Sugestões para projetos de interface que usam rastreamento de cabeça e comandos de voz

Alexandre A. Freitas, Carlos G. R. Santos, Marcelle P. Mota, Bianchi S. Meiguins
Universidade Federal do Pará
Belém, Pará
{aafreitas, carlosresque, mpmota, bianchi}@ufpa.br

ABSTRACT
Multimodal Interactions have been used in many fields of application, such as medicine, manipulation of assistive technologies, interactions in public environments, among others. It is important to not only develop technologies (hardware and software) but to study and project optimization possibilities for the usage of these innovative interfaces as well. This work aims to identify which are the major problems when using head tracking interactions combined with voice commands, this being a multimodal interaction. This evaluation is focused in the lowest level of interaction, which are actions more physical and less cognitive, such as click, drag and drop, scroll a page, among others. Therefore, as consequence of this research, there were proposed some suggestions to improvement of interface projects that use this form of interaction.

Author-Keywords
Multimodal Interfaces, Head Tracking, Voice Commands, Usability

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces; Input devices and strategies and Interaction styles.

INTRODUÇÃO
Interação multimodal é uma área de estudo que tem crescido em hardwares e softwares, devido ao avanço de recursos computacionais (como miniaturização de componentes eletrônicos, sensores, e computadores portáteis) e melhorias de algoritmos computacionais que realizam o processamento de imagem e áudio [1].

Neste sentido, Turk [2] identificou como um dos desafios da área, a avaliação dessas interações tanto em âmbito unimodal - que utilizam apenas um modo de interação - quanto em âmbito multimodal, a qual se refere à avaliação de várias combinações das interações unimodais.

As interações multimodais tem o objetivo de oferecer múltiplos modos de comunicação entre humanos e máquinas, de forma que essa comunicação pareça mais mais semelhante com a comunicação que o ser humano já está habituado [3].

Duas formas promissoras de interação multimodal são o rastreamento de cabeça e comandos de voz, uma vez que a informação está cada vez mais pervasiva, gerando a oportunidade das pessoas interagirem com essas informações de forma mais colaborativa e em ambientes públicos [4], o uso do mouse e teclado dificulta a interação. Além disso, Valkanova et al. [5, 6] mostra resultados que indicam que as formas não convencionais de interação em ambientes públicos despertam o interesse e a atenção do usuário em discutir e compartilhar pontos de vista.

Contudo, este trabalho tem o objetivo de avaliar a interação combinada de rastreamento de cabeça e comandos de voz, visando identificar os principais problemas de usabilidade, nas interfaces, e sugerir propostas para o design dessa modalidade de interação, baseadas nos resultados obtidos na avaliação.

Este estudo tem como escopo a avaliação das interações de baixo nível (mais física e menos cognitiva), que são comumente utilizadas nas aplicações, como clique, arrastar e soltar, rolar uma página entre outros. Para entender a hierarquia da interação referida neste trabalho, a Figura 1 apresenta uma pirâmide adaptada do trabalho de Sedig et al. [7] para interações na área de visualização da informação. O nível mais baixo da base da pirâmide contém as interações físicas – denominadas de interações de baixo nível – estas permeiam as aplicações de níveis mais altos.

Para avaliar essas interações de baixo nível foram realizados: a) um teste quantitativo medindo tempo e erro dos participantes em 62 tarefas simples em aplicações de visualização de Informação; b) uma entrevista com cada participante ao final de seus respectivos testes.

Este trabalho está organizado da seguinte forma: estudos que utilizaram comandos de voz e/ou rastreamento de cabeça; ambiente e métricas utilizadas nos testes de cabeça; resultados dos testes; sugestões de melhorias para o design desse tipo de interação; conclusões adotadas e trabalhos futuros.

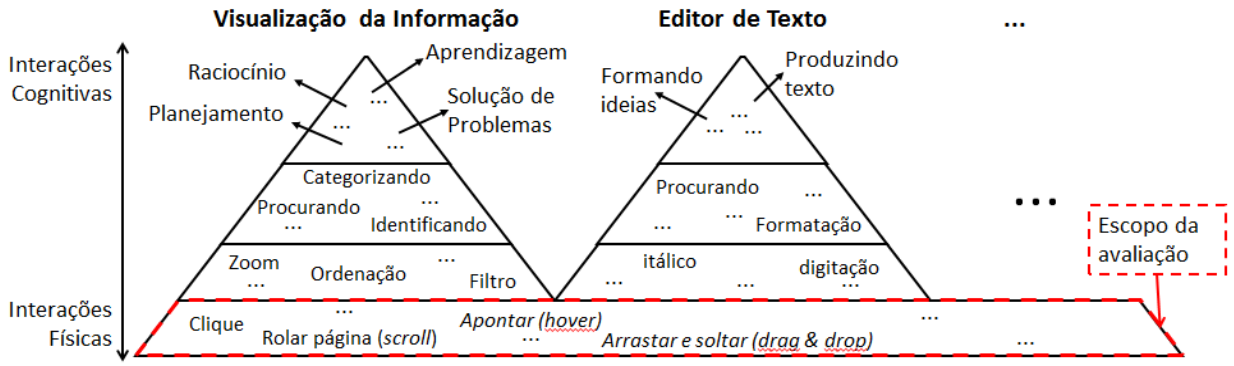

Figura 1. Pirâmide da interação de baixo nível em várias áreas de aplicação. Adaptação de Sedig et al [7].

INTERAÇÃO COM RASTREAMENTO DE CABEÇA E/OU COMANDOS DE VOZ

Esta seção apresenta o contexto em que este trabalho está inserido, ressaltando os estudos relacionados que auxiliaram no desenvolvimento da pesquisa e no conhecimento das tecnologias de rastreamento de cabeça e comandos de voz.

Krapic et al. [8] propuseram uma versão onde é estabelecido um rastreamento da cabeça do usuário para o movimento do ponteiro do mouse, a ação do ponteiro é feita a partir do piscar dos olhos do usuário. Os autores informaram que, interações que simulem o *mouse* não são muito aceitas, pois demandam um alto consumo de tempo para realizar as ações. O presente trabalho visa identificar quais são as possíveis causas desse consumo de tempo (identificado em [8]) e propor algumas sugestões que possam amenizar esse problema.

Em outra perspectiva, Sugai et al. [9] e Roig-Maimó [10], usaram o rastreamento de cabeça para otimizar a tarefa de localizar componentes gráficos dentro de uma aplicação de dispositivos móveis e desktop respectivamente. No presente trabalho a interação por rastreamento de cabeça foi utilizada para mover um ponteiro dentro da aplicação, ou melhor, para navegação.

A voz como modo de interação entre humanos e máquinas pode ser utilizada de diversas formas, como, buscar um texto através da voz [11], a comunicação de dois usuários que estejam interagindo com um ambiente virtual colaborativo [12], ou utilizar pequenos comandos para que o sistema realize a ação determinada [13].

Elepfandt [13] usou um ambiente de *pointing and speech*, ou seja, apontar e falar. Onde o usuário poderia interagir com uma tela, apontando para o item e falando a ação respectiva para a interação. Esse é um modelo presente na comunicação humana e a ideia deste trabalho desenvolvido é explorar essa capacidade.

Alguns trabalhos utilizam o rastreamento de cabeça combinado com comandos de voz para desenvolver e avaliar aplicações assistivas [14, 15]. Embora este trabalho seja inicial, o escopo se encaixa somente nas sugestões de melhorias, que podem auxiliar o desenvolvimento e design dessas aplicações.

A interação por rastreamento de cabeça pode encontrar utilidade tanto em um contexto específico, como na área da medicina [16], ou em âmbito mais geral, como mover o mouse de um computador e utilizar comandos de voz para o clique e outros [17]. Este trabalho visa à avaliação neste âmbito mais geral.

AMBIENTE DE TESTE

Considerando o objetivo e o escopo deste trabalho, não será realizada uma avaliação da precisão dos hardwares utilizados.

Para o rastreamento da cabeça foi usado o *Tracker Pro* [18]. Esse dispositivo utiliza um sensor infravermelho que rastreia um ponto reflexivo colocado na testa do usuário (caso o usuário use óculos, é aconselhável colocar no centro da armação). Esse dispositivo simula o movimento do ponteiro do mouse através do movimento da cabeça do usuário, sendo possível interagir com componentes gráficos.

O modo de interação com os comandos de voz é feito a partir da tecnologia *Google Web Speech API* [19]. Foi desenvolvido um serviço que envia *streams* (envio contínuo de dados) de voz constantemente ao servidor e dependendo do comando emitido a ação é realizada no computador.

A Figura 2 mostra um diagrama que ilustra de forma geral o funcionamento do reconhecimento da voz, onde o usuário fala uma palavra, e o áudio é enviado ao servidor. Após um processamento, é retornada uma lista de texto com as possíveis frases que foram ditas, de posse dessa lista o sistema identifica qual a ação desejada pelo usuário e realiza a mesma.

A gramática utilizada para o comando de voz é apresentada na Tabela 1, mostrando tanto a palavra que o usuário pode falar, quanto à ação que a palavra executa no sistema. Cada ação é uma interação de baixo nível [7], apontar (*Hover*) e deslizar (*Sliding*) são interações usadas a partir do rastreamento de cabeça, por isso, não apresentam uma gramática para realizar o comando.

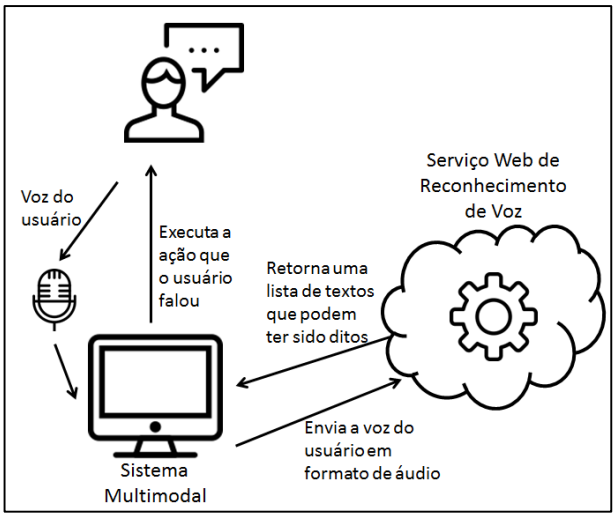

Figura 2. Modelo geral do funcionamento do reconhecimento da voz.

A Figura 3 apresenta uma imagem do ambiente em que foram realizados os testes com os usuários, e a disposição dos hardwares no mesmo.

Tabela 1. Gramáticas e ações do trabalho.

Gramática para Falar	Ação respectiva
Selecionar	Clicar (*Click*)
Segurar	Clicar e Segurar (*Drag*)
Soltar	Soltar o componente gráfico (*Drop*)
Agarrar	Navegar na página (*Scroll*)

Foi usado um monitor com 17 polegadas e resolução de 1280x1024 pixel (Figura 3 [B]). Utilizou-se um microfone (Figura 3 [A]) acoplado a um headphone, o qual possibilitava ao usuário ajustar a distância entre ele e sua boca.

O dispositivo para o rastreamento da cabeça utilizado está localizado acima da tela (Figura 3 [C]). Esta localização foi estabelecida para melhor alinhar a câmera do dispositivo e o ponto reflexivo (que o dispositivo rastreia) localizado na testa do usuário.

AVALIAÇÃO

Nesta seção é descrita como a avaliação do rastreamento de cabeça combinado com comandos de voz foi realizado, as métricas que foram utilizadas para avaliação das interações e o processo da avaliação,

As avaliações foram realizadas a partir dos cenários identificados por Lam et al. [20] [21], avaliação do Desempenho do Usuário (DU) e a avaliação da Experiência do Usuário (EU).

No cenário de DU as métricas para análises são aplicadas com base em duas perguntas: Quais os limites da proposta de interação? E quais os melhores e piores tipos de interações dentro dos componentes gráficos?

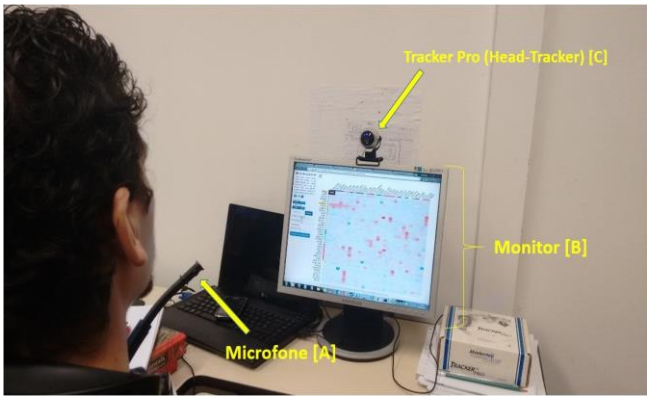

Figura 3. Foto do ambiente de teste.

Para isso foi avaliado o tempo de conclusão de cada tarefa e a quantidade de erros relacionados à interação com os componentes gráficos. O tempo foi registrado com o auxílio de um cronômetro. Enquanto que os erros eram contabilizados no momento em que o mesmo ocorria, através de um registro escrito.

Para o cenário de EU foi feita uma entrevista guiada por cinco perguntas ao final de todas as tarefas realizadas no cenário de DU. O objetivo das perguntas foi encontrar as dificuldades do usuário e obter os feedbacks. Estas entrevistas foram gravadas em áudio, e analisadas posteriormente. Abaixo segue a lista com as perguntas que serviram de roteiro para a entrevista.

Pergunta 1. Quais das formas de interações que você utilizou podem ser realmente úteis?
Pergunta 2. Que tipo de interação está faltando?
Pergunta 3. Como as interações utilizadas podem ser melhoradas?
Pergunta 4. Existe alguma limitação nessas interações que possam prejudicar sua utilização em larga escala?
Pergunta 5. As interações são fáceis de aprender e utilizar?

Processo da Avaliação

Para realizar a avaliação de rastreamento de cabeça e comandos de voz, foram apresentadas 62 tarefas para cada usuário, totalizando 620 tarefas. Participaram do teste 10 indivíduos, com ensino superior em andamento e estudantes de pós graduação.

As tarefas foram elaboradas em aplicações de visualizações da informação (InfoVis) que estavam disponíveis na internet [22]. Essas aplicações foram selecionadas por conterem uma quantidade variada de *widgets* (componentes interativos).

Os participantes foram submetidos a um período de cinco minutos para habituação aos hardwares e treino de tarefas semelhantes às aplicadas no teste. Desta forma o usuário poderia achar uma melhor posição da cabeça, se acostumar com o atraso no reconhecimento da voz (atraso causado pelo processamento e reconhecimento da voz, melhor observado na Figura 2) e entender de maneira geral o funcionamento da interface.

Cada tarefa tinha 1 minuto e 30 segundos de tempo máximo para sua conclusão. Este tempo foi estabelecido pela equipe de avaliação após um teste piloto. Apesar da quantidade de tarefas, estas eram concluídas em poucos segundo, em virtude de seu grau de dificuldade ser baixo. Um exemplo comum de tarefa era apontar o ponteiro (*hover*) sobre um componente gráfico presente na aplicação.

As tarefas foram desenvolvidas pensando em interações de baixo nível, então não poderiam demandar um processo cognitivo do usuário, mas apenas a sua ação física, por exemplo, clicar em um botão, arrastar um componente de um *slider*, selecionar uma caixa de *checklist*, entre outros.

Na Figura 4 é ilustrada uma das tarefas que foi proposta para os participantes, onde o usuário teria que interagir com o componente *slider* (Figura 4 [A]) e arrasta-lo até o ponto determinado (Figura 4 [B]). A seta indica o sentido da interação com o componente gráfico.

O critério adotado para registrar o erro consistia na falha da

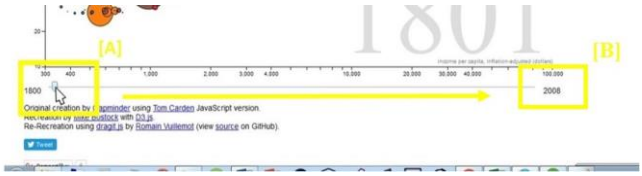

Figura 4. Exemplo de aplicação dos testes.

interação com o componente gráfico. Como, por exemplo, em um cenário onde o usuário tenta clicar em um botão e o clique é efetuado fora do componente, isto sugere um erro na interação com o botão.

RESULTADOS

A avaliação para interação de rastreamento de cabeça e comandos de voz foi feita a partir da análise do DU (na qual foram utilizadas métricas quantitativas) e da análise da EU (que foram utilizadas métricas qualitativas).

Resultados Quantitativos

A Figura 5 mostra os resultados da avaliação quantitativa (tempo e erro) com relação ao tipo de interação de baixo nível presentes na aplicação. É possível visualizar que as interações que utilizam agarrar (*Scroll*) demandaram mais tempo de execução, e também tiveram maior taxa de erro na interação. Segurar e soltar (*Drag and Drop*) foi a segunda interação que mais levou tempo para ser concluída, assim como uma taxa de erro mais elevada que as outras interações. As demais interações não tiveram uma taxa de erros ou consumo de tempo expressivo. Apontar (*Hover*) foi o que menos consumiu o tempo da tarefa e juntamente a isso, não apresentou taxa de erro na interação. Na análise da Figura 5, não foi identificada uma relação direta entre o tamanho do componente e o tempo que se leva para concluir as tarefas.

A barra de erro apresentada nos resultados são calculadas pelo desvio padrão do erro e do tempo adquirido com a interação do usuário. Estas barras representam a variabilidade dos valores, servindo principalmente como escala de mínimo e máximo.

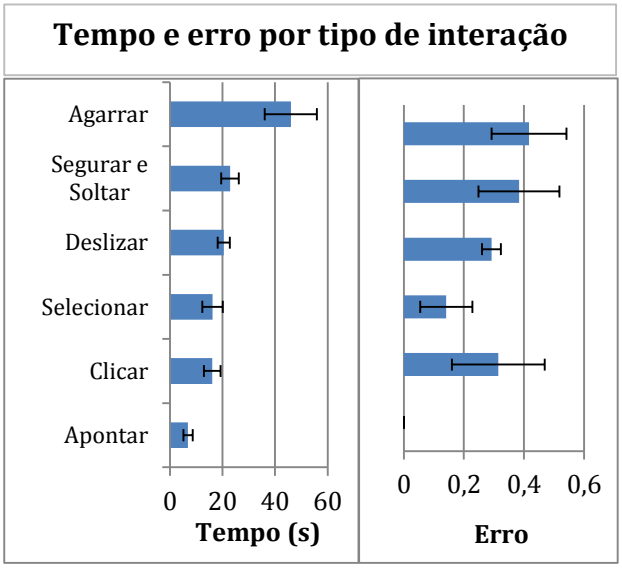

Figura 5. Gráfico de tempo e erro relacionado ao tipo de interação.

As interações são compostas por etapas para serem concluídas, como por exemplo, para o usuário clicar em um botão: 1. Deve-se navegar até o componente; 2. Falar a palavra respectiva do comando, totalizando 2 etapas.

Na Figura 6 é apresentado a relação entre o tamanho do componente gráfico existente na aplicação. Na análise, percebe-se que componentes com tamanhos classificados (pela equipe do trabalho) como grandes (menor medida em pixel maior ou igual à 25) e médios (menor medida em pixel maior ou igual à 15) possuem taxa de erro baixa. E em componentes que se classificam em pequenos (menor medida em pixel maior ou igual à 5) e mínimos (menor medida em pixel menor que 5) obtiveram taxa de erro elevadas.

A Figura 7 apresenta a relação entre tempo e erro da quantidade de etapas necessárias para concluir cada tarefa. O tempo para tarefas que possuíam 6 etapas foi maior, o que é esperado, pois a quantidade etapas e o tempo são diretamente proporcionais. As tarefas que possuíam 2, 3, 4 e 5 etapas variaram de forma crescente, se diferenciando pouco uma das outras.

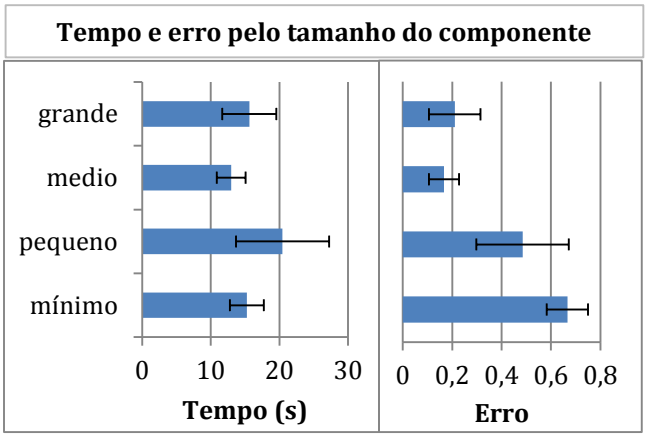

Figura 6. Gráfico de tempo e erro com relação ao tamanho do componente.

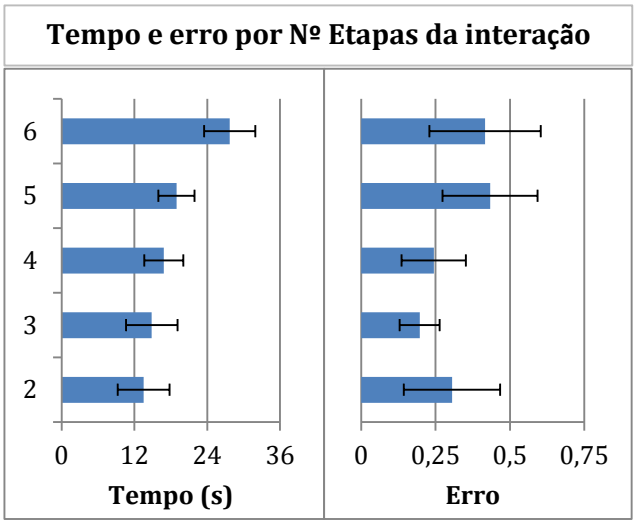

Figura 7. Gráfico de tempo e erro com relação à quantidade de etapas da tarefa.

Resultados Qualitativos

A avalição qualitativa foi realizada a partir de entrevista com os usuários do experimento contendo 5 perguntas ao fim das tarefas.

Das respostas foram extraídas as principais ideias e classificadas em categorias. Estas categorias são organizadas dentro das perguntas presentes no roteiro da entrevista e são apresentadas abaixo.

Quais das formas de interações que você utilizou podem ser realmente úteis?

Categoria 1. Selecionar e *drag & drop*.
Categoria 2. Interação com rastreamento de cabeça, exceto para itens pequenos.

Que tipo de interação está faltando?

Categoria 1. Escrever texto em caixas de texto.
Categoria 2. *Feedback* visual indicando quais comandos foram reconhecidos.
Categoria 3. Melhorar a precisão.
Categoria 4. Atalhos, como o desfazer (CTRL+Z).

Como as interações utilizadas podem ser melhoradas?

Categoria 1. A precisão do rastreamento de cabeça nas posições extremas da tela (ex. cantos da tela).
Categoria 2. O rastreamento de cabeça deve atender tanto a movimentos curtos quanto a movimentos rápidos com precisão.
Categoria 3. Travar o ponteiro quando o usuário começar a falar.
Categoria 4. Diminuir o atraso do reconhecimento da voz.
Categoria 5. Reposicionamento automático da posição do ponteiro.

Existe alguma limitação nessas interações que possam prejudicar sua utilização em larga escala?

Categoria 1. Interação com itens pequenos e pouca precisão nos cantos da tela.
Categoria 2. Parar o ponteiro e falar é difícil.
Categoria 3. O atraso do reconhecimento da voz e a pouca diversidade de comandos.
Categoria 4. A falta de uma interface gráfica adaptada.

As interações são fáceis de aprender e utilizar?

Categoria 1. Depois de um tempo os comandos se tornam autoexplicativos.
Categoria 2. O click e a centralização do ponteiro são fáceis.
Categoria 3. A interação é fácil de aprender, mas não de usar.
Categoria 4. Depois de mostrado os comandos (gramática) a interação fica fácil.

Com base nas respostas que os participantes deram, é possível analisar alguns comentários, como por exemplo, os comandos de voz que realizam as ações de selecionar, segurar e soltar (*drag & drop*) são mais uteis para serem usadas em uma aplicação. Assim como as interações com itens pequenos, que segundo os participantes foram desconfortáveis.

Com a análise qualitativa e quantitativa dos dados, sugestões para projetos de melhorias de interface de software foram propostas, objetivando maior ergonomia e um bom nível de usabilidade para o usuário.

SUGESTÕES PARA INTERAÇÃO DE RASTREAMENTO DE CABEÇA E COMANDOS DE VOZ

As sugestões de projeto de designer podem servir de guia para uma possível aplicação que tenha pelo menos essa combinação (rastreamento de cabeça e comandos de voz) de interação multimodal. A seguir são apresentadas as sugestões propostas deste trabalho.

Evitar itens selecionáveis pequenos

Os resultados obtidos com os usuários demonstraram que itens pequenos ou mínimos afetam a conclusão de uma tarefa. Neste ponto é aconselhável que a interface apresente itens selecionáveis médios ou grandes.

A Figura 8 apresenta a ideia da utilização de componentes gráficos adaptados em relação ao seu tamanho e ocupação de espaço na tela. São apresentados os componentes

gráficos que não seguem essa diretriz, que normalmente são configurados com tamanhos e interações padrões do sistema operacional ou da linguagem de programação utilizada e são apresentados exemplos de como adaptar os componentes estes respectivos componentes.

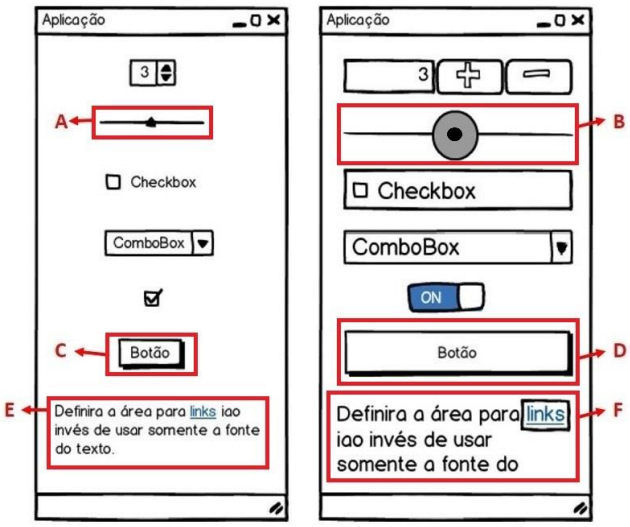

Figura 8. Exemplo de adaptação para tamanho e área de interação dos componentes gráficos.

O item de *slider* apresentado na área A da Figura 8 exemplifica o tamanho padrão deste componente em muitos sistemas operacionais. Na área B da Figura 8 o *slider* é representado de uma maneira adaptada, onde o componente selecionável é maior.

Na área D, assim como os demais componentes apresentados, a alteração foi em proporção e área de interação, para promover o uso de todo espaço disponível que for possível.

Na área E o usuário interage com um *link* clicando na fonte das letras, dependendo da fonte utilizada e do tamanho da mesma a interação pode ser prejudicada. Na área F, é apresentada uma adaptação onde o texto possui uma fonte maior e o *link* está dentro de um retângulo interativo que contém a palavra selecionável.

Interações com *scroll* devem ser evitadas

O *scroll* requer que o usuário tenha um controle da velocidade do movimento da cabeça. A precisão da interação se torna ainda maior, quando a página que está sendo navegada é grande.

A Figura 9 (1) apresenta a forma de scroll utilizada na grande maioria das aplicações, e a Figura 9 (2) apresenta a sugestão de melhoria na interface para facilitar a interação por rastreamento de cabeça e comando de voz.

A grande dificuldade para realizar o scroll com o componente da Figura 9 (1) está em quando a página de conteúdo é grande, o usuário precisa ter um controle mais preciso de velocidade, movimento e posicionamento da cabeça. Já o componente da Figura 9 (2) permite ao usuário controlar a navegação no texto de forma mais confortável e eficiente.

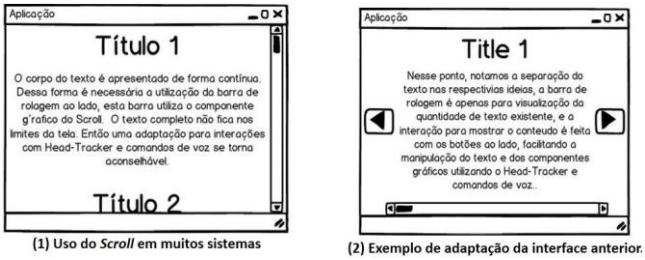

Figura 9. Exemplo de interface que utiliza scroll. (1) Exemplo do uso comum do scroll nos sistemas. (2) Uma sugestão de adaptação para interação.

Oferecer *feedbacks* visuais para o reconhecimento de comandos de voz

Em algumas tarefas o ponteiro do mouse não oferece um *feedback* visual da determinada ação. Os participantes dos testes observaram que há uma ausência de *feedbacks* visuais em algumas interações.

Na Figura 10 estão dispostos alguns exemplos deste tipo de *feedback*, como quando o comando emitido é o de clicar, uma animação de círculo crescente aparece envolta do ponteiro por um curto período de tempo, indicando o sucesso do reconhecimento da ação. Outras ações podem oferecer feedbacks para o usuário através de ícones como o caso de atalhos do sistema operacional.

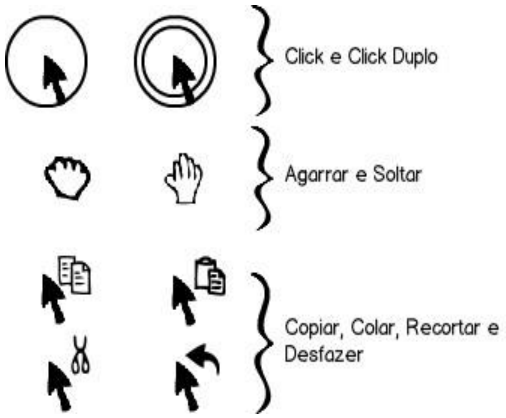

Figura 10. Exemplos de feedbacks visuais para o reconhecimento do comando de voz.

Utilizar atalhos na interação

Os atalhos são meios rápidos de realizar determinada ação que seria feita com três ou mais etapas. Geralmente, os sistemas e aplicações disponibilizam os atalhos como uma combinação de teclas, fixando sua utilização para o teclado.

A Figura 11 (1) demonstra a ideia da quantidade de etapas que são realizadas com a interação do *mouse*. Assim como, na Figura 11 (2), é ilustrada os atalhos através do comando de voz.

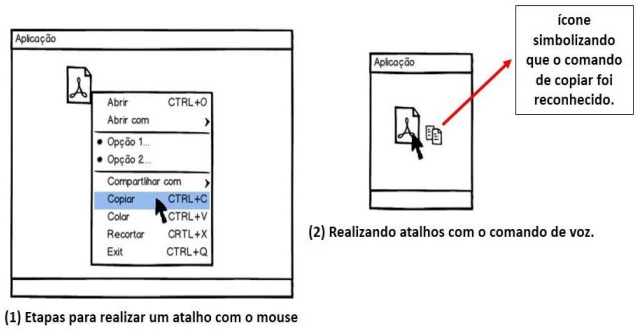

Figura 11. Interação com atalhos. (1) Etapas para realizar um atalho com o mouse. (2) Realizando atalhos com o comandos de voz.

A adaptação da interface mostrada na Figura 11 (2), é feita para que os atalhos possam ser efetuados através do comando de voz. O ícone que aparece acima do ponteiro, é para que seja oferecido um *feedback* visual de que o comando de voz foi reconhecido com sucesso.

Utilizar o potencial da voz para navegação e interação

A voz pode oferecer uma grande variedade de interações, tais como navegação, cliques, seleção de objetos, entre outros. O ideal é que usuário possa fazer uso da voz na forma mais natural possível, e a interface deve auxiliar o usuário em quais comandos estão disponíveis para fala. A Figura 12 mostra um exemplo de interface com componentes como botões e abas, cujas características podem ser lidas tais como rótulos, cor, posição, etc.

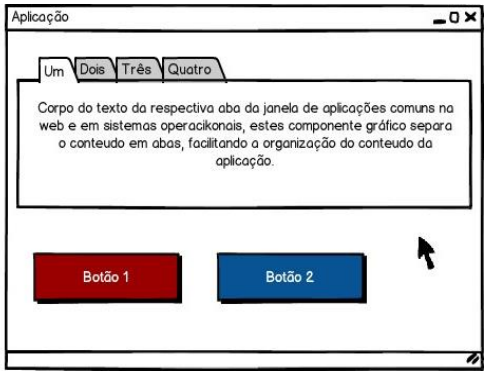

Figura 12. Interface que possa oferecer um sistema de comando de voz complexo.

Os componentes gráficos existentes na Figura 12 podem ter interação apenas com o comando de voz. Um exemplo seria o usuário falar o comando "Clicar no botão vermelho", e assim o botão respectivo faria a ação. Neste cenário, o rastreamento de cabeça iria auxiliar o usuário, por exemplo, extraindo uma área de interesse do usuário (para onde ele está olhando) servindo de filtro para o processamento inteligente da voz.

Dar dicas contextualizadas para utilização dos comandos de voz

Dar dicas contextualizadas aos usuários para interagirem com determinado componente gráfico, é muito importante, pois desta maneira o usuário não precisa lembrar constantemente o comando específico que precisa ser ditado para gerar a ação correspondente.

Na Figura 13 é ilustrado uma possível maneira de como dar dicas sobre as interações de determinados componentes gráficos que ponteiro se localiza.

Figura 13. Exemplo de dicas através de *tooltip*. (1) Dicas contextualizadas em botões. (2) Dicas contextualizadas em Slider.

Neste exemplo são usados *tooltips*, onde o usuário posiciona o ponteiro no componente gráfico, e, uma caixa de diálogo aparece por um tempo determinado. Na Figura 13 (2) o usuário está segurando o marcador do *slider*, e, o mesmo, dá uma dica, mostrando o que o usuário pode falar para finalizar a interação.

Utilizar uma quantidade de sinônimos e variações das palavras para o reconhecimento de voz

Os comandos de voz devem possuir uma quantidade adequada de sinônimos que representem a mesma ação para facilitar a interação sem a exigência de memorização [23].

A Tabela 2 demonstra alguns sinônimos que podem ser utilizados para uma ação do comando de voz, onde a primeira coluna apresenta a ação, e a segunda coluna mostra alguns possíveis sinônimos que podem ser usados para realizar a respectiva ação.

Tabela 2. Tabela com possíveis sinônimos para serem implementados.

Ação	Possíveis Sinônimos
Clique	Selecionar, clicar, pressionar, apertar, acionar.
Duplo clique	Duplo, clique duplo, dois, duas vezes.
Drag (segurar)	Segurar, pegar, agarrar, prender.
Drop (soltar)	Soltar, largar, desprender, desapertar.

Sinônimos que fazem parte do cotidiano dos usuários podem ser mais intuitivos para uma interação mais natural [23]. Na Tabela 2, por exemplo, o evento de segurar um componente pode ser representado pelos comandos:

"agarrar", "segurar", "pegar", "prender", e variações na conjugação destes verbos.

Porém, a utilização de muitos sinônimos pode causar uma confusão no usuário, algumas palavras podem ser sinônimas em outro contexto e podem mudar o sentido daquela ação.

Evitar interações nos cantos da tela

Interações que são efetuadas nos extremos da tela se mostraram menos precisas. Estas interações que se encontram nos cantos da tela, tornam-se imprecisas pelo fato do *hardware* perder o rastreamento quando se inclina demais a cabeça.

Logo, esta sugestão se aplica a hardwares baseados em visão que possuem apenas uma câmera, uma vez que, esse problema pode ser evitado com hardwares vestíveis ou redundância de câmeras. Entretanto, o custo aumenta com adição de equipamentos que evitam esse problema.

A Figura 14 apresenta duas interfaces, sendo que na Figura 14 (1) mostra um gráfico com um *slider*, o título do gráfico e dois botões que se localizam comumente em diversas aplicações. E na Figura 14 (2) mostra a mesma interface, porém aplicando a ideia de que as posições dos componentes interativos da interface não estão nos extremos da tela.

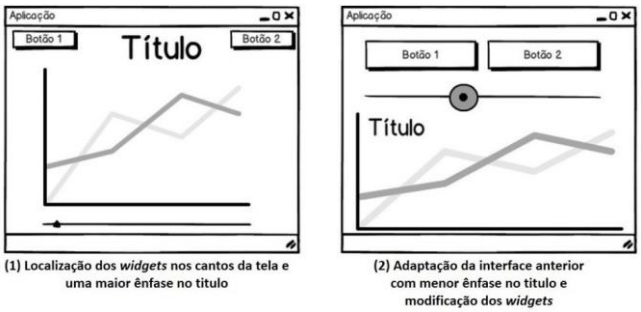

Figura 14. Adaptação da localização dos componentes gráficos. (1) Localização comum dos widgets nos sistemas. (2) Adaptação da localização do widgets para interação.

Utilizar técnicas para travar/manter o ponteiro em regiões de interesse

As regiões de interesse podem ser entendidas, neste contexto, como os componentes gráficos que permitem alguma interação. Estes componentes dentro de uma interface podem ser botões, *sliders*, caixas de seleção, etc.

Na Figura 15 é apresentada um exemplo para manter o ponteiro em uma região de interesse. Através de técnica de campo magnético [24], onde, uma vez próximo ao objeto de interesse, o ponteiro é atraído lentamente para mesmo.

A elipse envolta do botão apresentado na Figura 15 é uma espécie de campo magnético. Este campo atrai o ponteiro para o centro do componente gráfico. Essa atração pode ser feita em uma velocidade constante para que o usuário perceba a atração e também possa sair do campo caso não tenha interesse em interagir com o componente.

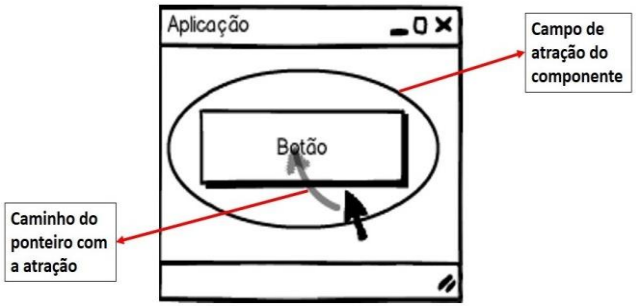

Figura 15. Exemplo de atração do ponteiro para o componente gráfico.

Vale ressaltar que o campo magnético é apenas um exemplo de como travar o ponteiro do mouse, outras técnicas para o mesmo propósito podem ser adotadas.

Utilizar técnicas para reposicionar a posição do ponteiro

Uma das questões levantadas nas entrevistas após os testes é a necessidade realizar um reposicionamento (calibração) automatizado do ponteiro. O rastreamento da cabeça pode em alguns momentos não corresponder para onde o usuário está olhando e onde está o ponteiro na tela. Alguns motivos podem provocar essa descalibração, dentre eles estão as interações com componentes nos cantos da tela e a velocidade do movimento da cabeça.

Nos testes os usuários podiam falar um comando de voz que centralizava o ponteiro na tela. Porém, este comando não faz parte da interação em si e é apenas uma medida corretiva para o problema da descalibração do rastreador de cabeça. O fato de o usuário recalibrar o ponteiro várias vezes se torna cansativo.

Uma técnica que pode ser utilizada para manter uma consistência automática entre a direção que o usuário está olhando e a posição do ponteiro na tela é estimativa da direção da cabeça (conhecido como *Head Pose*) [25]. Esta técnica tenta estimar qual a direção que o usuário está olhando (considerando que seus olhos estão sempre para frente), sendo assim, esta técnica pode ser utilizada para corrigir durantes alguns momentos esse problema.

Sugestões extras

As sugestões abaixo são comuns em vários projetos, mesmo que não sejam comandos de voz ou rastreamento de cabeça. A presença dessas sugestões se refere à construção e adequação de uma boa experiência do usuário durante um projeto. Neste caso, estas sugestões são voltadas para a interação como rastreamento de cabeça combinados com comandos de voz.

Oferecer treinamento para os usuários que interagem pela primeira vez

Os comandos de voz oferecidos são compostos por palavras pequenas e intuitivas que são rapidamente aprendidas pelo próprio contexto em que são utilizadas. Entretanto, o usuário não conhece o vocabulário utilizado pela aplicação quando utiliza o mesmo pela primeira vez. Sendo assim, o

usuário deve ter alguns minutos utilizando a interface para seu treinamento. O usuário deve falar alguns comandos, interagir com os componentes gráficos existentes na interface e se adaptar com o sistema.

Utilizar outras tecnologias para melhorar a precisão da interação

Os modos de interações variam em gestos, características físicas (por exemplo, piscar dos olhos), rastreamento dos olhos, entre outros modos que podem suprir algumas desvantagens de utilizar essa combinação da interação utilizada neste trabalho.

CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho, foram realizados testes com a interação de rastreamento de cabeça combinados com comandos de voz. A partir da análise quantitativa e qualitativa dos testes, foi possível gerar sugestões de melhorias que podem auxiliar os projetistas a construir uma interface que evite os problemas identificados nas avaliações. Estas sugestões atuam diretamente no modo de como os componentes gráficos, suas localizações e suas estruturas devem ser organizadas em uma interface multimodal.

A avaliação realizada neste trabalho teve um objetivo explorativo, no sentido de identificar os principais problemas existentes, nessa combinação de interação multimodal em um ambiente comumente utilizado. Assim foi possível identificar sugestões que podem melhorar a interação nestes ambientes.

Como o objetivo era explorativo não foram utilizadas métricas mais robustas, sendo assim, identifica-se como trabalho futuro a implementação dessas sugestões com a finalidade de validá-las, com testes positivos (implementação dessas sugestões) em relação ao cenário sem as sugestões e uma comparação com interação considerada convencionais, que neste caso é o mouse e teclado.

Também é identificado com trabalho futuro utilizar métricas conhecidas na literatura para medição de apontadores, como a ISO 9241-9, juntamente com métricas de experiência do usuário.

REFERÊNCIAS

1. Alejandro Jaimes and Nicu Sebe. 2007. Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding*. v. 108, p. 116-134.
2. Matthew Turk. 2014. Multimodal interaction: A review. *Pattern Recognition Letters*. v. 36, p. 189-195.
3. Bruno Dumas, Denis Lalanne and Sharon Oviatt. 2009. Multimodal Interfaces: A Survey of Principles, Models and Frameworks. In *Human Machine Interaction.* 3-26. Volume 5440 of the series Lecture Notes in Computer Science.
4. Rener B. Silva, Jessica Colnago & Junia Anacleto. 2014. Design de Aplicações Para Interação em Espaços Públicos: Formalizando as Lições Aprendidas. *IHC'14, Brazilian Symposium on Human Factors in Computing Systems.* Foz do Iguaçu, PR, Brasil.
5. Nina Valkanova, Sergi Jorda and Andrew V. Moere. 2015. Public visualization displays of citizen data: Design, impact and implications. *International Journal of Human-Computer Studies.* Baltimore, MD, USA.
6. Nina Valkanova, Robert Walter, Andrew V. Moere, Jörg Müller. 2014. MyPosition: Sparking Civic Discourse by a Public Interactive Poll Visualization. *Conference on Computer Supported Cooperative Work*. p. 1323-1332.
7. Kamran Sedig, Paul Parsons, Mark Dittmer, and Robert Haworth. 2014. *Human-Centered Interactivity of Visualization Tools: Micro- and Macro-level Considerations*. Springer New York
8. Luka Krapic, Kristijan Lenac, and Sandi Ljubic. 2013. *Integrating Blink Click interaction into a head tracking system: implementation and usability issues*. Springer-Verlag Berlin Heidelberg.
9. Ryosuke Sugai and Masamitsu Kurisu. 2015. Position Determination of a Popup Menu on Operation Screens of a Teleoperation System Using a Low Cost Head Tracker. 13-16. In *15th International Conference on Control,Automation and Systems.* Busan, Korea.
10. Maria F. Roig-Maimó, Javier V. G´omez and C. Manresa-Yee. 2015. Face Me! Head-Tracker Interface Evaluation on Mobile Devices. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (CHI '15). 1573-1578. http://dl.acm.org/citation.cfm?doid=2702613.2732829
11. Petar Aleksic, Cyril Allauzen, David Elson, Aleksandar Kracun, Diego M. Casado and Pedro J. Moreno. 2015. Improved Recognition of Contact Names in Voice Commands. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 5172 – 5175.
12. Bianchi S. Meiguins, Aruanda S. Gonçalves, Denis N. A. Santos, Marcelo B. Garcia, Rosevaldo D. S. Jr. 2003. Interação em Ambientes Virtuais Tridimensionais Utilizando Comandos de Voz. VI *Symposium on Virtual Reality*, Ribeirão preto. Oct. 15-18.
13. Monika. Elepfandt. 2012. Pointing and Speech - Comparison of Various Voice Commands. In *Proceedings of the 7th Nordic Conference on Human - Computer Interaction: Making Sense Through Design* (NordiCHI '12). 807-808. http://dl.acm.org/citation.cfm?doid=2399016.2399158
14. Eiichi Ito. 2001. Multi-modal Interface with Voice and Head Tracking for Multiple Home Appliances. *Kanagawa Rehabilitacion Center*.

15. Vikram Jeet, Hardeep S. Dhillon and Sandeep Bhatia. 2015. Radio Frequency Home Appliance Control based on Head Tracking and Voice Control for Disabled Person. In *Communication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on.* 559 – 563. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7279981

16. Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Graeme Penney, Andreas Varnavas, Helena Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur and Tom Carrell. 2014. Touchless interaction in surgery. *Magazine Communications of the ACM*. New York, NY, USA.

17. Devanand G. Khandar; Manteand, R. V., Prashant N. Chatur. 2015. Vision Based Head Movement Tracking for Mouse control. *International Journal of Advanced Research in Computer Science*.

18. AbleNet Inc. Tracker Pro. Retrieved May 18, 2016 from https://www.ablenetinc.com/trackerpro.

19. Google Inc. Web Speech API. Retrieved May 18, 2016 from https://www.google.com/intl/pt/chrome/demos/speech.html

20. Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant and Sheelagh Carpendale. 2012. Empirical Studies in Information Visualization: Seven Scenarios. In *IEEE Transactions on Visualization and Computer Graphics* (TVCG). 1520-1536. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6095544

21. Tobias Isenberg, Petra Isenberg, Jian Chen, Michael Sedlmair and Torsten Moller. 2013. A Systematic Review on the Practice of Evaluating Visualization. In *IEEE Transactions on Visualization and Computer Graphics* (TVCG). 2818-2827. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6634108

22. D3.Js. 2016. *D3 Js*. Retrieved May 18, 2016 from https://d3js.org/

23. Jakob Nielsen and Rolf Molich. 1990. Heuristic Evaluation of User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '90). 249-256. http://dl.acm.org/citation.cfm?doid=97243.97281

24. Elise V. D. HOVEN and Ali Mazalek. 2011. Grasping gestures: Gesturing with physical artifacts. In *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*. 255 – 271.

25. Erick Murphy-Chutorian and Mohan M. Trivedi. 2009. Head Pose Estimation in Computer Vision: A Survey. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (TPAMN). 607 – 626. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4497208