouroupetroglou et al., 2012; Pino et al., 2013; Raynal et al., 2013), immersive VR with different technologies (Bérard et al., 2009; Teather and Stuerzlinger, 2011; Bruder et al., 2013; Lubos et al., 2014,

Reference	Input Method	Display	Targets' Shape & Arrangement	Procedure	Fitts' Law	Main Metrics
(Béard et al., 2009)	Handheld controller (tracked Wii Remote)	VR with projectors	Cubes randomly placed in a cubic space	Move the white cube to enclose the target	Not reported	MT, ER
(Teather et al., 2010)	Handheld controller (tracked stylus)	CAVE	Spheres in a multi-directional ISO test in a frontal plane	Reciprocal multi-directional ISO test	ISO 9241-9 (Section 2.1)	TP, MT, ER
(Teather and Stuerzinger, 2011)	Handheld controller (tracked stylus)	Desktop VR	Spheres over pillars in a multi-directional ISO test in an angled plane	Reciprocal multi-directional ISO test	ISO 9241-9 (Section 2.1)	TP, MT, ER
(Kouroupetrolou et al., 2012; Pino et al., 2013)	Handheld controller (Wii Remote) and bare hand (Kinect)	2D TFT Monitor	8 spheres at the vertices of a cube	Serial selection between diagonally opposite targets	ISO 9241-9 (Section 2.1)	TP, MT, MCL, TRE, DT, others
(Bruder et al., 2013)	Index Finger (WorldViz PPT X4)	VR on tabletop	Spheres in a multi-directional ISO test in a horizontal plane with varying heights	Reciprocal multi-directional ISO test	ISO 9241-9 (Section 2.1)	TP, MT, ER, ED
(Raynal et al., 2013)	Handheld controller (Polhemus Patriot)	2D LCD Monitor	Spheres are homogeneously distributed in a sphere's shell	Discrete movements between diagonally opposite spheres	Adapted ISO 9241-9	TP, MT, ER, IR
(Lubos et al., 2014)	Index Finger (Optitrack VI20; Trio)	VR HMD	Spheres in a multi-directional ISO test in nine plane orientations	Serial movements, from a fixed sphere to all targets	Not reported	TP, MT, ER, ED
(Lubos et al., 2015)	Index Finger (WorldViz PPT X4)	VR HMD	Spheres in a multi-directional ISO test in three plane orientations	Reciprocal multi-directional ISO test	ISO 9241-9 (Section 2.1)	TP, MT, ER
(Young et al., 2017)	Wearable IMU	Desktop VR	Spheres over pillars in a multi-directional ISO test in a vertical plane at varying depths	Reciprocal multi-directional ISO test	ISO 9241-9 (Section 2.1)	TP, MT, ER
(Özacar et al., 2017)	Index Finger (Infrared cameras)	AR HMD	Spheres randomly distributed in a cubic interaction space	Serial selection of targets, measuring time to selection	Not applied	MT, ER

Table 1: Summary of related works that evaluated hand interaction in 3D environments. The meaning of the previously unmentioned acronyms are Missed Clicks (*MCL*), Target Re-Entry (*TRE*), Distance Travelled (*DT*), Error Distance (*ED*) and Inefficiency Ratio (*IR*).

2015; Young et al., 2017), CAVE system (Teather et al., 2010) and AR headset (Özacar et al., 2017).

155 We were specially interested in direct interaction evaluation in AR OST-HMDs but, to the best of our knowledge, only Özacar et al. (2017) have reported such study. However, their evaluation method was not based on the ISO standard and they did not apply Fitts' Law.

160 Other works employed VR systems in a way that resembled AR experience in the sense that the virtual content was presented and directly interacted in the real world instead of inside the VR monitor or HMD. Teather et al. (2010), for instance, conducted a user study to evaluate direct interaction in a CAVE. Teather and Stuerzlinger (2011) and Bruder et al. (2013), on the other hand, presented targets outside the screen of 2D stereo-capable monitors, allowing 165 users to direct interact with the virtual content.

2.3.2. Targets' shape and arrangement

With the exception of (Bérard et al., 2009), all surveyed work used spheres to represent targets and cursor. This is a natural choice because the width of a spheric target is the same in any direction of approach. Common colors chosen 170 for targets and cursor were white, gray, blue, green and red. Semi-transparent targets were used to improve depth perception and reduce cursor occlusion by (Raynal et al., 2013). The use of opaque targets with semi-transparent cursor is evident in (Bérard et al., 2009) and (Teather and Stuerzlinger, 2011) and is supported by the work of Zhai et al. (1994), which has demonstrated that a 175 semi-transparent volumetric cursor allows better performance than an equivalent wire frame cursor when acquiring opaque 3D objects in both stereo and mono displays.

Targets were randomly positioned in a cubic space by (Bérard et al., 2009; Özacar et al., 2017), positioned in the vertices of a cube by (Kouroupetroglou 180 et al., 2012; Pino et al., 2013) and homogeneously distributed in the shell of a sphere by (Raynal et al., 2013). The preferred approach, however, was to arrange them in a plane conforming to the ISO multi-directional test, as in

Figure 1. This was the choice of (Teather et al., 2010; Teather and Stuerzlinger, 2011; Bruder et al., 2013; Lubos et al., 2014, 2015; Young et al., 2017). Although
 185 one could argue that the interaction was not entirely 3D since movement was constrained to 2D, this approach has the advantage of improving comparability between studies, including the large literature body of 2D interaction evaluation (Teather and Stuerzlinger, 2011; Young et al., 2017).

2.3.3. Procedure

190 Discrete movements were used by (Bérard et al., 2009; Raynal et al., 2013) and consist in starting a task at a point and ending it when selecting the target. Serial movements were used by (Kouroupetroglou et al., 2012; Pino et al., 2013; Lubos et al., 2014; Özacar et al., 2017) and consists in a sequence of discrete movements, one starting immediately after the last one. The remaining works
 195 employed reciprocal movements, which are a specific type of serial movements where sequential movements resembles one another but in the opposite direction, such as in Figure 1.

The typical procedure of a HCI evaluation study was commonly used: to begin a test, participants must acquire an initial target; the next target is high-
 200 lighted and should be acquired as fast and accurately as possible using the selection method provided; the acquisition may or may not be successful; for a discrete test, task will end after the acquisition event and a new task may start by selecting another starting target; for a serial and reciprocal test, the next target is highlighted until all desired targets are presented.

205 When clearly reported, cursor interaction occurred when its volume collided with targets' shapes. Highlighting a target by changing its color was commonly used to indicate when interaction was possible.

2.3.4. Fitts' Law models

With the exception of (Bérard et al., 2009; Özacar et al., 2017), which only
 210 reported *MT* and *ER*, all other works in Table 1 used the Fitts' Law formulation from ISO standard to evaluate 3D interaction and compute *TP*.

Raynal et al. (2013) modified the *ID* formulation by adding the cursor's width to the target's width. This way, task difficulty is better represented because performing the same task with a larger or smaller cursor will impact performance. Despite this being valid for all other studies that applied the ISO 9241-9 formulation, they were the only ones to adopt this modified version for *ID*.

Extended 3D versions of Fitts' Law were not used by any of the works in Table 1. Teather and Stuerzlinger (2011) concluded that the ISO formulation explained most of the interaction techniques they had employed.

2.3.5. Reported metrics

Zhai (1998) enumerated six aspects to evaluate a 6 DOF input device: speed, accuracy, coordination, easy of learning, fatigue and device persistence and acquisition.

Speed and accuracy measures the device's efficiency to acquire a target. They were the main aspects studied in the surveyed literature and were reported as *MT*, *ER* and *TP*.

Coordination measures the capacity to control the virtual cursor to achieve the most efficient trajectory in a task. It may be evaluated by the Inefficiency Ratio (*IR*), which measures how much the actual Distance Travelled (*DT*) deviated from the ideal distance (Zhai and Milgram, 1998). *IR* was reported by (Raynal et al., 2013) and *DT* by (Kouroupetroglou et al., 2012).

Easy of learning and fatigue evaluates the psychological and physical efforts needed in order to use the device. Three instruments were used to evaluate fatigue: the comfort assessment questionnaires from Annex C of ISO 9241-9, the Borg Scale (Borg, 1998) and the NASA Task Load Index (Hart and Staveland, 1988). The first was applied by (Raynal et al., 2013) and the last two by (Özacar et al., 2017). Ease of learning was indirectly evaluated by (Raynal et al., 2013) through the application of a SUS questionnaire.

Device persistence and acquisition are usability aspects that should measure the cursor's capacity to stay in place after releasing the device and how easy it is

to acquire it and to start controlling the cursor. Raynal et al. (2013) suggested that the System Usability Scale questionnaire (Brooke, 1996) and interviews could be used to evaluate these aspects.

245 **3. A method for evaluating mid-air arms' reach 3D hand interaction**

The previous section indicated that the main approach to evaluate 3D interaction is to adapt the multi-directional test by using spheric targets and keeping the interaction constrained to a plane (Teather et al., 2010; Teather and Stuerzlinger, 2011; Bruder et al., 2013; Lubos et al., 2014, 2015; Young et al., 2017).
 250 Teather (2013) has validated such method in experiments with 2D and 3D input devices, and recommended its use. But the results cannot be directly extrapolated to generic interaction in the 3D space because movement is constrained to the tested plane.

Raynal et al. (2013) proposed to use discrete movements to select spheric
 255 targets that were homogeneously distributed in a spheric shell. While these expressive changes allow the evaluation of generic 3D movements, they also make results less comparable to the existing literature on 2D interaction. Moreover, their work did not analyze if their experimental data conformed to the underlying Fitts' Law model used to compute TP .

260 We decided to adopt the first approach because it is closer to a well established standard, it allows comparison with the 2D literature and it has been used and validated by more researchers. But, while most authors have studied interaction mainly in only one plane orientation, we propose that the evaluation process should include more orientations in order to sweep a volumetric interaction
 265 space in front of the participant. By doing this, the final result will better represent what a generic interaction in the 3D space is when using the chosen input device.

Figure 2 illustrates the obvious plane candidates to represent generic 3D interaction: the frontal, horizontal and longitudinal planes. The first two are a
 270 natural choice for near hand interaction because humans are used to interacting

with objects disposed in walls, white boards or tablespots. Evaluation on the longitudinal plane, however, is challenging with the traditional multi-directional test mechanics because targets closer to the participant will occlude the ones located further away, effectively degrading the final results. Furthermore, each new plane greatly increases test duration and demands more from participants and researchers. Fortunately, with only two planes one can already study movement performance in all three dimensions. Therefore, we suggest that at least the frontal and horizontal planes should be investigated for any 3D interaction evaluation study, with the longitudinal plane being a third option that should be analyzed with care.

This method is flexible enough to allow finer evaluation of a device or interaction technique. In addition to the frontal and horizontal planes, one could also evaluate plane orientations at 45° angles from the main planes in order to obtain greater details of how performance varies in different directions. It is possible, for instance, to evaluate if a hand tracking device has uniform tracking performance in the whole 3D space or if some particular regions present worse response. Because the two main planes were evaluated, these in-depth

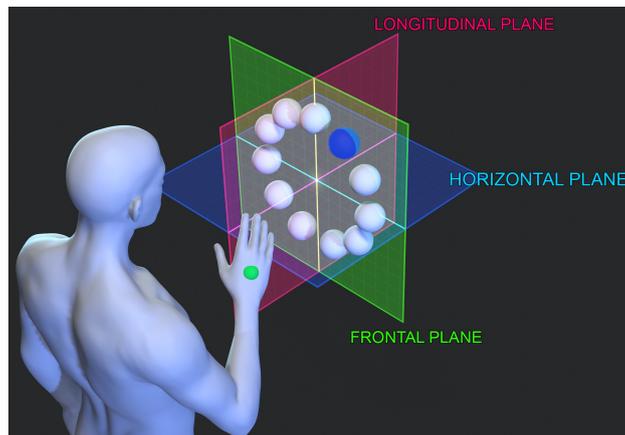


Figure 2: Representation of the multi-directional test on the frontal plane. The same test should be performed in the horizontal plane to describe how performance is affected by each dimension. The sphere on the hand of the human model represents the virtual cursor.

studies can be compared to other studies that were only interested in a coarse evaluation of a 3D interaction technique.

290 It is strongly recommended that any study evaluating 3D input devices also include a mouse in the experiment design in both 2D and 3D tests, in order to improve external validity and to increase comparability between studies (Teather and Stuerzlinger, 2011; Stuerzlinger and Teather, 2014; Young et al., 2017). For the 3D condition, it is sufficient to test the mouse only in the frontal
295 plane, which is the common way one would use and evaluate it.

Semi-transparent targets are usually employed because opaque ones can occlude the virtual cursor (Stuerzlinger and Teather, 2014). Here, we suggest to use semi-transparent targets together with an opaque cursor. This improves visual depth clues because participants' will be able to see a change of color
300 when the cursor is inside or behind the targets, which provides a similar visual experience to the Silk Cursor (Zhai et al., 1994). Additionally, it has been shown that target highlighting has a strong effect on performance and its use is recommended (Teather and Stuerzlinger, 2014).

Based on the above discussion and on the literature review presented in
305 Section 2, we summarize our main recommendations when evaluating within arms' reach interaction in a 3D immersive environment:

1. Use semi-transparent spheres as targets;
2. Arrange the targets in a plane as in the ISO 9241-411 multi-directional test (see Figures 1 and 2);
- 310 3. Use an opaque sphere as the virtual cursor;
4. Use volumetric collision to identify when the cursor is interacting with a target;
5. Highlight the target that is interacting with the cursor by changing its color;
- 315 6. Use reciprocal movements as defined by the ISO 9241-411 multi-directional test (Figure 1);
7. Use sound clues to indicate if target acquisition failed or succeeded;

8. Evaluate the test conditions at least in the frontal and horizontal planes, including more planes as needed for the study;
- 320 9. Compute and report the mean TP , MT and ER for each evaluated plane and input method;
10. Evaluate a mouse in the traditional multi-directional 2D test and on the frontal plane of the 3D test.

4. The 3DFitts software

325 The GoFitts software (MacKenzie, 2020) is commonly used by researchers to conduct 1D and 2D interaction evaluation based on the ISO 9241-411 standard. We have used it as the main reference to develop the 3DFitts software, which allows evaluating 3D interaction following the recommendations enumerated in the previous section. The software was developed using Unity engine, because it
330 is the main platform for developing applications for 3D environments nowadays and manufacturers of HMDs and 3D input devices usually offer a SDK that supports it.

The 3DFitts software has a core algorithm that executes the mechanics of the traditional multi-directional test. The targets and the cursor are spheres
335 and the collision between the volumes of the cursor and the targets defines when interaction is possible. With this design, it is possible to run both 2D and 3D interaction tests using the same algorithm: by displaying the content on a HMD, one can run the 3D test; on the other hand, the 2D test can be run by displaying an orthogonal view of the spheric targets on the computer
340 screen and by constraining the center of a 1-pixel cursor to the test plane. Additionally, the software allows mouse evaluation in a 3D environment. In this case, cursor's coordinates are constrained to the test plane and mapped to the mouse's coordinates in the screen. As the main algorithm is the same, confounding factors due to different mechanics in 2D and 3D tests are reduced.

345 Figure 3 illustrates the 3DFitts' user interface. In the main screen (Figure 3a), the researcher is able to configure the test, choosing its mode (2D or 3D),

the input and display methods, the number, distance and size of targets, the width of the cursor and the plane orientation. Some calibration procedures are also available for input and display methods that requires it. Figures 3b and 3c illustrates the execution a 2D test and Figure 3d, the final results.

In a formal test, we have used an auto-clicker software to replicate the same mouse inputs in an equivalent 2D test using the 3DFitts and the GoFitts software. Results indicated that both programs presented equivalent responses, that is, the observed differences were not statistically significant. Therefore, 2D evaluations performed with the 3DFitts software may be compared to the literature. Moreover, because 3DFitts can evaluate both 2D and 3D interactions with the same core algorithm, this result imply that 3D performance will be

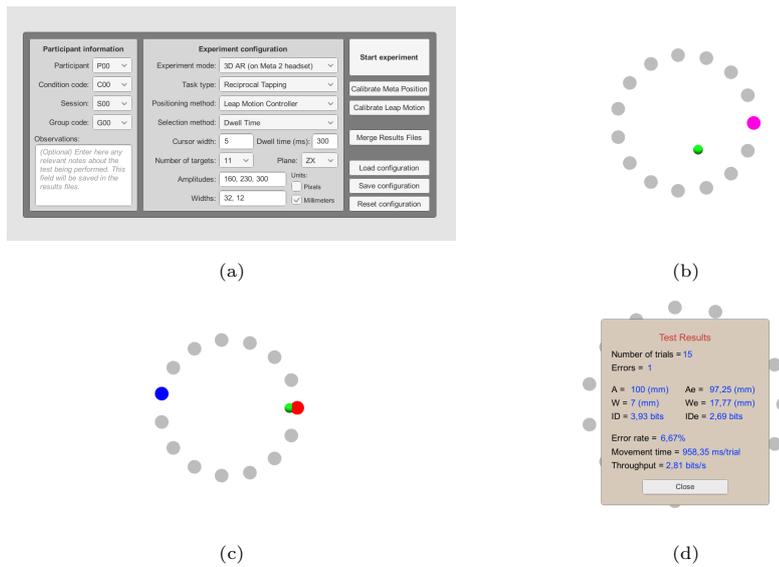


Figure 3: User interface of the 3DFitts software: (a) Main screen; (b) Starting view of a 2D multi-directional test, with the starting target in pink; (c) Target selection (highlighted in red) and next target (blue); (d) Results screen. The green sphere is the cursor, here presented with large width to illustrate the underlying implementation. For 2D tests, it would have only 1 pixel to resemble the mouse behavior. When running a 3D test, the participant will see the spheric targets positioned in the 3D space through the stereoscopic headset. Please refer to the article’s digital version for the colored figure.

computed by an algorithm that is equivalent to the 2D literature.

The next two sections present two user studies that were conducted to validate the 3DFitts algorithm and to evaluate 3D interaction in AR using the proposed method in Section 3. Both user studies were submitted to and approved by the Ethical Committee of University Hospital of University of São Paulo (HU-USP), receiving approval number 3.461.869.

5. User Study 1

Participants evaluated a mouse in a traditional 2D test using both GoFitts and 3DFitts applications. The primary goal was to compare the GoFitts' and 3DFitts' results when they were used in a formal HCI user study. The secondary goal was to measure the performance of the mouse in a traditional 2D evaluation that can be directly compared to the literature.

5.1. Methods

5.1.1. Participants

Twelve volunteers (7 man, 4 women and 1 trans woman) from a campus of the University of São Paulo were recruited for this study. Mean age was 28.1 years (SD = 6.2), ranging from 21 to 45 years. All participants were regular users of computer, with average usage of 42 hours/week (SD = 18).

5.1.2. Apparatus

A Dell G7 notebook (15.6" screen, NVIDIA GeForce GTX 1060 Max-Q, Intel Core i7-8750H, 16GB RAM) running Windows 10 was used to perform this study. GoFitts (MacKenzie, 2020) and 3DFitts programs were used to run the 2D multi-directional ISO test (Figure 1). The cursor size was configured to be 1 pixel in the 3DFitts software.

The test was performed using the Microsoft Basic Optical Mouse v2.0 (Model 1113), which is a corded mouse. All participants executed the test in the same

environmental conditions, sitting in the same desk and chair. The same applica-
 385 tions were running in the operational system for the period of each participant's
 test.

5.1.3. Design

The experiment was fully within-subjects with four independent variables:

- Software: { GoFitts, 3DFitts }
- 390 • Amplitude: { 400, 575, 750 } pixels
- Width: { 30, 80 } pixels
- Session: {1, 2, 3, 4, 5}

The software variable allows comparing how both programs evaluates the
 same interaction method. Amplitudes and widths were selected to allow the
 395 testing of a reasonable range of *ID*s (2.58 to 4.70 bits). Session variable was
 added to collect repetitive data to improve statistical analysis. The dependent
 variables were error rate, movement time and throughput.

The mouse's left button click was used as selection method. Participants
 were randomly assigned to one of two groups, which determined with which
 400 software the tests should start. Each test condition was performed with 15
 targets, generating 15 trials/sequence. Sequences were presented always in the
 same order (from the lowest *ID* to the highest). A total of 10800 trials were
 collected in this experiment (15 targets * 3 amplitudes * 2 widths * 5 session *
 2 programs * 12 participants).

405 5.1.4. Procedure

Participants received an oral explanation of the objective of the test and
 signed the consent form. They were introduced to both programs and to the
 multi-directional test and were asked to complete one session of the test in each
 software as a practice session. They were instructed to select targets as fastest
 410 and accurately as possible, to continue the test if a target was missed and to
 avoid too many misses.

Each participant completed five sessions using each software in an alternated manner. Breaks were allowed between sessions, but during a session participants were asked to be focused. The whole experiment lasted about 30 minutes per
415 participant.

5.2. Results and discussion

A total of 184 trials (1.7%) presented a movement time or an effective amplitude of movement that deviated more than 3 standard deviations from their respective mean. They were deemed outliers and removed from further analysis. The values were computed across all sessions for each task condition and
420 for each participant, and, therefore, outliers indicate a participant's uncommon behavior in a particular sequence. The remaining trials were used to compute the *MT*, *ER* and *TP* for each test sequence, resulting in 720 data points (6 test sequences * 5 sessions * 2 programs * 12 participants) that were used in the
425 following analysis.

5.2.1. Dependent variables

The grand means for the dependent variables were computed for each software and are shown in Table 2. Difference for *ER* was not significant ($F_{1,11} = 0.009, ns$). The grand mean for GoFitts' movement time was 4.4% lower than the
430 one for 3DFitts and the difference was statistically significant ($F_{1,11} = 52.52, p < .001$). Throughput was 4.4% higher for GoFitts and the difference was also statistically significant ($F_{1,11} = 108.6, p < .001$). Cohen's effect sizes were small for both *MT* ($d = 0.22$) and *TP* ($d = 0.31$), which suggests differences were trivial.

435 Previous experiment with an auto-clicker showed that both applications behave alike when no human factors are involved. Therefore, the presence of these significant effects may be explained by human factors. In fact, there are some visual and auditive discrepancies between the programs: targets are displayed with higher contrast on GoFitts, while the sounds for correct and incorrect
440 selections are more pronounced in 3DFitts. These differences may have affected

the participants' physiological responses, resulting in a 4.4% performance drop due to longer movement times.

The overall throughput obtained for the mouse was 5.32 bits/s, which is higher than the observed range of 3.7-4.9 bits/s in (Soukoreff and MacKenzie, 2004). The higher value may be explained by the improvement of the technology surrounding mouse interaction and by participants' greater computer experience.

Software	Error Rate (%)	Movement Time (ms)	Throughput (bits/s)
GoFitts	4.5 (SD = 3.6)	743 (SD = 101)	5.44 (SD = 0.55)
3DFitts	4.4 (SD = 3.7)	777 (SD = 109)	5.21 (SD = 0.52)

Table 2: Grand mean values and standard deviations for *ER*, *MT* and *TP*. The means were computed first for each participant, across all test sequences and sessions, and then across all 12 participants.

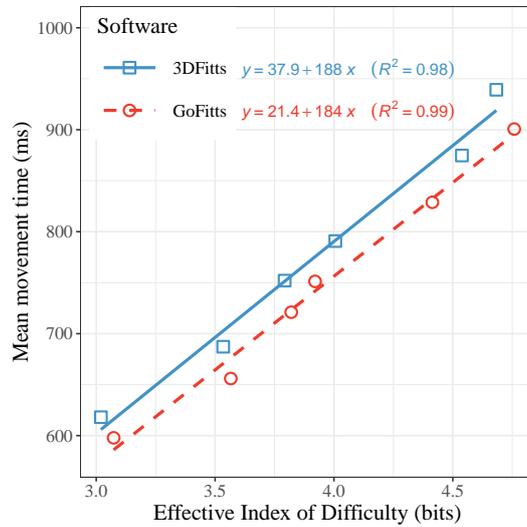


Figure 4: Fitts' Law models for both GoFitts and 3DFitts applications.

5.2.2. Fitts' law models

Participants' performance for each software was modeled using Equations
450 2 and 4. Data points were computed by averaging MT , W_e and A_e for each
sequence and software, across all sessions and participants (60 values per data
point). The models are presented in Figure 4. Participants' movements fairly
conformed to Fitts' Law, presenting $R^2 > 0.95$. Both models present similar co-
efficient values, indicating that GoFitts and 3DFitts behave similarly under the
455 test conditions, with the first consistently delivering slightly higher performance
results.

5.2.3. Distribution characteristics

The movement time and selection coordinates of the collected trials (without
outliers) were analyzed to check if they conformed to the normal distribution.
460 The skewness, kurtosis and a Shapiro-Wilk test ($\alpha = 0.05$) were computed for
each of the 720 test sequences, which contained between 13 and 15 trials. For
movement time, mean skewness was 0.50 ($SD = 0.61$), mean kurtosis was 2.99
($SD = 1.15$) and the null hypothesis for normality was not rejected for 81%
of the sequences by the Shapiro-Wilk test. For selection coordinates, mean
465 skewness was -0.001 ($SD = 0.60$), mean kurtosis was 2.80 ($SD = 0.96$) and
the null hypothesis for normality was not rejected for 93% of the sequences.
Skewness indicate that movement time distributions are slightly right-tailed.
Overall, these results suggest the distributions fairly conforms to the normal
distribution for the majority of the collected sequences.

470 5.3. Conclusion (User Study 1)

This user study was conducted with two objectives: to verify if 3DFitts and
GoFitts results are equivalent for formal HCI evaluation and to objectively eval-
uate a mouse in a traditional 2D test. The results indicate that both applications
behave similarly, although performance measured by 3DFitts was consistently
475 lower, albeit by small magnitude. Mouse's TP was higher about 10% higher
than that commonly reported in the literature.

6. User Study 2

The main objective of this user study was to apply our proposed evaluation method to measure the throughput of two hand tracking devices inside an AR environment. A secondary objective was to validate these results by comparing them to participants' subjective impressions. Our hypothesis is that both objective and subjective evaluations will converge. Other objectives are to analyze the effects of plane orientation on interaction performance and to study how well Fitts' Law model hand interaction on an AR headset.

6.1. Methods

The recommendations from Section 3 were applied to evaluate two hand tracking devices on the frontal and horizontal planes. The dwell time technique was chosen as a selection method to simplify interaction and to reduce errors since the act of clicking a button might displace the participants' interacting hand in space, moving the cursor unpredictably.

A mouse was evaluated in the frontal plane in order to allow comparisons between this study and the previous one and, through it, with the broader literature on 2D interaction evaluation. Two selection methods were studied: left button click and dwell time. The first condition can be compared to the previous user study since it also used it to select targets. The second condition will reveal the impact the dwell technique has in participants' performance, allowing a fairer comparison between the performance of the hand tracking devices and the mouse.

6.1.1. Participants

The 12 volunteers from User Study 1 were invited to contribute to this study. One could not participate and another one could not complete the experiment due to cyber-sickness. Therefore, data was gathered for a total of 10 participants. All were right-handed. Four declared to have myopia, one to have hypermetropia and one to have presbyopia.

505 6.1.2. Apparatus

The notebook from previous User Study was used to run the tests. The Meta 2 headset was used as OST-HMD to provide immersive AR. It has a display with 2550x1440 resolution, 60 Hz refresh rate and a field-of-view of 90°. The tests were performed with the 3DFitts standalone application.

510 Three devices were evaluated: the corded mouse from User Study 1, the hand tracker system embedded in the Meta 2 headset (from now on referenced as MetaTracker), and the Leap Motion Controller (LMC) (Ultraleap, 2020), a hand tracker sold by the late LeapMotion, Inc. (now incorporated by UltraLeap).

The Meta 2 headset was integrated into 3DFitts using the version 2.0.7.38
515 of the MetaSDK, and the LMC was integrated using the Leap Motion Core Version 4.4.0. The virtual cursor's width was set to 5 mm and was configured to be located in the center of the hand when the hand trackers were active. The AR occlusion feature was deactivated because preliminary tests indicated the cursor could disappear inside the hand, confounding the user.

520 Figure 5 illustrates the environment where tests were conducted and high-

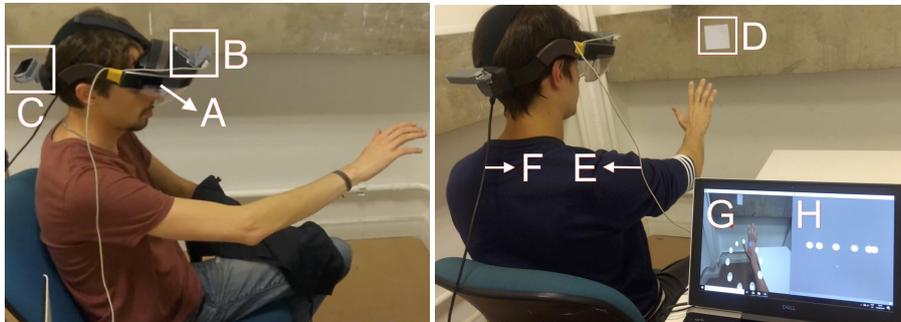


Figure 5: The experiment's setting. The Meta 2 OST-HMD (A) was adapted to include two LMCs: the first one (B) is responsible to track the participants' hand and the second one (C) was added to act as counterweight. A paper square (D) was glued to the wall and was used as reference to calibrate the Meta 2 anchor to the physical world. A USB cable (E) and a USB and HDMI cables (F) connected (B) and (A), respectively, to the computer. The Meta 2 Viewer (G) displayed the augmented world. The 3DFitts software (H) controlled the tests. The desk on participants' right side was used as support for the mouse.

lights the key elements of the apparatus. The LMC was fixed to AR headset using a 3D-printed support glued to its case with a double-sided tape. The added weight increased the feeling of discomfort in the forehead and, to mitigate this issue, another support with an LMC was glued in back of the headset to act as counterweight.

The Independent Rating Scale questionnaire from Annex C of ISO9241-411 was applied to measure participants' subjective perception on interaction quality. It evaluates participants' perceived comfort and effort for a interaction experience and is comprised by twelve 7-point Likert scale items.

6.1.3. Design

The experiment was fully within-subjects with four independent variables:

- Test Condition: { MC, MD, LMF, LMH, MTF, MTH } (*see Table 3*)
- Amplitude: { 160, 230, 300 } mm
- Width: { 12, 32 } mm
- Session: { 1, 2 }

To keep the design fully crossed, the Test Condition independent variable was defined. Table 3 presents its values and their relation to the input device, plane orientation and selection method. Amplitudes and widths were chosen to

Test Condition	Device	Test Plane	Selection Method
MC	Mouse	Frontal	Click (Left Button)
MD	Mouse	Frontal	Dwell Time (300 ms)
LMF	LMC	Frontal	Dwell Time (300 ms)
LMH	LMC	Horizontal	Dwell Time (300 ms)
MTF	MetaTracker	Frontal	Dwell Time (300 ms)
MTH	MetaTracker	Horizontal	Dwell Time (300 ms)

Table 3: Nomenclature and description for each test condition.

allow interaction in the biggest volumetric space possible without requiring head
540 or body movement, while keeping the resulting *IDs* proportional to the ones
from User Study 1. Session variable was added to collect repeated measures and
improve statistical analysis. Preliminary tests indicated that one entire session
would last about 50 minutes and make participants feel tired and demotivated,
so we decided to collect only two sessions. The dependent variables were error
545 rate, movement time and throughput.

Each test condition was performed with 11 targets, generating 11 trials/sequence.
Test order followed a 6x6 balanced latin squared design, with participants be-
ing randomly attributed to one of six groups. Two groups received only one
participant due to the absence of two participants. A total of 7920 trials were
550 collected (11 targets * 3 amplitudes * 2 widths * 2 session * 6 test conditions *
10 participants).

6.1.4. Procedure

Participants were welcomed, received an overall explanation of the experi-
ment and were asked to use their correction glasses every time they used the
555 AR headset if they had vision problems.

After this introduction, a calibration phase took place. The Meta 2 optical
calibration process was executed to make the headset display suit each partici-
pants' vision. Once finished, the 3DFitts software was initialized.

One of the challenges in evaluating interaction in AR is to guarantee that the
560 mapped space and visual background is alike for every participant. To reduce
such errors, the mapping process of the Meta 2 headset was initialized with
participants looking at a fixed point in the wall (Figure 5-D). Subsequently, the
LMC was also calibrated to adjust the orientation of its virtual hand model to
participants' hand.

565 The spatial position of the test plane was calibrated at the beginning of
every test condition to reduce performance variability between test conditions
and participants'. This was done by asking participants' to place their right
hand at a fixed distance from the headset (point A in Figure 6) and capturing

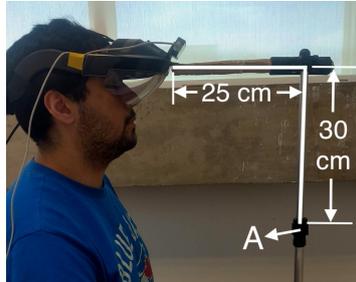


Figure 6: The desired position of the center of the test plane was defined to be on point A. The adapted banner stand allowed fast recalibration between test conditions.

its position using the MetaTracker.

570 Afterwards, participants' performed a practice session with test conditions MTF and LMH. This assured they became familiar to the hand trackers and plane orientations. The formal tests began and each test condition was evaluated following participants' group order. Participants were told which device, plane and selection method they would test. The test sequences were always presented
575 from lowest to highest *ID*. Optional breaks were allowed between sequences, but during a sequence participants were asked to be focused.

After finishing a test condition, the headset was removed and participants were asked to fill the comfort questionnaire related to the test they just concluded. Questionnaires were applied for both sessions and for all test conditions
580 but MC, resulting in 10 questionnaires/participant. When filling the second questionnaire for a test condition, participants were asked to used their first one as reference to improve their final judgment.

Mandatory breaks of two minutes between conditions and five minutes between sessions were enforced to mitigate the fatigue and discomfort caused by
585 prolonged hand interaction and headset usage. Participants were encouraged to extend the rest period if feeling tired.

All participants used their right hand for interaction. To avoid tracking error, the left hand was hidden from the hand trackers using a piece of cloth. The average duration of the experiment was 108 minutes per participant.

590 6.2. Results and discussion

Movement time or effective amplitude for 91 trials (1.2%) deviated more than 3 standard deviations from their respective mean, which was computed across all sessions for each task condition and for each participant. They were considered outliers and were removed from further analysis. The remaining trials
595 were used to compute MT , ER and TP for each test sequence, resulting in 720 data points (6 test sequences * 2 sessions * 6 test conditions * 10 participants) that were used in the following analysis.

6.2.1. Fitts' law models

The same procedure from User Study 1 was used to model participants' interaction performance for each test condition. The models' data points are
600 listed in Table 4. The regression lines (Figure 7) suggest that device strongly impacted participants' performance, with mouse presenting the best performance and MetaTracker, the worst. Plane orientation and selection method had smaller effects in the model's response.

605 The y-intercepts for the hand tracking devices were negative. Negative values are common in literature (Guiard and Olafsdottir, 2011) and have been object of intense debate. Soukoreff and MacKenzie (2004) suggested that values outside the range (-200 ms, 400 ms) might indicate some flaw with the methodology or data collection. Guiard and Olafsdottir (2011) rebated this argument, stating that the y-intercept does not carry any meaning per se and its
610 value simply reflects the overall result of the regression. We agree with the later view: negative values will arise for any interaction where performance degrades fast enough with increasing task difficulty, for a given ID range. In another words, the higher the slope of the model, the lower the linear coefficient. MetaTracker performance models offer a remarkable example: its slope values were
615 about four times greater than the mouse, pushing its y-intercept below -1000 ms.

The linear models explained more than 90% of the variability of the data for all six models ($R^2 > 0.9$), suggesting that Fitts' Law still holds for direct

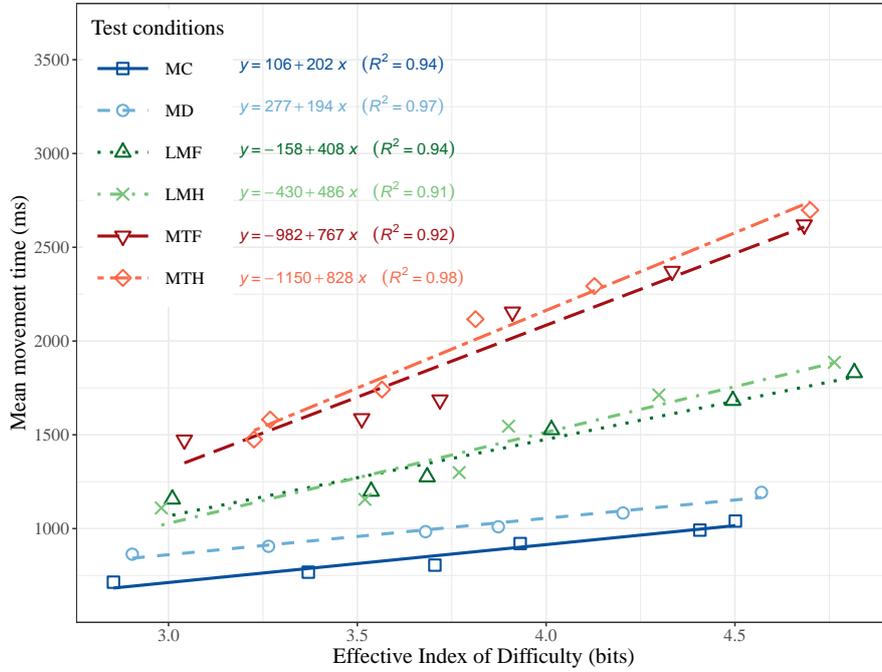


Figure 7: Regression lines for Fitts' Law models obtained for each test condition.

Test Sequence		Test Condition													
		MC		MD		LMF		LMH		MTF		MTH			
A (mm)	W (mm)	ID (bits)	IDe (bits)	MT (ms)											
160	32	2.58	2.85	714	2.90	864	3.05	1157	3.00	1110	3.01	1473	3.18	1475	
230	32	3.03	3.40	768	3.26	906	3.56	1199	3.55	1156	3.51	1587	3.26	1581	
300	32	3.38	3.73	805	3.70	984	3.71	1276	3.80	1299	3.72	1687	3.57	1742	
160	12	3.84	3.95	920	3.89	1010	4.04	1527	3.92	1546	3.93	2155	3.82	2116	
230	12	4.33	4.43	992	4.22	1083	4.51	1683	4.32	1713	4.35	2372	4.14	2293	
300	12	4.70	4.52	1040	4.59	1193	4.84	1831	4.78	1887	4.70	2620	4.71	2698	

Table 4: Data points used to model test conditions according to Fitts' Law. Each data point was obtained by averaging MT , W_e and A_e across 20 test sequences (2 sessions * 10 participants).

620 3D hand interaction in an AR OST-HMD when targets are positioned in the
frontal or horizontal planes.

6.2.2. Dependent variables

Figure 8 shows the grand means for the dependent variables for each test con-
dition. They were computed first for each participant, across all test sequences
625 and sessions, and then across all participants.

Effect of selection method. The dwell time technique had a negative impact
on the measured performance, as it can be seen by comparing MC and MD
conditions: MT for MC was 13% lower than for MD and this difference was
statistically significant ($F_{1,9} = 24.6, p < .001$). TP for MC was 18% higher than
630 for MD, which is also a statistically significant difference ($F_{1,9} = 52.5, p < .001$).

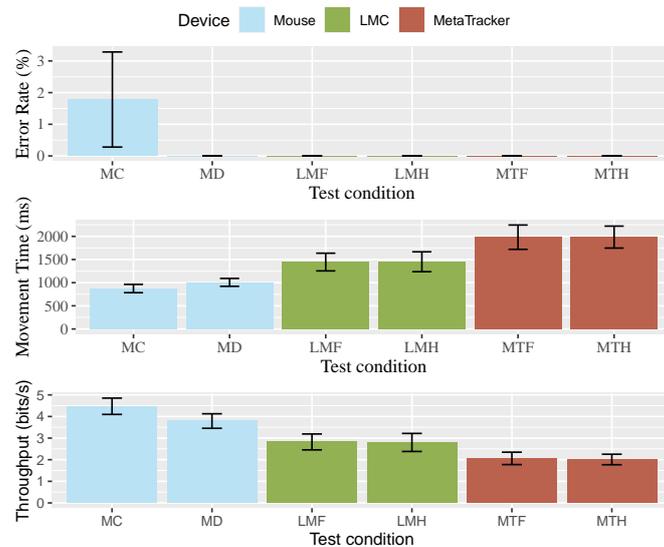


Figure 8: Error rate, mean movement time and throughput for each test condition in User Study 2. Error bars represent one standard deviation. ER was always zero for the dwell time technique because it only activated selection if the cursor stayed inside the target for the required time.

Effect of plane orientation. Results from LMF, LMH, MTF and MTH conditions revealed that plane orientation did not impact participants' performance significantly. Mean MT was 1714 ms (SD = 205 ms) for the frontal plane and 1718 ms (SD = 212 ms) for the horizontal plane. The difference was not statistically significant ($F_{1,9} = 0.01, ns$). Mean TP was 2.44 bits/s (SD = 0.30 bits/s) for the frontal plane and 2.41 bits/s (SD = 0.30 bits/s) for the horizontal plane, also a not statistically significant difference ($F_{1,9} = 0.53, ns$).

Effect of device. Grand means for TP were 4.13 bits/s (0.49 bits/s) for mouse, 2.81 bits/s (SD = 0.38 bits/s) for LMC and 2.04 bits/s (SD = 0.26 bits/s) for MetaTracker. Conditions MD, LMF and MTF were compared to analyze the effect of device in the frontal plane. MT for LMC condition was 1445 ms, 43% greater than that of MD (1007 ms) and 27% lower than the observed for MTF (1982 ms). Likewise, TP for LMC was 2.82 bits/s, 26% lower than the 3.79 bits/s for MD and 37% greater than the 2.06 bits/s obtained for MTF. An analysis of variance revealed a statistically significant difference for both MT ($F_{2,18} = 97.5, p < .001$) and TP ($F_{2,18} = 96.9, p < .001$). Two Tukey's test further indicated that the differences between all devices are statistically significant for each of these dependent variables. Therefore, mouse presented the best interaction performance and MetaTracker, the worse. Besides, there was a statistically significant difference between conditions LMH and MTH, indicating that LMC was faster than MetaTracker also on horizontal plane: MT was 27% lower for LMC ($F_{1,9} = 112.7, p < .001$) and TP was 39% higher ($F_{1,9} = 60.68, p < .001$).

6.2.3. Questionnaires

Questionnaires were applied in the first session only to increase participants' awareness of interaction quality for each test condition. Only answers given in the second session were analyzed and are summarized in Figure 9.

Effects of test condition on score were analyzed for each question using and Friedman test and were statistically significant for questions 1 to 11 ($p < 0.05$).

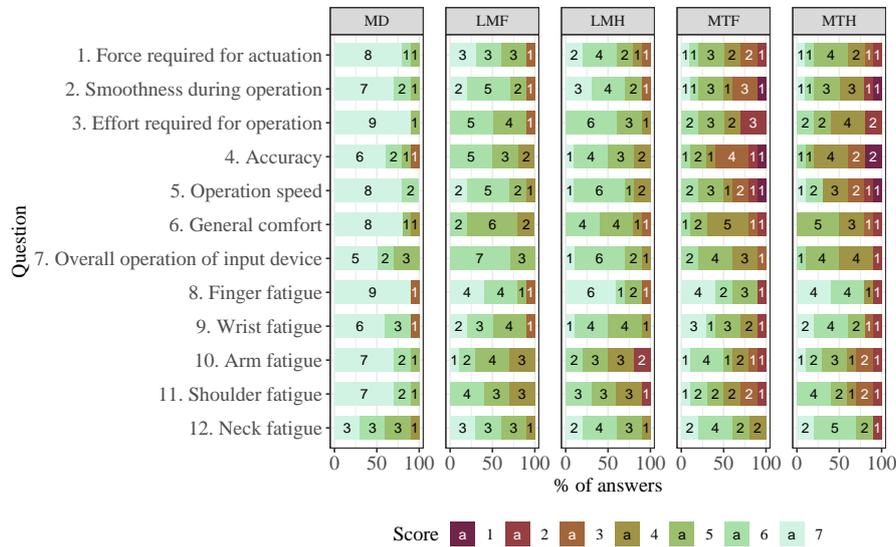


Figure 9: Answers for the comfort questionnaires applied during the second session. Questionnaires were not applied for MC condition. The number of participants who rated a question with a specific score is indicated inside the bars.

660 A post-hoc analysis was performed using Conover test with Bonferroni correction. Score differences between pair (MD, MTF) were statistically significant for questions 1 to 11. Between (MD, MTH), it was significant for questions 1 to 7, 10 and 11. The only significant effects found for all other pairs were: question 4 on (LMH, MTF) and (LMH, MTH); question 7 on (LMF, MTH) and (LMH, MTH); question 10 on (MD, LMH).

665

These results indicate that the mouse was clearly perceived as a better device than the MetaTracker. On the other hand, although LMC received consistently lower scores than the mouse for almost every condition and aspect studied, the statistical test did not have enough power to accuse significance except for “arm fatigue”. The same applies to MetaTracker and LMC: the first was consistently rated lower than the second, but significant differences were only observed for “accuracy” and “overall operation of input device”.

670

The physical effort demanded by hand tracking interaction is evidenced by

the significantly lower scores for arm and shoulder fatigue in relation to the
 675 mouse. Furthermore, MetaTracker accuracy was clearly perceived as the worst.
 In fact, during the tests some reported that it was laggy and tracking failures
 were common.

Comparisons for pairs (LMF, LMH) and (MTF, MTH) resulted in $p = 1.00$
 for all questions, which is a strong evidence that plane orientation had negligible
 680 effects on participants' perception of interaction quality and corroborates the
 objective results presented in Section 6.2.2.

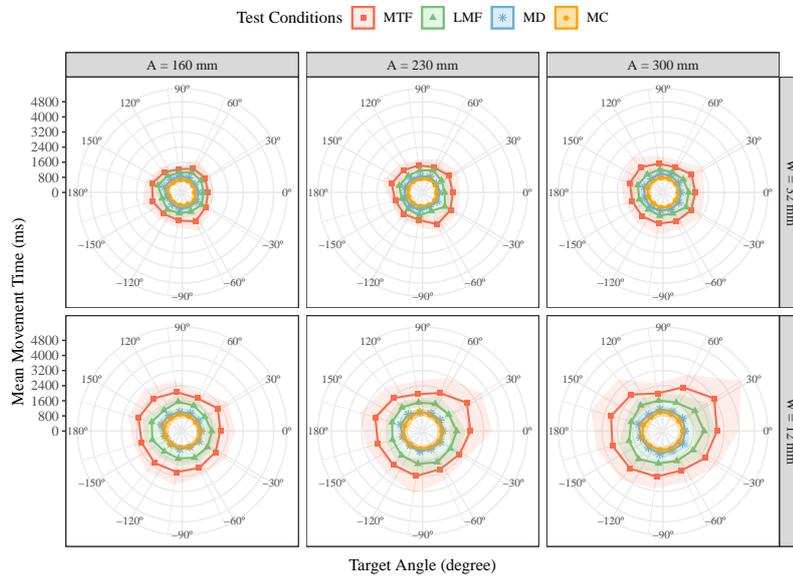
When performing ANOVA and Tukey's test, results indicate statistical sig-
 nificance at $\alpha = 0.05$ for all pairwise comparisons between mouse and Meta-
 Tracker conditions on questions 1 to 11; between mouse and LMC conditions
 685 on questions 6, 10 and 11; and between all pairs for LMC and MetaTracker
 conditions for questions 2 to 7, and for some pairs in questions 1, 10 and 11.
 These results feel more aligned to the data presented in Figure 9 and suggest
 that the scores may be interpreted as interval data. Under this analysis, it is
 possible to state that interaction comfort was clearly perceived to be the best
 690 with mouse and the worst with MetaTracker, with LMC in between.

6.2.4. *Impact of movement direction on movement time*

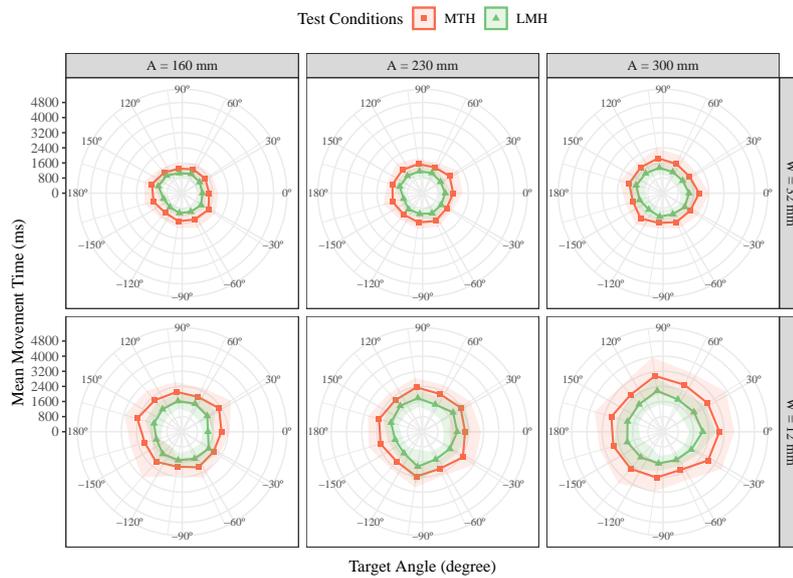
We analyzed how targets' angle affected the time to acquire a target. The
 mean movement time of all valid trials arriving at a specific target were com-
 puted across all sessions and participants, grouped by test condition and test
 695 sequence, and are illustrated in Figure 10.

A visual inspection of Figure 10a suggest there is an effect of target po-
 sition on movement time for MetaTracker, with lateral targets getting harder
 to select as task difficulty increases and top targets presenting lower selection
 times. However, the variability of our data does not allow us to draw any strong
 700 conclusion on the subject. In fact, a within-subjects ANOVA analysis of each
 one of the 36 test sequences did not reveal any statistically significant effects of
 target position on movement time.

Based on these results, we strongly recommend that any work interested in



(a) Frontal plane



(b) Horizontal plane

Figure 10: Mean movement time to reach a specific target. The shaded area around the lines represent the standard deviation. Imagining (a) on a wall and (b) on a tabletop helps interpreting the results.

characterizing the spatial response of visual trackers should use higher number
 705 of sessions and participants in order to increase the total number of samples
 collected for each target. Even so, eventual significant effects must be ana-
 lyzed with care because they might arise not only due to tracker’s response
 but also due to perceptual constraints of stereo displays, such as the vergence-
 accommodation conflict (Hoffman et al., 2008). For instance, even if our data
 710 did provide statistical evidence that top targets were faster to acquire, other re-
 searchers have also found such relation when studying movement performance
 in VR systems using long-range trackers, which suggests they were caused by
 perceptual constraints of stereo displays (Lubos et al., 2014; Clark et al., 2020).
 Thus, with the present experiment design we would not be able to decouple the
 715 actual effects of display and interaction device.

6.2.5. *Distribution characteristics*

The 720 sequences, which contained between 9 and 11 trials, were analyzed
 using the same procedure from User Study 1. For movement time, mean skew-
 ness and kurtosis were, respectively, 0.56 ($SD = 0.65$) and 2.84 ($SD = 1.15$),
 720 and Shapiro-Wilk test could not reject the null hypothesis for normality for
 80% of the sequences. For selection coordinates, mean skewness was 0.07
 ($SD = 0.56$), mean kurtosis was 2.44 ($SD = 0.82$) and the null hypothesis
 for normality was not rejected for 95% of the sequences. These results are
 analogous to the ones obtained for User Study 1, suggesting that participants’
 725 behavior was maintained and follows a normal distribution for the majority of
 the collected sequences.

6.3. *Conclusion (User Study 2)*

A user study was designed to demonstrate how the method proposed in Sec-
 tion 3 and the 3DFitts software may be used as a tool to evaluate hand tracking
 730 interaction with an AR OST-HMD. The objectives were to evaluate two hand
 tracking devices (LMC and MetaTracker) with both throughput (objective met-
 ric) and questionnaires (subjective metric), and to study how plane orientation

affects performance and if hand interaction in an AR environment conforms to Fitts' Law. A mouse was also evaluated to provide external validity and improve
735 comparison between studies.

Throughput results indicated that the mouse provided the best interaction while MetaTracker performed worst, with LMC ranked in between. Overall, comfort questionnaires corroborated these results, with MetaTracker receiving significantly lower ratings than the mouse and with LMC consistently receiving
740 intermediate scores. These results suggest that the 3DFitts multi-directional tests were effective in discriminating performance differences between different test conditions in an AR environment. Moreover, they evince that LMC is a better hand tracking device than MetaTracker. Notwithstanding, mouse provided the best interaction performance and the poor performance of hand tracking
745 devices may hinder their widespread use in near future.

Because hand tracking performance usually deteriorates when increasing the distance between hand and sensor, we expected to see some impact of plane orientation on tracking performance. However, no statistically significant effect was found neither on *MT* nor on *TP*. The applied questionnaires corroborated
750 this finding, with participants rating both plane interactions with similar scores for the same input device.

Fitts' Law models for all six test conditions presented $R^2 > 0.90$, suggesting that the law holds for interactions within arms' reach with targets arranged on the frontal or horizontal planes.

755 7. General Discussion

Figure 11 summarizes throughput results for each test condition in User Studies 1 and 2. We analyzed conditions B and C to reveal what impact the AR experience had on performance. Only trials from who participated in both studies were considered. Grand means for *MT*, *TP* and *ER* were computed first
760 by participant, then across all participants. The resulting values for condition B were, respectively, 778 ms (SD = 1120 ms), 5.21 bits/s (SD = 0.57 bits/s)

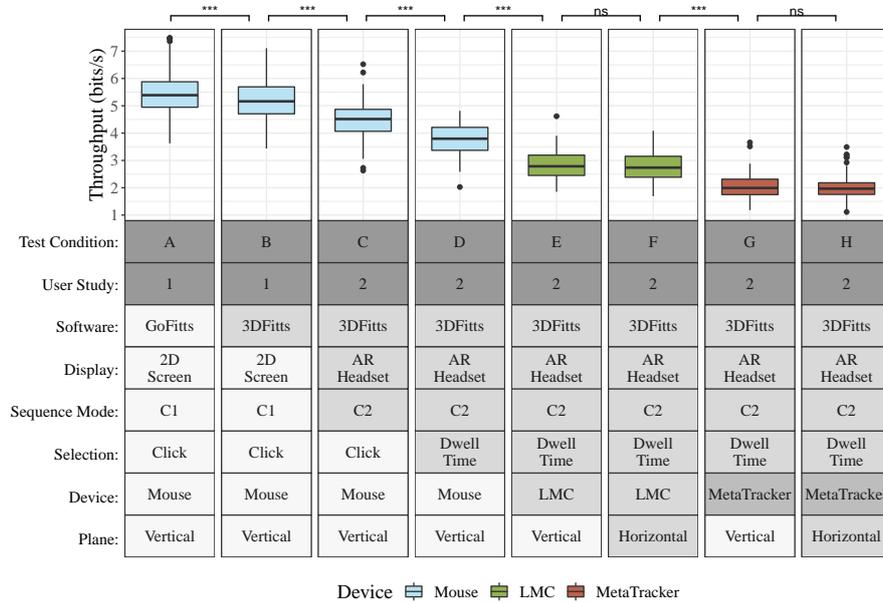


Figure 11: Throughput for every test condition on User Study 1 and 2. The values of the independent variables are displayed in different colors across test conditions to make it easier to identify what changed between the test conditions. For instance, from test condition A to B, only software changed; from B to C, only display and sequence mode; and so on and so forth. Levels of significance between consecutive conditions are shown on top. Sequence Mode variable denotes the pairs of A and W used in User Study 1 (C1) and User Study 2 (C2).

and 4.5% (SD = 4.0%), while for condition C they were 873 ms (SD = 90 ms), 4.47 bits/s (SD = 0.38 bits/s) and 1.8% (SD = 1.5%). Hence, with relation to C, condition B presented *MT* 11% lower, *TP* 17% higher and *ER* 150% higher.

765 A within-subjects ANOVA revealed the differences were statistically significant for *MT* ($F_{1,9} = 10.02, p < .05$) and *TP* ($F_{1,9} = 28.53, p < .001$), but not for *ER* ($F_{1,9} = 4.7, ns$). Therefore, user performance apparently drops about 10~15% when interacting with an AR headset. This drop might be attributed to perceptual constraints of current stereo displays, such as the vergence-accommodation

770 conflict (Hoffman et al., 2008). Nevertheless, other confounding factors might have had significant impact on performance. For instance, the second study presented greater targets' sizes and distances and was more tiresome due to its

design.

8. Summary

775 While VR systems have reached commercial maturity, immersive AR technology still struggles to stand on its own as a fluid computational system. Many past challenges have been solved, but quality of interaction still limits its applications. In particular, short-range hand trackers are usually embedded in headsets because they reportedly allow intuitive 6 DOF manipulation in the
780 virtual world. In the present work, we have studied the quality of such interaction from the perspective of user performance by conducting a formal Fitts' Law study.

Our work offers three main contributions: we proposed a method to evaluate mid-air hand interaction in immersive VR/AR systems, which may help
785 improve comparability between studies of interaction in both 2D and 3D settings; we implemented such method in the 3DFitts software and validated its core algorithm against the GoFitts program (MacKenzie, 2020), which is commonly used in the literature to evaluate 2D interaction; and we provided a use case for it by conducting an user study to evaluate interaction performance of
790 two hand tracking systems in an AR OST-HMD, providing external validity with 2D research by also evaluating a mouse. The 3DFitts is available as an open source software at <https://github.com/jlcouto/3DFitts>.

The proposed method was adapted from the standard multi-directional ISO 9241-411 test and from previous studies in 2D and 3D interaction on mono and
795 stereo displays. A mouse and two hand tracking systems were evaluated using the software and questionnaires. Participant's perception of each interaction method corroborated the objective throughput metric computed by 3DFitts, suggesting that the proposed method was able to meaningfully measure hand interaction performance in an AR environment.

800 Our results on throughput for MetaTracker (2.0 bits/s) and LMC (2.8 bits/s) add to a growing literature that has been evaluating hand interaction in both 2D

and 3D environments. Other reported values for LMC were 2 and 2.7 bits/s, but for 2D evaluation (Bachmann et al., 2014; Burno et al., 2015). Other studies also support the 2-3 bits/s throughput range for various hand tracking systems on
 805 both 2D and 3D environments (Kouroupetroglou et al., 2012; Pino et al., 2013; Lubos et al., 2015; Burno et al., 2015). Due to different evaluation methods, however, direct comparisons are not straightforward.

The intuitiveness of hand tracking interaction made it an auspicious topic of research. Yet, our results show that the present technology still lacks the
 810 accuracy and precision that made the mouse the universal interaction device it is today. Moreover, the questionnaires revealed that hand interaction was perceived as somewhat tiresome, specially for arm and shoulder. Even with ideal tracking technology, arm fatigue is probably one of the main reasons why its widespread adoption is not likely to occur (Özacar et al., 2017). Nevertheless,
 815 improving hand tracking technology is a worthy challenge per se and may usher in innovative ways of interacting with computers.

8.1. Key Findings

- Mouse’s throughput in an AR headset was 14% lower than in equivalent 2D interaction;
- 820 • Mouse’s throughput using dwell time as selection method was 15% lower than using button click for interaction in an AR environment;
- Frontal and horizontal planes had negligible impact on hand tracking interaction performance for both LMC and MetaTracker;
- Throughput for the LMC and for the MetaTracker are respectively 26%
 825 and 47% lower than the mouse in an AR environment;
- Mid-air interaction in an AR headset conformed to Fitts’ Law.

8.2. Future Works

One of the arguments for testing multiple planes when evaluating a hand tracker device is to allow characterization of its spatial response, which would

830 aid engineers to think about ways to improve the underlying devices' technology
and inform interface designers which spatial regions to avoid when developing
softwares to use with it. However, as discussed in Section 6.2.4, display deficien-
cies complicates this analysis. Our experiment design does not allow to pinpoint
the effect that input device has on movement direction. One possible solution
835 for this problem is to replicate the AR study using real targets and no display.
This could be done only once to evaluate the spatial response of a specific 3D
input device that could become the standard to be used in all future evaluations
of 3D interaction in VR/AR headsets.

Additionally, two-hands interaction in VR/AR systems seem to be an over-
840 looked aspect in HCI studies. Designing a solid experimental method to evaluate
simultaneous interaction with both hands would help humanity to get one step
closer to Sutherland's Ultimate Display.

9. Acknowledgements

We sincerely thank Professor Scott Mackenzie for kindly providing us the
845 GoFitts source code and for the useful documentation made available on his
website (MacKenzie, 2020), both of which were used as reference to develop
the 3DFitts' core algorithm. This work was supported by the Interdisciplinary
Center in Interactive Technologies from University of São Paulo (CITI-USP) and
was funded by the Brazilian National Council for Scientific and Technological
850 Development (CNPq) [grant 167205/2017-0].

10. Declarations of interest

None.

References

Al-Shamayleh, A.S., Ahmad, R., Abushariah, M.A.M., Alam, K.A., Jomhari,
855 N., 2018. A systematic literature review on vision based gesture recognition

- techniques. *Multimedia Tools and Applications* 77, 28121–28184. doi:10.1007/s11042-018-5971-z.
- Atsuo Murata, Hirokazu Iwase, 2001. Extending Fitts' law to a three-dimensional pointing task. *Human Movement Science* 20, 791–805. doi:10.1016/S0167-9457(01)00058-6.
- 860 Azuma, R.T., 1997. A Survey of Augmented Reality. *Presence: Teleoper. Virtual Environ.* 6, 355–385. doi:10.1162/pres.1997.6.4.355.
- Bachmann, D., Weichert, F., Rinkeauer, G., 2014. Evaluation of the leap motion controller as a new contact-free pointing device. *Sensors (Basel, Switzerland)* 15, 214–233. doi:10.3390/s150100214.
- 865 Barrera Machuca, M.D., Stuerzlinger, W., 2019. The Effect of Stereo Display Deficiencies on Virtual Hand Pointing, in: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, Glasgow, Scotland Uk. pp. 1–14. doi:10.1145/3290605.3300437.
- 870 Bérard, F., Ip, J., Benovoy, M., El-Shimy, D., Blum, J.R., Cooperstock, J.R., 2009. Did “Minority Report” Get It Wrong? Superiority of the Mouse over 3D Input Devices in a 3D Placement Task, in: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (Eds.), *Human-Computer Interaction – INTERACT 2009*, Springer Berlin Heidelberg. pp. 400–414.
- 875 Billinghamurst, M., Clark, A., Lee, G., 2015. A Survey of Augmented Reality. *Found. Trends Hum.-Comput. Interact.* 8, 73–272. doi:10.1561/1100000049.
- Borg, G., 1998. *Borg's Perceived Exertion and Pain Scales*. Borg's Perceived Exertion and Pain Scales, Human Kinetics, Champaign, IL, US.
- 880 Brooke, J., 1996. SUS: A “quick and dirty” usability. *Usability evaluation in industry* , 189.

- Bruder, G., Steinicke, F., Stürzlinger, W., 2013. Effects of visual conflicts on 3D selection task performance in stereoscopic display environments, in: 2013 IEEE Symposium on 3D User Interfaces (3DUI), pp. 115–118. doi:10.1109/3DUI.2013.6550207.
- 885
- Burno, R.A., Wu, B., Doherty, R., Colett, H., Elnaggar, R., 2015. Applying Fitts' Law to Gesture Based Computer Interactions. *Procedia Manufacturing* 3, 4342–4349. doi:10.1016/j.promfg.2015.07.429.
- Cha, Y., Myung, R., 2013. Extended Fitts' law for 3D pointing tasks using 3D target arrangements. *International Journal of Industrial Ergonomics* 43, 350–355. doi:10.1016/j.ergon.2013.05.005.
- 890
- Clark, L.D., Bhagat, A.B., Riggs, S.L., 2020. Extending Fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies* 139, 102413. doi:10.1016/j.ijhcs.2020.102413.
- 895
- DeLong, S., MacKenzie, I.S., 2018. Evaluating Devices for Object Rotation in 3D, in: Antona, M., Stephanidis, C. (Eds.), *Universal Access in Human-Computer Interaction. Methods, Technologies, and Users*. Springer International Publishing, Cham. volume 10907, pp. 160–177. doi:10.1007/978-3-319-92049-8_12.
- 900
- Esteves, A., Shin, Y., Oakley, I., 2020. Comparing selection mechanisms for gaze input techniques in head-mounted displays. *International Journal of Human-Computer Studies* 139, 102414. doi:10.1016/j.ijhcs.2020.102414.
- Fitts, P.M., 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47, 381–391.
- 905
- Grossman, T., Balakrishnan, R., 2004. Pointing at trivariate targets in 3D environments, in: *Proceedings of the SIGCHI Conference on Human Factors in*

- Computing Systems, Association for Computing Machinery, Vienna, Austria.
 910 pp. 447–454. doi:10.1145/985692.985749.
- Guiard, Y., Beaudouin-Lafon, M., 2004. Fitts' law 50 years later: Applications and contributions from human–computer interaction. *International Journal of Human-Computer Studies* 61, 747–750. doi:10.1016/j.ijhcs.2004.09.003.
- Guiard, Y., Olafsdottir, H.B., 2011. On the Measurement of Movement Difficulty in the Standard Approach to Fitts' Law. *PLoS ONE* 6, e24389.
 915 doi:10.1371/journal.pone.0024389.
- Hansen, J.P., Rajanna, V., MacKenzie, I.S., Bækgaard, P., 2018. A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display, in: *Proceedings of the Workshop on Communication by Gaze Interaction - COGAIN '18*, ACM Press, Warsaw, Poland. pp. 1–5.
 920 doi:10.1145/3206343.3206344.
- Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, in: Hancock, P.A., Meshkati, N. (Eds.), *Advances in Psychology*. North-Holland. volume 52
 925 of *Human Mental Workload*, pp. 139–183. doi:10.1016/S0166-4115(08)62386-9.
- Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S., 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision* 8, 33.1–3330. doi:10.1167/8.3.33.
- 930 International Organization for Standardization, 2000. ISO 9241-9:2000 - Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) - Part 9: Requirements for on-Keyboard Input Devices. Technical Report. International Organisation for Standardisation. Geneva.
- International Organization for Standardization, 2012. ISO/TS 9241-411 - Ergonomics of Human-System Interaction - Part 411: Evaluation Methods for
 935

the Design of Physical Input Devices. Technical Report. International Organisation for Standardisation. Switzerland.

- Kouroupetroglou, G., Pino, A., Balmpakakis, A., Chalastanis, D., Golematis, V., Nikolaos Ioannou, Ioannis Koutsoumpas, 2012. Using Wiimote for 2D and 3D Pointing Tasks: Gesture Performance Evaluation, in: Efthimiou, E., Kouroupetroglou, G., Fotinea, S.E. (Eds.), *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*, Springer Berlin Heidelberg. pp. 13–23.
- LaViola, J.J., 2013. 3D Gestural Interaction: The State of the Field. ISRN Artificial Intelligence 2013, 1–18. doi:10.1155/2013/514641.
- Lubos, P., Bruder, G., Steinicke, F., 2014. Analysis of direct selection in head-mounted display environments, in: 2014 IEEE Symposium on 3D User Interfaces (3DUI), pp. 11–18. doi:10.1109/3DUI.2014.6798834.
- Lubos, P., Bruder, G., Steinicke, F., 2015. Influence of Comfort on 3D Selection Task Performance in Immersive Desktop Setups. *Journal of Virtual Reality and Broadcasting* 12, 15.
- MacKenzie, I.S., 2015. Fitts' Throughput and the Remarkable Case of Touch-Based Target Selection, in: Kurosu, M. (Ed.), *Human-Computer Interaction: Interaction Technologies*. Springer International Publishing, Cham. volume 9170, pp. 238–249. doi:10.1007/978-3-319-20916-6_23.
- MacKenzie, I.S., 2020. GoFitts. URL: <http://www.yorku.ca/mack/FittsLawSoftware/doc/index.html?GoFitts.html>.
- Magic Leap, 2020. Magic Leap 1. URL: <https://www.magicleap.com/magic-leap-1>.
- Matney, L., 2016. Hands-on with the \$949 mind-bending Meta 2 augmented reality headset — TechCrunch. URL: <https://techcrunch.com/2016/03/02/hands-on-with-the-949-mind-bending-meta-2-augmented-reality-headset/>.

- Microsoft, 2020. HoloLens 2—Overview, Features, and Specs — Microsoft HoloLens. URL: <https://www.microsoft.com/en-us/hololens/hardware>.
- 965 Özacar, K., Hincapié-Ramos, J.D., Takashima, K., Kitamura, Y., 2017. 3D Selection Techniques for Mobile Augmented Reality Head-Mounted Displays. *Interacting with Computers* 29, 579–591. doi:10.1093/iwc/iww035.
- Pino, A., Tzemis, E., Ioannou, N., Kouroupetroglou, G., 2013. Using Kinect for 2D and 3D Pointing Tasks: Performance Evaluation, in: Kurosu, M. (Ed.), *Human-Computer Interaction. Interaction Modalities and Techniques*, 970 Springer Berlin Heidelberg. pp. 358–367.
- Raynal, M., Dubois, E., Schmitt, B., 2013. Towards Unification for Pointing Task Evaluation in 3D Desktop Virtual Environment, in: Holzinger, A., Ziefle, M., Hitz, M., Debevc, M. (Eds.), *Human Factors in Computing and Informatics*, Springer Berlin Heidelberg. pp. 562–580. 975
- Sambrooks, L., Wilkinson, B., 2013. Comparison of Gestural, Touch, and Mouse Interaction with Fitts’ Law, in: *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, ACM, New York, NY, USA. pp. 119–122. doi:10.1145/2541016. 980 2541066.
- Soukoreff, R.W., MacKenzie, I.S., 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts’ law research in HCI. *International Journal of Human-Computer Studies* 61, 751–789. doi:10.1016/j.ijhcs.2004.09.001.
- 985 Stuerzlinger, W., Teather, R.J., 2014. Considerations for targets in 3D pointing experiments, in: *Proceedings of HCI Korea*, Hanbit Media, Inc., Seoul, Republic of Korea. pp. 162–168.
- Sutherland, I.E., 1965. The Ultimate Display, in: *Proceedings of the Congress of the International Federation of Information Processing (IFIP)*, pp. 506–508.

- 990 Teather, R.J., 2013. Evaluating 3D Pointing Techniques. Ph.D. thesis. York University. Toronto, Ontario.
- Teather, R.J., Natapov, D., Jenkin, M., 2010. Evaluating haptic feedback in virtual environments using ISO 9241-9, in: 2010 IEEE Virtual Reality Conference (VR), pp. 307-308. doi:10.1109/VR.2010.5444753.
- 995 Teather, R.J., Pavlovych, A., Stuerzlinger, W., MacKenzie, I.S., 2009. Effects of tracking technology, latency, and spatial jitter on object movement, in: 2009 IEEE Symposium on 3D User Interfaces, pp. 43-50. doi:10.1109/3DUI.2009.4811204.
- Teather, R.J., Stuerzlinger, W., 2008. Assessing the Effects of Orientation and Device on (Constrained) 3D Movement Techniques, in: 2008 IEEE
1000 Symposium on 3D User Interfaces, IEEE, Reno, NV, USA. pp. 43-50. doi:10.1109/3DUI.2008.4476590.
- Teather, R.J., Stuerzlinger, W., 2011. Pointing at 3D targets in a stereo head-tracked virtual environment, in: 2011 IEEE Symposium on 3D User Interfaces
1005 (3DUI), pp. 87-94. doi:10.1109/3DUI.2011.5759222.
- Teather, R.J., Stuerzlinger, W., 2013. Pointing at 3D Target Projections with One-eyed and Stereo Cursors, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 159-168. doi:10.1145/2470654.2470677.
- 1010 Teather, R.J., Stuerzlinger, W., 2014. Visual aids in 3D point selection experiments, in: Proceedings of the 2nd ACM Symposium on Spatial User Interaction, Association for Computing Machinery, Honolulu, Hawaii, USA. pp. 127-136. doi:10.1145/2659766.2659770.
- Ultraleap, 2020. Tracking — Leap Motion Controller — Ultraleap. URL: <https://www.ultraleap.com/product/leap-motion-controller/>.
1015

- Van Krevelen, R., Poelman, R., 2010. A Survey of Augmented Reality Technologies, Applications and Limitations. *International Journal of Virtual Reality* (ISSN 1081-1451) 9, 1.
- Ware, C., Lowther, K., 1997. Selection Using a One-eyed Cursor in a Fish Tank VR Environment. *ACM Trans. Comput.-Hum. Interact.* 4, 309–322. doi:10.1145/267135.267136.
- Young, T.S., Teather, R.J., MacKenzie, I.S., 2017. An arm-mounted inertial controller for 6DOF input: Design and evaluation, in: 2017 IEEE Symposium on 3D User Interfaces (3DUI), IEEE, Los Angeles, CA, USA. pp. 26–35. doi:10.1109/3DUI.2017.7893314.
- Zhai, S., 1998. User Performance in Relation to 3D Input Device Design. *SIGGRAPH Comput. Graph.* 32, 50–54. doi:10.1145/307710.307728.
- Zhai, S., Buxton, W., Milgram, P., 1994. The “Silk Cursor”: Investigating transparency for 3D target acquisition, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94, ACM Press, Boston, Massachusetts, United States. pp. 459–464. doi:10.1145/191666.191822.
- Zhai, S., Milgram, P., 1998. Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM Press/Addison-Wesley Publishing Co., Los Angeles, California, USA. pp. 320–327. doi:10.1145/274644.274689.

11. Vitae



1040

1045

John Lennon Couto is currently pursuing a master's degree in Electrical Engineering at the Polytechnic School of the University of São Paulo (EP-USP), Brazil, under the supervision of Professor Roseli de Deus Lopes. He received his bachelor's degree in Control and Automation Engineering from the Federal University of Minas Gerais (UFMG), Brazil, in 2013.

During his undergraduate course, he participated as apprentice in four scientific researches in three main areas: biomedical engineering, evolutionary algorithms and fault-tolerant electronic systems. He has a technical degree in Electronics from the Federal Center for Technological Education of Minas Gerais (CEFET-MG), Brazil.

1050



1055

Marcelo Archanjo José received the bachelor's degree in Electrical Engineering from the São Judas Tadeu University, Brazil, in 1997, the master's degree in Electrical Engineering in 2008 and the doctor's degree in 2014 from the Polytechnic School of the University of São Paulo (EP-USP), Brazil. He is a researcher with an academic and industry carrier. He is specialized mainly in: human-computer interaction, computer graphics, image processing, stereo

vision, stereo matching, 3-D reconstruction, computer vision, accessibility, machine learning, internet of things, and games. He has been a member of the IEEE since 2007.

1060

1065



1070

1075

Roseli de Deus Lopes is an Associate Professor at the Electronic Systems Department from the Polytechnic School (EP) of the University of São Paulo (USP). She received her bachelor's, master's and doctorate's degrees in Electrical Engineering from EP-USP. She is a researcher at the Laboratory for Integrated Systems (LSI), where she is a principal investigator of the Interactive Electronic Media research group, which studies computer graphics, techniques and devices for HCI, virtual reality and augmented reality). She is the vice-chair of the Interdisciplinary Center in Interactive Technologies (CITI) and the vice-president of the Institute of Advanced Studies (IEA) of USP.