



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

NIKOLAS JORGE SANTIAGO CARNEIRO

**Um Processo para Definição e Avaliação de um Vocabulário
de Gestos para Aplicações de Visualização de Informação**

Belém
2016

Nikolas Jorge Santiago Carneiro

Um Processo para Definição e Avaliação de um Vocabulário de Gestos para Aplicações de Visualização de Informação

Dissertação de Mestrado apresentada como requisito para obtenção do grau de Mestre em Ciência da Computação. Programa de Pós-Graduação em Ciência da Computação. Instituto de Ciências Exatas e Naturais. Universidade Federal do Pará.

Orientador Prof. Dr. Bianchi Serique Meiguins

Belém

2016

Dados Internacionais de Catalogação - na - Publicação (CIP)
Biblioteca de Pós-Graduação do ICEN/UFPA

Carneiro, Nikolas Jorge Santiago

Um processo para definição e avaliação de um vocabulário de gestos para aplicações de visualização de informação/ Nikolas Jorge Santiago Carneiro; orientador, Bianchi Serique Meiguins. – 2016.

73 f.: il; 29 cm

Inclui bibliografias

Dissertação (Mestrado) – Universidade Federal do Pará,
Instituto de Ciências Exatas e Naturais, Programa de Pós-Graduação em Ciência da Computação, Belém, 2016.

1. Interação homem-máquina. 2. Visualização da informação. 3. Interfaces de usuário (Sistemas de computação). 4. Gestos-Vocabulário-Metodologia. 5. Sistemas de comunicação-Inovações tecnológicas. I. Meiguins, Bianchi Serique, orient. II. Título.

CDD – 22 ed. 004.019

Nikolas Jorge Santiago Carneiro

Um Processo para Definição e Avaliação de um Vocabulário de Gestos para Aplicações de Visualização de Informação

Dissertação de Mestrado apresentada como requisito para obtenção do grau de Mestre em Ciência da Computação. Programa de Pós-Graduação em Ciência da Computação. Instituto de Ciências Exatas e Naturais. Universidade Federal do Pará.

Data da defesa: Belém, 1 de novembro de 2016

Conceito: _____

Prof. Dr. Bianchi Serique Meiguins – Orientador
Programa de Pós-Graduação em Ciência da Computação – UFPA

Prof. Dr. Jefferson Magalhães de Moraes – Membro Interno
Programa de Pós-Graduação em Ciência da Computação – UFPA

Prof^a. Dr^a. Marcelle Pereira Mota – Membro Externo
Faculdade de Computação – ICEN/UFPA

*Dedico aos meus pais, Katia Rosa Santiago e Carlos Jorge Freire
Carneiro, e a minha esposa, Ilma Leida da Silva Costa Carneiro, pela
inesgotável paciência e incondicional apoio.*

AGRADECIMENTOS

Agradeço ao meu amigo e orientador Prof. Dr Bianchi Serique Meiguins pela confiança, pelos conselhos, pela dedicação, pela preocupação e acima de tudo, pela orientação dentro e fora do ambiente acadêmico, os quais me ajudaram a vir até este ponto e certamente continuarão comigo.

Agradeço a equipe do Laboratório de Visualização da Informação (LABVIS) da UFPA, tanto alunos quanto professores – em especial aos meus amigos Gustavo Resque, Brunelli Miranda e Tiago Araújo – por literalmente “vestirem a camisa”.

Agradeço a minha mãe pelo cuidado, ao meu pai pela confiança e a minha esposa pelo carinho, sem os quais nenhum trecho do caminho seria possível.

Agradeço ao meu irmão Jean Marcel dos Reis Costa, por usar o Facebook para me cobrar o término desta dissertação.

Agradeço à banca avaliadora formada pelos professores Jefferson Magalhães de Moraes e Marcelle Pereira Mota, pelas contribuições dadas a este trabalho.

Agradeço a todos que direta ou indiretamente contribuíram para a realização deste trabalho.

“Quanto mais eu trabalho, mais minha sorte melhora.” (Petra Isenberg)

RESUMO

O desenvolvimento tecnológico para sensores tem permitido a concretização do uso da interação não convencional nos mais variados produtos, tais como: jogos, aparelhos de TV, automação residencial e industrial, medicina, etc. Entretanto a utilização destas interfaces introduz uma série de novos e atuais desafios, tanto técnicos (fácil de instalar, adaptável a diversos produtos, precisão, etc.) quanto subjetivos (intuitivo, confortável, etc.), bem como questionamentos que acompanham esses desafios, como: é possível definir um padrão de interação para um determinado tipo de interação não convencional? O padrão encontrado pode ser generalizado a problema de outros domínios? Existem muitas outras perguntas e esta dissertação não pretende responder todas elas, mas contribuir na elaboração de algumas respostas, uma vez que em muitos casos não há padrões de interação para modos não convencionais. Assim, o objetivo desta dissertação de mestrado é apresentar um processo para seleção de gestos para interação com ferramentas de Visualização da Informação 3D (InfoVis 3D). O vocabulário de gestos adaptado para InfoVis 3D deve considerar tarefas comuns de visualização, como: filtros contínuos e discretos, navegação, configurações, entre outros. O processo de seleção de gestos é descrito abordando sua elaboração, levantamento e benchmark, bem como uma proposta de tradução do vocabulário selecionado para uma implementação utilizando sensores livremente comercializados e acessíveis. O levantamento do vocabulário e seu benchmark foi realizado com usuários que atendem ao perfil de ter alguma experiência acadêmica com Visualização da Informação e ter utilizado um número mínimo de ferramentas de análise de dados. Um protótipo foi desenvolvido em Unity 3D e os gestos implementados considerando os sensores Myo e Leap Montion. Ao final foram realizados ensaios de interação com usuários onde os cenários de testes foram compostos por tarefas típicas de InfoVis 3D.

Palavras-chave: Interação Não-Convencional, Gestos, Visualização da Informação.

ABSTRACT

The development of hardware and software technology for sensors has allowed the use of unconventional interaction in various products such as games, television sets, home automation, industrial automation, medical, etc. However the use of these interfaces introduces a number of new and current challenges, both technical (how to install, it is adaptable to other products, precision, etc.) and subjective (intuitive, comfortable, etc.) as well as questions that accompany these challenges, as: can you set a pattern of interaction for a particular type of unconventional interaction? May the pattern found be generalized to other areas? There are many other questions and this thesis does not intend to answer them all, but contribute to the development of some answers, since in many cases there is no interaction standards for unconventional ways of interaction. The objective of this master thesis is to present a process for selection and benchmark of gestures to interact with 3D Information Visualization (3D InfoVis) Tools. The vocabulary of gestures adapted to 3D InfoVis should consider common visualization tasks, such as continuous and discrete filters, navigation, settings, among others. The gestures selection process is described addressing its development, survey and benchmark as well as a proposal for the selected vocabulary translation into an implementation using consumer level sensors. Gathering the vocabulary and performing the benchmark to it was conducted with users with the profile of having some academic experience with Information Visualization and have used a minimum number of data analysis tools. A prototype was developed using Unity 3D and gestures were implemented to the Myo and Leap Motion devices. At the end we conducted an interaction test with users, test scenarios were composed of typical tasks of 3D InfoVis and considerations on the results are presented.

Key-words: Unconventional Interaction, Gesture, Information Visualization.

LISTA DE ILUSTRAÇÕES

Figura 1: Processo adaptado de levantamento e benchmark dos gestos.	33
Figura 2: Pirâmide de abstração de interações.	34
Figura 3: Número de erros por funcionalidade.	42
Figura 4: Número de reinícios por usuário.	44
Figura 5: número de reinícios por gesto.	44
Figura 6: Resultado da avaliação subjetiva de ergonomia.	46
Figura 7: Desvio Ulnar e Radial.	49
Figura 8: Gesto da funcionalidade pan sendo executado com o Myo.	54
Figura 9: Gesto da funcionalidade zoom, pinçar com polegar e indicador.	55
Figura 10: Posicionar a mão sobre uma seta de direção para rotacionar.	55
Figura 11: Configuração do limite de filtro contínuo, pressionar o thumb e arrastar.	56
Figura 12: Arrastar o range para mover o intervalo de seleção, com o Myo.	57
Figura 13: Gesto de navegar na lista de atributos com agarrar, arrastar e soltar.	57
Figura 14: Aprofundar a mão virtual a tocar um botão para selecionar.	58
Figura 15: Pressionar um botão e mover a mão verticalmente para selecionar múltiplos.	58
Figura 16: Comparativo entre as médias do teste subjetivo de Stress entre gestos implementados e levantados.	61

LISTA DE TABELAS

Tabela 1: Funcionalidades exploradas neste trabalho.	35
Tabela 2: Tabela de descrição de gesto observado.	35
Tabela 3: Resultado da Etapa 1, frequência por gesto.	37
Tabela 4: Notas médias esperadas e registradas em cada fase.	41
Tabela 5: Intervalo, média e desvio padrão encontrado por fase.	41
Tabela 6: Pontuação individual dos gestos para o teste de stress.	47
Tabela 7: Média dos resultados obtidos para avaliação subjetiva dos gestos implementados.	60

LISTA DE SIGLAS

API	<i>Application Programming Interface</i>
CAD	<i>Computer Aided Design</i>
EMG	<i>ElectroMyoGraphy</i>
GUI	<i>Graphical User Iterface</i>
HCI	<i>Human Computer Interaction</i>
SDK	<i>Software Development Kit</i>

SUMÁRIO

1. INTRODUÇÃO	16
1.1 Justificativa	17
1.2 Objetivos.....	18
1.3 Metodologia.....	18
1.4 Organização do Trabalho	19
2. CONTEXTUALIZAÇÃO	20
2.1 Visualização de Informações	20
2.2 Interação Multimodal	21
2.3 Trabalhos Relacionados	24
2.4 Trabalhos Anteriores	28
3. SELEÇÃO DE GESTOS	29
3.1 Avaliação Ergonômica	48
3.2 Vocabulário Final	50
4. MAPEAMENTO DO VOCÁBULÁRIO.....	51
5. TESTE COM O VOCABULÁRIO IMPLEMENTADO.....	59
CONCLUSÃO	64
5.1 Trabalhos Futuros.....	65
REFERÊNCIAS	67

1. INTRODUÇÃO

A popularização de diferentes formas de interação entre humanos e sistemas computacionais tem crescido em larga escala nos últimos tempos, esses dispositivos inovadores tendem a propiciar uma comunicação mais natural com sistemas computacionais, seja para com dispositivos pelos quais o ser humano manda informação para o sistema (como o Kinect, o Wiimote ou a Power Glove), seja para o sistema enviar informações para o ser humano (como experiências de cinema 4D ou roupas de resposta tátil). Essa popularização se deve muito a áreas de entretenimento, como jogos, que acabam fazendo a demanda por estes dispositivos multimodais ficar mais próxima do usuário comum, gerando uma gama maior de dispositivos a escolha e preços mais acessíveis.

Entretanto, por mais que estes dispositivos estejam se popularizando e se tornando cada vez melhores, tanto os sensores disponíveis como os métodos de análise (bem como o poder computacional para processar em tempo real) dos dados providos por estes sensores ainda não permite que a comunicação por estes dispositivos seja uma cópia fiel da comunicação humana realizada no dia-a-dia, fazendo com que seja necessário estudar maneiras de melhorar esta comunicação antes de poder extrair dela seu potencial. Essa busca por novas e melhores formas de interagir com computadores é um objeto de estudo recorrente na computação e é aplicado para as mais diversas áreas dentro desta, como a robótica e a interação humano-computador.

A utilização destas interfaces multimodais introduz uma série de novos desafios, tanto de caráter técnicos quanto não técnicos, uma vez que estas interfaces não tem um padrão estabelecido para seu desenvolvimento e estabelecer um padrão amplo pode comprometer a naturalidade e intuitividade da comunicação. Dessa forma, o objetivo deste trabalho é apresentar um processo para seleção de gestos, uma das três formas de comunicação mais utilizadas por seres humanos, adaptado para seleção de um vocabulário de gestos para interação com ferramentas de análise de dados e que considere os conceitos existentes na área de Visualização da Informação, bem como

a experiência do perfil de utilizadores de ferramentas de Visualização. O processo é descrito abordando sua elaboração, levantamento e benchmark do vocabulário de gestos, bem como uma proposta de tradução do vocabulário selecionado para uma implementação utilizando sensores livremente comercializados e acessíveis. O levantamento do vocabulário e seu benchmark foi realizado com usuários que atendem ao perfil de ter alguma experiência acadêmica com Visualização da Informação e ter utilizado um número mínimo de ferramentas de análise de dados.

1.1 Justificativa

Interações multimodais não possuem uma padronização em si. Esse tipo de interação faz uso de preceitos existentes em outros contextos, como ergonomia e projeto de interfaces, para resolver algumas situações do projeto de interação não convencional, sem efetivamente oferecer uma padronização na implementação destas.

Utilizar interação não convencional em ambientes que demandem elevado ou constante esforço mental por parte do usuário pode ser de grande valia, uma vez que estas interações forem ergonômicas e naturais, elas podem vir a reduzir o esforço físico e mental (no que se refere a demanda cognitiva de aprendizado e lembrança) do usuário.

Como os sensores de entrada não são perfeitos e a definição do que é natural depende de fatores culturais e sociais, é necessário definir o que é natural e intuitivo para um público alvo, ou seja, o que aquele conjunto de usuários espera de um determinado modo de interação no intuito de poder propor um conjunto de interações para tarefas conhecidas daquele grupo, que possa efetivamente propiciar os benefícios de se realizar aquela interação via um ou mais modos inovadores.

Devido a estes fatores, um processo que oriente na escolha de um conjunto de gestos intuitivos para um determinado grupo, público alvo, é de grande valia, ainda mais se este oferece um mecanismo que permita compara-lo com outros conjuntos de interações na busca pelas melhores formas de representar cada uma destas interações.

1.2 Objetivos

O objetivo principal deste trabalho é apresentar um processo adaptado que permita descobrir e avaliar um conjunto de gestos naturais para usuários de visualização de informação, propor uma implementação para estes e por fim, analisar se os gestos implementáveis são confortáveis para a utilização tanto quanto os gestos descobertos. Em partes estes objetivos podem ser especificados em:

- Definir um processo reaplicável e comparável entre diferentes conjuntos de usuários ou gestos.
- Gerar um vocabulário para interação com gestos ao ar livre em ferramentas de Visualização de Informação.
- Implementar o vocabulário definido.
- Avaliar o vocabulário implementado em relação ao definido.

1.3 Metodologia

A realização deste trabalho pode ser especificada nas seguintes etapas:

- Pesquisa de trabalhos com foco no levantamento de vocabulários de gesto.
- Escolha de um processo de referência para o levantamento de gestos.
- Adaptações necessárias ao processo de referência para se adequar ao levantamento do vocabulário de gestos para interações de visualização.
- Aplicar o processo adaptado para levantar e avaliar os gestos com participantes que tenham experiência em visualização.

O processo apresentado neste trabalho é adaptado de (NIELSEN, STÖRRING, *et al.*, 2004), que levanta um conjunto de gestos para interação específico para um cenário de aplicação. O que se quer desenvolver a partir deste trabalho é um conjunto de gestos intuitivos para interação, independente do cenário e da tecnologia utilizados. Além disso, este trabalho ainda apresenta uma proposta de mapeamento do conjunto de gestos encontrados para uma implementação específica.

Para os experimentos realizados foram selecionados participantes com experiência acadêmica em Visualização de Informação e que tenham utilizado pelo menos 5 ferramentas distintas de Visualização.

O conjunto de gestos encontrados será implementado em uma ferramenta de Visualização de Informação 3D, sobre a técnica de dispersão de dados, e com os dispositivos de entrada Myo e Leap Motion. A implementação destas além das diretrizes para cada um dos dispositivos de entrada faz uso das características dos sensores envolvidos e tem como base iterações prévias do protótipo apresentado que são relatadas na Seção 2.

O desenvolvimento deste trabalho envolveu a interação com diferentes alunos e professores que desenvolvem trabalhos em áreas correlatas, assim como pesquisa em repositórios científicos e a publicação de artigos.

1.4 Organização do Trabalho

Este trabalho está organizado em cinco seções: 1) Introdução, que apresenta de maneira geral o trabalho; 2) Contextualização, que introduz uma série de conceitos importantes para o trabalho, assim como os trabalhos relacionados; 3) Seleção de Gestos, onde é descrito o processo de descoberta dos gestos e o benchmark dos mesmos; 4) Mapeamento do Vocabulário, onde é descrito as decisões de implementação do vocabulário levantado; e 5) Testes com o Vocabulário Implementado, onde é apresentado os testes da avaliação subjetiva de ergonomia para o vocabulário implementado.

2. CONTEXTUALIZAÇÃO

Esta seção descreve o contexto em que o trabalho está situado e o embasamento para a realização das etapas definidas na Introdução. Aqui são contextualizados as áreas de Visualização de Informação e Interação Não Convencional, bem como são citados alguns trabalhos que ajudaram na realização deste e é brevemente relatado o histórico deste projeto.

2.1 Visualização de Informações

O ser humano adquire mais informação através da visão do que através de todos os outros sentidos combinados. A visão é um sentido que representa não apenas uma função sensorial, mas uma genuína função cognitiva (HAMPSON, 1996). Há uma estreita ligação entre o que o ser humano vê e o que ele pensa.

Esta função está associada ao flexível mecanismo humano de descoberta de padrões, que é essencial para a atividade cognitiva. A partir da visão, é desencadeada a visualização, um processo interno caracterizado pela formação de uma imagem mental (WARE, 2012), que possui como valor o ganho de entendimento e compreensão (SPENCER, 2007). A visualização é independente de computadores, é um processo inato ao ser humano.

Embora a princípio o conceito de visualização se refira a um processo interno, (WARE, 2012) e outros autores impõem uma sobrecarga ao termo. Sob esta perspectiva, visualização é uma representação gráfica de dados ou conceitos, um artefato externo, que funciona como uma ferramenta cognitiva e que possui como objetivos principais: descoberta, tomada de decisão e entendimento.

Visualização de Informação (InfoVis) é o ramo da visualização que mais tem crescido na última década e pode ser definida como a visualização aplicada a dados abstratos com o objetivo de aumentar a percepção do usuários sobre os dados (CHI, 2002). O que caracteriza a especificidade da InfoVis e a diferencia de outros ramos da visualização, como visualização

científica e geovisualização é a natureza abstrata dos dados. A Infovis objetiva auxiliar os usuários a formar modelos mentais para esta classe de dados que não possuem representação física, como forma ou cor, por exemplo, nem espacialidade óbvia (CARD e SHNEIDERMAN, 1999). Como exemplo, podemos citar os dados armazenados no departamento de recursos humanos de uma empresa.

Utilizar o computador não apenas para gerar visualizações, mas também como uma forma de permitir ao usuário adaptar a visualização as suas necessidades, permite explorar melhor o flexível detector de padrões e o adaptativo mecanismo de tomada de decisão humano (WARE, 2012). A união desses dois sistemas através das ferramentas de InfoVis visa proporcionar o aumento da capacidade de comunicação e tomada de decisão.

Para que o usuário possa fazer isso é ideal que as interações sejam simples, não aumentem o esforço cognitivo do usuário, não demandem demasiado esforço físico ou tempo, nem distraiam o usuário da visualização. Preferencialmente, estas devem ser fáceis de aprender e lembrar, precisas e bem definidas o suficiente para que o usuário tenha segurança em sua interação.

2.2 Interação Multimodal

Sistemas multimodais são sistemas computacionais com capacidade de interpretar informação a partir de vários sensores e canais de comunicação distintos (DUMAS e OVIATT, 2009) e interfaces multimodais são interfaces capazes de processar dois ou mais modos de entrada combinados, sendo interfaces inovadoras com objetivo de reconhecer formas e comportamentos mais naturais do ser humano, fazendo uso de tecnologias de reconhecimento (OVIATT, 2003).

(DUMAS e OVIATT, 2009) ressaltam que interfaces multimodais tendem a usar modos de comunicação mais ricos e naturais, como fala e gestos, e que o objetivo destas interfaces é suportar e acomodar as capacidades perceptivas e comunicativas do usuário, bem como integrar ainda mais a computação no mundo real, oferecendo maneiras mais naturais de interação. (DUMAS e OVIATT, 2009) apontam duas características à serem observadas em sistema de interação multimodal: a fusão de diferentes tipos de dados e o processamento em tempo real, assim como o limite de tempo para o processamento da informação.

A priori, a utilização de interfaces multimodais não deve ser justificada unicamente pela eficiência, uma vez que esta depende muito do projeto da interface e a literatura apresenta

resultados pouco expressivos de eficiência de tempo na comparação multimodal vs. unimodal (OVIATT, 2006). O foco na utilização destas interfaces deve ser a possibilidade do usuário realizar tarefas da maneira que se sinta mais confortável.

As vantagens de interfaces multimodais estão baseadas em estudos das áreas de HCI e psicologia cognitiva, como pode ser visto em (BADELLEY, 1992), (OVIATT, 2003) e (OVIATT, 2006). Características positivas da utilização multimodal incluem:

- O fato de que os seres humanos processam diferentes modalidades de interação de maneira paralela, fazendo com que apresentar informações utilizando múltiplas modalidades aumente a efetividade da memória de curto prazo;
- Humanos tendem a reproduzir padrões de interação interpessoal durante a interação multimodal com um sistema;
- A interação multimodal apresenta melhor performance devido a maneira como a percepção, comunicação e memória humana funcionam.

Entretanto, os benefícios acima citados dependem de como a interação multimodal é projetada e implementada na interface multimodal que o usuário vai utilizar. A interface deve procurar reconhecer um conjunto bem definido de interações, que se assemelhem a interação natural e propiciar que o usuário possa ter conhecimento dessas interações sem que isso atrapalhe a sua atividade fim (o motivo pelo qual ele está utilizando aquele sistema), preocupando-se com questões como o cansaço do usuário, a clareza na interação definida e a precisão no reconhecimento destas interações.

Voz e gestos estão entre os meios mais comuns de interação multimodal e equipamentos para o reconhecimento destes meios de interação podem ser encontrados comercialmente com facilidade, a exemplo dos dispositivos Leap Motion (LEAP MOTION, 2016), que faz o rastreamento das mãos baseado em visão, e o Myo (THALMIC LABS, 2016), que realiza o reconhecimento de gestos a partir dos músculos do antebraço, bem como recursos de software, como o Google Speech API (GOOGLE INC., 2013), para reconhecimento de voz.

Tecnologias de rastreamento de gestos podem ser divididas em dois grupos baseados nos sensores utilizados, os que são baseados em visão e não baseados em visão (NOWICKI, PILARCZYK, *et al.*, 2014). Os sensores baseados em visão utilizam imagens para extração de informações. Estas imagens podem ser obtidas por câmeras comuns, por projeção de alguma forma

de luz estruturada ou com apoio de marcadores, como luvas coloridas, luzes de LED ou *qr-codes*. Os sensores não baseados em visão utilizam recursos, como por exemplo acelerômetros, superfícies capacitivas e sensores vestíveis para o reconhecimento de gestos.

As atividades de interação estão agrupadas em 4 categorias principais, que são: dar instruções; conversação; manipulação; e navegação e exploração (LARSEN, 2007). Dependendo somente da GUI para essas interações pode levar o usuário sempre a escolher uma interação em favor de outras, fazendo com que seja interessante apresentar diferentes GUI com a mesma funcionalidade no processo de definir uma interação. Por exemplo, se um botão indica algo e pode ser clicado ou tocado, então o usuário pode entender esse *widget* somente com essa interação, sem a necessidade de usar comandos de voz se estiver disponível. Escolher os gestos e palavras corretos para utilizar uma certa função é necessário para que o usuário se sinta à vontade para alternar interações se necessário.

Meios tradicionais de interação por teclado e mouse são ineficientes em cenários de computação pervasiva (JIANG, 2006). O uso de voz e gestos são uma alternativa ideal para uso dessas aplicações, já que o ambiente natural desse tipo de computação é bem aplicável a interações naturais. O uso de comandos de gestos e voz para jogos também tem ganhando popularidade nos últimos anos bem como tem contribuído para experiências mais imersivas e engajantes, com uso de dispositivos que combinam diferentes sensores para mais de um tipo de modo de interação, como o Kinect.

O trabalho de (BLINOV, DERU e SONNTAG, 2011) utiliza uma mistura de gestos e voz para interações móveis e multimídia. A ferramenta desenvolvida aborda várias interações que utilizam gestos feitos com o polegar do usuário em uma tela de celular em conjunto com comandos de voz, para realizar operações em conteúdo multimídia. Interações que foram realizadas no protótipo são por exemplo, ativar o serviço de reconhecimento de voz e ativar os controles multimídias.

Para esse tipo de interface, hierarquias de menus baseadas em botões na tela de toque não é a melhor opção para uso em carros, que cada vez mais contém funções de assistência, de conforto e entretenimento. O uso de vários menus esconde funções do usuário, que não tem acesso simplificado a elas, em (PFLEGING, SCHNEEGASS e SCHMIDT, 2012) encontrou-se uma solução propondo o uso misto de voz e gestos, neste trabalho um teste foi realizado em um protótipo e dois cenários, um utiliza menus em uma tela de toque, e o outro somente os comandos

de voz e uso de gestos. O resultado mostrou que apesar de interações mais lentas que com o menu, o uso de voz e gestos não atrapalhou o desempenho na direção e a carga visual sentida era menor. Esse fator é muito importante para uma visualização. Esconder *widgets*, criar atalhos, diminuir menus e alternar modos na visualização são ações que se feitas com o uso da voz podem diminuir a carga visual do usuário, deixando mais espaço em tela para focar somente na visualização e possivelmente, se essa combinação for natural para o usuário, permitindo que ele não gaste esforço mental navegando em grandes hierarquias de menus e possa-se concentrar mais na visualização.

Para isso é importante saber como determinado público se sente mais confortável para realizar tipos diferente de interação, seja verificando como os usuários idealizam essa interação antes de disponibiliza-la, ou disponibilizando mais de uma forma de interação para verificar qual a melhor para um determinado contexto, como no trabalho de (MULLER, 1998), onde usuários podem interagir com um sistema de visualização de dados anatômicos em três dimensões derivadas de exames feitos. O usuário pode escolher o modo de interação, se por voz, gestos ou mouse 3D, e então fazer as interações características da sua escolha.

2.3 Trabalhos Relacionados

Esta seção apresenta um conjunto de trabalhos encontrados ao longo desta pesquisa que contribuíram de alguma forma para elaboração do trabalho aqui apresentado. Alguns deles tiveram uma contribuição maior, como por exemplo propondo um método que serviu de base para o processo aqui utilizado na seleção de gestos, enquanto outros apresentam pesquisas correlatas em áreas como projeto de interfaces e interação em ambientes virtuais, que oferecem conceitos importantes e que servem a partes menores deste trabalho, como o projeto de uma interação específica ou a forma como se dá a troca entre funcionalidades no protótipo.

Em (KARAM, 2005) é reunido um conjunto de rótulos para gestos, no intuito de identificar estes e facilitar o tratamento dos mesmos. Estes rótulos são de um ponto de vista descritivo (que descreve o movimento do gesto) ou semântico (que se refere ao que o gesto comunica ou seu proposito), e são uteis na classificação dos gestos para posterior análise.

Os rótulos descritivos são: estáticos, posturas das mãos que não levam em conta o movimento feito por estas; dinâmicos, são gestos que consideram o movimento (como a trajetória das mãos e a mudança de uma postura para outra com o passar do tempo); e espaço-temporal, que é um

subconjunto dos gestos dinâmicos no qual a localização espacial das mãos é importante para a significância do gesto.

Os rótulos semânticos são: emblemas, são símbolos intencionais comunicativos que representam palavras e variam de cultura para cultura (a exemplo do emblema formado por um círculo feito pelo indicador e o polegar com os três demais dedos estendidos, que nos Estados Unidos da América significa OK e no Japão significa dinheiro); proposicionais, gestos intencionais que indicam locais no espaço e é geralmente utilizado para indicar tamanho e movimento; icônicos, etc.

Em (NIELSEN e GRANUM, 2003) é proposto um procedimento para encontrar um conjunto específico de gestos naturais e apropriados para que uma interface específica seja mais eficiente, o trabalho de (NIELSEN e GRANUM, 2003) é amplamente discutido mais a frente sendo mais interessante aqui apontar as diferenças entre ele e o trabalho aqui desenvolvido. Este trabalho serve como base para a pesquisa aqui desenvolvida, mas com algumas diferenças consideráveis. Ao invés de procurar um conjunto de gestos apropriados para uma interface específica, este trabalho visa abordar o procedimento para escolha de gestos de modo que vocabulário de gestos resultante seja o mais natural possível para um determinado perfil de usuário em um conjunto de funcionalidades conhecidas, com base na história do usuário, em tarefas que ele conhece e sabe quando devem ser realizadas em um determinado contexto, em ferramentas que este já tenha utilizado e principalmente em como estes usuários idealizam a interação destas funcionalidades através de gestos.

Em (SEIXAS, CARDOSO e DIAS, 2015) foi conduzido um estudo sobre a seleção utilizando o dispositivo Leap Motion, comparando esta interação entre uma e duas mãos, apontando que possivelmente a interação quando realizada com duas mãos é mais rápida e precisa. Esse achado foi relevante neste trabalho pois sugere que pode-se dar prioridade para mapeamentos de gestos que utilizem o Leap Motion a serem realizados com duas mãos, caso o levantamento com usuários não indique preferência contrária.

Em (VELLOSO, TURNER, *et al.*, 2015) ressaltasse que gestos manipulativos isomórficos são mais naturais ao interagir em ambientes 3D, no entanto estes são afetados por fatores limitantes tanto por parte do dispositivo de entrada (como por exemplo o alcance do campo de visão e a oclusão, para dispositivos baseados em visão) quanto por limitações humanas (como por exemplo

o alcance do braço). Estes fatores ressaltam a importância de favorecer gesto do vocabulário que não forcem os limites de movimentação e posicionamento das mão, pulsos e braços, bem como sugere um mapeamento de interação que preferencialmente não dependa de elementos possam estar fora do alcance da representação virtual da mão (ou seja, que não dependam da profundidade do elemento manipulado).

Em (CABREIRA e HWANG, 2015) o uso de gestos ao ar livre é comparado entre diferentes dispositivos. Este estudo demonstra que entre aplicativos comercialmente disponíveis há uma repetição nos gestos que compõem o vocabulário destas ferramentas e entre estes há predominância de gestos cujo reconhecimento fazem parte das APIs destes dispositivos, indicando uma inclinação para escolha de gestos que são mais fáceis de serem reconhecidos. Esses dados serve de indício de que mesmo havendo outras propostas para seleção de gestos, selecionar um vocabulário que seja de fácil reconhecimento e atribuir a estes gestos funcionalidades ainda é uma abordagem largamente utilizada, mesmo que está abordagem não tenha como foco a naturalidade do vocabulário.

Em (KULSHRESHTH e LAVIOLA JR, 2014) é apresentado um modo de seleção de menus para interação ao ar livre baseado no mapeamento entre uma numeração na interface gráfica com os número de dedos esticados na mão do usuário. Técnicas como estas são interessantes por serem fáceis de serem assimiladas, representar com as mãos o número do menu a ser selecionado, e devem ser fortemente consideradas ao mapear gestos para uma implementação específica. Entretanto estas abordagens não devem ter precedência sobre gestos que sejam bem avaliados como intuitivos, nem devem ser a primeira escolha em interfaces que sejam muito preenchidas com textos, uma vez que a inclusão dos valores de seleção podem vir a confundir ou atrapalhar o usuário de ler o texto normal da interface.

Em (SADANA e STASKO, 2014) é apresentado um sistema de visualização de dados abstratos em dispositivos com telas multi-toque, onde ressalta a presença da técnica de dispersão de dados em diversas ferramentas de Visualização de Informação e em outros contextos de análise de dados, além da simplicidade da técnica ele ressalta que ela representa cada item de dados individualmente e é uma técnica boa para ressaltar agrupamentos, pontos discrepantes e tendências dentro dos dados analisados. O trabalho apresentado também faz uso de interfaces conhecidas de visualização para escolher as funcionalidades e a apresentação destas na interface de usuário, reforçando também

que a familiaridade com o modo de interação tem impacto sobre a experiência do usuário com a ferramenta.

Em (BRATH, 2014) é apresentado casos de utilização de Visualização de Informação em terceira dimensão e discuti sobre a efetividade destes, indicando alguns pontos que devem ser trabalhados no uso de visualização 3D. A considerações de (BRATH, 2014) são importantes para este trabalho, especialmente no tocante a navegação e comunicação da interface, para trabalhos futuros sobre o protótipo apresentado neste trabalho.

Em (BRUDER, STEINICKE e STURZLINGER, 2013) é apresentado uma comparação entre toques 2D e interações 3D ao ar livre em um experimento baseado na Lei de Fitts. Os resultados indicam que toques ao ar livre são melhores para seleção de objetos muito grandes ou renderizados com alguma profundidade, sendo mais adequados para uma grande variedade de situações em ambientes 3D. Para um ambiente de InfoVis essa situação de interagir com objetos em profundidade é esperada, graças a interações espaciais como zoom, pan e rotação.

Em (JANKOWSKI e HACHET, 2013) é apresentado um *survey* sobre técnicas de interação em ambientes 3D, entre estas estão as técnicas primárias de movimentação de câmera para ferramentas 3D (pan, zoom e rotação). Entre outras questões é indicado o quão pouco otimizada é a interação por mouse para se navegar e trocar constante mente de modo de navegação, sugerindo gestos com o mouse para diminuir o esforço de navegação da interface. Entretanto, quando o modo de interação padrão é baseado em gestos, especialmente ao ar livre, esse tipo de “atalho” entre os modos de interação pode ocasionar colisões (interpretação errônea ou não planejada de um gesto), levando a considerar outro método (modo de interação) para realizar este “atalho”.

Em (BREHMER e MUNZNER, 2013) ressaltasse a necessidade a necessidade de classificar a interação em InfoVis, visando estabelecer um modo de projetar, avaliar e comparar diferentes tipos de interações com diferentes graus de abstração em diferentes ferramentas de InfoVis. Uma definição clara da interação e do grau de abstração desta é importante neste trabalho pois o processo aqui apresentado visa ser reutilizável não apenas para a definição de diferentes vocabulários, mas para a comparação entre estes, evitando, por exemplo, a comparação entre dois vocabulários de gestos que se refiram a graus de abstração distintos.

Em (REN e O'NEILL, 2013) são propostas algumas sugestões para o projeto de interações ao ar livre a partir da análise de três estudos conduzidos em interações ao ar livre em ambientes 3D.

Uma das sugestões apresentadas é a quebra de tarefas que exigem precisão em um menor grau de liberdade, como utilizado neste trabalho na quebra de interações em tarefas auto contidas e no rastreamento de movimentos apenas nas direções necessárias.

2.4 Trabalhos Anteriores

Em iterações prévias deste trabalho, um processo *ad-hoc* foi utilizado para seleção de gestos utilizando como dispositivo de entrada o Leap Motion em um protótipo inicial, a escolha dos gestos foi feita em múltiplas etapas, através do refinamento dos gestos e testes preliminares com usuários.

O Unity 3D foi utilizado para o desenvolvimento desse protótipo, para gerar a visualização 3D de um Scatterplot. A interação através de gestos foi realizada inteiramente pelo Leap Motion, sendo responsável pela captura e rastreamento dos gestos. As funcionalidades implementadas para o protótipo foram as seguintes: Zoom, Pan, Seleção, Rotação e Filtro Categórico. Cada funcionalidade era acionada através de um menu inicial, e gestos diferentes eram utilizados para cada interação.

O trabalho em (CARNEIRO, MIRANDA, *et al.*, 2015) apresenta a arquitetura desse protótipo, em conjunto com os gestos selecionados para essa interação. Em (MIRANDA, SANTOS, *et al.*, 2016) é realizada uma avaliação de usabilidade do protótipo, utilizando tarefas da visualização no protótipo, identificando a facilidade de utilização de cada gesto nas interações, e no trabalho de (MIRANDA, CARNEIRO, *et al.*, 2016) foi realizada uma classificação de problemas encontrados no uso de gestos em aplicação de InfoVis, sendo esse protótipo a ferramenta que foi testada.

3. SELEÇÃO DE GESTOS

A definição de uma interface de interação baseada em gestos em geral é realizada por uma de duas abordagens (NIELSEN e GRANUM, 2003), a abordagem técnica e a abordagem humana. A abordagem técnica se baseia na escolha de gestos que sejam simples de serem reconhecidos pelo computador, ou seja, essa abordagem considera mais importante a taxa de reconhecimento da interface do que a intuitividade e naturalidade dos gestos realizados, bem como não considera, ou considera em segundo plano, aspectos como a facilidade de se fazer ou manter um gesto, nem se existe uma correlação lógica entre um gesto e uma ação realizada na interface por aquele gesto (NIELSEN, STÖRRING, *et al.*, 2004). Ainda que pareça a pior abordagem do ponto de vista do conforto do usuário, está é a abordagem mais presente em interfaces de gestos no ar livre, o que é indicado em (CABREIRA e HWANG, 2015), com um levantamento que indica que em aplicativos disponíveis para 3 dispositivos de entrada por gestos no ar livre a maior parte da interação é dada por gestos padrões, previamente disponíveis nas APIs destes dispositivos e que tendem a reproduzir toques em tela.

A abordagem humana se baseia na investigação dos usuários que utilizarão a interface, ou seja, utiliza o comportamento do usuário para modelar interações que sejam semelhantes aquele comportamento para que o usuário tenha um menor esforço ao se familiarizar com a interface em questão, considerando também os fatores ergonômicos desta interação. Esta abordagem não considera, a priori, a facilidade/dificuldade de um sensor disponível em reconhecer os gestos determinados para interação. A abordagem humana pode ser top-down, partir tanto da geração de um conjunto de gestos possíveis seguidos da utilização de métodos para refinar a seleção destes gestos para as tarefas que devem ser mapeadas (SCHIELE e VAN DER HELM, 2006), ou bottom-up, que consiste em definir as funcionalidades que serão utilizadas através de gestos e utilizar métodos para definir os gestos apropriados para estes (NIELSEN, STÖRRING, *et al.*, 2004).

Ainda que a abordagem humana seja melhor em relação ao conforto do usuário, utilizar apenas esta abordagem pode levar a escolha de gestos que: não possam ser reconhecidos pelos dispositivos de reconhecimento disponíveis, que sejam constantemente confundidos com outros gestos significativos ou que são reconhecidos em taxa muito baixa, efeitos estes que possivelmente levariam ao desconforto, esforço excessivo e frustração por parte dos usuários.

Em (NIELSEN, STÖRRING, *et al.*, 2004) é definido um procedimento para a definição de um vocabulário de gestos e benchmark do vocabulário definido para uma aplicação. Este procedimento é baseado em 4 passos: encontrar as funções que serão executadas através de gestos, coletar gestos dos usuários, extrair o vocabulário de gestos e analisar o vocabulário. Estas etapas (E1 à E4) são descritas a seguir.

E1. Encontrar as Funções. Definir quais funções em um sistema serão comunicadas através de gestos. Para tal devem ser considerados:

- os padrões de interfaces para aquelas funcionalidades em ferramentas conhecidas para o grupo de usuários. Isso visa utilizar a familiaridade do usuário com ferramentas, ou modos de interação, pré-existentes para que este se sinta mais seguro ao interagir de uma nova forma, como por exemplo: ter em mente os componentes de interface das ferramentas de CAD ao definir as funções realizadas por gestos para uma ferramenta de engenharia civil;

- manter o vocabulário de gestos mínimo. Dar preferência por componentes de interface que funcionem ou sejam operados de maneira semelhante, para que um gesto possa realizar mais de uma funcionalidade em diferentes contextos, o que demandaria menor esforço por parte do usuário em lembrar os diferentes gestos.

E2. Coletar Gestos dos Usuários. De posse das funcionalidades que serão utilizadas através de gestos, é necessário definir quais são os gestos apropriados para cada funcionalidade. Isso deve ser realizado através de experimentos com usuários, onde estes serão apresentados a cenários em que deverão realizar os gestos que utilizariam para comunicar as funções definidas enquanto são gravados (em vídeo) para posterior análise dos gestos. Para isso é importante considerar:

- projetar o experimento de forma que a utilização dos gestos ocorra naturalmente, especialmente se o usuário for técnico, para evitar que este ajuste seus gestos em termos de algoritmos ou interfaces;

- se o cenário do experimento for de aspecto técnico, considerar o uso de experimentos Mágico de Oz (HUMMELS e STAPPERS, 1998) (BERINGER, 2001), que além de testar os gestos permitiria avaliar outros aspectos da interface como o sequenciamento e o feedback oferecido por ela;

- o número de usuários necessários para o experimento depende do quão grande e diverso é o público alvo da interface sendo desenvolvida e do quão diverso serão os resultados dos testes.

E3. Extração do Vocabulário de Gestos. De posse dos vídeos dos usuários deve-se fazer a análise e extração do vocabulário de gestos. Verificar que gestos foram utilizados, o quão consistente cada gesto foi relacionado a uma funcionalidade, se foi realizado com uma ou duas mãos e se este é uma pose estática ou a movimentação realizada nele é importante para a interpretação do gesto. Nesta etapa deve-se considerar:

- força requerida pela postura;
- frequência e duração dos gestos;
- considerar os efeitos no pulso, baseado nas posições do pulso e dedos.

E4. Análise do Vocabulário. Por fim é necessário testar o vocabulário de gestos extraídos na etapa 3, identificando se a escolha do vocabulário foi apropriada e, em caso negativo, fazendo ajustes ao vocabulário. Esta etapa é dividida em 3 fases onde cada uma gera uma pontuação para o vocabulário definido:

E4.1. Adivinhar a Função: prover ao usuário uma lista de funções disponíveis e apresentar o vocabulário de gestos pedindo que o usuário adivinhe (“chute”), para tal gestos que dependam de um contexto específico devem ser apresentados neste contexto. A pontuação nessa fase é o número de erros dividido pelo número de gestos.

E4.2. Memória: apresentar ao usuário para que este conheça o vocabulário de gestos. Depois disso apresentar o nome das funcionalidades em sequência e de forma rápida, dois segundos por função, intervalo no qual o usuário deve realizar o gesto referente aquela função. Cada vez que o usuário errar o gesto o vocabulário de gestos deve ser reapresentado e a sequência deve ser reiniciada até que todos os gestos sejam realizados corretamente. A pontuação desta fase é dada pelo número de reinícios.

E4.3. Stress: esta fase é uma avaliação subjetiva de ergonomia. Uma lista com os gestos do vocabulário deve ser apresentada aos usuários e este deve realizar os gestos na lista X vezes (onde $X = \text{floor}(200 / \text{Tamanho do Vocabulário})$), sempre retornando à posição neutra entre um gesto e outro. Encorajar que o usuário faça observações sobre a realização dos gestos e registra-las. Ao final pedir que o usuário de uma nota para cada gesto, bem como para a sequência, de acordo com o critério a seguir:

- 1) Confortável
- 2) Pouco Cansativo/Estressante
- 3) Cansativo/Estressante
- 4) Muito Irritante
- 5) Impossível

Como a descrição das etapas deixa aparente, uma pontuação menor indica vocabulário de gestos melhor. (NIELSEN, STÖRRING, *et al.*, 2004) propõem a utilização da Etapa 4 tanto para avaliação de um único vocabulário de gestos como para realizar a comparação entre 2 ou mais vocabulários. Entretanto (NIELSEN, STÖRRING, *et al.*, 2004) não propõem um valor definido para uma pontuação “boa” ou “ruim” na avaliação. Esse procedimento é baseado na abordagem humana e serve como base para o processo de seleção de gestos utilizada neste trabalho.

O processo apresentado neste trabalho diferencia-se do apresentado em (NIELSEN, STÖRRING, *et al.*, 2004) pois considera que para gerar um vocabulário de gestos natural para um grupo específico definido pelo seu passado acadêmico e principalmente com o uso de ferramentas de visualização de informação não se pode excluir o conhecimento destes em como tais interfaces funcionam, pois é esse conhecimento que define o que eles esperam quando pretendem interagir de determinada forma com uma visualização. A Figura 1 mostra o processo adaptado onde cada retângulo é uma etapa, ou sub etapa, do processo, retângulos azuis são reproduções do presentes em (NIELSEN, STÖRRING, *et al.*, 2004), os retângulos em vermelho são etapas que já existiam mas foram modificadas para se adequar a proposta deste trabalho e os retângulos em verde são etapas inteiramente novas, propostas no intuito de adequar o processo.

Além disso a proposta nesse trabalho é descobrir um vocabulário ideal baseado nas funcionalidades que se visa realizar e então propor uma implementação, que tente ser fidedigna

aos gestos ideais encontrados, nos dispositivos de entrada disponíveis, enquanto no referido trabalho o processo busca encontrar um vocabulário dependente de interface. Além disso, este trabalho propõem uma etapa documentada de mapeamento dos gestos levantados para gestos implementados, bem como critérios de escolha e desempate na avaliação dos gestos.

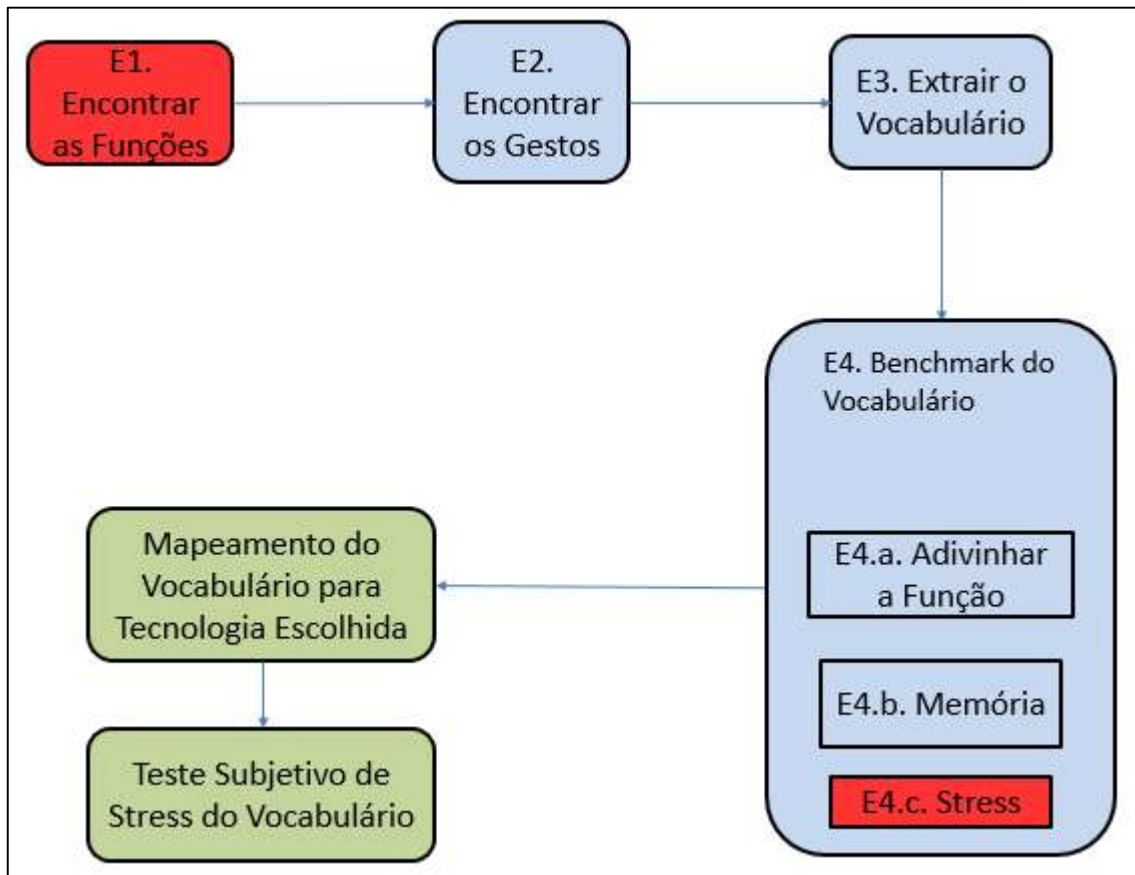


Figura 1: Processo adaptado de levantamento e benchmark dos gestos.

Para Etapa 1 (E1) as funções selecionadas se baseiam no nível 2 da pirâmide de (SEDIG e HAWORTH, 2014), vista na Figura 2, onde as interações com Visualizações são classificadas de acordo com seu grau de abstração indo do nível 1 (mais concreto) ao nível 4 (mais abstrato), ou seja, interações em nível micro que contém mais de um evento. As interações de nível 2 que serão tratadas são especificamente as interações de zoom, pan e rotação, pertencentes a *Explorar*, bem como filtro contínuo e filtro categórico, pertencente a *Filtrar*, com base nas categorias de (YI e JACKO, 2007).

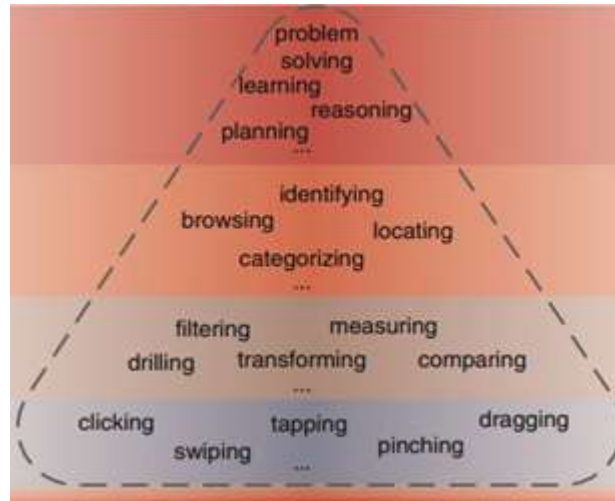


Figura 2: Pirâmide de abstração de interações.

Para a Etapa 2 (**E2**) o experimento será gravado em vídeo ao mesmo tempo em que um observador toma notas do usuário comunicando os gestos a um segundo observador, que se comunicará com o usuário ao longo do experimento para explicar como este acontecerá e para sanar dúvidas sobre determinada comunicação. Ao contrário do proposto por (NIELSEN, STÖRRING, *et al.*, 2004), de não ter uma identificação técnica dos gestos, o objetivo é identificar como os usuários imaginam que seria a interação ao ar livre com funcionalidades e *widgets* conhecidos de ferramentas de visualização de informação, visando evitar ou minimizar que o usuário imagine a interação que irá comunicar para um *widget* específico, serão apresentados ao usuário pelo menos 5 *widgets* ou mecanismos de interação para cada categoria, visando identificar como o usuário imagina gestos ao ar livre para aquela funcionalidade, e.g. realizar um filtro contínuo, ao invés de para uma implementação específica de uma funcionalidade.

A cada teste da **E2** o observador deve informar ao usuário que:

- serão apresentados ao usuário 5 *widgets* ou mecanismos de interação por funcionalidade, durante 30 segundos, e este pode pedir esclarecimentos sobre os apresentados;
- após a apresentação de cada funcionalidade o usuário deverá realizar o gesto ao ar livre referente aquela funcionalidade e modalidade, como por exemplo, o gesto para a funcionalidade zoom tem as modalidades *in (aproximar)* e *out (afastar)*;
- os testes serão gravados e que o observador pode requerer esclarecimentos sobre os gestos em qualquer momento do teste.

As funcionalidades e modalidades são listadas na Tabela 1 abaixo:

Tabela 1: Funcionalidades exploradas neste trabalho.

INTERAÇÃO	FUNCIONALIDADE
Explorar	Pan
Explorar	Zoom
Explorar	Rotação
Filtrar	Contínuo
Filtrar	Categórico

É importante ressaltar que testes Mágico de Oz não foram realizados, mesmo considerando o aspecto técnico do experimento, devido ao fato de que este trabalho não visa analisar uma ferramenta/interface em específico, mas sim encontrar o conjunto de gestos que representem de forma intuitiva funcionalidades bem definidas na interação com ferramentas de visualização de informação. Dessa forma um teste Mágico de Oz possivelmente faria o usuário se focar nos *widgets* apresentados na interface, em detrimento da funcionalidade daquele mecanismo.

Para a coleta de gestos foram utilizados 8 usuários, todos com experiência no uso de ferramentas de visualização de informação, experiência com pelo menos 5 diferentes ferramentas de visualização que apresentem as interações de *Filtrar* e *Explorar*, e tendo experiência acadêmica na área, seja participando de curso de 6 meses em Visualização de Informação ou tendo um ou mais artigo(s) científico(s) publicado(s) na área.

Na etapa 3 (E3) é realizada a análise dos vídeos gravados, considerando também as anotações dos observadores. Para tal, cada gesto diferente encontrado é preenchida a tabela abaixo:

Tabela 2: Tabela de descrição de gesto observado.

ATRIBUTO DO GESTO	VALOR DO ATRIBUTO
Interação	Explorar/filtrar
Funcionalidade	Pan/Zoom/Rotação/F. Contínuo/F. Categórico

Movimento	Estático/dinâmico
Movimento significativo	Sim/não
Número de mãos	1/2
Frequência	Valor inteiro.
Descrição	Descrição dos pontos principais que caracterizam o gesto.

Vale ressaltar que por vezes alguns participantes realizavam e descreviam diferentes gestos para uma mesma interação. Na ocorrência destes os mesmos eram questionados sobre se um dos gestos era o ideal na visão do participante, caso o participante indicasse um gesto como ideal apenas este era contabilizado, caso o participante afirmasse que todos os gestos apresentados eram ideais para funcionalidade, este era inquerido sobre que condições o fariam escolher por um ou outro gesto, e todos os gestos descritos eram contabilizados.

A descrição dos gestos utiliza referências espaciais de sentido (p.ex. à direita, à esquerda, acima e abaixo) que dizem respeito ao usuário e sua percepção, bem como de proximidade (p.ex. afastar, aproximar), que se referem à tela onde a visualização e os *widgets* são renderizados.

Uma vez preenchidos as tabelas para cada gesto, a maior frequência para cada funcionalidade foi selecionada como parte do vocabulário de gestos. Os casos onde a maior frequência encontrada ocorreu em 2 são descritos após a Tabela 3, que apresenta o resultado do processamento feita na Etapa 1.

Tabela 3: Resultado da Etapa 1, frequência por gesto.

		Gesto	sobre a ...	Taxonomia	Movimento	Frequência
Explorar	Pan	"agarrar, arrastar e soltar"	visualização	manipulativo	sim	3
		mover a mão sobre uma indicação da direção	widget/visualização	dêdético	não	3
		apontar com 4 dedos em uma direção	visualização	semafórico	não	1
		swipe na direção desejada	visualização	manipulativo	sim	1
	Zoom	"pinçar", aproximando (out) ou afastando (in)	visualização	manipulativo	sim	4
		aproximar (in) ou afastar (out) o indicador esticado	visualização	manipulativo	sim	2
		aproximar (in) ou afastar (out) a mão aberta	visualização	manipulativo	sim	1
		braço estendido com a palma da mão para baixo (in) ou para cima (out)	visualização	semafórico	não	1
		iconico de "segurar" uma bola e rotaciona-la	ambos	manipulativo	sim	7
	Rotação	apontar o eixo	widget e visualização	manipulativo	sim	2
apontar eixo e swipe para rotacionar		widget e visualização	manipulativo	sim	1	
Filtar	Contínuo	"agarrar" o range e arrastar horizontalmente	widget	manipulativo	sim	4
		"pressionar" o thumb e "arrastar"	widget	manipulativo	sim	3
		"agarrar" e mover o thumb	widget	manipulativo	sim	2
		"pinçar" e mover o thumb	widget	manipulativo	sim	2
		"arrastar" o thumb com a mão aberta	widget	manipulativo	sim	1
		"pressionar" o range e "arrastar"	widget	manipulativo	sim	1
		"pinçar" um widget radial onde o raio define o valor selecionado	widget	manipulativo	sim	1
		Categórico	"tocar" no componente de seleção	widget	dêdético	sim
	hand swipe, para cima e para baixo		widget	manipulativo	sim	3
	"tocar" em um componente e "arrastar" o dedo sobre outros para selecionar múltiplos		widget	dêdético	sim	3
	"agarrar" e mover sobre os componente selecionáveis		widget	manipulativo	sim	2
	posicionar a mão mais acima ou abaixo para indicar para onde os componentes devem se mover		widget	dêdético	não	1
	esticar os dedos sobre um selecionável para selecionar e punho fechado para desselecionar		widget	semafórico	não	1

Visando ter uma descrição menor e de legibilidade mais rápida, alguns termos foram utilizados na tabela para descrever um pequeno conjunto de gestos e/ou movimentos. O significado destes termos é detalhado a seguir:

- “agarrar”: é o gesto de, com a mão em posição neutra, fechar os dedos como um punho fechado.
- “soltar”: gesto de, com o punho fechado, abrir os dedos para posição neutra ou para posição de dedos esticados.
- “arrastar”: gesto de mover a mão, ou as mãos, tendo previamente ativado um componente a ser arrastado, geralmente precedido pelo gesto “agarrar” e sucedido pelo gesto “soltar”.
- “pinçar” é o gesto de aproximar ou afastar os dedos indicador e polegar, voltando para um estado neutro, polegar e indicador fechados/recolhidos, entre um gesto e o seguinte.
- “segurar” é o gesto de colocar os dedos da mão semiabertos, como se segurando uma bola imaginaria na palma da mão.
- “tocar” é o gesto de esticar um dedo sobre um objeto virtual para tocá-lo e em seguida recolher ou afastar o dedo esticado.
- “pressionar” é o gesto de manter o toque, ou seja, “tocar” e manter o dedo aproximado para uma interação contínua.

Baseado no conceito de gestos naturais e intuitivos (GRANDHI, JOUE e MITTELBERG, 2011), onde um gesto intuitivo e natural é um gesto espontâneo e feito com frequência para uma tarefa comum. Como pediu-se aos participantes da etapa um que demonstrassem os gestos ideais para cada funcionalidade na opinião dele(a), assume-se que os gestos coletados são espontâneos. A frequência de cada gesto é registrada para descobrir que gesto se repete mais para cada funcionalidade, visando aproximar-se do conceito de (GRANDHI, JOUE e MITTELBERG, 2011). É perceptível que esta abordagem não gerará um conjunto perfeito de gestos, uma vez que diferentes participantes podem demonstrar diferentes gestos, entretanto ao selecionar os gestos que mais se repetem existe a tendência de selecionar os gestos mais adequados para a maior parte do perfil de usuários, gerando um esforço de treinamento menor ao apresentar os gestos para os usuários.

Os gestos na Tabela 3 foram organizados de forma ordenada do gesto de maior frequência para o gesto de menor frequência para cada funcionalidade. Para cada funcionalidade foram escolhidos os gestos que apresentaram maior frequência, lembrando que algumas funcionalidades são representadas através de mais de um gesto significativo (como os filtros categóricos e contínuos), que representam diferentes interações, como a navegação em *widgets* e seleção múltipla.

Antes de listar os gestos selecionados é necessário eliminar as pluralidades geradas pelo critério da frequência. Estas ocorrem em duas situações, na funcionalidade de *pan* e na funcionalidade de filtro categórico. Para eliminar as repetições foram aplicados os critérios de repetição (o quanto aquele gesto aparecia em outras funcionalidades já definidas), visando minimizar o vocabulário de gestos. Definiu-se ainda que na ocorrência de novo empate entre dois gestos, utilizar ia-se a sequência de critérios de desempate a seguir: análise da ergonomia do gesto (onde o escolhido seria aquele que causasse o menor grau de stress), tipo de taxonomia ((KARAM, 2005) classifica os tipos de gestos considerando características da comunicação presentes nestes gesto, assim escolher gestos de uma mesma classificação para múltiplas funcionalidades faria o vocabulário como um todo ter características semelhantes) e por fim o objeto de interação (gestos podem ser realizados com o auxílio de *widgets* ou sobre/para a visualização, como interações diretamente sobre a visualização são uma tendência na área, gestos sem o auxílio de *widgets* tem preferência para a entrada no vocabulário). Caso a aplicação sequencial destes critérios não fosse capaz de definir um gesto a abordagem *ad-hoc* seria utilizada para o desempate.

Para a funcionalidade de *Pan*, dois gestos registraram a maior frequência (três), sendo eles “agarrar, arrastar e soltar” (gesto de pan 1 – gp1) e “mover a mão sobre uma indicação de direção” (gesto de pan 2 – gp2). Seguindo os critérios para eliminar repetições definidos acima, procurou-se similaridades entre estes gestos (gp1 e gp2), onde verificou-se que gp1 é similar ao gesto de maior frequência para funcionalidade de mover o intervalo no filtro contínuo e gp2 não tem similares entre os gestos mais frequentes, fazendo com que a duplicidade pudesse ser resolvida com o primeiro critério de desempate.

Dessa forma o vocabulário de gestos selecionado na Etapa 3 é apresentado a seguir, vale ressaltar que alguns gestos podem ser realizados sobre a visualização enquanto outros são realizados sobre componentes auxiliares, o termo “sobre” significa que a representação da mão do usuário fica (ou seja, é renderizada) numa área da interface, seja a da visualização ou dos *widgets* de controle:

- “Agarrar, arrastar e soltar”, para a funcionalidade de Pan. O gesto é realizado diretamente sobre a visualização.

- “Pinçar”, para a funcionalidade de Zoom, aproximando para zoom-out e afastando para zoom-in, analogamente a mesma interação com imagens em smartphones ou tablets. O gesto é realizado diretamente sobre a visualização.

- “Segurar” uma bola imaginária e rotacioná-la, seja sobre a visualização ou sobre um componente, sendo que este acompanha a rotação da mão.

- “Agarrar” um intervalo pré-definido e arrastá-lo para funcionalidade de filtro contínuo, para o posicionamento do intervalo. O gesto é realizado sobre um *widget* de configuração de intervalo (*range slider*).

- “Pressionar, arrastar e soltar” o *thumb*, para definir um valor de filtro contínuo. O gesto é realizado sobre um *widget*, seja de configuração de limite (*slider*) ou de intervalo (*range slider*).

- “Tocar” no componente, para selecionar ou des-selecionar um *widget* de filtro contínuo. O gesto é realizado sobre uma representação de um item em uma lista (*checkbox*).

- “*Swipe*” para cima ou para baixo, para navegar na lista de filtros categóricos respectivamente para cima e para baixo. O gesto é feito sobre um *widget* de lista navegável verticalmente (*scroll*).

- “Pressionar e arrastar” sobre a lista referida nos dois itens acima para selecionar múltiplos componentes de filtro categórico. O gesto é realizado sobre um *widget* ao “pressionar” e sobre outros *widgets* presentes na mesma lista ao arrastar.

Com o vocabulário de gesto definido passou-se para Etapa 4, onde utilizou-se um novo grupo de 8 participantes com o mesmo perfil do primeiro grupo para realizar um *benchmark* sobre o vocabulário.

Devido a nenhuma imagem de componente ser utilizada para ilustrar o contexto de nenhuma funcionalidade, apenas descrições faladas dos contextos, definiu-se como o score desejável para a fase de adivinhação do benchmark o valor menor ou igual a 0.3125 o que indica uma média de dois erros por usuário ao tentar adivinhar as funções. Esse valor foi definido a partir da premissa de que gestos realizados sobre *widgets* são mais difíceis de adivinhar uma vez que estes *widgets* não são mostrados para o usuário, apenas é indicado que aquela interação ocorre sobre um ou mais componente(s) auxiliar(es) e o comportamento destes (se é móvel, se muda de estado, etc.), o que compromete o contexto daquele gesto. Foi definido arbitrariamente que uma quantidade esperada de erros seria a metade dos gestos realizados sobre *widgets* e que a nota esperada para a fase de adivinhação, dado pela fórmula (número de gestos sobre *widgets* / 2) / número de gestos, que resulta em $(5/2)/8 = 0.3125$.

Para a fase de avaliar o quanto o vocabulário escolhido é fácil de ser lembrado, a meta do vocabulário foi estabelecida seguindo as seguintes hipóteses:

- o tempo de dois segundos para representar cada gesto causa alguma pressão sobre o usuário;

- funcionalidades sobre a visualização são menos afetadas pela pressão do tempo, pois não envolvem sequências de gestos para manipulação de *widgets*;

- mesmo que funcionalidades que sejam composta por mais de um gesto tenham mais tempo para serem expressas (2 segundos por gesto), estas serão mais propensas a gerar recomeços, devido à pressão do tempo para lembrar um maior número de gestos associados a funcionalidade.

Dessa forma, como duas funcionalidades utilizam mais de 1 gesto, é esperado que cada uma dessas funcionalidades gere um reinício, ou seja, a meta para esta fase é 2.

Por fim, para terceira fase foi definido um score meta de 2, devido a dois fatores: aos usuários foi pedido que realizassem os movimentos da forma ideal para eles (o que teoricamente seria um movimento confortável), ainda que os usuários não tenham um perfil de profissional ou estudante da área de ergonomia; e o processo escolhido requer a repetição dos gestos em sequência, 200 vezes, o que se espera torne gestos que são confortáveis em cansativos, devido a repetição.

Considerando a área de visualização de informação e a finalidade do vocabulário de gestos ser utilizado para a interação com ferramentas de visualização, onde o comportamento do usuário é em geral cíclico (formado pela análise da visualização, seguida de interação para buscar ou ressaltar uma informação, e novamente análise, interação, etc.) sendo que o tempo de gasto com a análise costuma ser maior que o dispendido com interação, fazendo com que nas fases 1 e 2, que testam intuitividade e memória, os scores meta sejam menos flexíveis, uma vez que a dificuldade em lembrar de um gesto pode atrapalhar o processo cognitivo do usuário na interpretação da visualização.

A Tabela 4 apresenta as médias esperadas e registradas para cada fase:

Tabela 4: Notas médias esperadas e registradas em cada fase.

	Média esperada	Média registrada
Fase 1	0,31	0,23
Fase 2	2	1,25
Fase 3	2	1,79

Em média, todas as fases registraram scores melhores que o esperado, no entanto a média verificada nas etapas prove informações limitadas para análise do resultado. A Tabela 5 mostra além da média para cada fase, o intervalo de valores registrados e o desvio padrão dos dados, evidenciando que os dados obtidos no teste estão bastante distribuídos no intervalo registrado, especialmente nas fases 1 e 2.

Tabela 5: Intervalo, média e desvio padrão encontrado por fase.

	Intervalo	Média	Desvio Padrão
Fase 1	0 - 0,37	0,23	0,14
Fase 2	0 - 3	1,25	1,16
Fase 3	1 - 4	1,79	0,5

Na Fase 1 os resultados encontrados em número de erros por funcionalidades são mostrados na Figura 3.

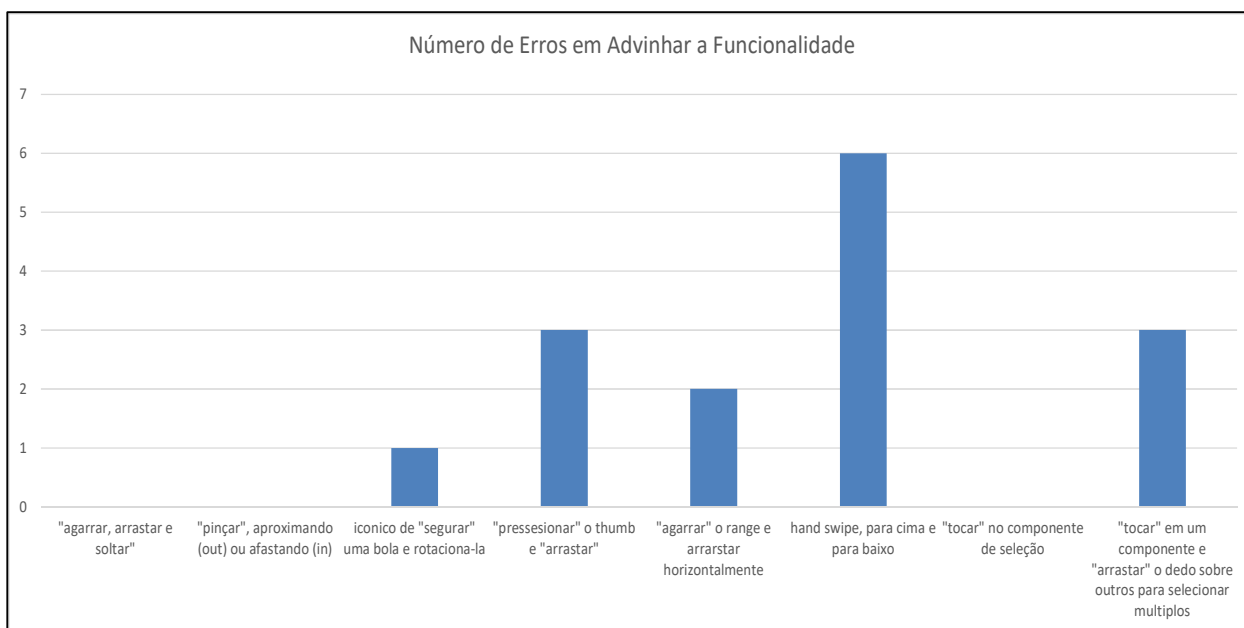


Figura 3: Número de erros por funcionalidade.

Os primeiros gestos, que são realizados inteiramente sobre a visualização, foram adivinhados corretamente 100% das vezes, indicando que a representação do gesto associado a descrição “Este gesto é realizado diretamente sobre a visualização.” é suficiente para os usuários participantes criarem uma relação entre o gesto e uma funcionalidade corretamente.

Outro gesto que foi corretamente adivinhado todas as vezes foi o gesto de esticar a mão e um dedo para “tocar” um widget de seleção, o que é associado a este gesto ser muito comum para o ato de selecionar, tanto em outras interfaces (como na seleção em tablets ou celulares, computadores com recurso de touchscreen e até mesmo a utilização do mouse de apontar e clicar é uma mimetização) quanto no dia a dia para indicarmos com o dedo, e até tocarmos, a escolha de uma coisa. Apesar deste gesto não ocorrer sobre a visualização ele é um dos gestos mais naturais no vocabulário, sendo encontrado, bem como algumas variantes, em outros estudos.

O gesto para a funcionalidade de *zoom*, que era acompanhado da descrição “Este gesto pode ser realizado diretamente sobre a visualização quanto sobre um *widget* auxiliar, tendo o mesmo efeito e funcionalidade.”, foi adivinhado erroneamente apenas uma vez, onde o participante indicou que o gesto seria de *pan*.

Dessa forma percebe-se que 92% dos erros estão em gestos nos quais as interações se dá diretamente sobre *widgets*, sendo que quase metade destes (46%) foram palpites errados sobre o

gesto de “*hand swipe para cima ou para baixo*”, para navegar na lista de filtro categórico, o que representa o dobro do próximo gesto com maior número de erros na Fase 1. Esse número de erros indica que o gesto não é intuitivo e deve ser analisada a substituição dele do vocabulário, principalmente no caso deste ser difícil de ser memorizado, ou seja, apresente resultado ruim também na Fase 2.

Ambos os gestos de “*pressionar o thumb e arrastar*” e “*pressionar um componente e arrastar o dedo sobre múltiplos componentes*” obtiveram 3 erros, sendo metade do gesto mais errado. Ambos os gestos foram confundidos com *zoom*, com maior ocorrência no “*pressionar o thumb e arrastar*”, e *pan*, com maior ocorrência no gesto “*pressionar um componente e arrastar sobre múltiplos*”, indicando que os *widgets* utilizados foram de difícil mapeamento no modelo mental associado. Pressionar e arrastar horizontalmente faz com que os *widgets* sejam associados a componentes comuns de *zoom*, encontrados com frequência em ferramentas de interação com mapas. Já para o movimento vertical de pressionar e arrastar sobre múltiplos componentes foi relatado por todos os usuários a desatenção para o fator de interagir com mais de um componente ao mesmo tempo, bem como entre os usuários que acertaram o chute para este gesto 2 indicaram pouca confiança sobre o chute, indicando que deve-se rever a descrição utilizada para este gesto e analisar a necessidade de definir melhor o que é um *widget*.

O último gesto realizado única e diretamente sobre *widgets* é o gesto de agarrar o range e arrastar horizontalmente, que teve dois chutes errados, sendo associado a funcionalidade de *pan*. Estes foram associados principalmente ao participante não atentar para a descrição “sobre um widget” e ao gesto ser uma versão mais limitada do gesto do *pan*. A similaridade entre os gestos não é um problema, ela na verdade diminui o tamanho do vocabulário propiciando que a repetição de gestos torne estes mais fáceis de serem lembrados e utilizados pelos usuários a longo prazo, enquanto a desatenção sobre onde este é realizado tende a ser resolvido com a utilização da ferramenta implementada e um contexto visual para o gesto, a medida em que o usuário for utilizando uma implementação deste vocabulário o significado do gesto vai ser mais fácil de ser lembrado de acordo com o contexto que o usuário estiver vendo (um *widget* com um intervalo, como uma barra por exemplo, para ser movido ou a própria visualização).

Tomando por base o vocabulário apresentado por (NIELSEN, STÖRRING, *et al.*, 2004), o vocabulário exposto neste trabalho, de tamanho oito, foi considerado relativamente pequeno, e a

média de reinícios durante a Fase 2 foi estimado em um máximo de 2 reinícios por usuário. A Figura 4, mostra quantos erros foram registrados por usuário.

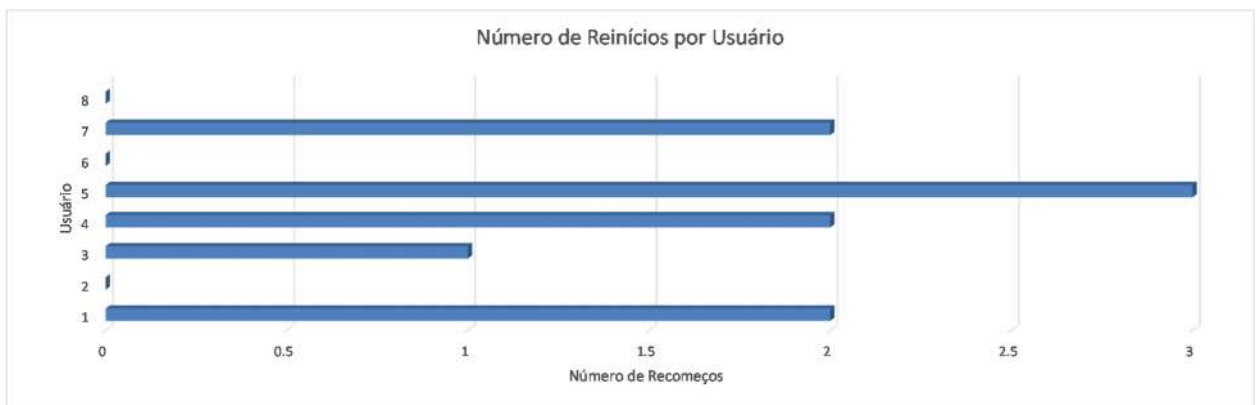


Figura 4: Número de reinícios por usuário.

Pode-se perceber que o maior número de reinícios registrados representa menos da metade do número de gestos no vocabulário (oito), exatamente 37%. Esse participante em específico reportou que a limitação do tempo, dois segundos para cada gesto, era um fator que causava muita pressão sobre ele e, exceto este usuário, os demais tiveram um número de reinícios dentro da meta estabelecida. A Figura 5 mostra quantos reinícios cada gesto ocasionou.

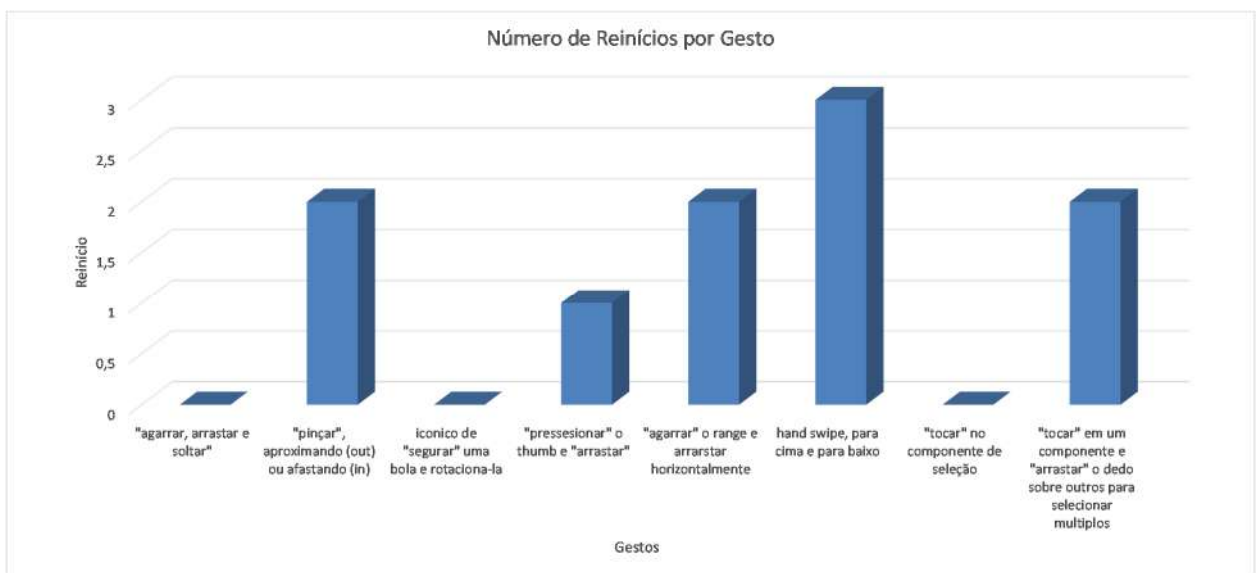


Figura 5: número de reinícios por gesto.

Pode-se perceber que o gesto que causou o maior número de reinícios foi o gesto de “*hand swipe para cima ou para baixo*”. Durante os testes, percebeu-se que os reinícios gerados por este gesto estavam relacionados a dificuldade dos participantes em associar a navegação em um widget

com a funcionalidade escrita no slide (Filtro Categórico). Associado o resultado desta fase ao grande número de erros presentes na fase de adivinhar a funcionalidade do gesto, leva a crer que este gesto não é apropriado para permanecer no vocabulário e um novo gesto deve ser selecionado para essa finalidade. Essa dificuldade de associação também ressalta uma falha na elaboração do teste, ao representar uma funcionalidade que não é exclusiva (navegação em lista vertical) do filtro categórico, dentro do conjunto de gestos deste. Para aplicações futuras deste processo de escolha de gestos, uma separação mais fina das funcionalidades deve ser realizada. O outro gesto da funcionalidade de Filtro Categórico, “*pressionar sobre um componente e arrastar sobre outros para selecionar múltiplos*”, gerou 2 reinícios e se manteve dentro da meta geral estabelecida.

O gesto de zoom, “pinçar, aproximando (out) ou afastando (in) os dedos”, está no limite definido como meta para esta etapa, tendo ocasionado dois reinícios. Estes reinícios não se deram devido aos participantes não conseguirem lembrar do gesto a tempo, mas devido a estes realizarem o gesto sem voltar para o estado de não interação entre os movimentos de zoom in e out, caracterizando o gesto realizado com errado e gerando o reinício. Dessa forma podemos assumir que de maneira geral o gesto, “pinçar” não é difícil de ser lembrado pois todos os participantes conseguiram recordar dele, no entanto, sua particularidade de ter um ponto de não interação (punho fechado) entre os movimentos de interação é que apresenta alguma dificuldade em ser lembrado. Isso sugere que uma possível implementação do reconhecimento deste gesto deva considerar a interação com e sem o estado de não interação.

Para a funcionalidade de Filtro Contínuo foram definidos dois gestos, “*pressionar o thumb e arrastar*” e “*agarrar o range e arrastar*”, que respectivamente servem para definir o limite ou borda de intervalo em um filtro contínuo e mudar o intervalo definido em um filtro contínuo, que geraram, também respectivamente, 1 e 2 reinícios. O participante que gerou reinício para o gesto “*pressionar o thumb e arrastar*” indicou, após o anúncio do reinício, que por aquele ser a primeira funcionalidade com mais de um gesto (para a execução deste usuário) ele teve mais dificuldade.

Nas duas ocasiões onde o reinício foi ocasionado pelo gesto “*agarrar o range e arrastar*” os usuários expressaram não recordar a priori deste gesto, um dos participantes realizou o outro gesto da funcionalidade e esperou o tempo para passar à próxima funcionalidade, enquanto o outro informou que lembrava ter outro gesto mas não recordava qual era este gesto. Um destes

participantes, ao ser reapresentado ao vocabulário, expressou o comentário “*Há! O mesmo gesto do Pan.*”, ressaltando a associabilidade entre os gestos.

Os demais gestos (“*agarrar, arrastar e soltar*” para a funcionalidade de zoom, “*segurar uma bola e rotaciona-la*” para a funcionalidade de rotação e “*tocar no componente de seleção*” para selecionar ou des-selecionar um valor de atributo) não geraram reinícios, o que indica a facilidade destes em serem lembrados, sendo os dois primeiros gestos realizados diretamente sobre a visualização e o terceiro é um gesto dêitico recorrente na literatura.

A terceira Fase é uma avaliação subjetiva de ergonomia, com uma meta media definida em 2, como discutido acima. A média das notas registradas para cada gesto é mostrada na Figura 6.

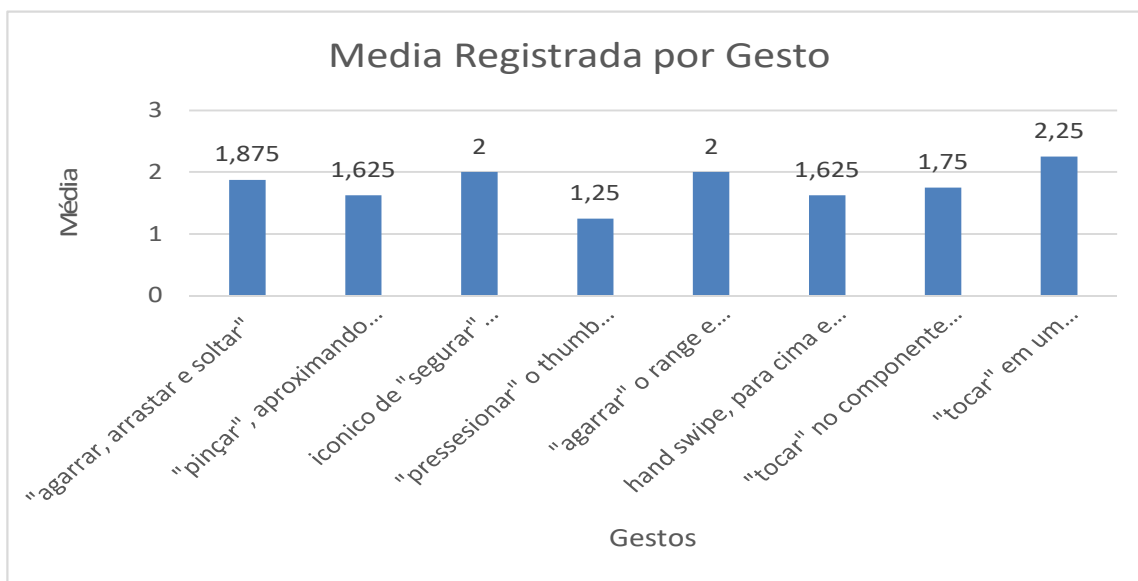


Figura 6: Resultado da avaliação subjetiva de ergonomia.

Percebe-se que o único gesto ater uma média acima (pior) que a meta estabelecida foi o gesto de “*tocar em um componente e arrastar o dedo sobre outros para selecionar múltiplos*”. Este gesto foi relatado por alguns usuários como “o mais cansativo”, bem como foi informado que realizar o gesto de baixo para cima é mais desagradável do que realiza-lo de cima para baixo, a média elevada para este gesto também está associada a amplitude do movimento realizado. Ainda que a média para este gesto tenha ultrapassado a meta definida, esta diferença pode não vir a ser relevante por dois fatores, a já mencionada característica de intercalar interação com análise e o fato de que a intenção de filtrar nem sempre significa filtrar múltiplos valores de atributos e nem que estes estejam em sequência, o que reduz a frequência com que este gesto seria realizado.

Dois gestos do vocabulário ficaram com médias no limite da meta estabelecida, sendo estes “*icônico de segurar uma bola e rotacioná-la*” e “*agarrar o range e arrastar horizontalmente*”. Para o primeiro gesto, foram comuns comentários sobre a dificuldade em realizar o gesto em todos os possíveis eixos, bem como 25% dos participantes reclamaram da necessidade constante de “agarrar” e “soltar” a bola imaginária para uma interação continuada (devido a rotação no pulso em relação cotovelo “encurtar” o movimento, gerando uma demasiada repetição), durante a repetição deste gesto não era incomum notar que os participantes por vezes ignoravam o “soltar” e um participante declarou “*não sei para onde estou rotacionando*”. Além disso, apesar de não ser representado pelas notas atribuídas, durante a observação dos testes percebeu-se que a rotação do pulso em relação ao cotovelo era um fator que causava desconforto aos participantes.

O gesto “agarrar, arrastar e soltar”, para a funcionalidade de pan, recebeu como notas individuais (dadas por cada participante) valores 1 e 2, tendo uma nota 3, conferida por um participante que realizou o gesto com um movimento bastante amplo. O gesto de pan, foi o gesto com score mais alto abaixo da meta 2 e com o segundo desvio padrão mais baixo (Tabela 6: Pontuação individual dos gestos para o teste de stress.).

Tabela 6: Pontuação individual dos gestos para o teste de stress.

GESTO	FUNCIONALIDADE	MÉDIA	DESVIO PADRÃO
"agarrar, arrastar e soltar"	pan	1,87	0,64
"pinçar", aproximando (out) ou afastando (in)	zoom	1,62	1,18
iconico de "segurar" uma bola e rotaciona-la	rotação	2	0,75
"pressionar" o thumb e "arrastar"	filtro contínuo	1,25	0,46
"agarrar" o range e arrastar horizontalmente	filtro contínuo	2	0,75
hand swipe, para cima e para baixo	filtro categórico	1,62	0,91
"tocar" no componente de seleção	filtro categórico	1,75	0,88
"tocar" em um componente e "arrastar" o dedo sobre outros para selecionar multiplos	filtro categórico	2,25	0,7

Os demais gestos todos marcaram abaixo do limite da meta, estando a contento pelo critério estabelecido para este vocabulário. Entretanto um gesto destes apresentou uma curiosidade, “*pinçar, aproximando (out) ou afastando (in)*”. Dois usuários deram nota acima da meta para este gesto, um afirmou que os estados de não interação (punho fechado) e as diferentes partidas do gesto (estender polegar e indicador juntos para zoom-in e estender afastados para zoom-out) eram prejudiciais ao gesto e atrapalhavam a realização repetitiva do mesmo (nota atribuída 3), o outro usuário afirmou ter uma conhecida dificuldade em realizar este gesto, o que foi perceptível durante a realização do teste e atribuiu uma nota 4. Vale ressaltar que o estabelecimento dos estados de

não interação e de partida foram definidos pelos próprios participantes, na fase de levantamento dos gestos (E1).

Baseado no benchmark realizado e resultados descritos acima, o vocabulário atende as metas estabelecidas, exceto pelo gesto de “hand swipe para cima e para baixo”, que registrou as maiores (piores) notas nas fases 1 e 2, ficando acima da meta estabelecida em ambos os casos e fazendo com que este precise ser trocado.

3.1 Avaliação Ergonômica

Para a seleção dos gestos é recomendável filtrar do vocabulário gestos que não sejam ergonômicos, ou seja, que devido a fatores físicos possam vir a gerar desconforto, dor ou lesões. A escolha de gestos baseada em interações no ar livre potencializa a escolha de um vocabulário de gestos que gere gestos menos estressantes, uma vez que manusear dispositivos (mesmo dispositivos leves ou que fiquem “apoiados”, como o mouse) pode causar esforço muscular excessivo durante tarefas que envolvam demandas mentais (ULLMAN, KANGAS, *et al.*, 2003) e devido a introdução de forças externas (NIELSEN e GRANUM, 2003), vale lembrar que forças externas são forças geradas por objetos manipulados no meio em relação a atividade realizada enquanto forças internas são caracterizadas pela força aplicada pelos músculos do indivíduo em relação a si próprio. Entre os princípios da ergonomia, alguns são recorrentemente listados como principais [(HEDGE, MUSS e BARRERO, 1999), (GRANT, 2002), (SHAW e HEDGE, 1997), (KORTUM, 2008)]:

- 1- Evitar posições limite.
- 2 - Evitar repetições.
- 3 - Relaxar os músculos.
- 4 - Utilizar posições neutras relaxadas, e que estejam no meio entre as posições limite.
- 5 - Evitar permanecer numa posição estática.
- 6 - Evitar forças externas ou internas na juntas, assim como posições que interfiram no fluxo de fluidos corporais.

Estes fatores devem ser considerados em cada gesto do vocabulário com caráter eliminatório em caso de um deles ser violado com frequência. Optou-se por eliminar gestos manipulativos que

dependam repetitivamente de movimentos das juntas da mão que tenham um intervalo curto de mobilidade (ângulo de movimento menor que 90), que tendem a gerar uma repetição maior, que é uma restrição comum encontrada em trabalhos de modelagem de movimentos da mão [(LIN, WU e HUANG, 2000), (LEE e KUNII, 1995)], pois estes tendem a gerar uma repetição mais frequente e com maior esforço durante atividades com demandas mentais acentuadas.

Considerando as restrições listadas acima o gesto de rotação (“segurar uma bola e rotacioná-la”) foi retirado do vocabulário devido a requerer movimento ulnar e radial do pulso (Figura 7), que geram um ângulo resultante de movimento de aproximadamente 55 graus, forçando o limite desta junta, principalmente no desvio radial. Nos testes foram observados também que este movimento, apesar das boas notas aferidas nos testes, era constantemente realizado de maneira a minimizar o desconforto, usuários não o executavam em todos os eixos, por vezes esqueciam de retornar à posição de não interação, reportaram confusão em relação a funcionalidade sendo realizada e quando realizavam o movimento de forma ampla (em geral causando movimentação do cotovelo) eles deixavam de realizar o movimento naquele eixo.

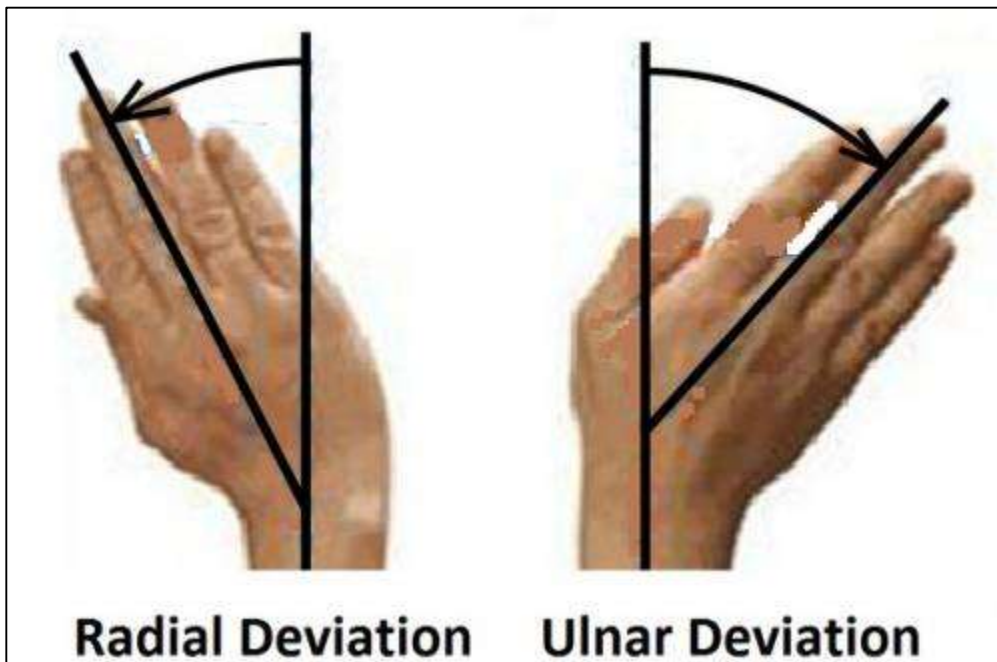


Figura 7: Desvio Ulnar e Radial

Para os demais gestos do vocabulário nenhum dos princípios é violado com frequência e não necessita ser excluído do vocabulário. Para o gesto filtrado devido a esta avaliação ergonômica,

adotou-se o próximo gesto de maior frequência que não viole os princípios referidos acima, “apontar a direção da rotação”.

3.2 Vocabulário Final

Após o benchmark e o filtro ergonômico, as seguintes alterações foram feitas ao vocabulário de gestos: “Segurar uma bola e rotaciona-lá” foi substituído por “apontar para um widget indicando a direção”, para funcionalidade de rotação, e “hand swipe, para cima e para baixo” foi substituído por “agarrar e mover a lista de componentes selecionáveis”, para funcionalidade de navegar no widget de filtro contínuo. Essas alterações terminam no seguinte vocabulário de gestos final:

- Gesto 1 (G1): “Agarrar, arrastar e soltar.”, para a funcionalidade de pan em qualquer direção no plano.

- Gesto 2 (G2): “Pinçar com indicador e polegar, aproximando (out) ou afastando (in).”, para a funcionalidade zoom.

- Gesto 3 (G3): “Apontar para um widget indicando a direção.”, para rotacionar na direção indicada.

- Gesto 4 (G4): “Pressionar um limite (widget) a arrastar horizontalmente.”, para definir o limite superior, ou inferior, do intervalo de filtro contínuo.

- Gesto 5 (G5): “Agarrar o intervalo (widget) e arrastar horizontalmente.”, para realizar o filtro contínuo com um intervalo pré definido.

- Gesto 6 (G6): “Agarrar e mover a lista de componentes selecionáveis (widget).”, para navegar no widget de filtro categórico.

- Gesto 7 (G7): “Tocar no componente de seleção (widget).”, para filtrar, ou recolocar, itens com um determinado atributo na visualização.

- Gesto 8 (G8): “Tocar sobre um componente selecionável e arrastar o dedo sobre outros para selecionar múltiplos.”, para interagir com múltiplos atributos com filtro categórico.

4. MAPEAMENTO DO VOCÁBULARIO

De posse do vocabulário final é hora de mapear este vocabulário para que seja representado de maneira mais fidedigna possível na tecnologia escolhida, tanto para a tecnologia de entrada disponível, ou seja, os sensores utilizados para o rastreamento das mãos, quanto para a apresentação desse rastreamento.

Como dispositivos de entrada foram escolhidos dois dispositivos comercialmente disponíveis, o Leap Motion, que é um dispositivo para rastreamento de mãos e reconhecimento de gestos baseado em visão, e a braçadeira Myo, que é um *wearable* capaz de detectar um conjunto de gestos da mão através de EMG (ElectroMyoGraphy - EletroMioGrafia) bem como dados espaciais fornecidos por sensores de acelerômetro e giroscópio. Ambos os dispositivos são acessíveis e fornecem SDKs para desenvolvedores, bem como possuem repositórios de exemplos de código e exemplos demonstrativos de suas funcionalidades.

Dada a disponibilidade destes dispositivos, o mapeamento considerou os guidelines, tanto do Leap Motion (LEAP MOTION, 2016) quanto do Myo (THALMIC LABS, 2014). Abaixo são sumarizadas características extraídas destes guidelines, primeiro para a braçadeira Myo e posteriormente para o Leap Motion.

Os guidelines do Myo baseiam-se principalmente na facilidade de reconhecimento do dispositivo, indicam a necessidade de ter estados de interação ativada e desativada e expressam pouco sobre o design de interfaces voltadas para interação usando Myo. Estas características são evidenciadas pelo fato de que os gestos apresentados nos guidelines são apenas os que são pré-definidos na SDK do dispositivo, de um gesto (*double tap*) ser expressamente indicado para ativar e desativar a interação pelo Myo e pelo fato de que são exemplificadas funcionalidades por gestos sem a descrição de interfaces, indicando que esta possam ser utilizadas conjuntamente com outros

métodos de entrada, sem a necessidade de widgets especiais para a interação pelo Myo (exceto os informativos de como os gestos devem ser executados).

Os guidelines do Leap Motion são mais focados na interface do que no reconhecimento de gestos, e esse foco pode ser sumarizado em duas características principais, tamanho e realce. As diretrizes de interface propõem que o tamanho de widgets utilizados com o Leap Motion seja maior que os widgets utilizados com o mouse, devido a questões de precisão, e que os componentes tenham um espaçamento maior entre eles do que interfaces WIMP (*Windows, Icons, Menus and Pointer* – Janelas, Ícones, Menus e Ponteiro), uma vez que além da precisão do dispositivo deve se levar em consideração a dificuldade do ser humano em manter a mão completamente parada e a percepção em relação a profundidade pode vir a introduzir ou aumentar erros de interação.

Em relação ao realce as diretrizes são: utilizar *widgets* com alto contraste entre fundo, bordas e, quando presente, texto; ao utilizar botões, manter o botão de sair de um menu afastado e destacado dos demais botões da interface; e utilizar realce de proximidade, ou seja, realçar de maneira gradativa os componentes baseados no quão próximos estes estão do componente sobre o qual a mão do usuário está.

Em relação a disposição dos menus no Leap Motion, são sugeridos duas padronizações que seguem a diretriz de ter botões que ocupem grandes áreas da tela e tenham pouco espaço entre eles, essas disposições são baseadas em *grid*, que divide a tela em áreas retangulares, e radial, que divide a tela em áreas a partir do centro. As diretrizes do Leap Motion também indicam as recorrentes orientações de ter *widgets* que informem os movimentos para interação com o usuário e sempre apresentar um feedback para o usuário de que determinada ação foi realizada.

Com base nos gestos levantados, nos guidelines e características dos dispositivos de entrada e na experiência previa adquirida nas primeiras etapas deste trabalho, foram definidas as seguintes diretrizes no mapeamento do vocabulário:

- Reduzir a hierarquia de menus tocáveis e quando um destes menus levar a outro, impedir que o tempo que o usuário leva para realizar um toque acabe gerando toques nos menus internos.
- Gestos baseados principalmente nos dedos do usuário (como por exemplo apontar, clicar, pinçar, etc.) devem ser tratados com o Leap Motion, devido o melhor rastreamento das juntas dos dedos, ao invés da detecção de gestos “fechados” oferecida pelo Myo.

- Gestos baseados principalmente na movimentação do braço, ainda que utilizem configurações da mão para indicar estados (como por exemplo “agarrar, arrastar e soltar”), devem ser tratados com a braçadeira Myo, pelo fato de que sendo um wearable baseado em EMG não há o risco de o gesto sair do campo de visão do dispositivo, o que provavelmente ocorreria com o Leap Motion, uma vez que gestos baseados na movimentação do braço tendem a ter amplitudes maiores que gestos baseados na movimentação apenas da mão. Para tal, deve existir uma etapa de calibração para utilização do Myo como entrada de informações espaciais.

- Para gestos onde o rastreamento da mão ou dedos movam componentes de interface sem a interação de agarrar, como por exemplo “pressionar um componente e arrastar”, deve-se utilizar a técnica de magnetismo do componente, ou seja, depois de uma determinada situação (seja pressionar, aguardar um determinado tempo ou outra) a representação do membro rastreado deve ser ajustada a posição do objeto.

- Quando a interação for baseada no Leap Motion, utilizar o campo de visão dele para diferenciar estados de interação e não interação, quando utilizando o Myo, deve haver um gesto para travar e destravar a interação. Em ambos os casos deve-se respeitar a descrição do gesto caso o mesmo já defina um estado de não interação.

- Utilizar um gesto que seja diferente dos gestos do vocabulário para alternar entre as funcionalidades de interação, e se possível até mesmo utilizar outro modo de interação, como por exemplo comandos de voz.

Vale ressaltar que essas diretrizes são baseadas nas tecnologias utilizadas (Leap Motion e Myo) e no vocabulário de gestos definido. Elas podem vir a servir para outros conjuntos de gestos baseados em tecnologias semelhantes e que compartilhem características deste vocabulário, como ser predominantemente manipulativo e ter granularidade baixa (três para este vocabulário), mas foram definidas dentro deste contexto limitado.

Considerando as diretrizes e como alguns gestos se repetem no vocabulário alguns gestos foram meados e reutilizados ao longo da implementação:

- “agarrar” foi mapeado para o gesto “*fist*” do Myo;
- “soltar” foi mapeado para o gesto “*finger spread*” do Myo;

- “arrastar” passou a ser movimento do braço quando precedido pelo gesto de agarrar e sucedido por soltar;

- “pinçar” é um gesto que é reconhecido através do Leap Motion pelo rastreamento dos dedos polegar e indicador;

- “apontar”, “pressionar” e “tocar” são todos reconhecidos através do Leap Motion e diferenciam-se entre si pela posição espacial do indicador do usuário dentro do campo de visão do Leap Motion. “apontar” considera a posição onde o indicador está representado na tela sem considerar a profundidade do mesmo, “pressionar” considera a posição do indicador nos três eixos sendo que a interação é marcada por manter o indicador numa profundidade onde este está em constante interação e, por fim, “tocar” é a interação onde o indicador no espaço virtual entra no espaço ocupado por um outro objeto (independentemente de permanecer ou não neste espaço).

Dessa forma o mapeamento dos gestos ficou como segue:

- G1: realizado utilizando o Myo. Como o rastreamento para este gesto não necessita do Leap Motion, ele pode ser utilizado fora do ângulo de visão do dispositivo, com a única diferença notável sendo o fato de que com a mão fora do campo de visão o usuário perde a representação da mão na interface. A Figura 8 apresenta a interface desenvolvida com esta funcionalidade ativada.

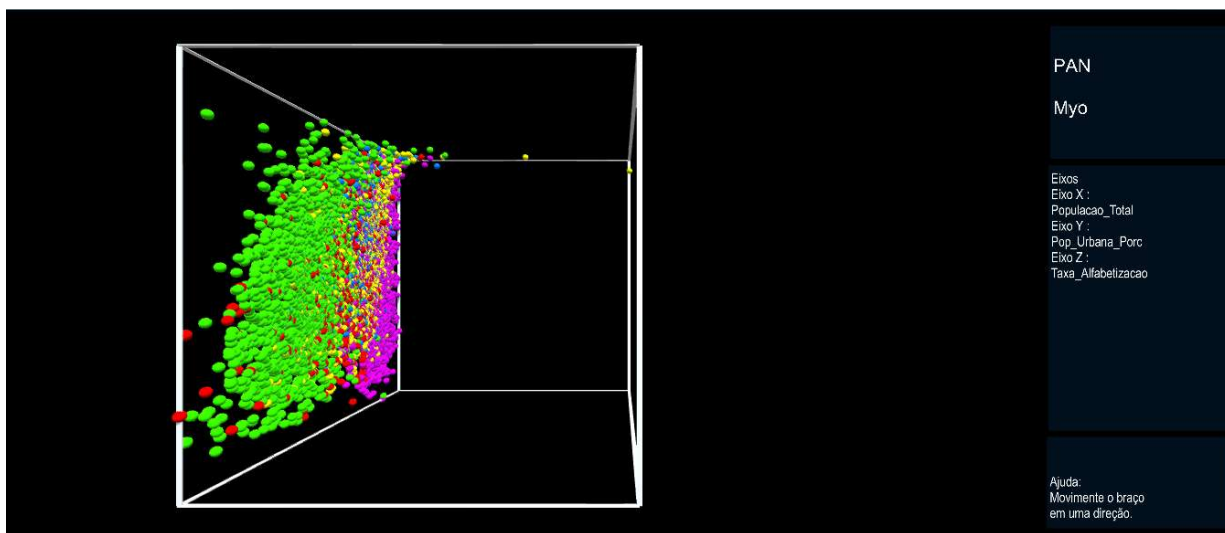


Figura 8: Gesto da funcionalidade pan sendo executado com o Myo.

- G2: gesto rastreado com o Leap Motion, como a mão do usuário precisa estar na área de visão do dispositivo ela sempre está representada na interface enquanto alguma interação de zoom

está sendo realizada, o feedback é dado pela própria visualização que se aproxima ou afasta do ponto de visão do usuário. A Figura 9 apresenta esta funcionalidade na interface.

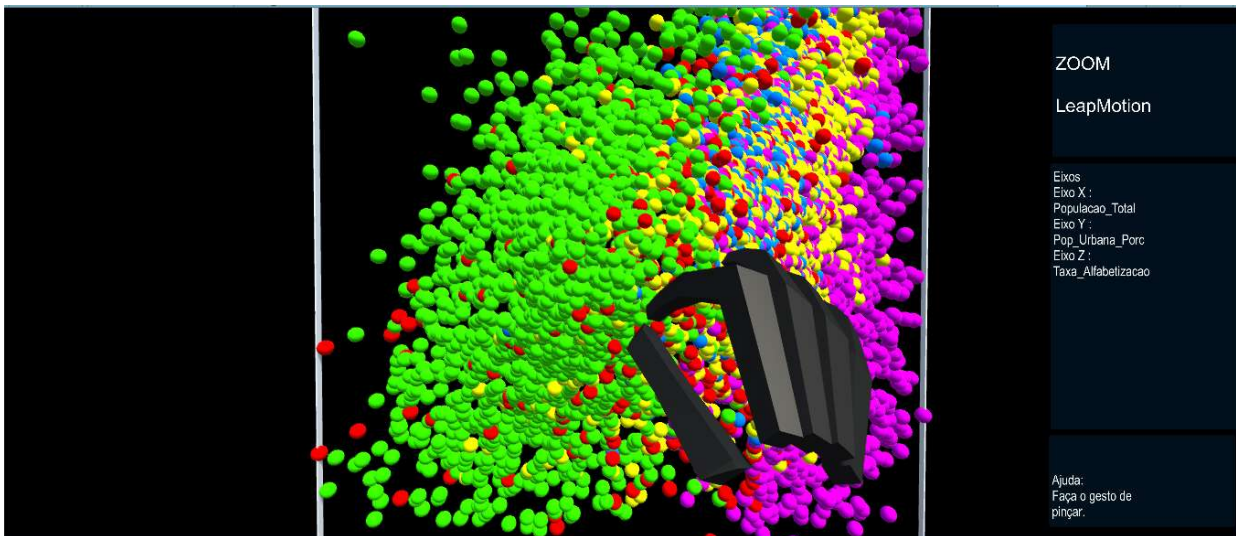


Figura 9: Gesto da funcionalidade zoom, pinçar com polegar e indicador.

- G3: gesto rastreado pelo Leap Motion, verifica se os dedos (efetivamente o dedo mais esticado) do usuário está na mesma área que uma das indicações de direção da rotação (desconsiderando a profundidade ao verificar área) e rotaciona na direção indicada, como pode ser visto na Figura 10.

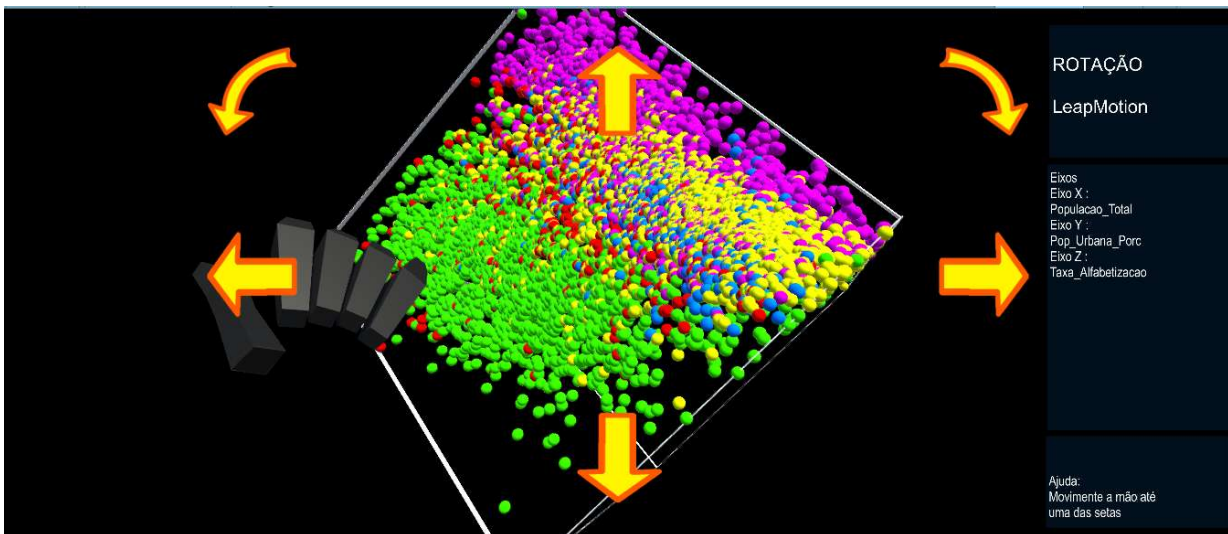


Figura 10: Posicionar a mão sobre uma seta de direção para rotacionar.

- G4: este é rastreado com o Leap Motion, através do toque do indicador em um dos limites do intervalo de filtro contínuo configurado, o limite tocado segue horizontalmente o indicador do

usuário até que este feche o punho ou saia do campo de visão do Leap Motion, um único limite do intervalo em azul indica que aquele é o limite sofrendo a interação. Funcionalidade é mostrada na Figura 11.

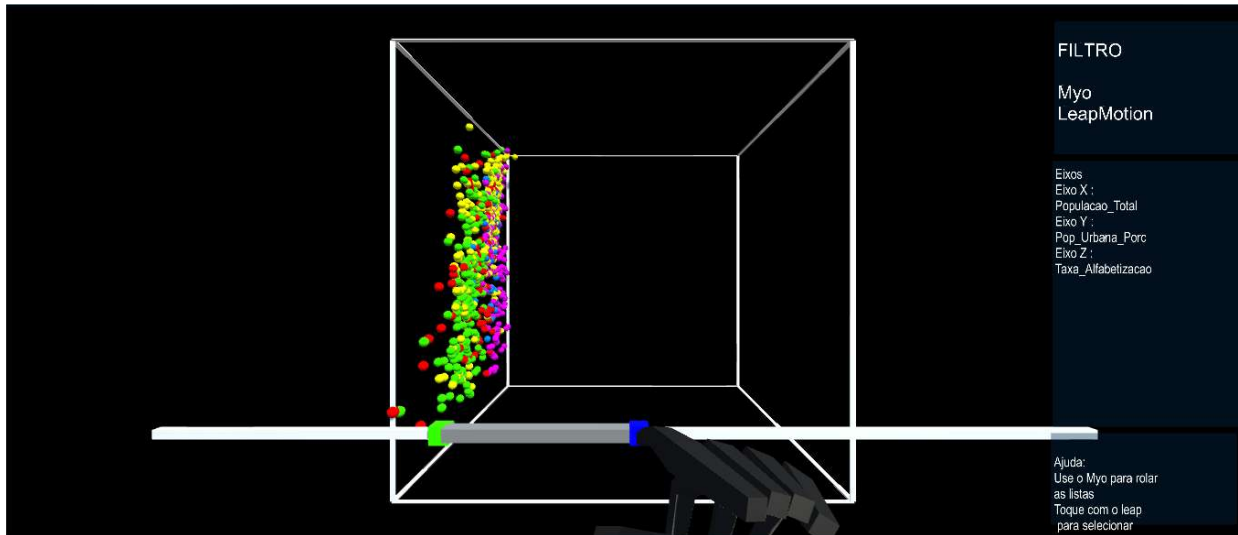


Figura 11: Configuração do limite de filtro contínuo, pressionar o thumb e arrastar.

- G5: gesto para funcionalidade de filtro contínuo que permite mover um intervalo pré definido dentro do eixo de configuração sem alterar a distância entre os limites deste intervalo, permitindo assim analisar diferentes intervalos iguais a cada interação (por exemplo, avaliar na visualização um período de 5 anos a cada interação), como mostrado na Figura 12, os limitadores na cor azul indicam que ambos (ou seja, o intervalo todo) está sofrendo interação. O gesto é realizado no Myo, tendo os gestos disponíveis na API do dispositivo como marcadores de início e fim da interação.

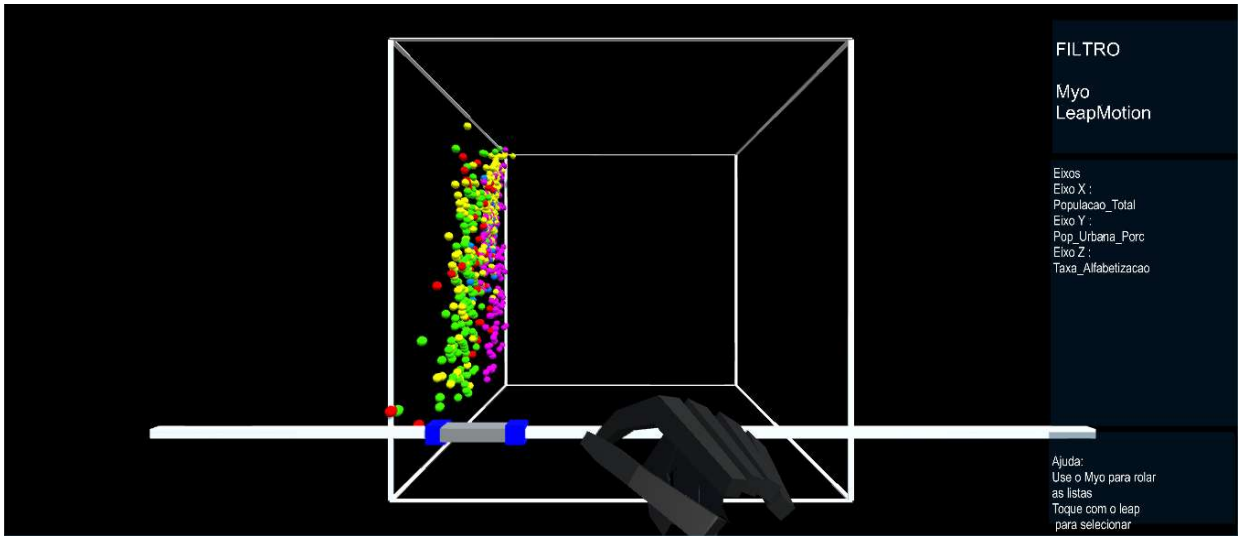


Figura 12: Arrastar o range para mover o intervalo de seleção, com o Myo.

- G6: gesto reconhecido e rastreado utilizando o Myo, os gestos da API do Myo de *Fist* e *Fingers Spread* marcam o início e o fim da interação, respectivamente. Entre o início e o fim da interação são utilizados os sensores espaciais do dispositivo (acelerômetro e giroscópio) para verificar se a lista deve ser arrastada para cima ou para baixo e atualizar o widget de lista de acordo com o movimento realizado. Esta funcionalidade é exibida na Figura 13.



Figura 13: Gesto de navegar na lista de atributos com agarrar, arrastar e soltar.

- G7: gesto reconhecido pelo Leap Motion, através da colisão do modelo virtual do indicador do usuário com objetos selecionáveis, ao indicador é ancorado um objeto que verifica e controla essa colisão. Os itens selecionados mudam de cor e possuem um tamanho bastante

superior a representação da mão, visando reduzir situações de seleção de objetos não planejada. Essa funcionalidade é mostrada na Figura 14.



Figura 14: Aprofundar a mão virtual a tocar um botão para selecionar.

- G8: reconhecido com o Leap Motion. Uma vez que um componente selecionável tenha sido tocado o usuário pode selecionar os próximos abaixo ou acima deste movendo a mão na vertical, causando a (des)seleção do componente tocado, pelo mesmo princípio do G7. Este gesto pode ser realizado tanto com 1 quanto com duas mãos. Exemplificado na Figura 15.

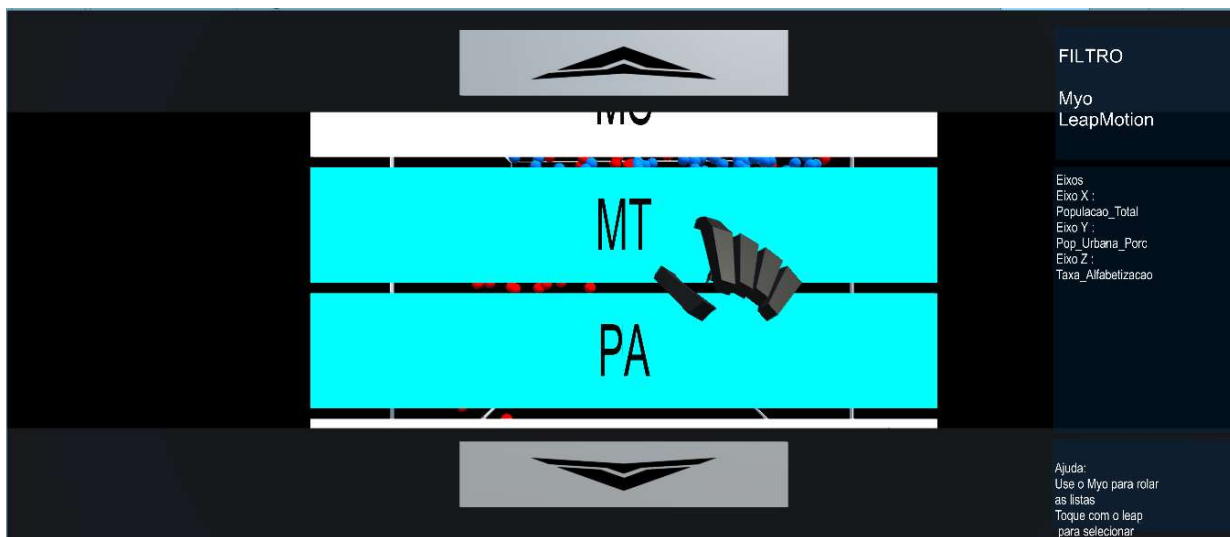


Figura 15: Pressionar um botão e mover a mão verticalmente para selecionar múltiplos.

5. TESTE COM O VOCABULÁRIO IMPLEMENTADO

Definida a implementação dos gestos do vocabulário foi elaborado um protótipo para aplicar essa definição e verificar a resposta dos usuários à implementação. Seis participantes, com experiência acadêmica e prática (pelo menos 5 diferentes ferramentas de InfoVis) com visualização de informação, realizaram uma serie de interações compostas por todos os gestos do vocabulário, alternando entre as funcionalidades ativas através de um modo de interação não baseado em gestos, a voz.

O protótipo foi implementado na plataforma Unity (UNITY, 2016), utilizando como dispositivos de entrada os sensores Myo e Leap Motion para rastreamento dos gestos, bem como os SDKs destes dispositivos para estas plataformas, e o Google Speech API (GOOGLE INC., 2013) para o reconhecimento dos comandos de voz (para alteração de modo de interação) utilizando o próprio microfone do laptop utilizado nos testes para a captura da voz. Para o reconhecimento de voz gerou-se um conjunto de sinônimos para ativar cada modo de interação, visando assim diminuir problemas de reconhecimento. Vale ressaltar que este teste não tem como foco avaliar as taxas de reconhecimento dos comandos de voz, nem fatores de comunicabilidade da interface (coisas que serão avaliadas futuramente, em outros trabalhos), o foco desta etapa é avaliar se o mapeamento dos gestos para um conjunto de tecnologias manteve o nível de stress dos gestos ideais mapeados, sendo que a intuitividade e memorabilidade dos gestos depende do mapeamento mais direto possível dos gestos levantados para a implementação.

Os participantes neste teste tinham o mesmo perfil dos participantes das etapas anteriores (experiência acadêmica com visualização de informação e experiência com ferramentas de visualização). Para o teste o participante ficava sentado em uma cadeira (giratória e com rodinhas) com apoio para costa e cotovelos, posicionada em frente à mesa onde foi colocado um laptop no qual a aplicação é executada. Os dispositivos de entrada foram colocados entre o participante e o

notebook (o Leap Motion) e no terço proximal do antebraço do braço dominante (o Myo), sendo realizada a calibração dos dispositivos antes de cada teste e criando-se um perfil para cada usuário no gerenciador do dispositivo Myo. Os dispositivos de entrada e uma breve explicação sobre seu funcionamento foram apresentadas aos participantes antes do teste, sendo que esta apresentação incluía características dos dispositivos, como a possibilidade de extrair dados espaciais da braçadeira Myo e uma demonstração dos limites do campo de visão do Leap Motion.

Este último teste tem como objetivo verificar se o mapeamento para tecnologia manteve o vocabulário condizente com o levantado, em relação as notas da avaliação subjetiva. Foi esperado que os gestos implementados tenham resultados mais elevados (piores) que os gestos levantados, mas que estes estejam com notas próximas, indicando que a tradução do gesto para algo reconhecível pelos dispositivos de entrada manteve o conforto do usuário na realização do mesmo, ainda que introduza nos gestos limitações impostas pelos dispositivos de entrada. Para aos gestos que foram modificados para implementação (os gestos levantados que seriam equivalente a G3 e G6, respectivamente “segurar uma bola e rotacioná-la” e “hand swipe para cima e para baixo”) essa avaliação também serve como uma realização tardia da etapa de stress do benchmark, que não foi realizada para estes gestos.

O vocabulário foi apresentado aos participantes e em seguida pediu-se que os mesmos realizassem repetidas vezes o mesmo gesto, alternando entre direção, sentido, ligar e desligar, e demais modos disponíveis para cada interação, de maneira semelhante ao realizado na avaliação subjetiva de stress do benchmark. Ao final, também pediu-se que os usuários fizessem comentários sobre a funcionalidade e dessem uma nota, também baseada na escala apresentada no teste de stress, para cada gesto implementado, como pode ser visto abaixo (Tabela 7):

Tabela 7: Média dos resultados obtidos para avaliação subjetiva dos gestos implementados.

G1	G2	G3	G4	G5	G6	G7	G8
2,17	1,67	1,67	2,50	2,00	1,83	1,50	4,67

Em comparação com o benchmark feito para os gestos levantados, em ordem respectiva a funcionalidade que estes representam, a Figura 16 demonstra como estão os gestos levantados em relação aos gestos implementados:

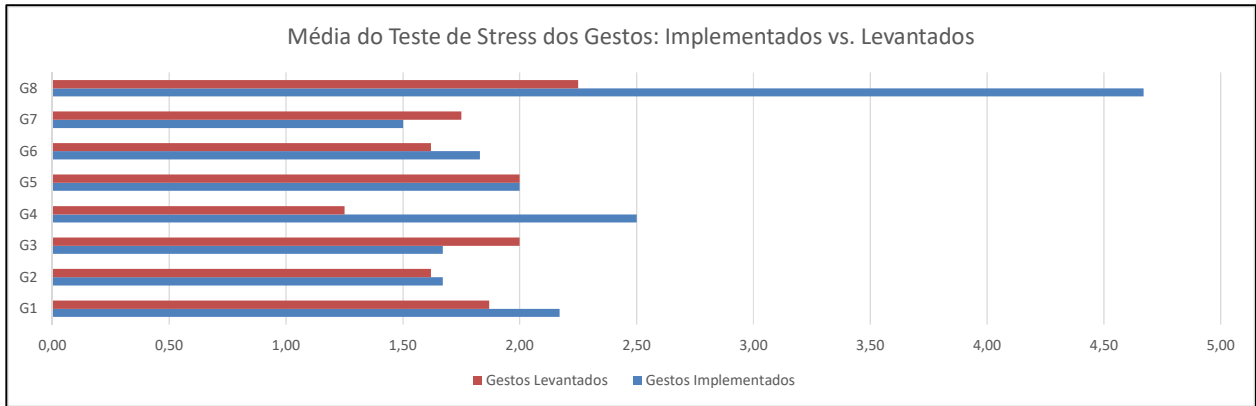


Figura 16: Comparativo entre as médias do teste subjetivo de Stress entre gestos implementados e levantados.

Para a maior parte do vocabulário verificou-se o esperado, gestos levantados obtiveram notas menores (melhores) do que gestos implementados. Em apenas dois gestos os implementados (G3 e G7) a nota foi melhor (menor) do que dos respectivos gestos levantados. Houve também um empate nas notas em G5.

A substituição em G3 parece ter agradado os participantes, nenhum requereu esclarecimentos adicionais sobre o funcionamento nem pareceu em dúvida ao realizar a interação, o que foi atribuído a controles (setas) intuitivos para a funcionalidade a ser realizada. Entre os comentários para este gesto os participantes ressaltaram a falta de um controle para travar a interação (indicando que a abordagem de ter a área central como ponto neutro não devesse ser a única), a proximidade entre as setas causando interações não planejadas com movimentos muito pequenos (indicando que deve-se considerar colocar espaços de não interação entre as setas ou utilizar um modelo de menu radial ao invés de grid, a exemplo do proposto em (LEAP MOTION, 2016)), e permitir que a velocidade de rotação seja setada pelo usuário (apesar da maioria dos usuários ter expressado satisfação com a velocidade codificada, essa possibilidade ajudar em algumas tarefas, como por exemplo realizar rapidamente um overview em 360 graus).

Para G7, mesmo com a calibração dos dispositivos e criação de um perfil para cada usuário, foi surpreendente a implementação ter alcançado uma média menor que o levantado. Esse resultado gera a hipótese de que o resultado foi melhor devido ao tempo de raciocínio e verificação da interação. Na etapa de benchmark os participantes realizavam os gestos na velocidade definida por eles, mas devido a não haver interface (ou seja, não havia feedback visual) estes em geral realizavam os gestos em sequência acelerada o que provavelmente gerava um esforço maior. Realizando o gestos na implementação, cada interação gerava um efeito sobre a visualização e

feedback visual dos widgets de interação, então a cada interação o participante tinha um conjunto de elementos visuais para checar (ex: toquei no widget correto, o efeito sobre a visualização foi o desejado, qual é o novo estado da visualização e como ela se modificou em relação ao último estado) tais verificações levam algum tempo, o que diminui a cadência de interação e pode ter contribuído para a melhor nota (sensação de menor stress). Entre os comentários dos participantes, ressalta-se o comentado aprendido do gesto (participantes reportaram ter mais dificuldade nas primeiras interações, mas logo se acostuma com está e não tem mais problemas para executar a mesma), participantes informaram que o uso de sombras poderia ajudar na percepção de profundidade (especialmente no início da utilização deste gesto) e um usuário reclamou da profundidade do toque, sugeriu-se que este aproximasse sua cadeira da mesa com o computador e o Leap Motion, mas este não se aproximou e reportou não sentir melhora na interação.

G5 obteve médias iguais tanto para os levantados quanto para os implementados. Como G5 é praticamente o mesmo gesto que G1, só que limitado a apenas um eixo, esperava-se que as medias destes gestos tivessem comportamentos semelhantes, ou seja, notas ligeiramente piores para implementação do que para o levantamento. Entretanto, a calibração individual do Myo (dispositivo que rastreia este gesto) e o fato do movimento manipulativo estar limitado a apenas um eixo pode ter colaborado para a melhora na nota, uma vez que o objeto virtual se movendo em apenas um eixo elimina a demanda por precisão no movimento por parte do usuário (ex: ao mover a mão para esquerda acabamos por move-la ligeiramente para cima ou para baixo, movimento indesejado que acaba sendo mais perceptível devido a erros, ou problemas de calibragem, do dispositivo de entrada). Em relação aos comentários dos participantes, este gesto teve sua velocidade (de movimentação do intervalo) elogiada por todos os participantes, entretanto ao interagir com uma área densa em itens os usuários perceberam a redução de velocidade na resposta da ferramenta, o que alguns usuários reportaram como comprometendo a experiência.

O último gesto com comportamento fora do esperado, foi o G8. Este gesto marcou uma média muito inferior a sua contraparte levantada, devido principalmente ao mapeamento ideal não considerar a limitação da tecnologia. G8 foi mapeado de maneira mista, o reconhecimento de toque reconhecido pelo Leap Motion e o reconhecimento de movimento do antebraço reconhecido pela braçadeira Myo. Ao realizar estas coisas em conjunto os erros de entrar e sair da área de toque dos botões virtuais em todos os testes, se concentrar em navegar na lista mantendo o braço esticado para alcançar uma determinada profundidade também se mostrou confuso (possivelmente por estar

com atenção dividida, por exemplo, estar tocando os botões e navegando na direção certa) e desagradável (participantes se mostravam frustrados e inclinados a não terminar estas interações depois de algum tempo). Outro fator problemático se deu por escolhas de implementação, seguindo diretrizes de desenvolvimento para o Leap Motion foram implementados botões grandes e devido a descrição do gesto levantado era necessário que estes rolassem verticalmente na tela, gerando um efeito de ter aproximadamente 3 botões visíveis por vez na interface, sem uma noção clara de que botão (valor de atributo) vinha a seguir, aumentando a incerteza do usuário na interação e piorando ainda mais a experiência dos usuários. Um participante sugeriu utilizar a mão não dominante para realizar os toque e a mão dominante para a navegação, o que perceptivelmente melhorou a interação deste usuário com a ferramenta. G8 é um gesto que claramente precisa ser revisto e modificado, tanto na interface quanto no gesto, talvez até mesmo levantado novamente, para uma implementação de ferramenta com esse vocabulário de gestos.

Nos demais gestos verificou-se o esperado, gestos implementados foram ligeiramente piores, indicando que o mapeamento dos gestos levantados para implementação não prejudicou muito a demanda requerida do usuário. Entre estes o mais notável é o G4, que apresentou a maior diferença entre médias. Participantes reportaram ter gostado do gesto para interagir assim como se mostraram animados com a possibilidade de utilizar este com duas mãos (um indicador em cada limite do intervalo), entretanto a velocidade do movimento foi rápida demais para alguns enquanto foi relatada como “ótima” por outros, indicando a necessidade de ter um controle personalizado para a velocidade dessa interação.

Esta etapa demonstrou que calibrar os dispositivos de entrada individualmente para cada usuário não é suficiente para gerar uma experiência adequada para este tipo de ferramenta, é necessário também um perfil individual de usuário que guarde parâmetros de ajuste para cada funcionalidade/gesto, como por exemplo um parâmetro para ajustar a velocidade do zoom e outro para ajustar o ângulo mínimo de rotação, que seja controlável pelo usuário.

CONCLUSÃO

Este trabalho apresentou um processo para seleção de gestos para interação com ferramentas de Visualização de Informação. O processo aqui exposto cobre a descoberta dos gestos que são considerados ideais pelo público alvo, o benchmark destes gestos para validar a escolha dos mesmos, uma proposta de implementações destes gestos (ou seja, que tenta tornar estes gestos reconhecíveis pelos dispositivos de entrada escolhidos gerando um mínimo de descaracterização em relação ao gestos selecionados após o benchmark) nas tecnologias escolhidas para interação e, por fim, uma avaliação subjetiva de ergonomia sobre os gestos implementados, visando verificar se estes proporcionam ao usuário um esforço semelhante do que o estimado para os gestos levantados.

O processo apresentado não pretende ser definitivo, nem aplicável em qualquer contexto. Ele serve mais como uma proposta que permite selecionar gestos através de uma abordagem lógica e iterativa, que gera notas que servem para avaliar o desempenho de um vocabulário de gestos em relação ao esperado do mesmo e também comparar diferentes vocabulários de gestos. Como o número de repetições de um gesto na etapa de avaliação subjetiva é dependente do tamanho do vocabulário definido, é possível até comparar subconjuntos de gestos de um vocabulário com outros vocabulários (ou subconjuntos) contanto que estes tenham passado pelas mesmas etapas de benchmark.

Este trabalho tem um escopo contido e limitações, sendo a primeira delas é o número baixo de participantes do experimento, devido ao perfil específico de usuário requerido nos experimentos e ao número reduzido de respostas ao e-mail de convite para participar do mesmo, pouco menos de 20 usuários divididos entre as diferentes etapas realizadas. Outra limitação é em relação a quantidade de funcionalidades de Visualização de informação abordadas neste trabalho, um subconjunto destas foi escolhido para dar mais foco em tarefas presentes em visualizações 3D de

propósito geral, uma vez que tratar visualizações específicas geraria um número imenso de diferentes funcionalidades e a experiência com ferramentas que representes estas visualizações específicas reduziria ainda mais o número de participantes. Outra limitação evidente neste trabalho é seu foco em visualizações 3D, ainda que se acredite que estas interações não seriam representadas de maneira muito diferente para um ambiente 2D, não é possível afirmar isto.

Outras ocorrências dizem respeito a ordem de execução das etapas do processo, que deveriam ser revistas, como realizar o filtro ergonômico antes da realização do benchmark. Bem como avaliar maneiras de testar não apenas a ergonomia subjetiva nos gestos implementados, como por exemplo definindo um método que indique o quanto a implementação se afasta do gesto levantado.

Por fim, este trabalho mostra uma investigação a mais na interação através de gestos ao ar livre com ferramentas de Visualização de Informação, de modo que as decisões, descobertas e diretrizes aqui expostas sirvam para que outros trabalhos aumentem ainda mais o conhecimento gerado nas áreas de interação multimodal e visualização de informação seja corroborando, refutando, extraindo significado das descobertas aqui realizadas, ou mesmo propondo melhorias ou modificações para o processo aqui apresentado, da mesma forma que este trabalho se baseia em muitos outros realizados previamente.

5.1 Trabalhos Futuros

Este trabalho realizou o todas as etapas do processo ao que foi proposto, no entanto uma série de atividades ainda podem ser realizadas para melhorar e expandir o trabalho, estas são listadas abaixo:

- Aumentar o número de funcionalidades de InfoVis consideradas no levantamento dos gestos;
- Verificar se o processo aqui representado pode ser replicado para outros modos de interação, como a interação realizada principalmente por comandos de voz ou através do rastreamento de olhos.
- Verificar se o processo apresentado pode ser reproduzido para representar visualizações de propósito específico.
- Aumentar o número de participantes em todas as etapas do processo.

- Melhorar a interface do protótipo, criar perfis de configuração individuais no protótipo e realizar testes de usabilidade, além dos ensaios de interação.
- Avaliar a utilização de diferentes ou novos dispositivos de entrada, wearables e baseados em visão, como luvas de realidade virtual ou múltiplos Leap Motion.
- Utilizar a interação em ambientes com alto potencial para usufruir deste tipo de interação, como projeções e *wall mounted displays*.
- Realizar um levantamento formal das taxas de acerto para diferentes interações utilizando diferentes meios de entrada, como uma forma de propor mais acertadamente como mapear que tipo de interação.

REFERÊNCIAS

BADELLEY, A. Working memory: The interface between memory and cognition. **Journal of cognitive neuroscience**, 1992. 281-288.

BERINGER, N. **Evoking Gestures in SmartKom – Design of the Graphical User Interface**. International Gesture Workshop. Berlin: Springer. 2001. p. 228-240.

BIMBER, O.; RASKAR, R. **Spatial Augmented Reality: Merging Real and Virtual Worlds**. Danvers, MA: A K Peters/CRC Press by Taylor & Francis Group, 2005.

BLINOV, M.; DERU, M.; SONNTAG, D. **Mobile thumb interaction and speech**. Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services. [S.l.]: ACM. 2011.

BRATH, R. 3d infovis is here to stay: Deal with it. **2014 IEEE VIS International Workshop on 3DVis** , 2014. 25-31.

BREHMER, M. .; MUNZNER, T. A multi-level typology of abstract visualization tasks. **IEEE Transactions on Visualization and Computer Graphics**, 2013. 2376-2385.

BRUDER, G.; STEINICKE, F. .; STURZLINGER, W. To touch or not to touch?: comparing 2D touch and 3D mid-air interaction on stereoscopic tabletop surfaces. **Proceedings of the 1st symposium on Spatial user interaction**, 2013. 9-16.

CABREIRA, A. T.; HWANG, F. An analysis of mid-air gestures used across three platforms. **Proceedings of the 2015 British HCI Conference, HCI 15**, Lincoln, 13-17 jul. 2015. 257-258.

CARD, S. K. . M. J. D. .; SHNEIDERMAN, B. **Readings in information visualization: using vision to think**. [S.l.]: [s.n.], 1999.

CARNEIRO, N. J. S. et al. Hand-based Interaction in a 3D InfoVis Environment. **Anais do XVI Workshop de Realidade Virtual e Aumentada**, Presidente Prudente, p. 132-137, 2015.

CARTER, M. E. A. **Player Identity Dissonance and Voice Interaction in Games**. Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play. [S.l.]: ACM. 2015.

CHEN, L.-W.; PENG, Y.-H.; TSENG, Y.-C. **An Augmented Reality Based Group Communication System for Bikers Using Smart Phones**. IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). Seattle, WA: IEEE. 2011. p. 325-327.

CHEN, Y. et al. **Research about Mobile AR System based on Cloud Computing**. 22nd Wireless and Optical Communication Conference (WOCC). Chongqing: IEEE. 2013. p. 355 - 359.

CHI, E. H. H. **A Framework for Visualizing Information**. [S.l.]: Springer Science & Business Media. 2002.

DUMAS, B. . L. D. .; OVIATT, S. **Multimodal Interfaces: a Survey of Principles, Models and Frameworks**. Human machine interaction. Berlin: Springer. 2009. p. 3-26.

GOOGLE INC. Voice Driven Web Apps: Introduction to Web Speech API | Web. **Developers Google**, 2013. Disponivel em: <<https://developers.google.com/web/updates/2013/01/Voice-Driven-Web-Apps-Introduction-to-the-Web-Speech-API>>.

GRANDHI, S. A. . J. G. .; MITTELBERG, I. **Understanding Naturalness and Intuitiveness in Gesture Production: Insights for Touchless Gestural Interfaces**. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. [S.l.]: ACM. 2011. p. 821-824.

GRANDHI, S. A.; JOUE, G.; MITTELBERG, I. Understanding Naturalness and Intuitiveness in Gesture Production: Insights for Touchless Gestural Interfaces. **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'11**, Vancouver, 7-12 maio 2011. 821-824.

GRANT, C. Ten Things You Should Know about Hand and Wrist Pain. **F-One Ergonomics**, Michigan, 2002.

HAMID, A. K. A. et al. **Design and Implementation of an AR-assisted Tool for Basic Home Network Management**. IEEE/IPSJ International Symposium on Applications and the Internet. Munich, Bavaria: IEEE. 2011. p. 208-213.

HAMPSON, P. J. . A. P. E. M. **Understanding Cognition**. Oxford: Blackwell, 1996.

HEDGE, A.; MUSS, T. M. .; BARRERO, M. Comparative study of two computer mouse designs. **Methods**, 1999.

HUMMELS, C. .; STAPPERS, P. J. **Meaningful gestures for human computer interaction: beyond hand postures**. Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition. [S.l.]: IEEE. 1998. p. 591-596.

JANKOWSKI, J. .; HACHET, M. A survey of interaction techniques for interactive 3D environments. **Eurographics 2013-STAR**, 2013.

JIANG, L.-F. E. A. **Design and implementation of speech interaction system in pervasive computing environment**. 2006 International Conference on Machine Learning and Cybernetics. [S.l.]: IEEE. 2006.

KARAM, M. **A taxonomy of gestures in human computer interactions**. [S.l.]: [s.n.]. 2005.

KORTUM, P. HCI beyond the GUI: Design for haptic, speech, olfactory, and other nontraditional interfaces, 2008.

KULSHRESHTH, A. .; LAVIOLA JR, J. J. Exploring the usefulness of finger-based 3D gesture menu selection. **Proceedings of the 32nd annual ACM conference on Human factors in computing systems**, 2014. 1093-1102.

LARSEN, L. B. E. A. **Affordance in mobile speech-based user interaction**. Proceedings of the 9th international conference on Human computer interaction with mobile devices and services. [S.l.]: ACM. 2007.

LEAP MOTION. Leap Motion | MAC & PC Motion Controller for Games, Design, Virtual Reality & More. **Leap Motion**, 2016. Disponivel em: <<https://www.leapmotion.com/>>.

LEAP MOTION. Menu Design Guidelines. **Developer Leap Motion**, 2016. Disponivel em: <https://developer.leapmotion.com/documentation/csharp/practices/Leap_Menu_Design_Guidelines.html>.

LEE, J. .; KUNII, T. L. Model-based analysis of hand posture. **IEEE Computer Graphics and applications**, 1995. 77-86.

LIN, J.; WU, Y. .; HUANG, T. S. Modeling the constraints of human hand motion. **Proceedings of Workshop on Human Motion** , 2000. 121-126.

LINS, C. et al. **Animar: Augmenting the Reality of Storyboards and Animations**. XVI Symposium on Virtual and Augmented Reality. Salvador, BA: [s.n.]. 2014. p. 106-109.

MAHRIN, M. N.; STROOPER, P.; CARRINGTON, D. **Selecting Usability Evaluation Methods for Software Process Descriptions**. Proceedings of the 16th Asia-Pacific Software Engineering Conference. Penang: [s.n.]. 2009. p. 523-529.

MATSUSHITA, Y. E. A. **Design and implementation of speech interaction system in pervasive computing environment**. 2015 18th International Conference on Network-Based Information Systems (NBiS). [S.l.]: IEEE. 2015.

MIRANDA, B. P. et al. Categorizing Issues in Mid-air InfoVis Interaction. **Information Visualisation (IV), 2016 20th International Conference**, 2016. 242-246.

MIRANDA, B. P. et al. Evaluation of Information Visualization Interaction Techniques using Gestures and Widgets in 3D Environments. **Proceedings of HCI International 2016**, Toronto, 2016.

MULLER, J. . E. A. **Speech interaction in virtual reality**. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing. [S.l.]: IEEE. 1998.

MUNTEANU, C.; PENN, G. **Speech-based interaction: myths, challenges, and opportunities**. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. [S.l.]: ACM. 2016.

NIELSEN, M. . S. M. . M. T. B. .; GRANUM, E. **A Procedure for Developing Intuitive and Ergonomic Gesture Interface for HCI**. International Gesture Workshop. [S.l.]: Springer. 2003. p. 409-420.

NIELSEN, M. et al. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In: CAMURRI, A.; VOLPE, G. **Gesture-Based Communication in Human-Computer Interaction**. Genova: Springer, 2004. p. 409-420.

NOWICKI, M. . P. O. . W. J. . Z. K. . & J. W. Gesture recognition library for Leap Motion controller, 2014.

NOWICKI, M. et al. **Gesture recognition library for Leap Motion controller**. Bachelor Thesis. [S.l.]: Poznan University of Technology. 2014.

OVIATT, S. **Advances in Robust Multimodal Interfaces**. IEEE Computer Graphics and Applications. [S.l.]: [s.n.]. 2003. p. 62-68.

OVIATT, S. **Humam-Centered Design Meets Cognitive Load Theory: Designing Interfaces that Help People Think**. Proceedings of the 14th ACM international conference on Multimedia. [S.l.]: ACM. 2006. p. 871-880.

PFLEGING, B.; SCHNEEGASS, S.; SCHMIDT, A. **Multimodal interaction in the car: combining speech and gestures on the steering wheel**. Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. [S.l.]: ACM. 2012.

RABINER, L. . & J. B. H. **Fundamentals of speech recognition**. [S.l.]: [s.n.], 1993.

REN, G. .; O'NEILL, E. 3D selection with freehand gesture. **Computers & Graphics**, 2013. 101-120.

REN, G.; O'NEILL, E. Touching the 3rd Dimension: 3D selection with freehand gesture. **Journal Computers and Graphics** , Elmsford, v. 37, n. 3, p. 101-120, maio 2013. ISSN 0097-8493.

SADANA, R. .; STASKO, J. Designing and implementing an interactive scatterplot visualization for a tablet computer. **Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces**, 2014. 265-272.

SCHIELE, A. .; VAN DER HELM, F. C. **Kinematic Design to Improve Ergonomics in Human Machine Interaction**. IEEE Transactions on Neural Systems and Rehabilitation Engineering. [S.l.]: [s.n.]. 2006. p. 456-469.

SEDIG, K. . P. P. . D. M. .; HAWORTH, R. **Human-Centered Interactivity of Visualization Tools: Micro- and Macro-level Considerations**. Handbook of Human Centric Visualization. New York: Springer New York. 2014. p. 717-743.

SEIXAS, M.; CARDOSO, J. .; DIAS, M. T. G. One Hand or Two Hands? 2D Selection Tasks with the Leap Motion Device, 2015.

SHAW, G. .; HEDGE, A. The effect of keyboard and mouse placement on shoulder muscle activity and wrist posture, 1997.

SOBOTA, B.; KOREČKO, Š.; HROZEK, F. **Mobile mixed reality**. 11th IEEE International Conference on Emerging eLearning Technologies and Applications. Stary Smokovec: IEEE. 2013. p. 355-358.

SPENCER, R. **Information Visualization – Design for Interaction**. [S.l.]: Person Education, 2007.

SRISUPHAB, A. et al. **Integrated ZooEduGuide with Multimedia and AR, From the largest living classrooms to wildlife conservation awareness**. TENCON 2014 - 2014 IEEE Region 10 Conference. Bangkok: IEEE. 2014. p. 1-4.

TELEA, A. C. **Data Visualization: Principle and Practice**. [S.l.]: CRC Press, 2014.

THALMIC LABS. Myo User Experience Guidelines 0.3. **Thalnic Downloads**, 2014. Disponivel em: <https://s3.amazonaws.com/thalnicdownloads/ux/MyoUserExperienceGuidelines-0.3.pdf>.

THALMIC LABS. Myo Gesture Control Armband | Wearable Technology by Thalnic Labs. **Myo**, 2016. Disponivel em: <https://www.myo.com/>.

ULLMAN, J. et al. A new approach to the mouse arm syndrome. **nternational Journal of Occupational Safety and Ergonomics**, 2003. 463-477.

UNITY. Unity - game engine, tools and multiplatform. **Unity**, 2016. Disponivel em: <https://unity3d.com/pt/unity>. Acesso em: 1 jul. 2016.

VASSELAI, G. T.; REIS, D. S.; GOMES, P. C. R. **A Case Study of Augmented Reality for Mobile Platforms**. XIII Symposium on Virtual Reality. Uberlandia: IEEE. 2011. p. 225-231.

VELLOSO, E. et al. An empirical investigation of gaze selection in mid-air gestural 3D manipulation. **Human-Computer Interaction**, 2015. 315-330.

WARE, C. **Information Visualization: Perception for Design**. [S.l.]: Elsevier, 2012. 2012 p.

YI, J. S. . A. K. Y. . S. J. .; JACKO, J. **Toward a Deeper Understanding of the Role of Interaction in Information Visualization.** IEEE transactions on visualization and computer graphics. [S.l.]: [s.n.]. 2007. p. 1224-1231.