

4 Estudo de Caso: Environ

De forma a avaliar a usabilidade da v-Glove elaboramos um estudo de caso exploratório baseado no programa Environ [Raposo et al 2009]. O Environ é uma ferramenta de visualização 3D de modelos CAD voltada para atividades de engenharia na indústria de petróleo e gás natural e foi desenvolvido pelo Grupo de Tecnologia em Computação Gráfica (Tecgraf) da PUC-Rio. O Environ foi concebido para ser utilizado tanto como aplicação desktop para um único usuário (Figura 4.1) como em modo imersivo para ambientes de realidade virtual 3D (conforme pode ser visto na Figura 3.12, que mostra a Cave da Petrobras).

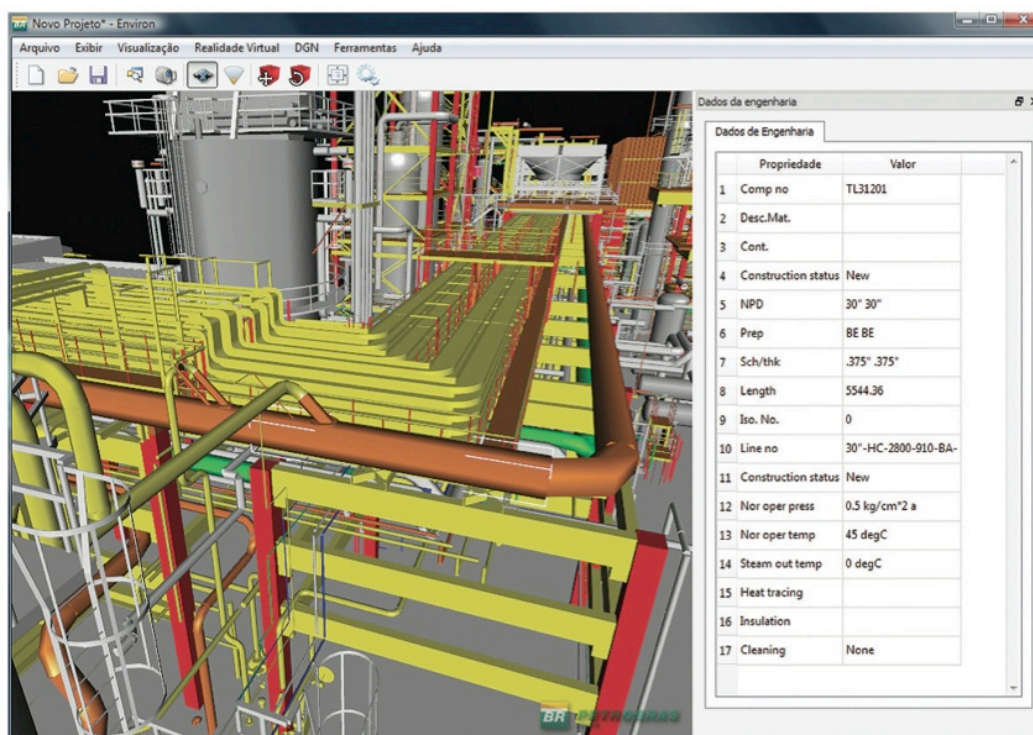


Figura 4.1: Tela principal do Environ em modo desktop [Raposo et al 2009]

E foi justamente por já prover um suporte ao modo imersivo, juntamente com o fato de já ser uma aplicação estável e com um número elevado de usuários que a usam diariamente, que o Environ foi escolhido para o estudo de caso deste trabalho. Espera-se com a utilização de uma aplicação já

consolidada a obtenção de resultados significativos que possam trazer contribuições reais a esse grupo de usuários no futuro.

Por ter sido uma aplicação concebida inicialmente para uso em desktops, o Environ ainda não foi totalmente adaptado para ambientes imersivos. Apesar de já contemplar suporte a visualização estereoscópica que permite sua utilização em ambientes de realidade virtual, a interação do usuário com o software é feita através de menus e controles baseados no padrão WIMP, como pode ser visto na imagem da Figura 4.2.

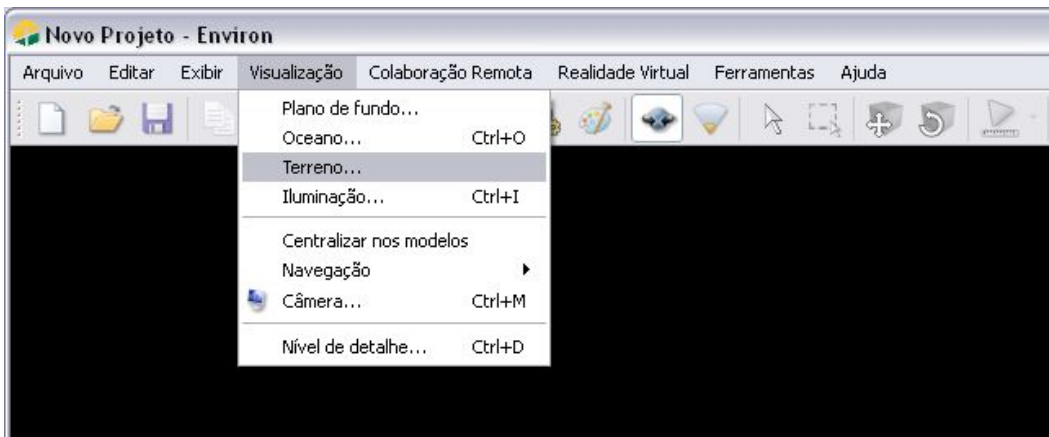


Figura 4.2: Interface de menus do Environ no padrão WIMP [Raposo et al 2009]

Por isso foi necessária a adaptação do sistema através da criação de novos menus e *widgets* com um formato mais adequado para a interação com a v-Glove. Essa adaptação ocorreu apenas nos componentes relevantes para os testes realizados durante este trabalho, uma vez que não havia tempo suficiente para um reprojeto mais abrangente de toda a interface do Environ. Mais detalhes sobre as adaptações feitas serão discutidos na seção 4.1.

Limitaremos nessa prova de conceito o escopo do Environ às seguintes tarefas:

- Navegação em modelos CAD
- Seleção de objetos em modelos CAD
- Navegação e seleção de opções em menus da aplicação
- Configurações e ajustes do programa

4.1. Adaptações no Environ

Com a finalidade de adaptar a utilização do Environ para a v-Glove foi criada uma nova interface de menus e *widgets* exclusivos para os testes. Além disso, foi feita uma adaptação no formato do cursor do mouse exibido durante a interação com os menus da aplicação.

4.1.1. Interface de menus e *widgets*

Foi feito inicialmente um levantamento dos componentes de interface gráfica utilizados no Environ. Esta listagem dos componentes foi então classificada de acordo com sua utilização na aplicação, sendo assim priorizados os componentes mais utilizados (Tabela 4.1). A partir destes componentes foram selecionadas as tarefas que fariam parte dos testes com os usuários da v-Glove e relacionados os componentes que deveriam sofrer algum tipo de adaptação.

Prioridade	Componente	Tipo
1	Botão	Simple
2	Checkbox	Simple
3	Slider	Simple
4	Combobox	Simple
5	Textbox	Simple
6	Menu	Composto
7	Integer Textbox	Simple
8	Toolbar	Composto
9	File Open	Composto
10	File Save	Composto
11	Radio Button	Simple
12	Color selection box	Composto

Tabela 4.1: Classificação dos componentes de interface gráfica do Environ

Com base nessa análise foram selecionados os cinco componentes com maior prioridade para terem suas interfaces adaptadas à utilização com a v-

Glove. A Figura 4.3 mostra como ficaram os componentes após a adaptação: em (a) pode ser observada a *checkbox*, em (b) o *slider*, (c) mostra a *combobox* e (d) a *textbox* utilizada em conjunto com o botão para prover a funcionalidade de entrada numérica.



Figura 4.3: *Widgets* criados neste trabalho

4.1.2. Cursor do mouse

Como complemento ao recurso de feedback tátil, foi realizada uma alteração no comportamento do cursor do mouse quando o usuário interage com o sistema. Dependendo da localização da v-Glove em relação à interface de toque, a cor do cursor na tela é alterada. A Figura 4.4 mostra as cores possíveis do cursor.



Figura 4.4: Variação das cores do cursor

O cursor permanece branco enquanto a luva encontra-se fora da área de cobertura da câmera responsável pelo eixo Z ou quando estiver a uma distância de mais de 400 pixels da área de toque. A cor muda para amarelo quando a distância entre a luva e a superfície de toque for maior que 200 e menor que 400 pixels, e muda para laranja quando essa distância for inferior a 200 pixels. Finalmente o cursor ficará vermelho quando for atingida ou ultrapassada a área de toque.

4.2. Experimento

Realizamos testes quantitativos e qualitativos com usuários divididos em grupos de acordo com seu perfil de utilização de aplicações 3D. No primeiro grupo foram selecionados usuários que fazem parte da equipe de desenvolvimento do Environ. No segundo grupo, usuários que tenham experiência no desenvolvimento de aplicações 3D, mas que não façam parte da equipe do Environ e nunca utilizaram este software. O terceiro grupo foi composto por usuários sem experiência no desenvolvimento ou uso de aplicações 3D. O objetivo dessa divisão em grupos foi permitir uma comparação no desempenho dos grupos dependendo do seu grau de familiaridade com o desenvolvimento de software 3D e especialmente do Environ.

4.2.1. Formato do Estudo

O experimento foi composto por três sequências de tarefas a serem realizadas por cada um dos participantes. Em cada uma das sequências foi utilizado um dispositivo de interação diferente, dentre os seguintes: v-Glove sem o recurso de feedback tátil, v-Glove com o recurso de feedback tátil e mouse com acelerômetro [Air Mouse 2011] (Figura 4.5). O objetivo de testes com diferentes dispositivos foi comparar inicialmente a luva com e sem o recurso de

feedback tátil e posteriormente compará-la a um dispositivo utilizado atualmente por usuários de aplicações imersivas de realidade virtual.



Figura 4.5: Mouse 3D com acelerômetro utilizado nos testes [Air Mouse 2011]

Em cada uma das sequências foram executadas 5 tarefas de interação. Inicialmente o usuário tinha um tempo de dois minutos para se familiarizar com o dispositivo em tarefas de seleção de botões e navegação no Environ. Essas tarefas iniciais eram diferentes da tarefa a ser executada nos testes para evitar efeitos de aprendizado antes da efetiva realização da sequência de tarefas. Ao final do tempo livre, o usuário deveria clicar em um botão que dava início aos testes. A partir desse momento iniciava-se a tomada de tempo e a filmagem das ações do participante.

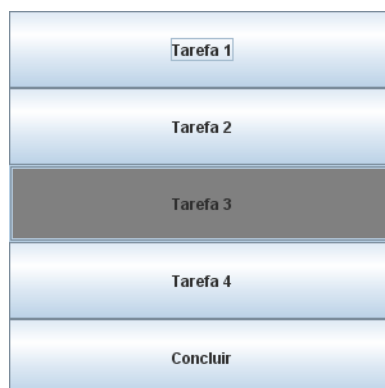


Figura 4.6: Menu principal da aplicação de testes

Na primeira tarefa o usuário era solicitado a navegar por uma cena do Environ e procurar por um determinado objeto, tendo que selecioná-lo. Ao selecionar o objeto correto, o sistema exibia um menu com as tarefas seguintes

(Figura 4.6), que deveriam ser executadas clicando-se no botão apropriado nesse menu de opções.

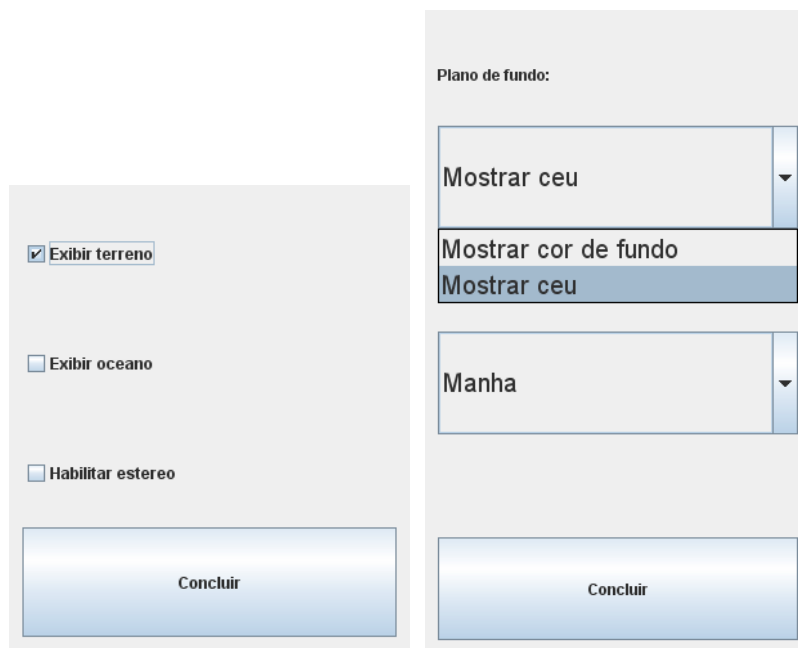


Figura 4.7: Opções da segunda e terceira tarefas

Na segunda tarefa o usuário deveria selecionar um dos componentes do tipo *checkbox* apresentados na tela e, em seguida clicar em um botão para concluir a tarefa (Figura 4.7 (a)). A terceira tarefa envolvia a seleção de opções em dois componentes do tipo *combobox*, seguidas da seleção do botão de concluir (Figura 4.7 (b)). Na quarta tarefa o objetivo era selecionar o valor solicitado em um controle do tipo *slider*, em seguida clicando no botão de concluir (Figura 4.8 (a)). Na quinta e última tarefa o usuário deveria corrigir o valor digitado em uma *textbox*, digitando a seguir um novo valor numérico e selecionando o botão de concluir (Figura 4.8 (b)). Ao final das cinco tarefas o participante deveria selecionar um botão para concluir aquele conjunto de tarefas.

The image shows two screenshots of a software interface. The left screenshot displays three sliders for terrain settings: 'Escala da textura do terreno: 20' (range 0-50), 'Altura das ondas: 600' (range 300-700), and 'Separacao entre os olhos: 7' (range 0-20). The right screenshot displays three input fields: 'Velocidade de navegacao' (value 1), 'Direcao do sol' (value 60), and 'Angulo de visao da camera' (value 45). Below these fields is a numeric keypad with buttons for digits 1-9, 0, and a left arrow, and a 'Concluir' button at the bottom.

Figura 4.8: Opções da quarta e quinta tarefas

Terminada a primeira sequência de tarefas o usuário tinha seu dispositivo trocado e a sequência era repetida. A única diferença entre duas sequências de tarefas eram as opções que deveriam ser selecionadas. Isso foi feito com a finalidade de evitar cansar o usuário tendo que executar exatamente a mesma sequência de tarefas por três vezes.

Para reduzir os efeitos de aprendizado entre uma tarefa e a seguinte, a ordem de utilização dos dispositivos não se repetiu entre usuários de um mesmo grupo. Considerando-se que havia três dispositivos e seis usuários em cada grupo, foi possível determinar uma ordem de utilização exclusiva para cada usuário dentro de seu grupo, já que a permutação de um universo de três elementos em grupos de três resulta em um total de 6 combinações possíveis de sequências.

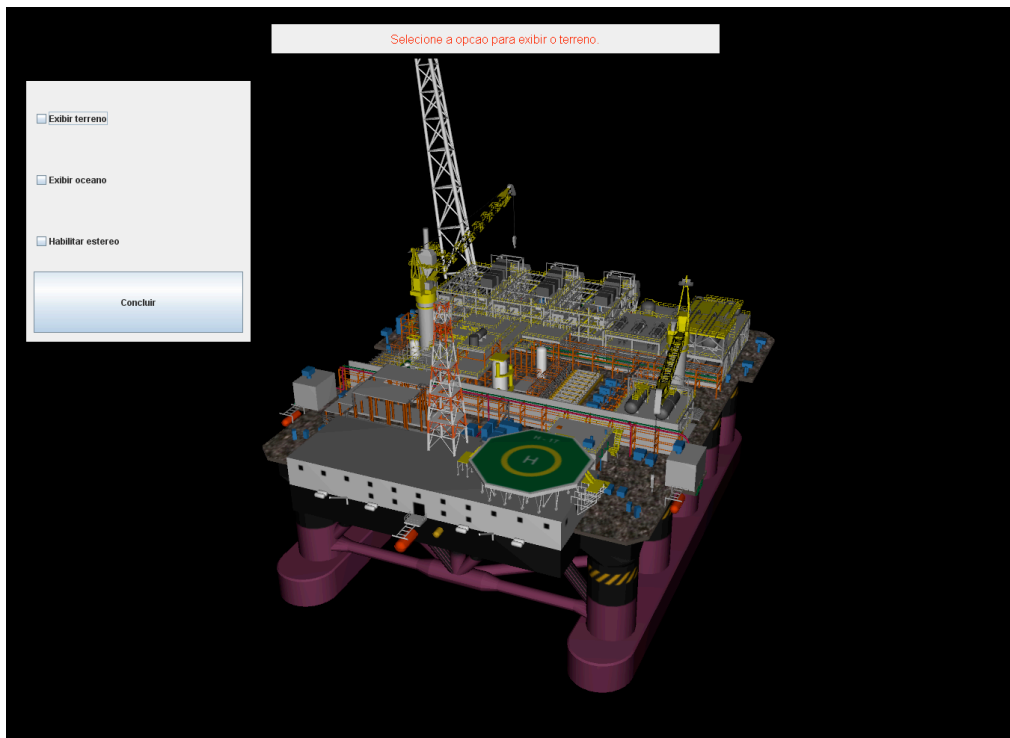


Figura 4.9: Instruções para o usuário durante a realização das tarefas

Para cada uma das tarefas o próprio sistema exibia as instruções do que deveria ser feito pelo usuário (Figura 4.9), evitando que este fosse obrigado a ler as instruções a partir de uma folha de papel ou mesmo que dependesse do avaliador para receber as instruções. Assim o avaliador pode dedicar-se integralmente à observação dos testes.

Fez parte do experimento o preenchimento de dois questionários, um antes e outro ao final dos testes, além de uma ficha de acompanhamento que ficava de posse do avaliador e onde deveriam ser anotados todos os problemas (*issues*) ocorridos durante os testes. O questionário inicial teve como objetivo verificar o perfil dos usuários e foi baseado em exemplos disponíveis em Mitchell (2007). O questionário pós-testes tinha como finalidade observar a opinião dos usuários em relação aos dispositivos utilizados. Para facilitar a análise estatística seguiu-se nesse questionário o padrão de respostas em uma escala de Likert com 7 pontos [Tullis e Albert 2008]. Todos os testes foram gravados em vídeo e as entrevistas de preenchimento dos questionários após os testes foram gravadas em áudio para permitir uma análise posterior mais detalhada.

Todos os participantes preencheram e assinaram um termo de consentimento livre e esclarecido antes do início dos testes. Este termo

explicava a finalidade dos testes, o papel do participante no processo e fornecia a opção de auto-exclusão em qualquer etapa do processo. Um exemplo do termo de consentimento pode ser encontrado no Apêndice C.

4.2.2.

Grupos de Usuários

Um total de dezoito pessoas participou dos testes. Os usuários foram divididos em três grupos de mesmo tamanho (6 participantes) de acordo com sua experiência prévia no uso e no desenvolvimento de aplicações 3D. O primeiro grupo foi chamado de “**Grupo E**” e foi composto por analistas de sistemas que trabalham há pelo menos 6 meses no desenvolvimento do software Environ. Do segundo grupo, denominado “**Grupo O**”, faziam parte analistas de sistemas com experiência no desenvolvimento de software para computação gráfica, mas que não tenham participado do desenvolvimento do Environ. O terceiro grupo, denominado “**Grupo N**”, foi composto por indivíduos sem experiência no uso de software 3D.

4.2.3.

Perfil dos Participantes

A faixa etária dos participantes selecionados para os testes variou de 21 a 60 anos, resultando em uma média de idade de 29 anos (a Tabela 4.2 mostra a média de idade dos participantes por grupos). Dezesesseis indivíduos eram do sexo masculino (89%) e dois do sexo feminino (11%). Todos os participantes eram destros e tinham por hábito utilizar mouse com a mão direita. Este era um ponto importante a ser observado, uma vez que o protótipo da v-Glove confeccionado para os testes funciona apenas na mão direita, de forma que os usuários canhotos poderiam ter seus resultados comprometidos pela utilização do dispositivo na mão não-dominante.

	<i>Média de idade</i>	<i>Menor idade</i>	<i>Maior idade</i>
Geral	29	21	60
Grupo E	25	22	28
Grupo O	31	26	37
Grupo N	31	21	60

Tabela 4.2: Média de idade dos participantes

Os participantes tinham em sua totalidade uma média semanal de mais de quinze horas de utilização de computador, sendo que 94% deles indicaram ter algum tipo de experiência com aplicações 3D. Exemplos de aplicações mencionadas foram jogos, ferramentas de modelagem 3D e visualizadores CAD. Todos os usuários do Grupo O indicaram conhecer o Environ e já ter tido algum tipo de contato com a aplicação, mas foi um contato superficial, não fazendo parte do grupo de desenvolvedores nem de usuários do sistema.

4.2.4. Equipamentos

Para os testes foi utilizada uma sala de visualização disponível em um dos prédios do Tecgraf (Figura 4.10). Na sala encontram-se dois projetores ProjectionDesign evo22sx+ [Projection Design 2011] configurados em modo estereoscópico passivo através da utilização de um filtro de polarização circular. A imagem é projetada em uma tela anti-depolarizadora (composta por um material que retém a polarização) de 120 polegadas. O microcomputador utilizado como estação de controle da aplicação possui uma placa gráfica NVIDIA Quadro FX 4600 [Nvidia Quadro 2011] com duas saídas de vídeo, processador Intel core i7 com 3.20 GHz e 12 Gigabytes de memória RAM.

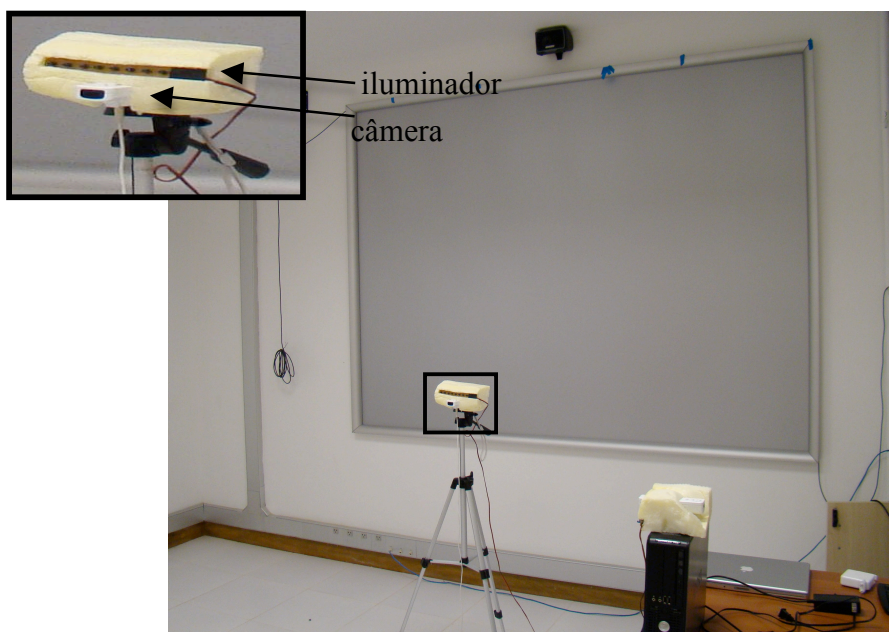


Figura 4.10: Sala utilizada nos testes com os usuários

Nesta sala foram instaladas as duas câmeras utilizadas no rastreamento da v-Glove e foi definida uma área no espaço a ser mapeada pelas câmeras como o plano de toque para interação com a luva.

4.2.5. Teste Piloto

Antes do início dos testes com os usuários foram efetuados dois testes piloto com o objetivo de validar o roteiro do experimento e corrigir eventuais problemas. O teste piloto reproduziu o roteiro completo de um teste normal, inclusive com o preenchimento dos questionários e a gravação em áudio e vídeo. O perfil dos participantes utilizados no teste piloto foi compatível com o dos usuários do Grupo N, uma vez que se espera desse grupo um maior grau de dificuldade na realização das tarefas.

Essa etapa foi muito importante, pois nela foi possível observar diversos pontos de melhoria no modelo inicial de experimento que havia sido desenhado. Durante o piloto verificou-se a necessidade de correção de falhas na aplicação construída para os testes, aperfeiçoamento das instruções apresentadas para os participantes, aperfeiçoamentos no formato dos *widgets* e até mesmo melhorias no processo de coleta dos dados.

4.3. Resultados dos Testes

Como já mencionado, os dados resultantes dos testes foram coletados através do preenchimento de questionários e formulários de acompanhamento, gravação em vídeo da realização das tarefas e gravação do áudio das entrevistas. Todas as informações coletadas foram concentradas em uma planilha eletrônica para facilitar o processo de análise e comparação dos dados. Os resultados foram então separados em quatro categorias de acordo com as métricas de usabilidade a serem observadas.

Em Tullis e Albert (2008) são apresentados dez cenários de estudo de usabilidade e os mesmos são relacionados com as métricas de usabilidade que se aplicam para cada tipo de estudo. De acordo com a classificação dos autores, nosso trabalho se encaixa no cenário de “comparação de designs alternativos” e a ele se aplicam cinco métricas: sucesso da tarefa (*task success*), duração da

tarefa (*task time*), métricas de situações de usabilidade (*issue-based metrics*), métricas auto-relatadas (*self-reported metrics*) e métricas combinadas e comparativas (*combined and comparative metrics*). As quatro primeiras métricas de usabilidade serão discutidas em mais detalhes nas seções a seguir. A quinta métrica será apresentada como parte da seção 4.4 (Discussões).

4.3.1.

Sucesso da Tarefa (*task success*)

A primeira análise que se fez nos dados capturados durante os testes foi em relação ao sucesso ou falha na realização das tarefas solicitadas. A maioria dos usuários não teve dificuldades em completar 100% das tarefas. A Figura 4.11 mostra o percentual de sucesso para cada uma das tarefas realizadas por dispositivo.

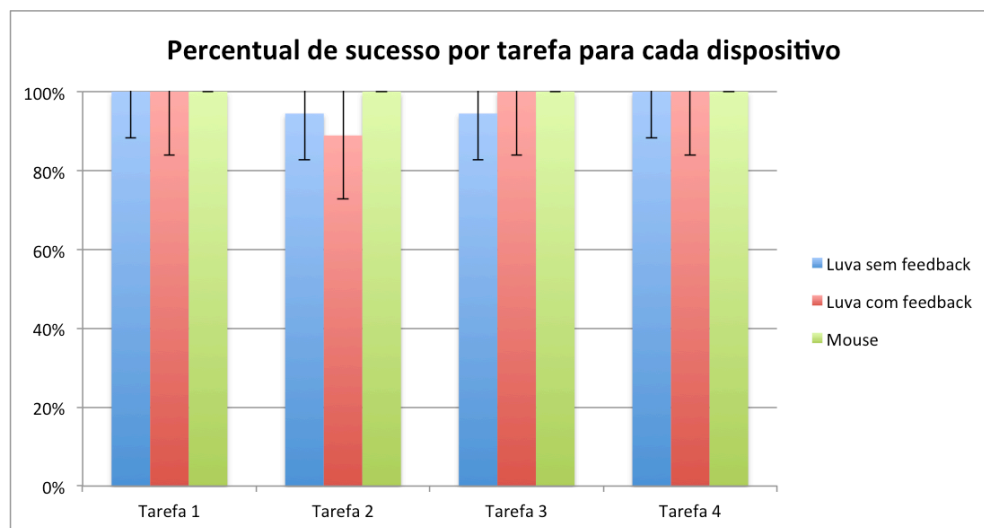


Figura 4.11: Percentual de sucesso por tarefa para cada dispositivo

Pode-se observar que 100% dos usuários completaram as tarefas realizadas com o mouse, enquanto na luva com feedback tátil o percentual de sucesso da tarefa 2 foi de 89%. Na luva sem feedback tátil verificou-se um percentual de sucesso de 94% nas tarefas 2 e 3.

4.3.2.

Duração da tarefa (*task time*)

A seguir realizou-se uma análise do desempenho dos usuários de cada um dos grupos na realização das tarefas. Para isso foram usadas as medidas de duração em segundos das sequências de tarefas. A Figura 4.12 mostra o gráfico

com a média de tempo de cada um dos grupos de usuários por dispositivo (os intervalos de confiança utilizados foram de 90%).

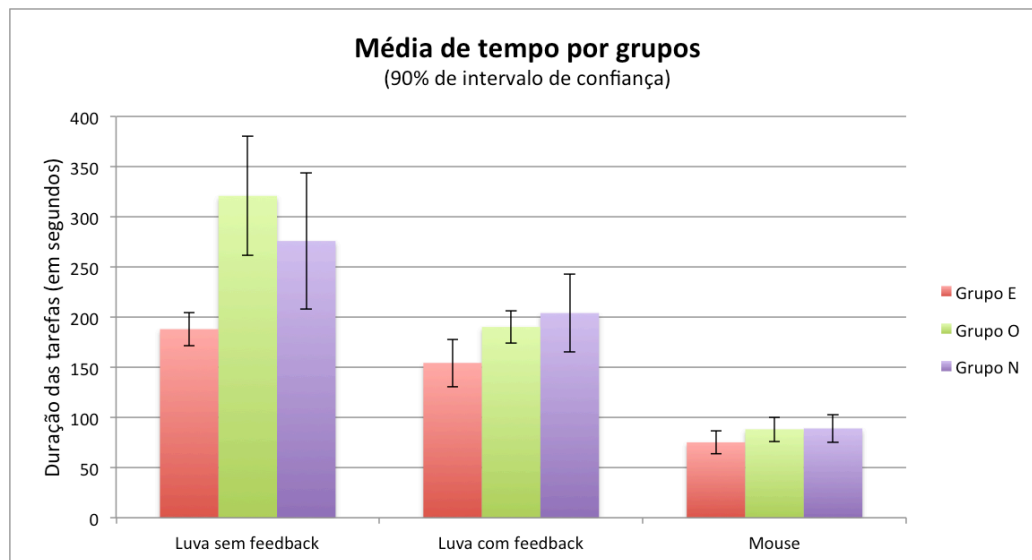


Figura 4.12: Média de tempo por grupo

Visualmente podem-se observar duas situações no gráfico. Primeiramente que a luva sem feedback tátil possui a pior média de desempenho, seguida pela luva com feedback e pelo mouse. Além disso, pode-se observar que o tempo de realização das tarefas parece diminuir de acordo com o nível de experiência do usuário com aplicações 3D. Mais à frente discutiremos se estas diferenças possuem de fato alguma relevância estatística.

4.3.3.

Métricas de situações de usabilidade (*issue-based metrics*)

A terceira análise efetuada nos dados dos testes foi sobre as métricas de problemas de usabilidade. Durante a realização dos testes os avaliadores coletaram informações de situações ocorridas durante a execução das tarefas. Posteriormente foi feita uma tabulação classificando os problemas encontrados por frequência, prioridade, *rating* e categoria.

4.3.3.1.

Frequência

Para o cálculo de frequência foi feita a contagem do número de usuários onde foi observada uma determinada situação.

4.3.3.2. Prioridade

A priorização das situações foi realizada de acordo com o modelo de três níveis proposto por Tullis e Albert (2008). A prioridade alta indica situações que levem à não conclusão de uma tarefa. Situações de prioridade média são aquelas que contribuem para o insucesso na realização de uma tarefa, mas que não são a causa direta desse insucesso. E a prioridade baixa indica situações que causam algum tipo de incômodo ou frustração no participante, mas que não estão relacionados a falhas na realização de alguma tarefa.

4.3.3.3. Rating

Para uma melhor classificação das situações Tullis e Albert (2008) indicam a adoção de um fator de classificação composto ao invés de considerar-se apenas a prioridade ou a frequência. Pode-se afirmar que uma situação com prioridade alta que ocorre apenas com um participante é mais importante do que uma de prioridade média, mas que ocorre com metade dos participantes?

Com a finalidade de se levar em consideração na classificação essas duas grandezas de forma equivalente é que o *rating* foi introduzido como grandeza. O cálculo do *rating* é feito somando-se o valor da frequência e da prioridade para uma determinada situação.

4.3.3.4. Categoria

Como forma de organizar as situações verificadas foram utilizadas quatro categorias: processo de testes (T), interface gráfica (G), operação de dispositivo (D) e infraestrutura (I).

Problema	Frequência	Prioridade	Rating	Categoria
Falta de precisão por deslocamento da mão em Z	8	3	11	D
Falta de precisão do <i>slider</i> na luva com feedback	6	3	9	G
Cansaço no braço ou na mão	5	3	8	D
Dificuldade em diferenciar cursor laranja do vermelho	4	3	7	G
Falta de precisão do <i>slider</i> na luva sem feedback	4	3	7	G

Dificuldades para selecionar a combo box em dois cliques	4	3	7	G
Dificuldades com a precisão dos botões na luva sem feedback	5	2	7	D
Dificuldades de compreensão das instruções	3	3	6	T
Selecionou opção errada em alguma tarefa	2	3	5	D
Problema de travamento do software (teve que ser reiniciado)	2	3	5	I
Dificuldades com a postura para utilizar o sistema (não conseguiu achar um posicionamento ideal)	2	3	5	D
Não executou alguma tarefa por clicar acidentalmente em 'Concluir' antes de executá-la	2	3	5	D
Dificuldades com a precisão da <i>checkbox</i> na luva sem feedback	3	2	5	D
Clicou direto no valor final do <i>slider</i> , não utilizou o conceito de arrastar	4	1	5	G
Usou a técnica do balanço do corpo para melhorar a precisão na luva sem feedback	2	2	4	D
Dificuldades com a precisão do <i>slider</i> no mouse	1	3	4	G
Dificuldades ao clicar nos botões na luva com feedback	1	3	4	D
Teve dificuldades inicialmente com a navegação do mouse por ter que manter o botão pressionado ao movimentar	2	1	3	D
Usou o scroll (zoom) do mouse mesmo sem ter sido pedido	2	1	3	T
Usou a mão esquerda para melhorar a precisão	2	1	3	D
Tentou orientar o movimento da mão pela tela e não pela câmera	1	2	3	D
Executou as tarefas fora da ordem porque teve dificuldades em clicar nos botões do menu	1	1	2	D
Tentou levar o cursor da luva para o canto para "resetar" como se faz com o mouse	1	1	2	D

Tabela 4.3: Situações de usabilidade observadas

A Tabela 4.4 mostra a listagem de situações observadas durante os testes e o valor observado para cada uma das grandezas mencionadas. Partindo-se dos dados da tabela foram selecionadas as oito situações com o maior *rating* para uma análise mais detalhada. Essas situações já permitem que se tenha um indicativo de pontos a serem melhorados na v-Glove.

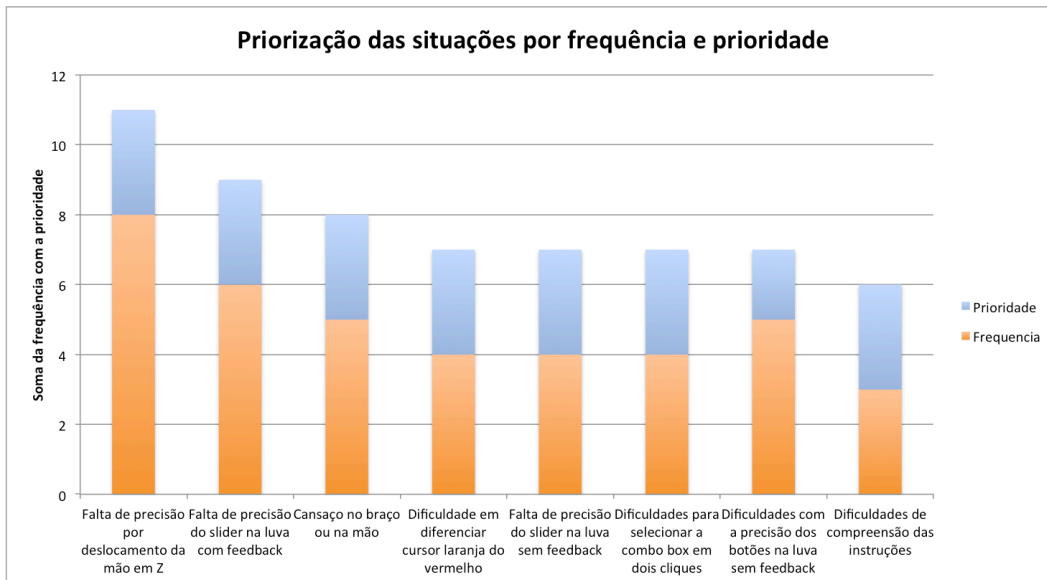


Figura 4.13: Classificação das situações de usabilidade por frequência e prioridade

O gráfico da Figura 4.13 mostra o *rating* das oito situações selecionadas para análise. É possível observar para cada uma delas a contribuição referente às grandezas de prioridade (em azul) e frequência (em laranja) na formação do *rating*. A seguir descreveremos cada uma dessas oito situações individualmente:

- Falta de precisão por deslocamento da mão em Z** – esta foi a situação que ocorreu com maior frequência durante os testes. Ao deslocar a mão de sua posição original até a superfície virtual de toque é muito difícil para o usuário manter a posição de forma a não variar a localização do cursor nos eixos X e Y. Isso faz com que ao atingir a área de clique a posição do cursor não seja mais a mesma de quando o movimento foi iniciado, fazendo com que o clique ocorra em uma região que não seja a desejada pelo usuário. Esta situação pode acarretar dois efeitos indesejados. O primeiro é a falha no clique e o insucesso na seleção da opção desejada ou no clique de um botão, o que faz com que a operação tenha que ser repetida até que se obtenha sucesso. O segundo efeito é a possibilidade do clique ocorrer em uma área onde exista outro componente de interface, ocasionando a seleção de uma opção incorreta. A consequência deste segundo efeito pode ser a não realização de uma tarefa ou a necessidade de repeti-la, caso isso seja possível.

- **Falta de precisão do *slider* na luva com feedback e falta de precisão do *slider* na luva sem feedback** – o *slider* foi o componente de interface gráfica mais afetado pelo problema da falta de precisão relatado no item anterior. Uma vez que o usuário selecionava o indicador do *slider* ele tinha dificuldades em colocá-lo na posição referente ao valor desejado na régua. Isso ocorria porque, dependendo do tamanho da escala, a distância entre dois marcadores na régua do *slider* era pequena a ponto de fazer com que uma oscilação mínima na mão do usuário impedisse seu posicionamento na localização adequada. Esse problema foi observado tanto na luva com o recurso de feedback tátil quanto na luva sem este recurso. Por terem nos dois casos a mesma causa optou-se por analisá-los em conjunto.
- **Cansaço no braço ou na mão** – o posicionamento da mão com a v-Glove para a interação exige que se mantenha o braço esticado e elevado um pouco à frente do corpo. Nos casos onde se realizou operações com a luva por um período de tempo mais elevado, alguns usuários relataram cansaço no braço e dores similares às de tendinite no pulso ou no antebraço. Relatos referentes a cansaço e fadiga já eram esperados levando-se em consideração resultados já verificados em trabalhos similares. Entretanto, entende-se que houve um agravamento nessas situações causado pelo problema da falta de precisão no clique com a v-Glove, que acarretava um número maior do que o esperado de repetições na execução das tarefas resultando em uma maior duração das tarefas e ocorrências de cansaço em alguns usuários.
- **Dificuldade em diferenciar cursor laranja do vermelho** – O feedback visual criado para auxiliar na percepção da profundidade da mão no eixo Z tinha como últimas cores o laranja (que indicava que a v-Glove estava próxima à área de clique) e o vermelho (indicando que a área de clique havia sido atingida). Alguns usuários tiveram dificuldades em perceber a diferença entre as tonalidades de laranja e vermelho utilizadas, causando uma confusão na percepção do clique especialmente no caso da luva sem o feedback tátil. Na v-Glove com feedback tátil esse problema ocorreu com menor intensidade, pois mesmo se não conseguisse diferenciar as cores do cursor o usuário tinha o auxílio do recurso de vibração para saber se havia atingido a área de clique. O problema da

pouca diferença na tonalidade também foi agravado por limitações no espectro de cores dos projetores utilizados, situação verificada ao comparar-se a imagem do cursor no projetor com a mesma imagem gerada em um monitor LCD convencional.

- **Dificuldades para selecionar a combo box em dois cliques** – alguns usuários tiveram dificuldades na seleção do componente *combobox*, tanto na luva com o recurso de feedback tátil quando na versão sem este recurso. Para selecionar uma opção na *combobox* o usuário deveria clicar inicialmente sobre o componente, fazendo com que fossem listadas as opções disponíveis, e depois clicar uma segunda vez na opção desejada, concluindo a operação. Esta operação em duas etapas ocasionou dificuldades porque ao clicar a primeira vez sobre o componente o usuário ficava em dúvida se deveria manter a mão posicionada na área de clique ou se deveria retornar a mão à posição inicial e fazer um novo clique.
- **Dificuldades com a precisão dos botões na luva sem feedback** – este problema está relacionado a dois itens já mencionados: a falta de precisão no deslocamento da mão em Z e a dificuldade de diferenciar cursor laranja do vermelho. Sem poder contar com o auxílio do recurso de feedback tátil, o usuário dependeria da indicação de cor do cursor para saber se atingiu a superfície de clique. Com o problema da tonalidade este recurso ficou comprometido, sendo ainda agravado pela falta de precisão no movimento, gerando assim consequências similares às já observadas no primeiro item, como o clique em áreas diferentes da desejada.
- **Dificuldades de compreensão das instruções** – observou-se ainda que alguns usuários tiveram dificuldades na compreensão de algumas instruções presentes no software. Tarefas não foram compreendidas adequadamente ocasionando a escolha de opções incorretas ou uma demora maior na realização da tarefa até que se tivesse o entendimento adequado. Alguns usuários solicitaram ajuda ao avaliador para compreender o que deveria ser feito na tarefa e um deles não entendeu que deveria executar apenas o que era solicitado nas instruções, acabando por interagir com todos os componentes de interface gráfica presentes na tela e não apenas com o que era indicado nas instruções.

4.3.4. Métricas auto-relatadas (*self-reported metrics*)

A partir dos dados coletados nas entrevistas pós-testes, tivemos condições de analisar métricas auto-relatadas de forma quantitativa e qualitativa. O preenchimento do questionários com perguntas na escala Likert é a base da análise quantitativa apresentada a seguir. A gravação do áudio das entrevistas em formato de bate-papo com os participantes permitiu o levantamento de uma grande quantidade de informações sobre problemas e sugestões de melhorias por parte dos usuários. Essas informações serão apresentadas na seção 4.3.4.2 (Análise qualitativa).

4.3.4.1. Análise quantitativa

A Figura 4.14 mostra a média das notas em resposta à primeira pergunta do questionário pós-testes, que pedia para o usuário indicar sua opinião sobre a utilização de cada um dos dispositivos com uma nota variando de 1 a 7, onde o primeiro valor significa “muito ruim” e o último “muito bom”. A seguir foram feitas perguntas solicitando a opinião dos participantes sobre cada uma das tarefas individualmente. Os resultados observados indicam que a opinião referente a cada questão individualmente segue a mesma tendência observada na questão inicial, que cobre a visão geral. Por isso optou-se por não analisar as questões de forma individual nesta seção, discutindo-se apenas os feitos da visão geral dos usuários sobre a realização das tarefas em cada um dos dispositivos.

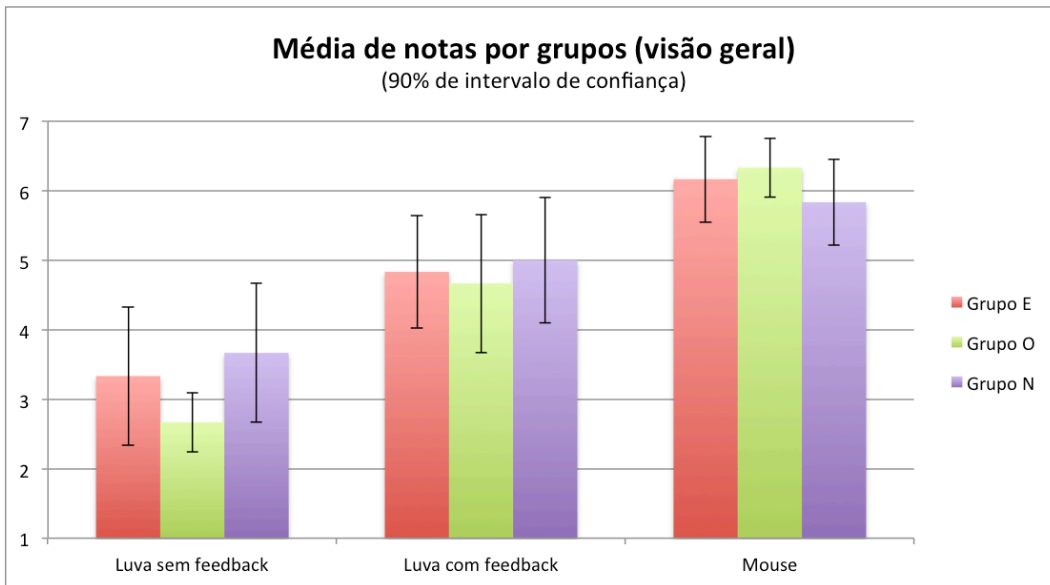


Figura 4.14: Média de notas por grupos por dispositivo (visão geral)

Os resultados das avaliações foram compatíveis com os resultados de desempenho observados nas métricas de duração das tarefas. A v-Glove sem o recurso de feedback tátil teve a pior avaliação, seguida pela luva com o recurso de feedback tátil e pelo mouse. Veremos mais à frente se essas diferenças representam alguma relevância em termos estatísticos.



Figura 4.15: Notas por *widget* para cada dispositivo

A Figura 4.15 mostra um gráfico de radar comparando a avaliação dos usuários por componente de interface gráfica utilizado em cada um dos dispositivos. Nesse caso a divisão dos grupos foi ignorada e o domínio de participantes considerado refere-se aos dezoito indivíduos que colaboraram com os testes. Pode-se observar neste gráfico que os componentes *checkbox* e de entrada numérica seguem um mesmo padrão ao se comparar os três dispositivos. Já no caso dos componentes *slider* e *combobox* observa-se que a distância entre as duas versões da luva e o mouse é maior, sendo perceptível pela deformação no gráfico tanto no setor sul (referente ao *slider*) quanto no setor leste (referente à *combobox*).

4.3.4.2.

Análise qualitativa - sugestões de melhorias

Por tratarem-se de usuários que, em sua maioria, já possuíam experiência no desenvolvimento de aplicações 3D, buscamos durante as entrevistas após os testes capturar críticas e contribuições dos participantes entrevistados. A Tabela 4.5 lista as sugestões dadas pelos usuários durante as entrevistas. As sugestões comuns a mais de um usuário são indicadas com o valor da frequência entre parênteses à direita da descrição.

Sugestões dos participantes
Aumentar a escala de movimento da mão, pois ela é pequena em relação ao cursor na tela
Considerar o clique no local do primeiro contato com a área de clique e não ao sair dela
Detectar o clique a partir de uma variação mínima no eixo Z após atingir a área de toque
Diminuir a intensidade da vibração
Fazer o clique de outra forma para ser menos cansativo para a mão
Fazer o clique em um movimento de pinçar entre o dedo indicador e o polegar (2)
Fazer o reconhecimento do clique em uma região relativa à localização do usuário e não fixa
Fazer o <i>slider</i> de forma vertical e não mais horizontal
Implementar o feedback tátil quando o cursor atingir algum <i>widget</i> que permita interação
Implementar um feedback tátil progressivo (2)
Implementar um filtro de ruídos de movimento similar ao que existe no mouse
Incluir um botão para clique na luva (2)
Manter a vibração (numa intensidade menor) enquanto a v-Glove estiver na área de clique
Melhorar o feedback visual sincronizando a mudança de cor do cursor com mudanças nos <i>widgets</i>

Tabela 4.4: Sugestões dos participantes

Algumas das sugestões mencionadas pelos usuários chegaram a ser avaliadas durante a construção do protótipo da v-Glove. Para cada uma delas apresentaremos a seguir os motivos que nos levaram a descartá-las durante o desenvolvimento deste trabalho.

- **Considerar o clique no local do primeiro contato com a área de clique e não ao sair dela** – essa abordagem foi testada durante a construção do driver de controle do cursor da v-Glove, mas acabou sendo descartada após a realização do primeiro teste piloto. O princípio de funcionamento da operação de clicar em um objetivo com a luva é o seguinte: ao atingir a área de clique da câmera o driver da v-Glove gera um evento do tipo *MOUSE_PRESS*, que corresponde ao botão do mouse pressionado. Esse estado é mantido até que o dedo do usuário saia da área de clique, quando então é gerado um evento de *MOUSE_RELEASE*. Apenas nesse momento é que o clique é efetuado, seguindo o mesmo princípio de um mouse convencional. O problema neste caso é que entre estes dois estados ocorrem variações involuntárias (nos eixos X, Y e Z) do posicionamento da mão do usuário. Por isso, no instante em que o clique é considerado é possível que o posicionamento da mão do usuário tenha sido alterado para outra área que não a desejada por ele para interagir. Por limitações do driver, para considerar o clique no instante em que o mouse é pressionado seria necessário reposicionar o cursor de volta para as coordenadas do momento em que ocorreu o evento de *MOUSE_PRESS*. Isso, entretanto, acaba por gerar uma mudança brusca na posição atual do cursor que prejudica a utilização da v-Glove, conforme pode ser observado no teste piloto. Além disso, essa mudança causa um problema quando o usuário ao invés de selecionar deseja arrastar o objeto, uma vez que ao iniciar a operação de arrastar o cursor retorna bruscamente à sua posição anterior causando um salto de posicionamento do objeto na tela.
- **Detectar o clique a partir de uma variação mínima no eixo Z após atingir a área de toque** – variações mínimas ocorrem naturalmente, uma vez que é muito difícil para o usuário manter sua mão completamente imóvel no ar. Por isso o algoritmo da aplicação filtra movimentos com intensidades menores do que um valor de *threshold* configurável, fazendo com que os movimentos sejam considerados apenas se forem maiores do que esse *threshold*. Para considerar essa sugestão, a variação no eixo Z deveria ser maior do que este valor de descarte, causando um impacto na precisão do movimento de arrastar do cursor.

- **Diminuir a intensidade da vibração** – o componente de vibração utilizado no protótipo (Lilypad Vibe Board) não suporta o ajuste de intensidade da vibração, o que impede a implementação dessa funcionalidade. Para obtermos este resultado teríamos que substituir este componente por outro com suporte a diferentes intensidades de vibração.
- **Implementar o feedback tátil progressivo** – esta funcionalidade não pode ser implementada pelo mesmo problema do tópico anterior. Sem o suporte a diferentes níveis de vibração no hardware utilizado não é possível a implementação do conceito de feedback tátil progressivo.
- **Incluir um botão para clique na luva** – a ideia de se acrescentar um botão na luva para fazer o clique foi considerada nos primeiros estudos sobre a operação de clique na v-Glove. Entretanto, ela acabou sendo descartada porque, no nosso entendimento, o fato do usuário ter que pressionar um botão reduz a naturalidade da interação com a v-Glove, deixando-a muito similar a um mouse como o que foi utilizado nos testes. O conceito de pressionar um botão para clicar acaba por ser intuitivo pela analogia que atualmente se tem com o mouse tradicional, mas distancia-se da proposta desse trabalho onde a intenção era propor formas de interação mais próximas do que se faz com objetos no mundo real.
- **Manter a vibração (numa intensidade menor) enquanto a v-Glove estiver na área de clique** – a ideia de manter a vibração enquanto o dedo do usuário estiver na área de clique chegou a ser testada durante a construção do protótipo, mas acabou sendo descartada porque a manutenção da vibração por um período prolongado causa desconforto no usuário. Como já foi comentado, a falta de suporte ao ajuste de intensidade da vibração no hardware utilizado não permitiu também que se testasse a ideia de manter a vibração com menor intensidade.

As sugestões que não haviam sido consideradas durante a construção da v-Glove serão relacionadas como alternativas para trabalhos futuros e serão discutidas em maiores detalhes no capítulo 5.

4.4. Discussões

Nesta seção faremos uma comparação dos resultados obtidos nas quatro métricas de usabilidade aplicadas. A partir dessa análise espera-se obter uma visão mais clara do potencial da v-Glove como dispositivo de interação para ambientes imersivos de realidade virtual e um direcionamento sobre os pontos a serem aprimorados em trabalhos futuros.

4.4.1. Comparação dos resultados das métricas de usabilidade e impacto do perfil dos usuários

A análise de sucesso das tarefas não permitiu descobertas muito relevantes, já que os testes foram desenhados de forma a que os usuários não tivessem dificuldades em completar as tarefas. Os poucos casos em que os participantes tiveram dificuldades em executar suas tarefas ocorreram porque o usuário selecionou uma opção errada por engano ou acidentalmente. Estes problemas foram identificados em algumas situações de usabilidade observadas e podem ser resolvidos com melhorias no software criado para os testes dos usuários.

Nas métricas de desempenho na realização das tarefas, por outro lado, foi possível observar uma diferença entre o tempo médio gasto em cada um dos dispositivos e no tempo médio de realização das tarefas em cada um dos grupos.

Os dados visualmente indicam uma leve tendência de um melhor desempenho de acordo com o nível de experiência do usuário com o desenvolvimento de aplicações 3D. Entretanto, realizando-se uma análise da variância (ANOVA) de dois fatores com replicação foi possível constatar que essa diferença não possui relevância estatística quando comparamos o desempenho de um dos grupos em relação aos outros ($F(3,20) = 2,56, p < 0,09$). A mesma técnica confirmou também que as diferenças de tempo entre os dispositivos são de fato relevantes ($F(3,20) = 19,19, p < 0,01$), o que comprova os resultados obtidos nas primeiras observações, onde projetamos que o mouse teria o melhor desempenho, seguido pela luva com feedback tátil e, por último, a

versão da luva sem o feedback tátil. Maiores detalhes sobre análise da variância (ANOVA) realizada neste trabalho podem ser encontrados no Apêndice D.

Os resultados da métrica de situações de usabilidade indicam que 58% das situações observadas durante os testes estão relacionadas a problemas no funcionamento dos dispositivos de interação. O gráfico da Figura 4.16 mostra a percentual dessas situações de usabilidade relacionadas a todos os dispositivos, apenas à luva e apenas ao mouse. No caso da luva consideramos em um mesmo grupo situações relacionadas à luva com e sem o recurso de feedback tátil.



Figura 4.16: Percentual de situações de usabilidade por dispositivo

Pode-se observar no gráfico que apenas 7% das situações observadas são exclusivas do mouse, enquanto 72% estão relacionadas à operação da luva com e sem feedback tátil. Os 21% restantes referem-se a situações observadas tanto na luva quando no mouse, como problemas de postura de uso dos dispositivos, dores ou cansaço em sua operação.

A partir desses resultados é possível projetar que um trabalho direcionado à correção e redução dessas situações e problemas de usabilidade na v-Glove

poderá resultar em uma melhora nos indicadores da luva em comparação ao mouse.

No caso das métricas auto-reportadas baseadas nas notas dadas pelos usuários tivemos um resultado similar ao obtido na análise de desempenho. As avaliações não seguem um padrão de acordo com os grupos, mas a análise da variância (ANOVA) de dois fatores com replicação indicou uma diferença com relevância estatística entre os dispositivos ($F(3,2) = 41,23$, $p < 0,01$).

4.4.2. Impacto do feedback tátil na v-Glove

O objetivo ao incluirmos a luva sem o recurso de feedback tátil na lista de dispositivos avaliados foi verificar se este recurso melhora a usabilidade da luva se comparado com uma versão da mesma luva sem ele. O resultado do ANOVA de dois fatores com replicação para a comparação entre o desempenho dos usuários indica que a diferença entre a luva com feedback tátil e sem feedback tátil possui relevância estatística ($F(4,17) = 5,23$, $p < 0,03$). As opiniões colhidas dos usuários reforçam ainda mais essa constatação, especialmente quando lhes foi solicitado que classificassem com uma nota de 1 a 7 a importância do recurso de feedback tátil na v-Glove.

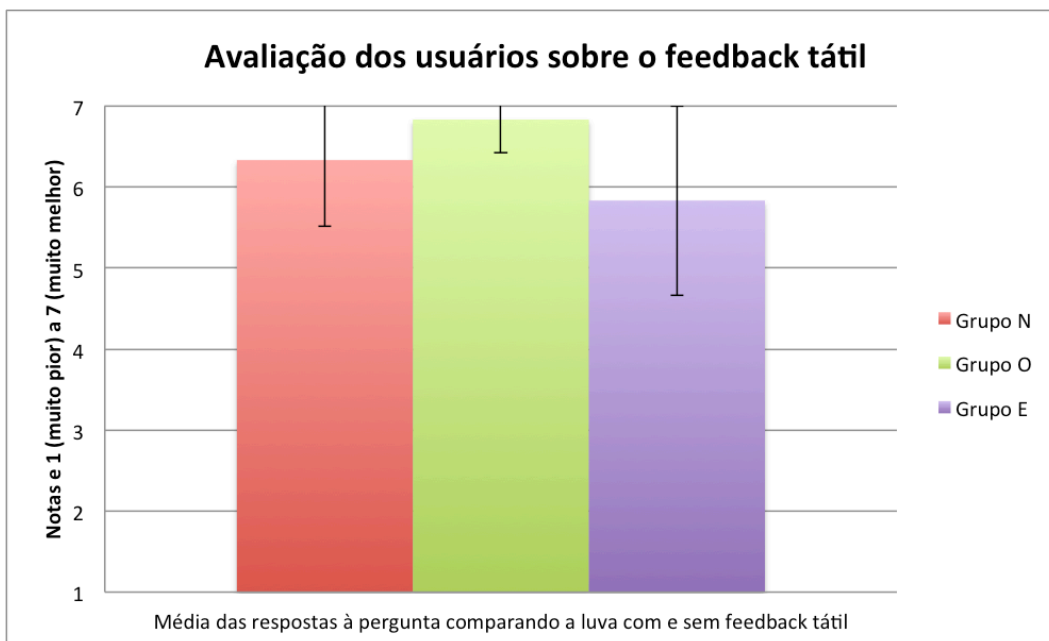


Figura 4.17: Opinião dos usuários sobre o recurso de feedback tátil na v-Glove

A Figura 4.17 mostra o gráfico com a média das respostas dos usuários à pergunta: “Comparando-se a luva com e sem o recurso de feedback tátil, qual nota na escala de 1 a 7 você atribuiria a este recurso, sendo 1 equivalente a muito pior e 7 indicando que o feedback tátil é muito melhor?”. A média geral das notas foi 6,3, variando de 5,8 no Grupo E a 6,8 no Grupo O.

4.4.3. Impacto do aprendizado na avaliação

Um efeito já esperado era uma melhora de desempenho natural na segunda e na terceira sequências de tarefas, resultado do efeito do aprendizado de uso da interface. No caso da luva havia ainda o efeito de aprendizado do funcionamento da própria luva, uma vez que entre a versão com e sem o recurso de feedback tátil não havia nenhuma outra diferença além da presença ou não da vibração. O gráfico da Figura 4.18 mostra a evolução dos tempos de duração das tarefas para as duas versões da luva de acordo com a ordem em que cada dispositivo era utilizado na sequência de tarefas. Neste caso não estamos considerando o agrupamento dos usuários por perfil, mas sim comparando todos os usuários que realizaram tarefas com cada uma das luvas antes ou depois de ter realizado a tarefa com a outra versão.

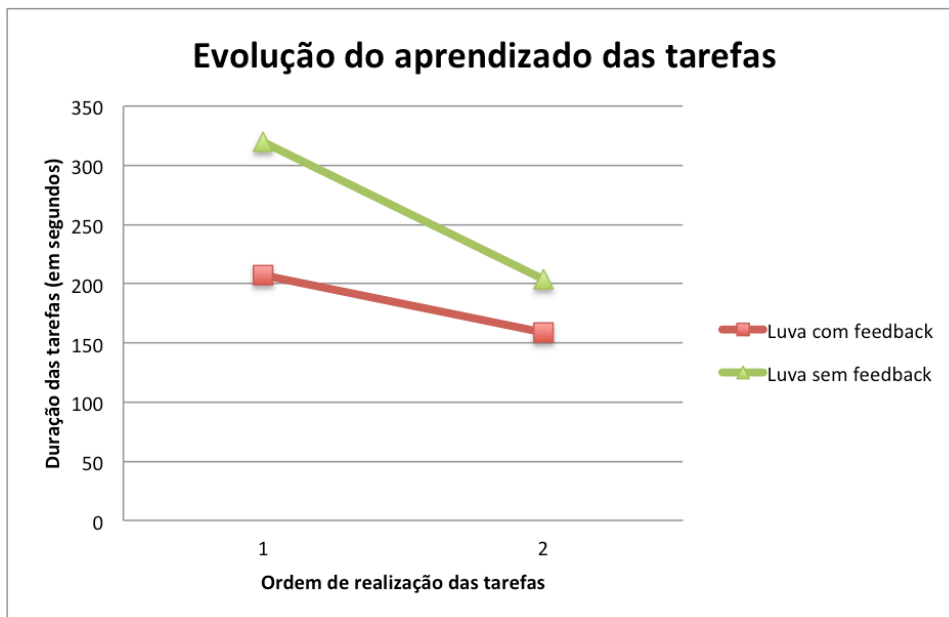


Figura 4.18: Evolução do aprendizado das tarefas

O gráfico confirma que o efeito de aprendizado influencia diretamente no desempenho dos usuários na realização das tarefas. A média de tempo de

duração dos usuários que utilizaram a luva sem feedback tátil antes da versão com feedback foi de 319 segundos, enquanto os usuários que utilizaram a versão sem feedback depois de já ter usado a luva com feedback tiveram média de 204 segundos. Isso indica uma melhora de 36% no desempenho dos usuários como efeito do aprendizado do dispositivo.

No caso da v-Glove com feedback tátil, os usuários que a utilizaram antes da versão sem este recurso tiveram um tempo médio de 207 segundos, enquanto a média de quem usou a luva com feedback após ter usado a versão sem este recurso foi de 159 segundos. Neste caso o efeito do aprendizado ocasionou uma melhora de 23% em média.