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

ANÁLISE DE ASPECTOS DE USABILIDADE EM INTERAÇÕES NATURAIS VIA INTERFACES MULTIMODAIS

Lucio Polese Cossio

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciência da Computação na Pontifícia Universidade Católica do Rio Grande do Sul.

Orientadora: Prof. Milene Selbach Silveira

Porto Alegre

2014

Dados Internacionais de Catalogação na Publicação (CIP)

C836a Cossio, Lucio Polese

Análise de aspectos de usabilidade em interações naturais via interfaces multimodais / Lucio Polese Cossio. – Porto Alegre, 2014.
125 f.

Dissertação (Mestrado) – Faculdade de Informática, PUCRS.
Orientador: Profª. Drª. Milene Selbach Silveira.

1. Informática. 2. Interface com o Usuário. 3. Computação Móvel. I. Silveira, Milene Selbach. II. Título.

CDD 004.019

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

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

TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO

Dissertação intitulada "Análise de Aspectos de Usabilidade em Interações Naturais via Interfaces Multimodais" apresentada por Lucio Polese Cossio como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, aprovada em 17/03/2014 pela Comissão Examinadora:

Profa. Dra. Milene Selbach Silveira – Orientadora — PPGCC/PUCRS

Prof. Dr. Márcio Sarroglia Pinho – PPGCC/PUCRS

Profa. Dra. Luciana Nedel – UFRGS

Homologada em 26/03/2015, conforme Ata No. 004 pela Comissão Coordenadora.

Prof. Dr. Luiz Gustavo Leão Fernandes
Coordenador.

PUCRS
Campus Central
Av. Ipiranga, 6681 – P32– sala 507 – CEP: 90619-900
Fone: (51) 3320-3611 – Fax (51) 3320-3621
E-mail: ppgcc@pucrs.br
www.pucrs.br/facin/pos

ANÁLISE DE ASPECTOS DE USABILIDADE EM INTERAÇÕES NATURAIS VIA INTERFACES MULTIMODAIS

RESUMO

A presença de dispositivos computacionais cresce a cada dia, tornando-os disponíveis nos mais diferentes cenários de uso. A interação com o sistema deve evoluir em conjunto com a tecnologia para prover aos usuários uma melhor experiência de uso nestes diferentes ambientes. Diversos estudos na área de interfaces multimodais defendem os benefícios das mesmas, pela disponibilidade de formas de interação mais naturais, permitindo aos usuários maior eficiência e satisfação na execução da tarefa do sistema. Nos últimos anos, dispositivos que possibilitam formas de interação consideradas mais naturais, começaram a estar amplamente disponíveis ao público e ser utilizados com mais frequência, demonstrando grande potencial para uso. Este trabalho apresenta uma pesquisa com objetivo de compreender e comparara preferência e aceitação de uso dessas tecnologias,a partir de sua implementação e análise em um sistema de apresentação, utilizando o dispositivo Kinect e um *smartphone* Android.O sistema permite aos usuários a execução de apresentações de slides e imagem para uma plateia, utilizando modos de fala, gestos de corpo e *smartphone* (gestos de toque) para interação. A primeira fase do trabalho se deteve na definição da interação através de entrevistas individuais e execução de grupos focais. Posteriormente a implementação do sistema foi feita com base nos resultados obtidos da fase anterior.Por fim a avaliação do sistema foi feita pelo uso e execução de tarefas com o sistema em um ambiente de sala de aula (mas sem plateia). Os resultados aqui presentes demonstram a opinião diversificada de usuários quanto a perspectiva de uso dos diferentes modos do sistema. O uso do *smartphone* foi a tecnologia mais precisa e preferida pela maioria dos usuários pelo fácil uso, no entanto, alguns participantes apresentaram grande interesse no uso das outras duas modalidades, demonstrando potencial de aceitação para as mesmas. O contexto de uso pretendido do sistema demonstra desafios, uma vez que as modalidades de gestos de corpo e fala são também utilizadas para os usuários do sistema se comunicarem com outras pessoas ao mesmo tempo, sendo consideradas por alguns dos participantes como pouco apropriadas para a situação. Os testes do sistema foram realizados em um

ambiente isolado, e futuramente devem ser aplicados para um contexto real para uma validação mais precisa.

Palavras Chaves: Modalidade, Multimodal,Interfaces, Dispositivos de Interação

USABILITY ASPECTS ANALYSIS IN NATURAL INTERACTIONS VIA MULTIMODAL INTERFACES

ABSTRACT

The presence of computing devices grows day by day, making them available at the most different scenarios of use. The system' interaction needs to evolve together with the technology to provide a better user experience in these distinct environments. Several studies in the multimodal interfaces area advocate the benefits of these interfaces, because of the availability of more natural ways of interaction, allowing the users more efficiency and satisfaction for the execution of system tasks. In the last years, devices that allow more natural interactions, started to become widely available to the public and to be used more often, showing great potential of use. This work presents a research with the goal of understand and compare the preference and use acceptance of these technologies, through the implementation and analysis of a presentation system, using a Kinect device and an Android smartphone. The system allows users to execute slide and image presentations to an audience, using speech, body gestures or the smartphone (touch gestures) for interaction. The first work phase focused in the interaction definition through individual interviews and focus groups execution. After that, the system was implemented following the results of the previous phase. At the end, the evaluation of the system was done through the use and execution of tasks in a class environment (without audience). The obtained results show the diverse users' perspective of use of the different systems interaction modes. The smartphone was the most precise and preferred technology by most of the users because is easy to use, although, some participants showed great interest in the use of the other modalities, showing a potential for acceptance of them. The intended system context of use have some challenges, since body gestures and speech are also used by system users to communicate with other people at the same time, considered by some of the participants as unsuitable for the situation. The system tests were executed in an isolated environment, and future tests should be applied in a real use context for more precise evaluation.

Keywords: Modality, Multimodality, Interfaces, Interaction Devices

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Em décadas anteriores já se previa o grande crescimento da utilização de dispositivos computacionais, de forma que estes iriam se tornar parte de nosso dia-a-dia e estar presentes em todo lugar [80,94]. Essa diversidade de contextos de uso iria requerer que as interfaces provessem formas de interação mais naturais [1,80,84], como toque, gestos e fala. Sugerido por alguns trabalhos, esta necessidade leva naturalmente ao desenvolvimento de interfaces multimodais [9,24], na tentativa de melhorar a interação do usuário com o sistema e tornar esta mais similar à forma com que as pessoas interagem umas com as outras.

Atualmente já vivenciamos o grande aumento de uso de dispositivos móveis com tela de toque e amplo poder computacional, como *smartphones* e *tablets*. Além dessas tecnologias multitoque, a indústria de jogos permitiu a disponibilidade em massa de novas tecnologias gestuais, por meio de dispositivos como o Nintento Wii [67], Microsoft Kinect [55], e Playstation Move [75], que permitem aos usuários utilizarem movimentos do seu corpo para jogar. O interesse e popularidade desses dispositivos cresceram e começaram a ser utilizados para aplicações além de jogos [29,54].

O crescimento de uso e disponibilidade dessas novas tecnologias torna necessário entender seu potencial de uso e características, para projetar interfaces que utilizem de suas capacidades de forma a melhorar a interação de seus usuários. Com o objetivo de estudar a capacidade de uso dessas formas naturais de interação em outras aplicações (que não apenas jogos), e comparar seus benefícios, foi efetuado o desenvolvimento e avaliação de uma interface multimodal para um sistema de apresentações, que permite ao usuário utilizar gestos de toque, gestos de corpo e fala para interagir com o sistema. O tipo de sistema foi escolhido em vista de ser relacionado a uma tarefa comum no dia-a-dia de professores e alunos da universidade, em sala de aula, ou em conferências, o que favoreceria sua análise neste ambiente.

Tal sistema foi planejado para utilizar de um dispositivo Kinect, para captura de gestos de corpo e fala, e um *smartphone* Android para interação de toque em tela.

1.1 Questão de Pesquisa

Visto que novas modalidades se tornam mais acessíveis para uso em sistemas computacionais, é importante compreender como aproveitar essas tecnologias de forma a beneficiar o usuário na interação com os sistemas que as utilizam.

O desafio é, portanto, compreender se, e de que forma, essas modalidades podem ser utilizadas para aprimorar a interação com sistemas computacionais.

1.2 Objetivos

O objetivo do trabalho é o desenvolvimento e avaliação de um sistema multimodal de apresentação que possui modalidades de fala, gestos de corpo, e gestos de toque, para compreender e comparar a satisfação de uso dos usuários em relação a essas modalidades.

De uma forma específica, o objetivo do trabalho é resumido nos seguintes pontos:

- Definir e desenvolver um sistema multimodal utilizando dispositivos de fácil acesso e em constante expansão de uso. Foi previamente definido o uso dos dispositivos Kinect e um *smartphone* Android;
- Executar testes com usuários para extrair dados de comparação entre as modalidades de voz, gestos (via Kinect) e toque (via *smartphone*);
- Analisar os dados coletados para compreender se, e como, as modalidades disponíveis são apropriadas para uso na tarefa proposta, comparando seu desempenho/preferência;
- Contribuir com a área de Interfaces Multimodais através de sugestões para o processo de desenvolvimento, e para o entendimento de uso e preferência das modalidades pelos usuários.

1.3 Metodologia Empregada

Através de uma revisão bibliográfica sobre como empregar gestos para controle de sistemas computacionais, decidiu-se pela execução de um estudo com usuários para especificação das técnicas de interação a serem empregadas, antes de sua implementação, visto que uma má decisão nessa fase poderia influenciar negativamente os resultados finais de avaliação

posteriores. Tal estudo foi baseado em trabalhos anteriores e foi composto da execução de entrevistas individuais, e também de entrevistas em grupo (grupos focais).

Uma vez tendo definido as técnicas de interação com o sistema, o desenvolvimento deste foi iniciado através do uso das tecnologias previamente estabelecidas, com o uso dos *kits* de desenvolvimento oficiais das mesmas. Nesta etapa, ainda existiram alguns ajustes necessários que deviam ser interpretados da fase anterior.

Uma vez que a fase de desenvolvimento foi finalizada, testes de avaliação do sistema foram realizados, compostos por uma etapa de treinamento e execução de duas pequenas tarefas com uso de cada modalidade, de forma a extrair métricas e opiniões de uso das mesmas. Uma análise dos dados coletados foi realizada ao fim para extrair informações relevantes ao objetivo do trabalho.

1.4 **Estrutura do Trabalho**

O trabalho esta dividido em três principais fases: a definição, a implementação, e a avaliação de um sistema multimodal. Nesse contexto, os capítulos neste trabalho se dividem da seguinte forma: revisão de interfaces multimodais e suas características (capítulo 2); escopo do trabalho (capítulo 3); definição da interação com o sistema (capítulo4); implementação do sistema (capítulo 5); avaliação do sistema (capítulo 6) e por fim as conclusões do trabalho (capítulo 7).

2 INTERFACES MULTIMODAIS

No processo de interação entre usuários e sistemas computacionais, o usuário envia informações para o sistema, este executa as requisições e apresenta o resultado destas ao usuário. As diferentes formas de entrada, envio de informações do usuário ao sistema, e saída de dados, envio de informações do sistema ao usuário, processo representado na Figura 1, são as modalidades, diferentes representações de informações, que possibilitam a interação entre o sistema e o usuário [70].

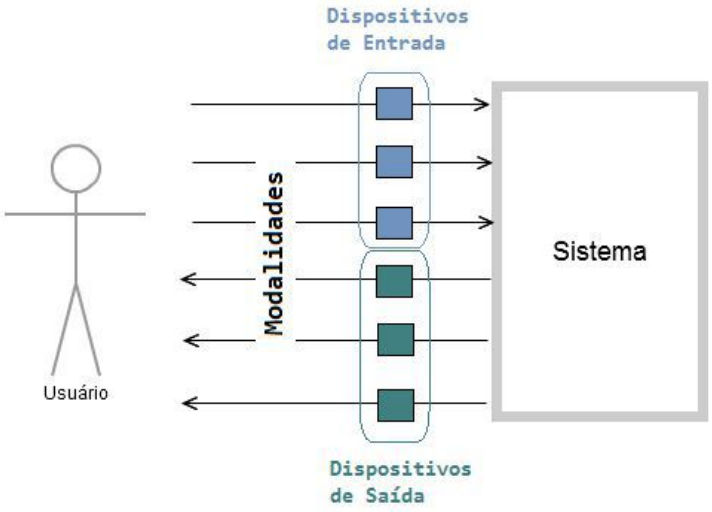


Figura 1 - Interação de usuários com o sistema.

Esta área de estudo é vasta e possibilita a exploração de diversas questões envolvidas. Alguns trabalhos focam-se, por exemplo, no entendimento e aplicação de uso das modalidades de saída [7,27], enquanto outros se focam nas modalidades de entrada [70,76].

Neste trabalho, o foco maior está nas modalidades de entrada e sua preferência/desempenho de uso pelos usuários. As modalidades disponíveis no sistema são a fala, toque (em tela) e gestos de corpo. Uma visão geral das características, vantagens e desvantagens desses sistemas é apresentada nas seções a seguir.

2.1 **Definição**

Os trabalhos existentes na área de interfaces multimodais apresentam diferentes interpretações para termos importantes utilizados, termos estes relacionados a própria

definição de modalidade e sistemas multimodais[1]. Essa divergência pode dificultar a discussão entre os pesquisadores da área, uma vez que não há o uso de uma linguagem comum.

Este capítulo tem como objetivo debater estas divergências analisando diferenças e semelhanças na abordagem de diferentes autores e consolidando os conceitos da forma como serão utilizados neste trabalho.

2.1.1 Modalidade

O objetivo de desenvolver interfaces computacionais para usuários em geral traz como necessidade o estudo do fator humano envolvido no processo de interação. Alguns trabalhos da área da psicologia apresentam pesquisas no uso da combinação de modalidades sensoriais para aumentar a capacidade de memória e aprendizagem de estudantes, focando-se nas modalidades visual e auditiva [30,47,63]. As modalidades são consideradas, nestes estudos, como as formas físicas de apresentação de informações que envolvem um determinado sentido humano para sua percepção. A modalidade visual é a forma de apresentação de informações visualmente, que utiliza, por exemplo, imagens e textos, enquanto a modalidade auditiva seria o uso da representação sonora da informação, como a fala.

Essa divisão de modalidades surge dos modelos teóricos baseados em evidências que apontam para o entendimento de que o processamento das informações de diferentes modalidades é feito por mecanismos que são, até certo ponto, independentes [10,47]. Isso significa que a apresentação de informações utilizando-se mais de uma modalidade possibilita um nível de processamento e retenção dos dados de forma paralela, resultando na memorização de uma quantidade maior de informações e em um melhor aprendizado, como é demonstrado em diversos experimentos [30,47,63], embora o ganho não possa ser predito pela simples soma dos resultados advindos do uso das modalidades de forma isolada.

A definição do termo modalidade apresenta algumas divergências na área de interfaces multimodais de sistemas computacionais, como também relatado em [16]. Alguns trabalhos

---

[1] O uso dos termos sistemas multimodais e interfaces multimodais são utilizados neste trabalho como sinônimos, uma vez que as interfaces fazem parte dos sistemas computacionais foco de estudo.

desta área se utilizam da definição de modalidade ligada aos sentidos humanos [12,82]. Já em [23] é utilizado o termo modalidade de forma abrangente e pouco explicativa, sendo “um método de interação que um agente utiliza para atingir uma meta” ou exemplificado de forma geral como “utilizando fala” ou “utilizando microfone”.

Mesmo que o campo da psicologia tenha utilizado inicialmente uma classificação baseada nos sentidos para diferenciação entre as modalidades, eram apontadas evidências da existência de subsistemas responsáveis pelo processamento de diferentes propriedades sobre as informações capturadas por um mesmo sentido. A modalidade visual apresenta as propriedades de forma e localização espacial que são processadas independentemente [10], sendo possível haver dificuldade do reconhecimento de uma destas propriedades sem comprometer o reconhecimento da outra [28]. Isso demonstra que mesmo informações obtidas por um mesmo sentido serão processadas para extração de determinadas propriedades por diferentes (sub)sistemas. Como exemplo, a utilização de texto em conjunto com imagens é capaz de aumentar a capacidade de aprendizagem em comparação com algum desses dois modos utilizado sozinho [46,48], embora ambos os modos de apresentação sejam capturados pelo mesmo sentido humano.

Torna-se claro que o conceito do termo modalidade deve levar em conta não apenas a representação física que a informação utiliza no processo de comunicação (luz ou ondas sonoras, por exemplo), e que é percebida por um particular sentido humano, mas também as diferentes propriedades físicas de apresentação que são distinguíveis, como no caso de uma representação visual que possui cor, forma, e posição. Essa classificação tem a vantagem de separar as diferentes propriedades da representação física que serão processadas por diferentes áreas cerebrais, possibilitando explorar quais propriedades são mais eficientes para troca de informações. Ainda, essas propriedades possuem contrastes entre si, possibilitando a codificação de informações em formas alternativas quando, por exemplo, alguma modalidade não possa ser utilizada.

A definição de modalidades a partir das propriedades existentes em determinada representação física é utilizada na taxonomia de Bernsen [14]. Uma determinada modalidade, como Bernsen [14] destaca, é definida por:

- **Um meio físico (ou mídia):** Toda informação deve ser instanciada fisicamente para ser transmitida e percebida. Os portadores físicos de informações podem ser a luz,

ondas sonoras ou forças mecânicas, se relacionando às mídias (meios de comunicação com os sentidos) gráfica, acústica e háptica, respectivamente, relacionada aos sentidos de visão, audição e tato.

- **Propriedades físicas**: Existem ainda diferentes formas que determinada informação pode assumir, mesmo que utilizando a mesma mídia física. Por exemplo, imagens e textos são informações visuais, mas possuem diferentes propriedades. Na mídia gráfica podem-se identificar propriedades de forma, tamanho, posição, cor e textura.

Nessa definição, uma modalidade é uma forma de representar informação em um determinado meio. A taxonomia de Bernsen [14] separa as modalidades em diferentes níveis, sendo estas definidas pelas propriedades físicas particulares, ou “canais de informação”, dentro de três principais mídias (embora outras existam, não são exploradas em seu trabalho): háptica, gráfica, e acústica.

Jaimes & Sebe [37] definem modalidade como sendo um modo de comunicação ligado a um sentido humano ou tipo de dispositivo de interação. Esta definição abrange formas de interação que não se relacionam diretamente aos sentidos humanos, como o mouse e teclado. Para Bernsen [14] os dispositivos de entrada, como o mouse e o teclado, estariam ligados a mídia háptica.

A definição de Bernsen [14], como apresentada acima, com consideração do meio físico e as propriedades existentes neste meio, é a definição mais correta de ser utilizada. As propriedades físicas particulares referentes a cada modalidade não serão relacionadas de forma detalhada aqui, podendo ser verificadas no trabalho do autor [14] e outros que o utilizam e ampliam suas propriedades [27,91].

É considerado neste trabalho que o detalhamento proposto por Bernsen é muitas vezes desnecessário, e a simples citação do dispositivo de interação abrange informações suficientes para compreender a interação que o sistema está considerando, e, portanto, as modalidades suportadas.

2.1.2 Interfaces Multimodais

O termo multimodal é utilizado de diferentes formas na literatura, e esta confusão é apontada em alguns trabalhos [14,86].

A partir do conhecimento sobre o conceito de modalidade é possível compreender o significado do termo interfaces multimodais como as interfaces que possuem múltiplas modalidades, utilizadas para troca de informação entre o sistema e o usuário no processo de interação.

A definição de Bernsen [14] para sistemas multimodais é a seguinte:

> *Um sistema interativo multimodal é um sistema que utiliza pelo menos duas modalidades diferentes para entrada e/ou saída. Assim, [IM_1, OM_2], [IM_1, IM_2, OM_1] e [IM_1, OM_1, OM_2], são alguns exemplos mínimos de sistemas multimodais, I significando entrada, O saída, e M_n uma modalidade específica n.*

E de forma correspondente:

> *Um sistema interativo unimodal é um sistema que utiliza a mesma modalidade para entrada e saída, i.e., [IM_n, OM_n]*

Divergente à definição apresentada por Bernsen [14], alguns trabalhos preocupam-se mais com as modalidades de entrada do que de saída para a definição do termo, e focam seus estudos em sistemas com diferentes modalidades de entrada que, embora utilizem diferentes formas de apresentação de informações, não fazem análises ou comparações entre diferentes configurações destas [37,70].

A partir da definição de Bernsen [14] fica claro que a grande maioria dos sistemas computacionais possuem interfaces multimodais. Alguns dos poucos sistemas unimodais seriam sistemas de conversação (com entrada e saída de fala), e sistemas que recebem informações gestuais do usuário e respondem com um personagem virtual da mesma forma.

Alguns autores, no entanto, desconsideram as GUI (*Graphical User Interfaces*) como interfaces multimodais [70,74]. O desenvolvimento dessas interfaces gráficas (GUIs) introduziu uma grande facilidade de uso que impulsionou a comercialização de computadores para a população em geral [33,65]. Uma interface GUI utiliza diversos objetos gráficos para apresentação, e o uso de dispositivos como teclado e mouse para manipulação. Neste caso, tanto os modos de entrada para o sistema, teclado e mouse, como os modos de saída, monitor (que permite uma variada combinação de modos de apresentação na mídia gráfica) e freqüentemente sons, sendo analisados de forma separada, já tornariam a classificação deste padrão de interfaces como multimodal.

Oviatt & Cohen [71] afirmam que “sistemas multimodais são radicalmente diferentes de GUIs padrão”. Oviatt [70], na sua definição de interfaces multimodais, expõe que estas interfaces processam dois ou mais modos de entrada de maneira coordenada com uma saída multimídia. As interfaces multimodais representariam um novo paradigma, sendo diferentes das interfaces convencionais que utilizam janelas, ícones, menus e dispositivos de apontamento (do inglês WIMP – *Window, Icon, Menu, Pointing device*), tendo como foco o reconhecimento de formas naturais de linguagem e comportamento humano, incorporando pelo menos uma tecnologia baseada em reconhecimento (como exemplo a fala, caneta, ou visão).

E, em relação à diferença entre interfaces multimodais e interfaces GUIs, este autor declara que:

- GUIs tipicamente assumem que um único fluxo de eventos controla o ciclo de eventos subjacentes. Por exemplo, a maioria das GUIs ignora entradas digitadas quando um botão de mouse é pressionado. Em contraste, interfaces multimodais tipicamente podem processar entradas contínuas e simultâneas vindas de fluxos de chegada paralelos;
- GUIs assumem que ações básicas da interface, como seleção de um item, são atômicas e não ambíguas. Em contraste, sistemas multimodais processam modos de entrada utilizando tecnologias baseadas em reconhecimento, que trabalham com incertezas utilizando métodos de processamento probabilísticos;
- GUIs freqüentemente são construídas para serem separáveis do software de aplicação que elas controlam, embora os componentes da interface geralmente residam centralmente em uma máquina. Em contraste, interfaces baseadas em reconhecimento tipicamente possuem requerimentos grandes de processamento e memória, que freqüentemente tornam desejável distribuir a interfaces através de uma rede para que máquinas separadas trabalhem com diferentes reconhecedores ou base de dados. Por exemplo, telefones celulares e PDAs podem extrair características da fala de entrada, mas transmitem-nas para um reconhecedor que reside em um servidor;

- Interfaces Multimodais que processam dois ou mais fluxos de entradas baseados em reconhecimento requerem a marcação do tempo de entrada, e o desenvolvimento de restrições temporais para modos de operação de fusão. A este respeito, elas necessitam de arquiteturas, com sensibilidade e gerenciamento do tempo de eventos, únicas.

As diferenças entre interfaces GUIs e interfaces multimodais apontam questões específicas da construção de aplicações, e características específicas de novas tecnologias. Como apresentado, a definição de Oviatt [70] restringe as modalidades que devem ser consideradas para determinar uma interface como multimodal, além de focar-se nos dispositivos de entrada de informações para o sistema. As modalidades que são utilizadas no processo de comunicação entre humanos seriam mais naturais, e de maior interesse de serem incorporadas nos sistemas computacionais.

Apesar de Oviatt [70] utilizar uma definição errônea em uma perspectiva teórica, sua definição está certa quanto ao foco de estudo que a área de interfaces multimodais apresenta, que é a verificação do uso de novas tecnologias de interação que despertam um grande interesse por serem consideradas formas mais naturais e apresentarem resultados promissores em muitos estudos. Como serão apresentados ao longo deste trabalho, os estudos na área de interfaces multimodais acabam utilizando-se em sua grande maioria de dispositivos de entrada com tecnologias de reconhecimento, sendo utilizadas de forma combinada em diferentes formas temporais, visando tornar a interação com o sistema mais natural, além de outras vantagens.

Em vista das diferentes interpretações na literatura, a definição considerada pertinente aqui neste trabalho é de que interfaces multimodais são aquelas que possuem mais de uma modalidade. No caso deste trabalho o sistema é considerado multimodal por possuir três modalidades de entrada com o sistema, que podem ser utilizadas de forma equivalente, e assim trocadas a qualquer momento durante a interação. Embora muitos considerem importante a forma de combinação temporal de tais modalidades, o foco aqui é quanto a sua naturalidade e satisfação de uso.

2.2 **Modalidades Disponíveis**

As modalidades de saída [82] existentes são as sonoras (fala, ícones auditivos e *earcons* [50,93]), visuais (menus, ícones, e animações), hápticas (variação de intensidade, frequência e ritmo de toques), e aromas (ícones olfativos e *smicons* [41]). Como modalidades de entrada [70] existem as visuais (gestos de mãos/dedos/corpo e direção do olhar), hápticas (botões, teclado, mouse, e toque de dedos) e a fala.

As características das principais modalidades de interesse desse trabalho são apresentadas a seguir.

2.2.1 Modalidades Hápticas

Os dispositivos mais populares para entrada de informações a um sistema interativo computacional são o teclado e o mouse. Estes dispositivos disponibilizam opções de modalidades hápticas (táteis) para interação em *desktops* assim como *joysticks* e *gamepads* o fazem para vídeo games, e telas de toque para dispositivos móveis e monitores.

Tais dispositivos são geralmente utilizados em conjunto com modalidades de saída visuais como menus e ícones, integrando as interfaces gráficas muito utilizadas atualmente. Esse tipo de interface é apropriado quando a tarefa tem um número limitado de ações e os objetos sobre o qual as ações são feitas em um dado tempo são visíveis na tela [19]. A interação com essas interfaces é intuitiva, as opções são claras, e permitem rápida e precisa identificação de localizações espaciais [69], embora muitas vezes exijam navegação longa entre menus para invocar as ações desejadas.

A utilização das telas de toques tem se tornado popular devido a *smartphones* e *tablets* que vem ganhando grande participação no mercado. É possível encontrar dispositivos que suportam o uso de canetas *stylus* ou dos dedos dos usuários para interação com toques na tela. As telas multi-toques permitem o uso de múltiplos dedos do usuário, sendo muito eficazes na manipulação de objetos na tela [87], e permitem uma interação mais rápida que o mouse em algumas tarefas [38].

2.2.2 Modalidades Visuais

Gestos com mãos e/ou outras partes do corpo são identificados por meio do uso de câmeras associadas ao sistema e utilizados com sucesso como forma de interação em diversas aplicações. Um estudo conduzido por Hauptmann [34] demonstra a naturalidade e conformidade dentro de um conjunto de gestos comuns que os usuários utilizaram para interagir com um cubo virtual. Os usuários foram instruídos a interagir da forma como bem quisessem com um suposto sistema computacional que reconhecia suas intenções de gestos e fala, a partir de uma técnica de simulação com um operador identificando os gestos e disparando respostas aos usuários pelo sistema.

É importante que os gestos de mãos e de corpo estejam em conformidade com o que os usuários acham natural, caso contrário confusões podem acontecer. Como relatado no estudo de McGlaun et al. [51], muitos usuários acabaram esquecendo a maioria dos possíveis gestos explicados no inicio dos testes, e acabaram utilizando seus próprios gestos, embora muitos não soubessem como expressar determinados comandos.

Sistemas que identificam a mão do usuário, mas não são capazes de identificar gestos mais finos (como os dedos), podem utilizar, para uma simples tarefa de seleção de opções na tela, diferentes técnicas, como apresentado por Schapira & Sharma [83]:

- *Point and Wait* - Uma vez que o cursor esteja sobre a opção desejada, o usuário deve mantê-lo em cima da opção por um determinado período de tempo. Dependendo do tempo de espera necessário, o usuário pode em alguns casos selecionar opções sem intenção;
- *Point and Shake* - Sobre a opção desejada, o usuário sacode a mão rapidamente. Esta técnica apresentou baixo desempenho no estudo de referência [83];
- *Point and Speak* - Com o auxílio da fala, a seleção é feita na opção em que o cursor se encontra no momento do comando de voz. É uma técnica difícil de ser utilizada caso mais de um usuário interaja com o sistema ao mesmo tempo. No estudo realizado [83], alguns usuários acharam tediosa a repetição do comando após certo tempo de interação.

Para facilitar a identificação, o usuário pode utilizar luvas de diferentes cores nas mãos; em situações em que os gestos são complicados, os usuários necessitam de ambas as mãos e

estes gestos são realizados nas três dimensões, facilitando o processo de identificação pelo sistema e permitindo o processamento em tempo real. Um exemplo desta situação é apresentado por Moustakas et al. [64] para a tarefa de reconhecimento de linguagem gestual.

Além do reconhecimento de gestos, o monitoramento da direção do olhar através de uma câmera é uma técnica que pode ser utilizada pelo sistema para ter conhecimento das áreas da tela de maior interesse do usuário, e como consequência amplificar ou disponibilizar um maior número de informações relacionadas a estas áreas [85].

A direção do olhar também pode ser utilizada como uma forma de interação similar ao mouse, aonde o usuário é capaz de navegar pela tela e selecionar opções com o uso do olhar. Comparado ao *mouse* pode resultar em maior velocidade na seleção de objetos na tela, mas possui menor precisão e a seleção precisa ser feita com o auxilio de algum botão auxiliar, uma vez que o uso de piscar dos olhos para disparo pode resultar em seleções indesejadas e a fixação do olhar por um tempo determinado aumentaria o tempo do processo [26]. Uma das grandes vantagens é a acessibilidade que este método traz para usuários que tem dificuldade de realizar movimentos motores.

A captura de emoções pelo sistema é outra informação que pode ser identificada através do processamento de imagens, a partir da análise de movimentação do corpo e postura do usuário. Essa informação permite que o sistema se adapte ao estado de humor do usuário fornecendo alternativas para melhor interação, por exemplo, em sistemas de aprendizagem *online* [8,32,37].

2.2.3 Modalidades Acústicas

A fala é uma opção de entrada no sistema que está frequentemente presente nos sistemas multimodais. Tal escolha deve-se particularmente no fato de a fala ser uma forma natural de comunicação entre humanos, e, portanto, poder permitir maior naturalidade na utilização de sistemas computacionais [2].

A fala pode ser utilizada em conjunto com outras formas de interação, como o mouse ou gestos, na seleção de opções e disparo de eventos, ocasionando um aumento de desempenho para execução de tarefas [25,83]. Ela pode ser preferida em situações em que as mãos do usuário estão ocupadas, quando apenas um teclado ou telas limitadas estão

disponíveis, quando o usuário tem limitações motoras ou quando a linguagem natural é preferida [19]. Assim como informações visuais, a fala pode ser utilizada para identificar emoções dos usuários e permitir ao sistema adaptar-se a eles [32,37].

Embora a linguagem natural não seja diretamente relacionada à mídia auditiva, ela é frequentemente utilizada em conjunto da fala para entrada de informações. Linguagem natural é particularmente apropriada para descrever objetos e períodos de tempo que não podem ser referidos diretamente. Ela é genérica e deixa aberto um escopo de interpretação, o qual pode ser incrementalmente estreitado através da adição de mais expressões linguísticas [15]. Uma das dificuldades que sistemas com linguagem natural trazem é que, as vezes, os usuários não sabem o que o sistema é capaz de entender, embora saibam que ele não é capaz de entender tudo [18].

O uso da fala não apresenta vantagens frente ao teclado quanto à tarefa de escrita de textos quando comparada com digitadores experientes, seja para entrada ou correção de informações [17,86]. Ela pode ser interessante em dispositivos móveis que não possuem um teclado para entrada rápida de dados, embora não seja uma forma boa de interação para utilização em público.

2.3 **Características de Sistemas Multimodais**

As propriedades CARE (Complementaridade, Atribuição, Redundância, e Equivalência) [23] apresentam uma forma de caracterizar a interação com interfaces multimodais, relacionando as noções de estado, objetivo, modalidade, e relação temporal. Um estado determina um conjunto de propriedades observadas em um dado momento que caracterizam uma situação. Um objetivo é o estado na qual o usuário deseja chegar. As quatro propriedades se resumem da seguinte forma:

- **Complementaridade:** Quando modalidades são usadas em conjunto dentro de uma janela de tempo para alcançar outro estado, podendo ser de forma paralela ou sequencial;
- **Atribuição**: Uma determinada modalidade é designada para ir de um estado a outro se apenas ela pode ser utilizada para isso, sem nenhuma modalidade alternativa que alcance o mesmo objetivo;

- **Redundância:** Modalidades de um conjunto são redundantes se para ir de um estado ao outro elas tem o mesmo poder de expressão (são equivalentes) e precisam ser usadas dentro da mesma janela de tempo. Em outras palavras, o agente apresenta comportamento repetido sem aumentar o poder de expressão;
- **Equivalência**: O conjunto de modalidades disponíveis que podem ser utilizadas de forma equivalente para alcançar o mesmo objetivo. Não impõe nenhuma forma de relação temporal entre modalidades.

Tais propriedades permitem a definição da interação com o sistema em termos das possibilidades existentes para cada comando ou tarefa. São essas conexões entre modalidades e suas diversas combinações que introduzem características únicas aos sistemas multimodais e exigem um design consistente e complexo.

2.4 Vantagens e Desvantagens

Os estudos na área de interfaces multimodais focam-se no uso de tecnologias de reconhecimento [70], como fala, e gestos, referidos geralmente como interações mais naturais. Como apresentado por Hauptmann [34], usuários tenderiam a utilizar gestos de mãos e fala de uma forma similar e consistente na interação com um sistema computacional, já que esta forma de interação é similar a forma com que estes interagem com outras pessoas. Tal naturalidade permitiria que o sistema fosse mais fácil de ser utilizado.

Por poder disponibilizar diferentes modalidades para interação, um sistema multimodal permite que o usuário escolha entre as diferentes formas de interação disponíveis, satisfazendo melhor a preferência do usuário ou eficácia de uso em diferentes situações [70].

Apesar disso, a disponibilidade de uma maior quantidade de modalidades torna o desenvolvimento do sistema mais complexo, e o uso de modalidades mais naturais nem sempre irão resultar em uma melhor experiência de uso para os usuários [68]. Desta forma é ainda importante compreender as características de cada modalidade e como seu uso e combinações afetam a interação com o sistema.

2.5 Avaliação de Interfaces Multimodais

O teste com usuários é a técnica mais utilizada para avaliação de interfaces multimodais [13]. O usuário é encarregado de executar tarefas simples com o sistema, e sua interação é

gravada e analisada. As medidas mais comuns de serem extraídas são as medidas de usabilidade [13], como tempo para realização da tarefa, percentual de erros do sistema e do usuário, e aceitação subjetiva. A grande diferença na avaliação destes sistemas é que é necessário analisar os resultados em comparação com as diferentes configurações possíveis de modalidades do sistema na execução de uma tarefa. O objetivo é identificar quais as modalidades são mais adequadas para execução da tarefa, como elas devem se relacionar umas com as outras, e de que forma o usuário deve estar ciente ou no controle da forma de interação, sendo utilizado como ponto de comparação uma interface com formas de interação padrões, já bem conhecidas e utilizadas na maioria dos sistemas.

2.5.1 Ambiente e Condições dos Testes

Embora a tendência seja realizar estes testes com usuários em ambientes totalmente controlados, como dentro de laboratórios fechados, é muitas vezes importante realizar os testes no ambiente real de uso do sistema, como em casos de interfaces para dispositivos móveis, pois a percepção da interface pelo usuário pode variar de acordo com o ambiente [11,39].

No trabalho de Jöst et al. [39] foram realizados testes de uma interface multimodal em ambiente interno de laboratório e em ambiente externo. Os resultados apontam que o grupo de usuários que realizou os testes em ambiente externo teve mais aceitação da interface multimodal do que aqueles que executaram o teste em ambiente interno (eram grupos com sujeitos diferentes). Embora os grupos possuíssem diferenças de idade e gênero significativos, um estudo de Baillie & Schatz [11] apresentou resultados similares. Neste segundo estudo, a avaliação da interface multimodal aconteceu também em ambiente de laboratório e em campo, com dois grupos de usuários que diferiram apenas na ordem em que realizaram os testes em ambos ambientes (o primeiro grupo fez os testes no laboratório, e, depois, em campo, e o segundo grupo fez primeiro os testes em campo). Em campo, os usuários se apresentaram mais relaxados, cometeram menos erros, resolveram as tarefas mais rapidamente, e utilizaram mais a interação conjunta de modalidades. Ainda, usuários do primeiro grupo disseram não achar útil a opção de mistura de ambas as modalidades após os testes em laboratório, mas mudaram de opinião após o uso em campo.

A influência do ambiente no resultado é uma característica que parece ser comum em sistemas de dispositivos moveis. O trabalho de Kjeldskov & Stage [42] apresenta que a descoberta de um maior número de problemas de usabilidade em um sistema para dispositivos móveis ocorreu em condições de ambientes de laboratório, uma vez que os usuários estariam concentrados na tarefa de interação com o sistema (e não com outras tarefas do ambiente), e assim pensavam em voz alta (*thinking aloud*) com maior frequência. Os problemas identificados a mais, no entanto, foram de severidade baixa. Em condições de movimentação foram encontrados problemas de usabilidade mais relacionados ao *layout* da interface, tamanho e localização de elementos.

2.5.2 Técnicas de Simulação

Alguns trabalhos utilizam técnicas de simulação do sistema para verificar como os sujeitos comportam-se na interação com interfaces multimodais [5,20,34,72,73], referenciadas comumente como *Wizard of Oz* [36,51]. Estas técnicas são úteis para auxiliar no desenvolvimento de interfaces multimodais e suas avaliações sem ter um sistema funcional implementado, como em casos da etapa de *design* do sistema para decidir as modalidades a serem utilizadas e seu comportamento com a tarefa a ser executada, e quando as tecnologias a serem simuladas não existem ou possuem limitações que poderiam comprometer os resultados.

Nos estudos que utilizam simulações, operadores são responsáveis por intermediar a etapa de reconhecimento de informações de entradas para o sistema, visualizando a interação do usuário com o sistema e disparando eventos de resposta de forma a parecer que um sistema computacional estivesse interpretando e respondendo aos comandos do usuário diretamente.

A técnica de simulação foi utilizada por Hauptmann [34], para analisar o comportamento de sujeitos na interação com um sistema utilizando comandos de voz e gestos para operações em um cubo virtual. Os usuários tinham que realizar tarefas utilizando (1) somente voz, (2) somente gestos ou (3) utilizando voz e gestos da forma que preferissem. A pesquisa demonstra uma tendência na utilização de palavras dentro de um pequeno vocabulário e padrões de gestos comuns. A maioria dos usuários teve preferência em interagir de forma multimodal (com possibilidade de utilizar fala e gestos como preferissem).

No trabalho de Oviatt et al. [73], usuários foram observados na interação com um sistema simulado multimodal na situação de ocorrência de erros, e uma análise foi feita de suas estratégias e comportamento nesta situação. Ainda, Anthony, Yang & Koedinger [5] utilizaram a técnica de simulação do sistema para estudar a aceitação dos usuários em diferentes modalidades para entrada de equações matemáticas em um sistema.

A simulação permite, portanto, que o usuário interaja de forma natural, utilizando uma linguagem que seja familiar a este e ao que ele está acostumado a usar na comunicação com outras pessoas (e que por estas é inteligível), e a análise de sua interação auxilia a criação de sistemas que se adequem melhor a seus comportamentos.

2.6 **Design Definido por Usuários**

A interação com modalidades naturais como fala e gestos precisa ser estudada para geração dos melhores princípios de *design* a serem aplicados. Existem diversas características que precisam ser consideradas para prover o melhor *design* ou essas interfaces podem causar problemas por más decisões [70].

Dependendo do contexto na qual o sistema será usado, é importante gerar gestos, como movimentos de dispositivo ou de mãos, que sejam socialmente aceitáveis [78]. Usuários podem sentir-se desconfortáveis executando alguns gestos em determinados locais ou na frente de certos tipos de audiência.

É também importante definir gestos que sejam intuitivos de utilizar. Por exemplo, Mc Glaun et al. [51] avaliaram um sistema multimodal incluindo gestos de mãos e cabeça em um sistema para ser utilizado dentro de um carro, utilizando a técnica de *Wizard of Oz*, descrita anteriormente. O sistema teve os gestos definidos pelos projetistas e no contexto proposto os usuários utilizaram menos estes do que as opções de fala, teclado ou tela de toque. Além disso, 13, de 15 participantes, esqueceram os gestos que lhes foram apresentados no início do teste e tentaram utilizar seus próprios, embora em alguns casos eles não conseguissem descobrir como expressar o comando.

Uma solução para mitigar estes possíveis problemas de *design* é a execução de estudos com usuários antes da implementação do sistema. Como afirmado por Nielsen et al. [66], essa abordagem pode levar a gestos que são fáceis de executar, lembrar, intuitivos e mais ergonômicos. Morris et al. [62] compararam gestos para superfícies de toque criados por

usuários e por pesquisadores, e concluíram que participantes preferiram gestos criados por um grupo grande de pessoas, como aqueles criados por usuários finais, ou propostos por mais de um pesquisador.

Existem outros trabalhos que propuseram estudos com usuários para geração de gestos naturais para diferentes sistemas. Nielsen et al. [66] apresentam um trabalho com uma abordagem para geração de gestos de mãos livres por usuários. O trabalho utiliza diferentes cenários com o objetivo de fazer os participantes não pensarem tecnicamente e então extrair os gestos sugeridos. Estes gestos foram depois avaliados por outros participantes quanto a sua intuitividade e facilidade de memorização com os comandos existentes.

Vatavu [90] utilizou uma abordagem similar, pedindo a usuários que propusessem gestos de mãos livres para ativar comandos em um cenário para controlar a TV. Os gestos foram analisados utilizando uma medida de índice de concordância. Essa abordagem de forma muito similar foi utiliza por Ruiz et al. [79] para gerar gestos de movimento para interação móvel, e Wobbrock et al. [95] para gerar gestos para superfícies de toque.

Henze et al. [36] derivaram, em seu trabalho, gestos de mãos livres para comandos de um tocador de músicas. O trabalho também utilizou a técnica de *Wizard of Oz* para *feedback* durante a geração e avaliação dos gestos em diferentes fases.

Com base nesses trabalhos da literatura, e devido a dificuldades encontradas em conseguir definir os gestos de corpo para interação com o sistema, nos propomos a executar um processo para derivação da interação com o mesmo. Embora o modo de gestos de corpo seja o que mais apresenta dificuldades de definição, e que se beneficiaria mais deste processo, todos os modos foram gerados seguindo as mesmas etapas, para poder-se compará-los ao final, de forma consistente.

3 DEFINIÇÕES INICIAIS

Nos dias de hoje, dispositivos computacionais como *smartphones* e *tablets* estão se tornando comuns em nosso cotidiano. Estes dispositivos introduzem uma nova forma de interação em dispositivos móveis, e são utilizados pelos usuários em diferentes situações. Também, novas tecnologias, antes de pouca disponibilidade, começam a estar presentes na nova geração de vídeo games como o Nintendo Wii, Playstation Move e Microsoft Kinect, que permitem o uso de gestos e movimentos do corpo para interação com o sistema e possuem um potencial de uso em outros ambientes e tarefas, além das áreas de jogos e entretenimento em geral.

Existe a previsão de que modalidades consideradas como mais naturais irão ser amplamente utilizadas [1,9,80], permitindo o uso de dispositivos computacionais nos mais diversos ambientes. No entanto, é importante definir formas para garantir um bom *design* das aplicações, para que estas modalidades sejam úteis para o sistema planejado.

A área de sistemas multimodais tem trabalhado no entendimento de uso das modalidades, de forma a compreender suas características individuais e conjuntas para melhorar a interação do sistema, trazendo sempre formas de interação mais naturais como possibilidades de interação (fala, e gestos). Neste trabalho foi decidido fazer uma análise da aceitação do uso de modalidades de fala, gestos de corpo e gestos de toque em uma tarefa comum a um grupo de usuários.

Neste capítulo é descrito em maior detalhe o escopo do trabalho, os comandos iniciais e os modos do sistema, e uma introdução a trabalhos relacionados ao *design* de interface com interações naturais, referências que serão utilizadas no próximo capítulo.

3.1 Escopo do trabalho

Tendo em vista a questão de pesquisa e objetivos propostos, apresentados na Introdução deste trabalho, serem compreender e comparar a satisfação de uso de modalidades de fala, gestos de corpo e gestos de toque, foi definido para desenvolvimento um sistema de apresentação, uma vez que contempla uma tarefa comum no dia-a-dia de membros da Universidade. Os comandos iniciais foram definidos para uma simples apresentação de slides, e apresentação de imagens. Ainda, a delimitação das modalidades contempla a disponibilidade dos dispositivos Kinect e um *smartphone* Android, sendo, portanto, as

tecnologias já previamente escolhidas para o sistema. Tais tecnologias contemplam formas de interação que começam ser utilizadas em tarefas similares, como para manipulação de mídia digital na televisão, por meio de um sistema dedicado [81], ou uma integração como o Xbox One [61], tornando interessante seu melhor entendimento e uso.

Como o objetivo do trabalho é a análise da interação com o sistema, no uso das diferentes formas disponíveis, foi decidido que não seria prioridade explorar ou implementar algoritmos de reconhecimento, e sim utilizar ao máximo as funções existentes nos kits de desenvolvimento oficiais dos dispositivos. Embora pudesse haver alternativas livres, possíveis problemas de compatibilidade, e a aprendizagem necessária para uso das mesmas, poderiam ser um fator de risco para o tempo disponível para desenvolvimento do trabalho.

O tipo de sistema escolhido para desenvolvimento e avaliação possui determinadas características. Quanto às propriedades CARE, como apresentadas na Seção 2.3, foi pensado em fornecer a propriedade de equivalência, de forma a que todos os comandos do sistema pudessem ser executados com qualquer uma das modalidades. O uso das outras propriedades não entrou no escopo deste trabalho.

Comparado a outras aplicações de uso, na qual um usuário interage isoladamente com um computador, a tarefa escolhida depende do uso do sistema em um contexto em que o usuário interage com este e com outras pessoas ao mesmo tempo. Embora sejam importantes testes em um ambiente real de uso para analisar esta característica com maior precisão, o que não estava previsto no escopo deste trabalho, a percepção dos usuários em um uso restrito do sistema, compreendendo o contexto planejado do mesmo, permite projetarmos um entendimento do uso real.

Para tal análise de aceitação das modalidades do sistema foram planejados, portanto, testes com potenciais usuários em um ambiente restrito, composto por simples tarefas de uso, com objetivo de fazer o usuário ter experiência com o sistema, extrair métricas de sua interação, e melhor entender a percepção de cada participante sobre a interação realizada.

3.2 **Lista Inicial de Comandos**

Um conjunto inicial de comandos foi definido e utilizado como referência para interação com o sistema.

Para a tarefa de apresentação de slides, os comandos disponíveis são:

- Iniciar apresentação;
- Avançar slide;
- Voltar slide;
- Fechar apresentação.

Para a tarefa de apresentação de imagens, os comandos disponíveis são:

- Abrir imagem;
- Aumentar/Diminuir Zoom;
- Rotacionar Imagem;
- Mover área de visualização;
- Fechar Imagem.

A descrição de cada comando foi definida de forma a deixar sua interpretação aberta até certo ponto, como por exemplo, a quantidade de zoom ou de rotação que a imagem irá sofrer. A razão disto é que as modalidades são muito diferentes na forma que expressam dados. Assim, torna-se possível utilizar cada modalidade de uma melhor forma na fase de definição da interação.

3.3 **Modos de Interação**

O sistema foi dividido em três possíveis modos de interação para derivação e análise:

I. Dispositivo Móvel (*Smartphone* Android);
II. Gestos de Corpo (Kinect);
III. Comandos de Fala (Kinect).

É utilizado o termo modo, para manter compatibilidade com a definição de modalidade que outros trabalhos utilizam [14], aonde se pode considerar que os modos utilizados aqui representam na verdade mais de uma modalidade. Aqui foi considerado que, no caso da interação por *smartphone* (modo I), o usuário pode interagir com o sistema através da percepção da tela de toque, acelerômetro ou giroscópio. Já o Kinect provê tanto a captura de 20 pontos do corpo como também reconhecimento de fala (modos II e III).

É importante mencionar que a multimodalidade aqui referenciada é em razão do uso das diferentes formas de interação disponíveis pelo sistema (toque de tela, gestos de corpo e comandos de fala) e não tem relação com a utilização de dois diferentes dispositivos computacionais. Tais formas de interação (modalidades) poderiam estar inseridas em um único dispositivo (um *smartphone*, por exemplo).

4 GERAÇÃO DA INTERAÇÃO

A fim de definir quais as técnicas de interação a serem utilizadas para cada modalidade, foi seguido um processo de estudos com usuários. A ideia de utilizar usuários para participar do processo de design de interfaces não é nova, e nem seu uso para gerar gestos [66]. Foi escolhida essa abordagem para o *design* do sistema pois o método pode criar uma interação mais aceitável para os usuários, como apontado por outros trabalhos [62,66].

A principal etapa deste processo, destacada pelos trabalhos relacionados, era o uso de entrevistas para geração de gestos pela sugestão de usuários. No entanto, os trabalhos que o fizeram para gerar gestos de mãos livres não apresentaram resultados de sua implementação [66,90]. Tendo em mente o dispositivo a ser utilizado (Kinect), suas limitações, e a necessidade de implementação de tais gestos, foi adicionada uma etapa posterior às entrevistas, o uso de grupos focais, para melhor discutir as técnicas geradas na etapa de entrevistas, avaliar, detalhar, e resolver possíveis conflitos que pudessem surgir.

O processo executado seguiu quatro fases seguintes à definição do sistema para derivar a interação nos três modos existentes:

1. Entrevistas Individuais;
2. Análise das Entrevistas;
3. Grupos de Foco;
4. Convergência de Interação.

Estas fases serão apresentadas nas subseções a seguir.

4.1 **Entrevistas Individuais**

Para extrair os gestos através das sugestões dos usuários, foram executadas entrevistas individuais, elaboradas de forma semiestruturada, com nove participantes.

4.1.1 Procedimento

As etapas seguidas em cada uma das entrevistas foram as seguintes:

(I) Introdução: Introdução e apresentação do objetivo da pesquisa e assinatura do termo de consentimento livre e esclarecido.

(II) Perfil do participante: Questões abertas sobre o perfil do participante e a experiência deste com tecnologias similares as que serão utilizadas pelo sistema, como diferentes dispositivos móveis com tela de toque, ou outros dispositivos de gestos como o Nintendo Wii, PS3 move ou *Smart* TVs.

(III) Propostas de interação: Apresentação da ideia do sistema e, para cada comando que o sistema provê, requisitado ao participante que propusesse qual ele acharia que seria a melhor forma de executá-lo utilizando cada um dos diferentes modos disponíveis, sem opções prévias.

(III) Questões finais: Usuários foram questionados sobre a preferência de utilização dos modos para interação com o sistema e de possíveis comandos adicionais.

Para todas as entrevistas, o áudio foi gravado durante todas as etapas, e na etapa de proposição os gestos foram filmados. Na etapa III, primeiro os participantes propuseram a interação com o sistema para o modo I, depois para o modo II, e por ultimo o modo III.

Os participantes foram incentivados a pensar em voz alta sobre suas decisões, e foram questionados a explicar melhor algumas destas. Possíveis conflitos e problemas relacionados com as propostas de interação foram brevemente discutidos no processo de entrevista para obter melhores detalhes ou alternativas. No entanto, a entrevista foi planejada para ser rápida (todas levaram menos de 40 minutos no total). Os usuários foram questionados de forma a propor as primeiras ideias que surgiam em suas mentes para o dado comando e não foi reservado muito tempo para que pensassem em alternativas. Devido à naturalidade que estas formas de interação se propõem a oferecer, as primeiras sugestões seriam as mais intuitivas de serem utilizadas. Um participante podia propor mais de uma técnica de interação para o comando, em cada modo.

4.1.2 Perfil dos Usuários

Nove participantes foram recrutados utilizando-se uma amostra por conveniência. Cinco destes eram do sexo feminino. Do total, dois tinham experiência como professores (um como professor de ensino superior e outro em cursos educacionais básicos de informática), embora atualmente eles não trabalhem na área. Todos os nove participantes estudavam, no momento das entrevistas, em cursos de pós-graduação (Doutorado, Mestrado e

Especialização, com um participante de cada categoria) e graduação (seis participantes), todos da área de computação. Os participantes tinham idades de 19 a 32 anos.

Dos nove participantes, seis possuíam *smartphones*, sendo que três possuem iPhones e três dispositivos Android. Do total três possuíam *tablets*. Apenas dois dos nove não possuem nem *smartphone* nem *tablet*, embora dissessem já terem utilizado e possuírem familiaridade com os mesmos.

A experiência dos participantes na utilização de gestos de movimentação e orientação de dispositivos móveis, com exceção de mudanças de orientação de tela, é quase que exclusivamente para jogos, sendo apontados dois casos diferentes como o gesto de sacudir o iPOD para passar uma música aleatória, e virar o *smartphone* de cabeça para baixo para desligar o alarme.

Dos nove participantes, apenas dois possuem um Nintento Wii, e quatro outros já utilizaram o mesmo. Nenhum deles possui algum outro dispositivo de gestos como o Kinect ou o PS3 Move, embora cinco já tenham utilizado o Kinect e um já tenha utilizado o PS3 Move. Apenas um dos participantes nunca havia utilizado nenhum destes três dispositivos e disse não gostar deles. O uso dos mesmos foi apontado como exclusivamente para jogos.

Todos participantes possuíam pelo menos a experiência de testar a utilização de comandos de voz em aparelhos. A maioria descreveu sua insatisfação com esta forma de interação por causa da grande presença de erros de reconhecimento, e nenhum dos participantes utiliza frequentemente essa função.

4.2 **Análise das Entrevistas**

As gravações da fase de entrevistas foram analisadas para extração de sugestões de interação e outros dados relevantes. Os resultados da etapa de propostas de interação foram utilizados para gerar o conjunto inicial de técnicas de interação do sistema.

4.2.1 Considerações Gerais

Para cada comando, as propostas dos usuários foram categorizadas por similaridade. Como o foco era a implementação destas sugestões, algumas propostas foram categorizadas levando em conta as limitações das tecnologias. O dispositivo Kinect não é capaz de perceber os dedos das mãos individualmente, mas apenas o ponto central da mão como um

todo, portanto, as propostas nesta etapa que utilizaram gestos com um ou mais dedos foram categorizadas em um conjunto de propostas similares que foram realizadas com a mão inteira. Essas limitações obrigam a modificação dos gestos para implementação, e foram deixadas para serem melhor analisadas na fase seguinte, com os grupos focais.

Os resultados das entrevistas apresentaram alguns comandos com propostas similares entre quase todos os participantes, como, por exemplo, os gestos para avançar um slide ou voltar, enquanto outros comandos, como o de fechar, tiveram varias diferentes propostas.

Frequentemente os participantes propuseram mais de uma forma de executar o comando para cada modo. Eles muitas vezes executavam um gesto, ou um comando de fala, e trocavam para uma segunda ou terceira alternativa. Todas as alternativas que não foram desqualificadas pelos participantes foram levadas em conta para análise.

Foi comum, nos resultados, que muitas das propostas fossem similares a técnicas já existentes de outras interfaces, como o caso de gestos de aplicações de dispositivos móveis, e também algumas de *desktops,* como utilização de um botão de fechar no canto superior da janela.

4.2.2 Definição da Interação

As duas categorias de propostas de interação mais frequentes, para cada modo de interação com o sistema, foram escolhidas para serem apresentadas na próxima fase como as opções principais do conjunto de interação para aquele modo a ser discutido com o grupo. No caso em que menos de três propostas foram feitas, e a diferença de frequência era muito alta para uma das opções, em comparação com as outras, apenas a categoria mais escolhida foi selecionada.

A seguir são apresentadas as propostas que surgiram e que apresentaram maior frequência para serem selecionadas para a próxima fase.

4.2.2.1 Iniciar Apresentação e Abrir Imagem

Uma vez que os arquivos a serem abertos estariam, de alguma forma, disponíveis para serem selecionados, os participantes propuseram a forma com a qual pudessem ser abertos. As propostas para cada modo são apresentadas na Tabela 1.

Tabela 1 - Propostas de Interação para Iniciar Apresentação.

Modo	**Interação**	**Frequência de citação**
Dispositivo Móvel	Um toque em uma lista.	7
	Dois toques em uma lista.	3
Gestos de Corpo	Empurrar mão para frente (em cima do arquivo, Figura 2).	3 - com um dedo. 3 - mão aberta.
	Iniciando com mãos juntas na frente do corpo, abrir braços.	3
Comando de fala	“Abrir”/”Iniciar” + nome do arquivo.	5
	Nome do arquivo + “Iniciar”/”Abrir”.	4

Figura 2 - Empurrar mão para frente.

4.2.2.2 Avançar Slide

Para este comando as propostas foram bem similares, oferecendo poucas opções para os gestos (duas) sendo as mais frequentes quase unânimes. As mais citadas nos diferentes modos são apresentadas na Tabela 2.

Tabela 2 - Propostas de Interação Para Avançar Slide.

Modo	**Interação**	**Frequência de citação**
Dispositivo Móvel	Lançar para esquerda (Figura 3).	8
Gestos de Corpo	Empurrar para esquerda (Figura 4).	8
Comando de fala	“Próximo”.	5
	“Avançar”.	4

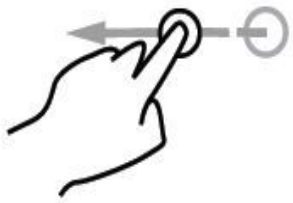


Figura 3 - Lançar para esquerda (adaptado de [31]).

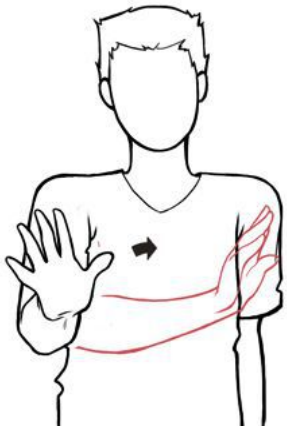


Figura 4 - Empurrar para esquerda.

4.2.2.3 Voltar Slide

As propostas para este comando foram as mesmas do comando avançar, apenas em direções opostas para os gestos (e utilizando as mesmas mãos/dedo). Para o caso dos

comandos de fala, os mais citados foram “Anterior” (citado duas vezes) e “Voltar” (Citado seis vezes).

4.2.2.4 Rotacionar Imagem

Os participantes foram convidados a propor um comando que provocasse a rotação da imagem que está sendo apresentada. A quantidade de rotação associada a resposta foi muitas vezes explicitada como 90º ou um pouco menos (uma quantidade fixa). As propostas para cada modo são apresentadas na Tabela 3.

Tabela 3 - Propostas de Interação para Rotacionar Imagem.

Modo	**Interação**	**Frequência de citação**
Dispositivo Móvel	Rotacionar com dois dedos (Figura 5).	4
	Rotacionar com três dedos.	2
	Duplo clique abre menu com opções para rotação de 90º.	2
Gestos de Corpo	Rotacionar mão, com braço esticado para frente (Figura 6).	2 - mão semi-fechada, como agarrando algo, Figura 6. 1 - mão aberta.
	Rotacionar com as duas mãos, como se girasse algo (como um volante de carro) para um dos lados.	3
Comando de fala	Girar/Rotacionar direita/esquerda (uma quantidade fixa).	6

Figura 5 - Rotacionar com dois dedos (adaptado de [31]).

Figura 6 - Rotacionar mão.

4.2.2.5 Aumentar/Diminuir Zoom

Este comando apresentou propostas similares dos diferentes participantes, com exceção do modo de fala que possui diversidade na forma de manipular a quantidade de zoom esperada. As propostas são apresentadas na Tabela 4.

Tabela 4 - Propostas de Interação para Modificar Zoom.

Modo	**Interação**	**Frequência de citação**
Dispositivo Móvel	Aperto/Escala (*pinch*) com dois dedos (Figura 7).	8
Gestos de Corpo	Separar/Aproximar mãos, como se apertassem ou esticassem algo (Figura 8).	7 – horizontal 1 - diagonal
Comando de fala	“Aumentar” ou ”Ampliar” e “Diminuir” ou ”Reduzir”, quantidade fixa.	4
	“zoom mais”, “zoom menos”, opcionalmente com percentagem.	3

Figura 7 - Aperto com dois dedos (adaptado de [31]).

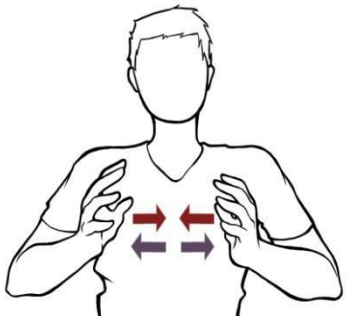


Figura 8 - Separar/aproximar mãos.

4.2.2.6 Mover Área de Visualização

Este comando apresentou propostas similares dos diferentes participantes, e são apresentadas na Tabela 5.

Tabela 5 - Propostas de Interação para Mover Área de Visualização.

Modo	**Interação**	**Frequência de citação**
Dispositivo Móvel	Arrastar com um dedo (como se estivesse empurrando a imagem).	8
Gestos de Corpo	Empurrar com uma mão em qualquer direção.	7 – mão toda. 1 – apenas com um dedo.
Comando de fala	“Mover direita”, “Mover esquerda”.	4

4.2.2.7 Fechar apresentação/imagem

Diferente do comando anterior, “fechar” teve várias propostas diferentes. As mais frequentes que foram escolhidas para serem apresentadas e discutidas na próxima fase são apresentadas na Tabela 6.

Tabela 6 - Propostas de Interação para Fechar Apresentação.

Modo	**Interação**	**Frequência de citação**
Dispositivo Móvel	Pressionar botão voltar do *smartphone*, citando dispositivos Android.	2
	Toque simples/duplo para abrir menu e mostrar opção para fechar.	2
	Aperto multitoque (Figura 9).	2
	Botão de fechar no topo da tela, como um 'X'.	2
	Lançar para cima/baixo.	2
Gestos de Corpo	Juntar mãos (Figura 9).	4
	Braço esticado para frente, fechar mão.	2
Comando de fala	"Fechar" (opcionalmente incluir "apresentação/imagem/arquivo" no final).	8

Figura 9 - Aperto multitoque (adaptado de [31]).

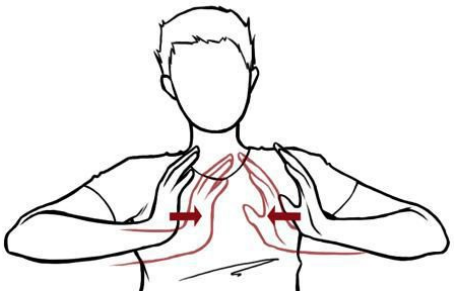


Figura 10 - Juntar mãos.

4.2.3 Comandos Adicionais

Os participantes foram questionados para destacar possíveis comandos adicionais para as tarefas existentes do sistema. Algumas sugestões foram novos comandos que poderiam fazer parte das tarefas propostas, enquanto outras foram um pouco além criando novas tarefas (como manipulação de vídeos). As sugestões que surgiram foram: passar slides automaticamente (com tempo definido), minimizar, editar imagem, escrever em cima, clicar/selecionar links da apresentação, ir para determinado slide, utilizar apontador, abrir arquivos Word e PDF e abrir arquivos de vídeo.

4.2.4 Preferência dos Modos

Nesta etapa os participantes também foram questionados sobre a preferência de uso dos diferentes modos para o sistema proposto.

Em geral, a maioria disse que usaria a opção de *smartphone* mais que as outras (seis participantes), por motivos como menor quantidade de erros e menor esforço. Um participante disse que se sentiria mais seguro tendo algo físico em mãos. Dos outros três participantes que disseram não preferir o *smartphone*, dois apontaram os gestos de corpo como a melhor opção, e o último disse que usaria tanto fala como gestos de forma igual, pois cada um tem seu uso nos diferentes comandos, aonde utilizaria a fala para comandos simples como ‘próximo’, ‘fechar’, e gestos para manipulação da imagem como ‘mover’ e ‘zoom’.

Entre as vantagens de gestos de corpo, um dos participantes disse que estaria com as mãos livres, para poder, por exemplo, escrever no quadro sem precisar largar o aparelho. Outro destacou que a tela que estaria utilizando para manipulação seria maior que a do *smartphone*. Entre as desvantagens, dois participantes disseram que se sentiriam constrangidos em ficar fazendo gestos. Outros comentários negativos vindos de três diferentes participantes foram que não seria prático, por exigir se movimentar muito, não se sentiria seguro inicialmente em como executá-los, pois não é algo que está acostumado a fazer, e que poderia ser difícil de identificar caso estivesse gesticulando muito.

O modo de fala foi o que apresentou maior rejeição. Durante a etapa de proposição de ideias este modo foi difícil de ser gerado e definido em manipulações físicas, como zoom, e mover

imagem. Esta maior dificuldade de indicar questões físicas já era esperada, uma vez que a modalidade de fala apresenta características diferentes das outras. Por outro lado, algumas vantagens e facilidades aparecem em outros comandos, apontado por alguns participantes, como o comando para fechar, e ir para um determinado slide. Ainda assim, alguns participantes disseram que este modo geralmente apresenta muitos erros e quase não o usariam, e que a fala já estaria sendo utilizada como principal canal de comunicação em uma apresentação. Ainda um dos participantes disse que seria um pouco constrangedor e estranho usar a fala para controlar o sistema no meio de uma apresentação.

4.3 **Grupo Focal**

O uso do método de estudo de grupos focais,uma entrevista realizada com um grupo de participantes, já foi utilizado em outras pesquisas na área de IHC [35,49], e pode revelar sentimentos e opiniões que se beneficiam da discussão em um ambiente social. É sugerido na literatura que sejam utilizados grupos com tamanho de três a dez participantes [49], e pelo menos dois grupos.

Neste trabalho foram utilizados dois grupos focais com propósito de validar, reduzir e melhorar o conjunto de propostas derivadas das entrevistas individuais. Era de interesse discutir possíveis conflitos ou problemas de reconhecimento que podem aparecer na implementação para o dispositivo pretendido, e descobrir alternativas ou uma melhor definição da interação. Ao contrário das entrevistas nas quais os participantes fizeram propostas a partir do seu próprio entendimento e opinião, no grupo focal houve a oportunidade de diferentes participantes compartilharem suas opiniões e discutir melhor os benefícios e motivos de cada escolha, abrangendo as decisões com mais informações vindas um dos outros.

Um moderador[2] ficou responsável em encorajar a discussão sobre as opiniões dos participantes para definir como estes gostariam que o sistema fosse.

---

[2]Papel comum em grupos focais, responsável por liderar a discussão e manter os participantes dentro do tópico de pesquisa.

4.3.1 Procedimento

Os passos executados para os dois grupos focais foram os seguintes:

(I) Introdução à pesquisa: Apresentação aos participantes do objetivo da pesquisa e assinatura dos termos de consentimento.

(II) Perfil dos participantes: Questões sobre o perfil dos participantes e suas experiências com tecnologias similares, seguindo uma abordagem similar a das entrevistas individuais.

(III) Explicação do sistema: Explicação aos participantes do conceito geral do sistema a ser desenvolvido.

(IV) Discussão da interação: para cada comando do sistema,

(a) O comando e seu efeito foram explicados aos participantes;

(b) As propostas de interação selecionadas das entrevistas individuais foram apresentadas aos participantes, através da explicação e demonstração feita pelo moderador;

(c) Os participantes foram questionados sobre, dentre as opções apresentadas, qual consideravam a melhor, e, caso não concordassem que estas fossem boas, foram incentivados a sugerir novas opções de interação;

(d) Os participantes foram questionados para melhor descrever as técnicas de interação para possível implementação, levando em consideração possíveis limitações da tecnologia ou contexto, quando necessário.

A discussão de cada comando se deu na mesma ordem apresentada nas entrevistas individuais, primeiro todos os comandos para o *smartphone* (modo I), depois para os gestos de corpo (modo II) e ao fim para os comandos de voz (modo III).

Gestos de corpo que não eram possíveis de ser implementados, foram apresentados para os participantes em conjunto com as outras opções. A opinião dos participantes sobre os gestos foi registrada e depois eles foram informados que a captura fina de gestos, como a orientação da mão, não eram detectadas pelo dispositivo. Outros gestos que foram agrupados juntos, como por exemplo, o movimento de empurrar com o dedo ou com a mão, foram ambos apresentados aos participantes. Novamente, eles foram questionados sobre suas opiniões e preferências, e depois foram informados que o dispositivo não reconhece os dedos individualmente. O objetivo foi apresentar aos usuários as propostas originais, mesmo

que não fossem possíveis de ser implementadas, e discutir possíveis alternativas tendo conhecimento destas.

Embora as categorias mais frequentes tenham sido apresentadas aos participantes como escolhas iniciais, durante a discussão em grupo algumas vezes as principais técnicas selecionadas não eram aceitas pela maioria. Para tornar a discussão mais abrangente, o moderador introduziu algumas vezes opções não inicialmente selecionadas, surgidas das entrevistas individuais, ou então novas ideias derivadas das discussões do momento.

A fala possui uma natureza linguística com mais distinção e semântica do que os gestos, por este motivo a disponibilidade de mais de uma palavra para disparar o mesmo comando apresenta menor possibilidade de conflitos. Devido à limitação de tempo disponível (uma hora e meia, requisitado como limite por alguns participantes), e a ordem de apresentação dos modos para derivação, durante a interação foi preferido abordar esta modalidade de forma mais breve que as outras, considerando de uma forma geral como esta deveria ser utilizada em conjunto da tarefa principal de apresentação, ao invés de uma revisão mais minuciosa para cada um dos comandos.

4.3.2 Perfil dos Participantes

Os participantes dos dois grupos compõem amostras recrutadas por conveniência, e são alunos da Universidade aonde este trabalho está sendo desenvolvido. O primeiro grupo continha oito estudantes de graduação em cursos de computação. As idades dos participantes variavam entre 17 e 20 anos. Apenas um dos participantes possuía um Nintendo Wii e nenhum dos outros possuía qualquer uma de outras tecnologias baseadas em gestos. Quatro destes não possuíam *smartphones*, e apenas dois possuíam *tablets*. Nenhum destes utilizava a fala como forma de interação, embora já tivessem a experiência de testá-la em algumas aplicações.

O segundo grupo era composto de seis estudantes de graduação de cursos de computação, com idades variando entre 19 e 26 anos. Quatro possuíam *smartphones*, dois possuíam *tablets*. Dois destes possuíam um Nintendo Wii, mas nenhum outro possuía algum dispositivo de tecnologias baseadas em gestos. Nenhum destes utilizava a fala como uma forma de interação, embora já houvessem testado esta opção em algumas aplicações.

4.4 **Convergência da Interação**

Os grupos focais foram uma boa fonte de ideias e reflexões sobre a interação em um cenário real. Uma grande preocupação foi sobre como o uso do sistema é afetado pelo contexto. Por exemplo, se o apresentador está se movendo, gesticulando ou falando muito, a ativação não desejada de um comando não pode acontecer.

O uso de dois grupos focais acrescentou diversidade ao estudo. Alguns comandos foram aceitos de forma geral assim como vieram das entrevistas individuais, ou apenas levemente modificados, enquanto outros tiveram diferentes respostas para cada grupo.

A seguir serão apresentadas as discussões para os comandos, organizadas entre os comandos para o dispositivo móvel e via gestos de corpo, separando os comandos de fala em uma seção diferente.

4.4.1 Gestos de Corpo e Dispositivo Móvel

A **Erro! Fonte de referência não encontrada.** apresenta os gestos escolhidos para mplementação no sistema a partir das considerações dos grupos sobre os comandos de Iniciar Apresentação de Slides e Imagem, para interação via dispositivo móvel e via gestos de corpo.

Tabela 7 - Comandos de Iniciar Apresentação.

Forma de Interação	**Comandos Escolhidos**
Dispositivo Móvel	Um toque simples e o gesto de arrastar para cima.
Gestos de corpo	Empurrar a mão para frente sobre o arquivo, e fechar a mão para navegar entre a tela de arquivos.

Para o modo de smartphone, os dois grupos concordaram que gostariam da possibilidade de visualizar o arquivo pelo *smartphone* antes de iniciar a apresentação. O primeiro grupo definiu a interação como um toque para abrir localmente o arquivo para visualização, e outro toque no primeiro slide/imagem iniciaria a apresentação. Um dos participantes deste primeiro

grupo citou que gostaria de usar o gesto de arrastar o slide para cima como forma de iniciar a apresentação, mas tal gesto não foi aceito pelo restante do grupo.

O segundo grupo sugeriu inicialmente o uso de um menu com a opção de inicializar a apresentação, que poderia existir na tela ou ser aberto através de um toque. Com o decorrer da discussão entre os participantes, foi sugerido, e aceito pela maioria, que uma boa escolha seria arrastar o slide/imagem para cima para iniciar a apresentação, com a possibilidade de um menu auxiliar. Um dos participantes disse desgostar de ter que tocar na tela para o menu aparecer.

Para o modo de gestos de corpo, o primeiro grupo sugeriu uma nova forma de interação, que seria agarrar e arrastar o arquivo para uma área de visualização. No entanto, após demonstração do protótipo do sistema, eles gostaram da forma apresentada, que utilizava a mão fechada para indicar o deslizamento da tela da pasta de arquivos, e o empurrar para selecionar o arquivo.

O segundo grupo disse acreditar que o gesto de empurrar para selecionar o arquivo pudesse provocar certa dificuldade para pessoas que não sabem utilizar direito a tecnologia e que, por exemplo, poderiam ficar com o braço esticado já de inicio. Como alternativa eles decidiram por definir o movimento de fechar a mão para selecionar o arquivo. Para rolagem entre a pasta, foi definido o posicionamento da mão sobre uma das bordas de limites superior ou inferior, provocando o deslocamento naquele sentido.

Devido às diferentes escolhas dos grupos, ambas as opções foram consideradas. No entanto, elas entram em conflito direto uma com a outra, não sendo possível manter a implementação de ambas ao mesmo tempo, uma vez que o gesto de fechar a mão não pode ser utilizado para selecionar o arquivo e deslizar a tela ao mesmo tempo. Uma vez que o gesto de empurrar a mão para frente, como forma de selecionar o arquivo, foi proposto com maior frequência nas entrevistas individuais, este foi escolhido para ser implementado, em conjunto com o uso do gesto da mão fechada para navegar entre a tela de arquivos.

A Tabela 8 apresenta as considerações dos grupos sobre os comandos de Avançar e Voltar Slide, para interação via dispositivo móvel e via gestos de corpo.

Tabela 8 - Gestos para Avançar e Voltar Slide.

Forma de Interação	**Comandos Escolhidos**
Dispositivo Móvel	Arrastar para os lados, para avançar e voltar os slides.
Gestos de corpo	Empurrar com a mão fechada para os lados.

O gesto apresentado na opção de *smartphone* foi rapidamente aceito pelos dois grupos. Um participante do segundo grupo sugeriu a disponibilidade adicional da utilização do botão de volume para avançar/voltar slides (presente na lateral de alguns modelos de *smartphone*). Um outro participante disse que talvez dessa forma a função pudesse ser erroneamente ativada durante a apresentação, e os demais participantes não demonstraram preferência pela opção.

O gesto apresentado, empurrar com a mão para o lado, foi aceito em ambos os grupos. O uso da mão fechada, como se estivesse agarrando algo, foi uma opção que surgiu em ambos como uma possível modificação, de forma a não confundir com movimentos naturais de gestos do usuário durante uma tarefa de apresentação, e indicar início e fim do movimento. Devido ao consenso entre os grupos, este gesto foi escolhido para ser implementado.

A Tabela 9 apresenta as considerações dos grupos sobre o comando de Rotacionar Imagem, para interação via dispositivo móvel e via gestos de corpo.

Tabela 9 - Gestos para Rotacionar Imagem.

Forma de Interação	**Comandos Escolhidos**
Dispositivo Móvel	Rotação com dois dedos na tela.
Gestos de corpo	Rotação de ângulo entre as mãos fechadas.

No modo de smartphone, para o primeiro grupo, o gesto de rotação com dois dedos foi preferido pela maioria dos participantes, embora dois dissessem preferir utilizar três dedos. Uma opção alternativa ainda seria o uso de botões acessados através de um menu.

O segundo grupo teve um resultado similar. Primeiramente, o uso de botões para rotação à direita e esquerda, 90 graus, foi considerado. Ao final, eles disseram preferir utilizar o gesto de rotação com dois dedos, por ser de mais rápido acesso, mas de forma que um dedo de uma das mãos ficasse parado, e um segundo dedo, da outra mão, fizesse o movimento circular. Eles disseram que dessa forma seria mais simples de ser executado.

Uma vez que ambos os grupos consideraram boa a utilização de dois dedos para rotação, este foi escolhido para ser implementado. Embora o segundo grupo tenha demonstrado o comando de uma forma um pouco diferente, esta, a princípio, não acarretaria em diferenças na implementação. Ainda, assim como nas entrevistas individuais, a opção de botões para rotação em ambas as direções, acessível em um menu, foi citada como uma possível opção, e poderá ser considerada para implementação com uma mais baixa prioridade.

Para o modo de gestos de corpo, o primeiro grupo definiu o gesto de rotação com as duas mãos de forma similar ao apresentado, com uma das mãos iniciando em uma altura maior que a outra, e, seguindo um movimento circular, trocarem de alturas. O grupo foi questionado como o sistema deveria identificar o início do movimento. A forma aceita pelo grupo foi que o movimento iniciaria, seguindo a posição do gesto como definido, quando as mãos estivessem fechadas.

O segundo grupo definiu um gesto similar utilizando duas mãos, mas em posições diferentes. Pensando em utilizar gestos um pouco mais sutis, eles definiram uma mão parada, de forma similar ao *smartphone*, e a segunda mão realizando um movimento circular em volta, ambas fechadas.

Devido à similaridade entre os gestos definidos pelos dois grupos, foi escolhido implementar um gesto que considere a mudança de ângulo entre as duas mãos, seguindo ambas as formas de execução apresentadas.

A Tabela 10 apresenta as considerações dos grupos sobre os comandos de Aumentar e Diminuir Zoom, para interação via dispositivo móvel e via gestos de corpo.

Tabela 10 - Gestos para Aumentar e Diminuir Zoom.

Forma de Interação	**Comandos Escolhidos**
Dispositivo Móvel	Aperto com dois dedos para aumentar/diminuir ampliação e dois toques na tela para voltar para a ampliação inicial.
Gestos de corpo	Separar e juntar mãos fechadas, por movimento horizontal ou diagonal.

Para o modo de gestos de corpo, o primeiro grupo preferiu que o gesto fosse feito na diagonal utilizando ambas as mãos. As mãos em posição fechada foram uma solução para indicar início e duração do movimento.

O segundo grupo considerou útil o gesto apresentado, com a separação das mãos, e definiu o uso das mãos fechadas também como forma de demonstrar o início do gesto. Um dos participantes sugeriu que o gesto pudesse ser utilizado em conjunto com o de rotação, levando em conta a diferença de distância das mãos, aceito pelo restante do grupo.

Este já é um gesto amplamente utilizado com dois dedos, e foi aceito por ambos os grupos. Ambos também gostaram da opção de tirar o *zoom* utilizando dois toques sobre a imagem. Devido ao consenso, esse gesto foi escolhido para implementação.

Devido às diferentes escolhas de cada grupo, ambas as opções foram escolhidas para serem implementadas. Os gestos de separar ou juntar as mãos, enquanto fechadas, horizontalmente ou diagonalmente, farão a alteração do *zoom* da imagem. O uso conjunto de *zoom* e rotação, que requer capturar a distância das mãos em qualquer ângulo, potencialmente pode provocar a mudança de *zoom* quando não desejado pelo usuário e vice-versa, assim, poderá ser considerado somente após testes preliminares, sendo não prioritário.

A Tabela 11 apresenta as considerações dos grupos sobre o comando para Mover a Área de Visualização, para interação via dispositivo móvel e via gestos de corpo.

Tabela 11 - Gestos para Mover Área de Visualização.

Forma de Interação	**Comandos Escolhidos**
Dispositivo Móvel	Arrastar com um dedo.
Gestos de corpo	Mover a mão fechada, como se agarrasse a tela.

O gesto apresentado para uso no dispositivo móvel já é amplamente utilizado em outras aplicações, e envolve simplesmente a movimentação de um dedo sobre a tela como se estivesse empurrando um pedaço de papel em uma superfície, na qual uma movimentação para cima envolve a visualização de uma parte mais abaixo da imagem. Ambos os grupos aceitaram esta forma de interação para este modo, e portanto será utilizado no sistema.

O gesto de corpo apresentado aos grupos, advindo das entrevistas individuais, foi aceito como forma de interação. No entanto, os grupos foram questionados em que momento este iniciaria o movimento. Como solução, o uso da mão fechada foi aceito por ambos os grupos para ativar este comando. Com consenso de ambos os grupos, este gesto será utilizado para implementação.

A Tabela 12 apresenta as considerações dos grupos sobre os comandos de Fechar Apresentação de Slides ou Imagem, para interação via dispositivo móvel e via gestos de corpo.

Para este comando, no modo de *smartphone*, diversas propostas foram feitas nas entrevistas individuais sobre este modo de interação. Ambos os grupos concordaram que o gesto de aperto multitoque poderia trazer dificuldades em telas pequenas, após uma discussão entre os participantes.

O primeiro grupo preferiu as opções da utilização de um botão externo para voltar, como o disponível em dispositivos Android, seguido pela opção de um menu, o qual seria aberto após um toque na tela. Uma opção alternativa foi a de um toque longo na tela, que questionaria o usuário para fechar a apresentação.

No segundo grupo, foi decidido que, uma vez que o gesto para iniciar a apresentação aceito foi o de arrastar para cima, o gesto no sentido inverso seria o de fechar. Para o caso da imagem, o gesto para fechar poderia ser confundido com o gesto de mover para baixo, e os participantes concordaram que ele poderia estar habilitado apenas quando a imagem estivesse na escala original (sem zoom e portanto sem possibilidade de mover sua área de foco). A vantagem levantada pelos participantes sobre uma opção de botão seria a de não precisar olhar para a tela.

Para o modo de gestos de corpo, o primeiro grupo aceitou a opção apresentada, utilizando as duas mãos separadas até se juntarem. O segundo grupo aceitou duas opções como válidas: a primeira similar a apresentada, aonde o usuário poderia levantar e esticar o braço para frente até que a mão fosse identificada, e fechá-la para disparar o comando. A segunda opção foi utilizar um gesto similar ao do *smartphone*, com a mão fechada movimentar ela para baixo como se agarrasse a tela para baixo.

Seguindo as decisões dos grupos focais, os gestos escolhidos para serem implementados foram o de juntar as mãos, e fechar a mão na frente do usuário, gestos que também apareceram nas entrevistas individuais. O segundo, no entanto, pode apresentar conflitos com o gesto de mover ou passar/voltar slides, e, como sugerido na interação do segundo grupo, poderá ser ativado após alguns segundos de pose estática do usuário com a mão levantada.

Tabela 12 - Gestos para Fechar Apresentação.

Forma de Interação	**Comandos Escolhidos**
Dispositivo Móvel	Botão de voltar do aparelho, um menu com a opção de fechar, e o gesto de arrastar para baixo.
Gestos de corpo	Juntar as mãos, e fechar a mão com o braço na frente do usuário (após alguns segundos aberta).

Na execução das entrevistas individuais e grupos focais os participantes disseram que o modo de *smartphone* deveria apresentar um *feedback* da apresentação na tela do

dispositivo, e não ser apenas uma superfície de toque. Isto foi um dos requisitos escolhidos para serem implementados no sistema na etapa seguinte.

4.4.2 Comandos de Fala

O modo de fala, como discutido anteriormente na subseção 4.3.1, foi abordado de forma diferente dos outros dois modos, nesta etapa de grupo focal. Devido a limitações de tempo, e por decisão do moderador, conhecendo a natureza linguística do modo de fala, foi preferido abordar outras questões e não exatamente quais as sentenças a serem utilizadas para ativar o comando neste modo. Diferente dos modos de gesto, é muito mais difícil haver conflitos entre as definições dos comandos de fala. As sentenças advindas das entrevistas individuais para os diferentes comandos do sistema foram brevemente apresentadas aos participantes, que concordaram com estas ou pelo menos não apresentaram comentários contrários, mas não foi dado tempo para maiores discussões quanto às melhores opções.

Uma das questões abordadas mais a fundo foi quanto à seleção do arquivo neste modo, no momento de iniciar uma apresentação de slides ou abrir uma imagem. Nas entrevistas individuais o nome do arquivo foi sugerido para indicar o arquivo a ser selecionado. Os participantes dos grupos foram questionados se esta forma seria adequada, uma vez que o nome dos arquivos poderiam ser longos e difíceis de serem pronunciados. Para este caso a opção de numeração dos arquivos foi introduzida. Ambos os grupos concordaram que o nome do arquivo poderia ser um problema, e que o uso de uma numeração seria mais simples para ser utilizada.

Uma outra questão discutida para este modo, levantada pelos participantes, tratava da preocupação de que a fala poderia ser erroneamente interpretada enquanto o usuário estivesse falando, durante a apresentação, e comandos poderiam ser ativados sem intenção. Neste sentido, foi sugerido, por um participante de um dos grupos, que o *smartphone* pudesse habilitar ou desabilitar algum dos modos para que não atrapalhassem o apresentador. O moderador sugeriu que uma forma de o sistema identificar a intenção de utilizar a fala para disparar um comando pelo apresentador poderia ser pelo uso de uma palavra chave, antes do comando pretendido, ou então pela direção do olhar do usuário, direcionado ao sistema. Entre estas opções os participantes dos dois grupos acharam a direção do olhar para o sistema uma boa escolha.

4.4.3 Comentários gerais

Durante a condução dos grupos focais foram apresentadas algumas considerações gerais sobre o sistema pelos participantes. Dois participantes, um de cada grupo, expuseram sua opinião de que o usuário pudesse controlar a ativação ou desativação de alguns modos. Uma das sugestões foi utilizar o *smartphone* para, por exemplo, ativar ou desativar o uso da fala e gestos. Outra sugestão foi o uso de comandos de fala para ativar ou desativar os gestos de corpo.

Uma outra sugestão, apresentada por um participante, foi que o usuário fosse capaz de definir como o comando seria ativado, indicando, por exemplo, qual seria o gesto que o ativaria. Embora esta opção pareça bastante interessante, ela foge um pouco do propósito para implementação deste trabalho.

Uma consideração apresentada por um dos participantes do segundo grupo foi manter gestos similares entre os modos de *smartphone* e gestos de corpo. Essa opinião surgiu enquanto era discutido o comando de fechar a apresentação utilizando os gestos de corpo, o movimento de arrastar com a mão seria um gesto similar ao do *smartphone* para a mesma função.

A aceitação entre os modos durante a discussão dos grupos apresentou resultados similares aos das entrevistas individuais. Em geral a fala não lhes pareceu uma boa opção para ser utilizada durante uma apresentação, mas talvez uma boa opção para determinados comandos, como para fechar. Os gestos de corpo, principalmente no momento da apresentação do protótipo, resultaram em comentários positivos a este modo por alguns participantes, embora o uso em uma situação real possa apresentar problemas diversos entre gesticulação normal do apresentador, e desconforto em executá-los na presença de uma plateia.

5 IMPLEMENTAÇÃO DO SISTEMA

Neste capítulo serão apresentadas os passos que foram seguidos para implementação do sistema, sua arquitetura e visão geral de desenvolvimento.

5.1 Tecnologias

Previamente já haviam sido escolhidos os dispositivos de controle do sistema a serem utilizados, sendo um dispositivo Kinect e um *smartphone* Android.

A Microsoft disponibiliza um Kit de Desenvolvimento de Software (*Software Development Kit*, ou apenas SDK) oficial para programação com o dispositivo Kinect, chamado de Kinect for Windows SDK [53]. A versão 1.7 foi utilizada para implementação.

Foi escolhida a linguagem C# para desenvolvimento do sistema no servidor. Esta linguagem apresenta bom suporte para desenvolvimento de aplicações com o SDK do Kinect, e serviços REST. Foi utilizado o .NET Framework 4, e plataforma de desenvolvimento Microsoft Visual Studio 2010. O serviço REST foi implementado utilizando o framework WCF.

Para desenvolvimento da aplicação Android foi utilizado o ADT Bundle [3], que possui a IDE de desenvolvimento Eclipse e ferramentas para desenvolvimento Android. Através da ferramenta de gerenciamento de SDKs do Android (SDK Manager) é possível fazer download da versão da API do Android que se deseja desenvolver. Para o sistema desse trabalho foi escolhida a versão 4.1.2 (API16). Foi utilizado a biblioteca HttpClient do projeto Apache HttpComponents [6] para possibilitar a comunicação do cliente Android com o serviço WCF.

Foi escolhido utilizar a ferramenta Power Point [60] para controle da apresentação de slides, através do uso da biblioteca *Microsoft Office Interop* versão 12.

5.2 Arquitetura do Sistema

Em um ambiente ubíquo com vários dispositivos é comum o uso de protocolos sobre a rede local, permitindo a comunicação através de serviços utilizando arquiteturas como REST Web Services [24,44]. Seguindo esta mesma ideia foi pensada a utilização da rede local, com o uso de um roteador Wi-Fi, para proporcionar a comunicação entre o *smartphone* e o servidor local. Embora esta tecnologia consuma mais energia do que, por exemplo, o Bluetooth, ela possibilita a comunicação entre um maior número de dispositivos ao mesmo tempo,

possibilitando a futura expansão do sistema para compartilhar dados entre um maior número de usuários.

A configuração do ambiente como planejado é listado a seguir e representado na Figura 11. Ele é composto dos seguintes componentes:

1. Projetor de imagens e uma tela
2. Computador Desktop ou Notebook, atuando como servidor
3. Microsoft Kinect
4. Smartphone Android

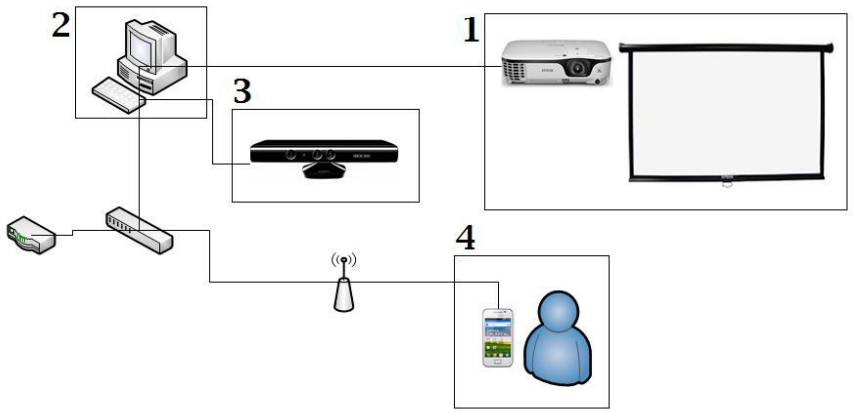


Figura 11 - Configuração do sistema

Nas salas de aula atuais da universidade em que este trabalho foi desenvolvido, os elementos 1 e 2 estão sempre presentes, possibilitando ao professor utilizar o computador e apresentar informações aos alunos. O dispositivo Kinect (3) ficará conectado diretamente ao computador no qual o sistema será executado, capturando os gestos do apresentador e identificando comandos de fala. A comunicação entre o *smartphone* (4) é feita através de uma rede local com o computador. A forma de comunicação entre os dispositivos de entrada, *smartphone* Android e o Microsoft Kinect, se dão através de um protocolo HTTP e a interface USB, respectivamente.

Uma imagem do sistema no ambiente de sala de aula e configuração que foi testado pode ser visualizada na Figura 12. Na região ‘A’ da figura está localizado tanto o dispositivo Kinect, quanto o notebook que foi utilizado para execução do sistema (este logo atrás do Kinect na imagem). A região ‘B’ mostra a posição da tela de projeção, em relação ao usuário na região ‘C’.

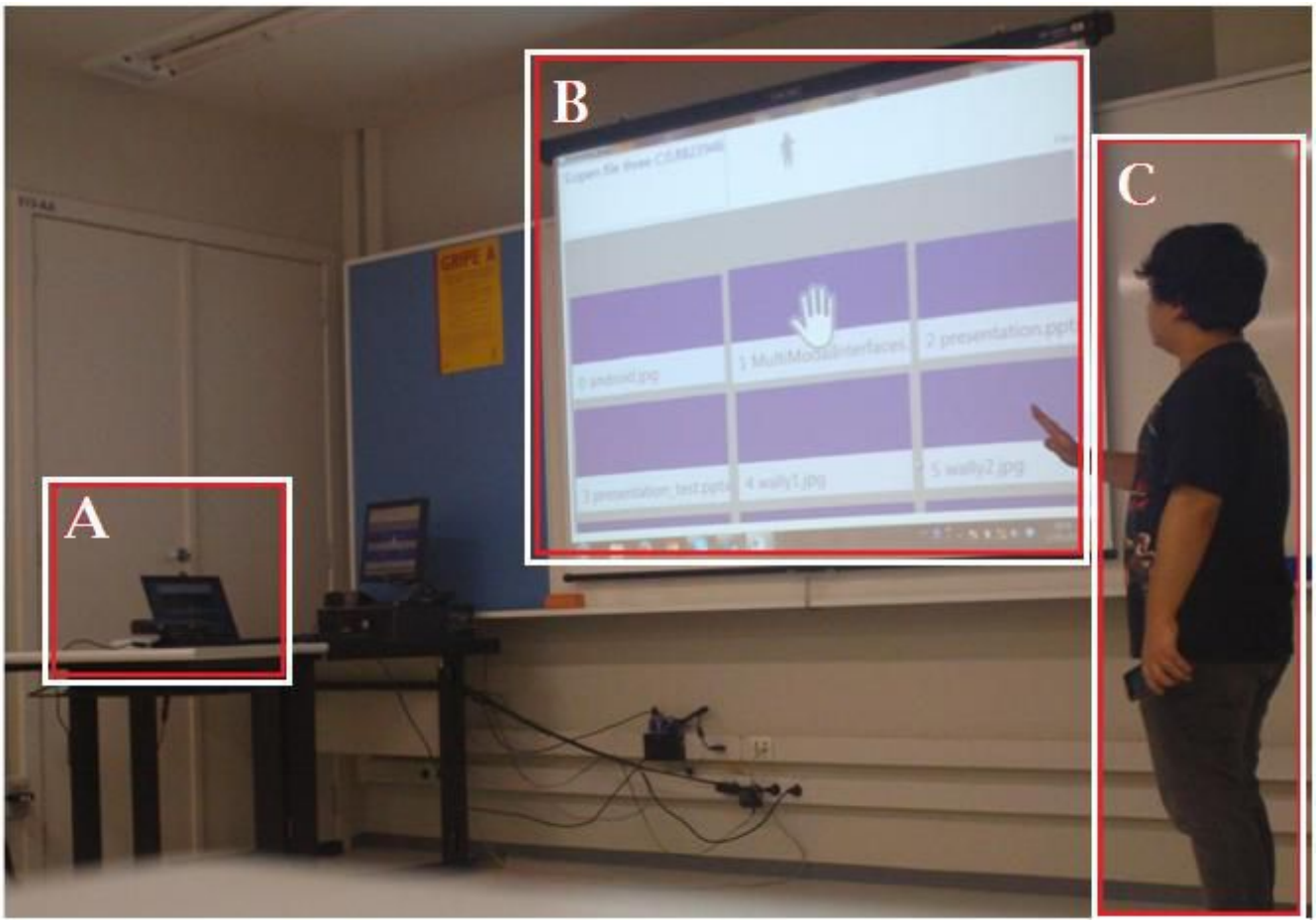


Figura 12 - Imagem da Configuração do Sistema no Ambiente Testado.

5.2.1 Servidor

Os componentes principais de interface são apresentados na Figura 13.O sistema do servidor é um único processo com diferentes *threads*, uma para o serviço REST, e uma para cada interface gráfica (Tela de imagem, Tela de Arquivos, e Interface Kinect). Ainda, é feita uma comunicação com o software Power Point, instalado na máquina, através da biblioteca *Microsoft Office Interop* versão 12.

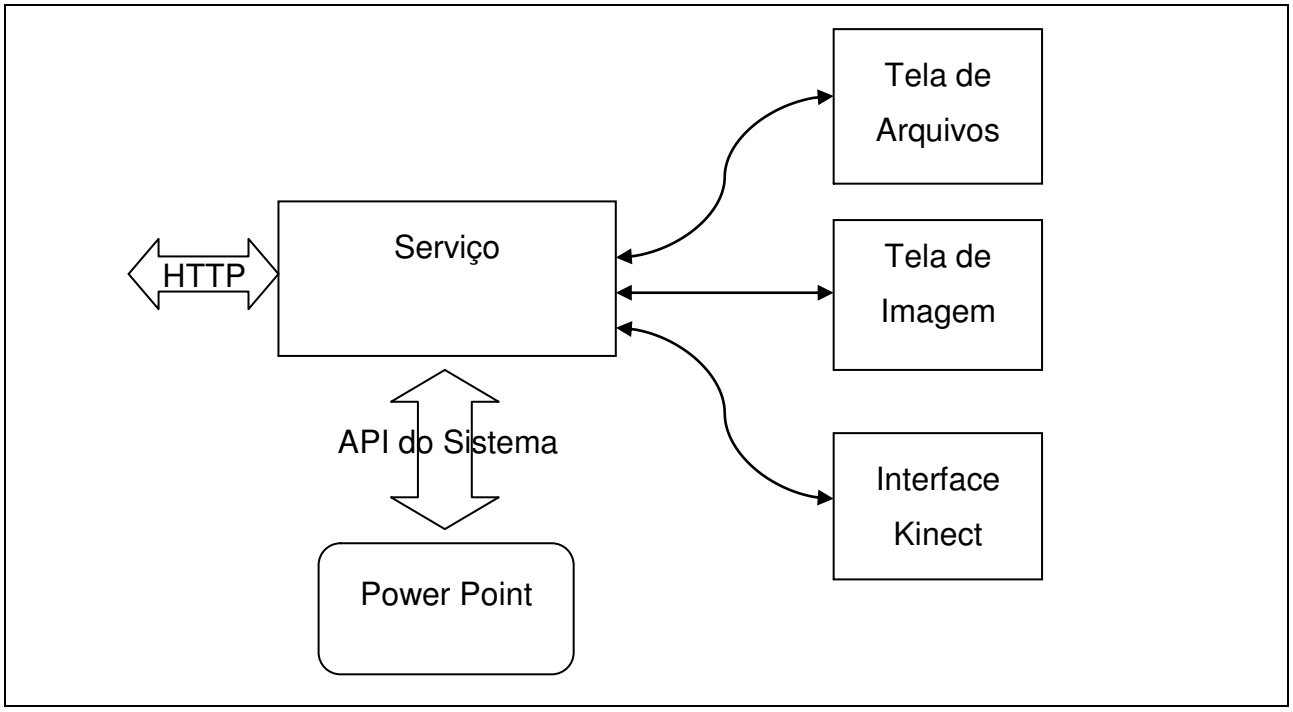


Figura 13 - Componentes de Controle e Execução de Eventos no Servidor.

Na tela inicial de escolha de arquivos, a interface apresentada é a Tela de Arquivos. Ela é responsável por receber eventos de gestos de corpo e fala do Kinect, e também apresentar *feedback* destes eventos, além de mostrar os arquivos acessíveis pelo sistema.

Uma vez que o usuário tenha aberto um dos arquivos por meio do uso de gestos de corpo ou fala, a Tela de Arquivos dispara um método de execução do componente Serviço, assincronamente, requisitando que o arquivo escolhido seja aberto. O Serviço é responsável por abrir o arquivo desejado, iniciando a tela ou o programa apropriado para o mesmo.

No caso de uma apresentação de slides, o componente do Serviço utiliza a API do sistema para controlar o programa Power Point, requisitando que este carregue e inicialize a apresentação do arquivo de slides desejado. No caso da escolha de uma imagem, o Serviço inicializa a Tela de Imagem, enviando o caminho do arquivo desejado. Para ambos os casos, seja quando uma apresentação de slides ou uma imagem estão sendo abertos, uma janela auxiliar é inicializada, responsável por fornecer a interface com o dispositivo Kinect durante as tarefas de apresentação, a Interface Kinect. Tal interface foi criada para prover um *feedback* adicional dos gestos de corpo e fala, feedback este que não seria oferecido uma vez que a aplicação Power Point fosse inicializada. Tal janela fica sobreposta a todos os outros componentes da tela do sistema, e foi utilizada também para o caso de apresentação

de uma imagem, sendo responsável, ainda, por receber os eventos de entrada e identificar os gestos de corpo ou fala desejados.

Durante o modo de apresentação, portanto, eventos de gestos de corpo ou fala são capturados pela Interface Kinect, e enviados para o Serviço. O Serviço é então responsável por executar as funções adequadas, comunicando-se com a Tela de Imagem para atualizar a mesma, ou enviando comandos ao Power Point.

Um processo similar é seguido na manipulação de eventos que surgem do *smartphone*. O Serviço recebe os eventos de requisições HTTP apropriados para abrir um arquivo desejado, e inicializa as janelas apropriadas. No caso de eventos de controle da apresentação, as requisições HTTP são redirecionadas da mesma forma para o controle dos componentes envolvidos.

Quando uma apresentação está ativa, e um comando é executado através de gestos de corpo ou fala, o *smartphone* precisa ser avisado de que uma mudança de estado ocorreu para atualizar sua tela com as informações corretas da apresentação. Para que isto aconteça, foi criado um simples sistema de notificação no servidor. O *smartphone* uma vez que abra uma apresentação, ou entre em sincronização com o servidor, através da opção de sincronização da tela principal, envia um pedido para registro de notificações com o servidor e inicializa um pequeno servidor HTTP rodando na porta 8080. Quando novos eventos de atualização ocorrem na apresentação corrente, que não tenham sido executados pelo IP registrado para receber notificações, uma requisição é enviada ao endereço registrado, de forma a avisar que uma atualização do estado é necessária. O *smartphone* pode também requisitar a remoção da lista de notificações.

5.2.1.1 Telas do Sistema

A Tela de Arquivos é apresentada na Figura 14, com os principais elementos marcados. Essa tela foi aproveitada de um exemplo advindo no SDK do Kinect (ControlsBasics – WPF). O elemento ‘A’, localizado no campo superior esquerdo da tela, apresenta a região de *feedback* de fala, com as hipóteses de reconhecimento de fala que o Kinect está tentando identificar, indicando o grau de confiança ao lado. Para a implementação corrente foi estipulado um grau de confiança de 0,7 para aceitar a sentença. O elemento ‘B’ apresenta um pequeno *feedback* do reconhecimento do usuário pelo Kinect, mostrando sua silhueta.

Quando o usuário está com a mão ativa, isto é, em determinada altura e em movimento sobre o elemento de interação da tela, um ícone de uma mão é apresentado e se move em conjunto com o movimento da mão ativa do usuário, elemento ‘C’. Por fim, a tela apresenta cada arquivo acessível pelo sistema, simbolizados por botões, com exemplo de um deles destacado como elemento ‘D’, com o nome do arquivo e uma numeração ao seu lado.

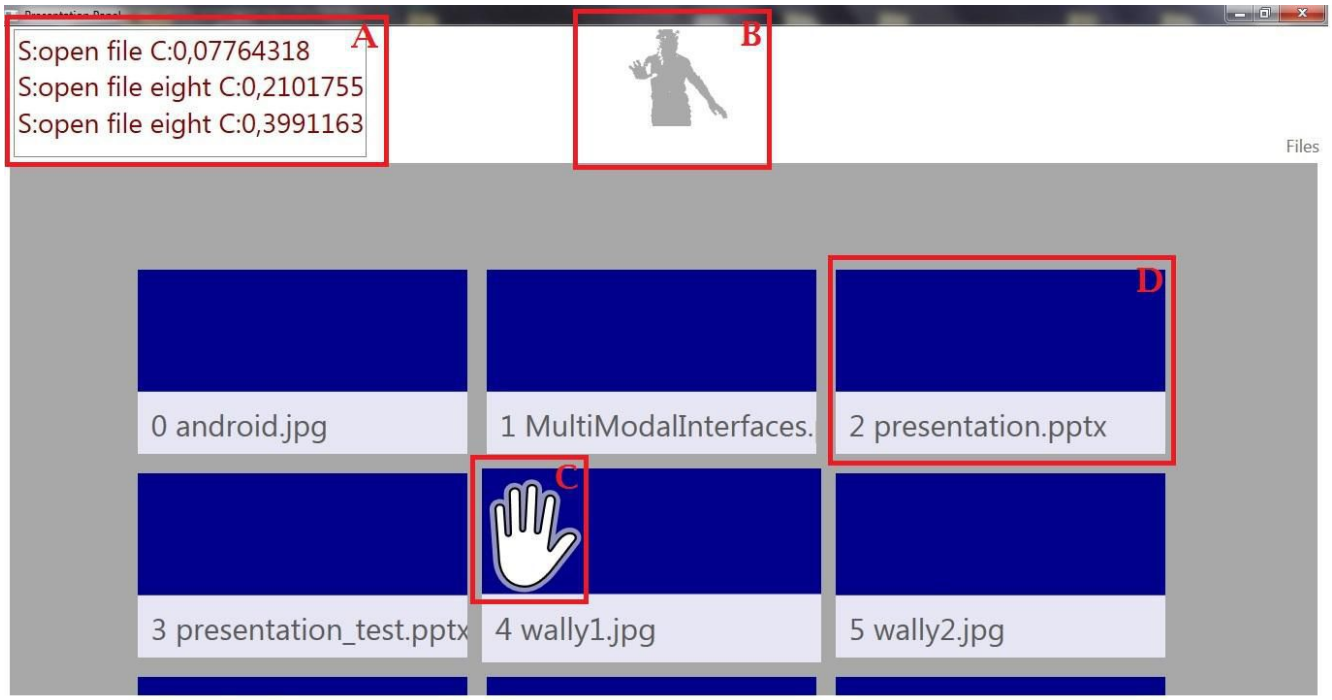


Figura 14 - Tela de Arquivos e Seus Diferentes Elementos.

A janela auxiliar de Interface Kinect pode ser visualizada na Figura 15 e na Figura 16. Quando uma apresentação de slides é inicializada (Figura 15), o slide corrente é mostrado no centro da tela (região ‘A’) e a Interface Kinect fica localizada na parte inferior direita da tela. A Interface Kinect apresenta um *feedback* da câmera em cores, sinalizando pequenos círculos vermelhos nos principais pontos do esqueleto do usuário, acima da cintura, que estão sendo identificados (região ‘B’). Logo abaixo, na região ‘C’, é reservado um espaço para as hipóteses e sentenças de fala a serem identificadas, que assim como na Tela de Arquivos, mostram o grau de confiança de reconhecimento. Quando uma imagem é aberta (Figura 16), a Tela de Imagem é inicializada (região ‘A’), e em conjunto da mesma, a Interface Kinect (região ‘B’). A Figura 16apresenta também o *feedback* dado quando as mãos do usuário são identificadas como fechadas, sinalizando as mesmas com um círculo amarelo.

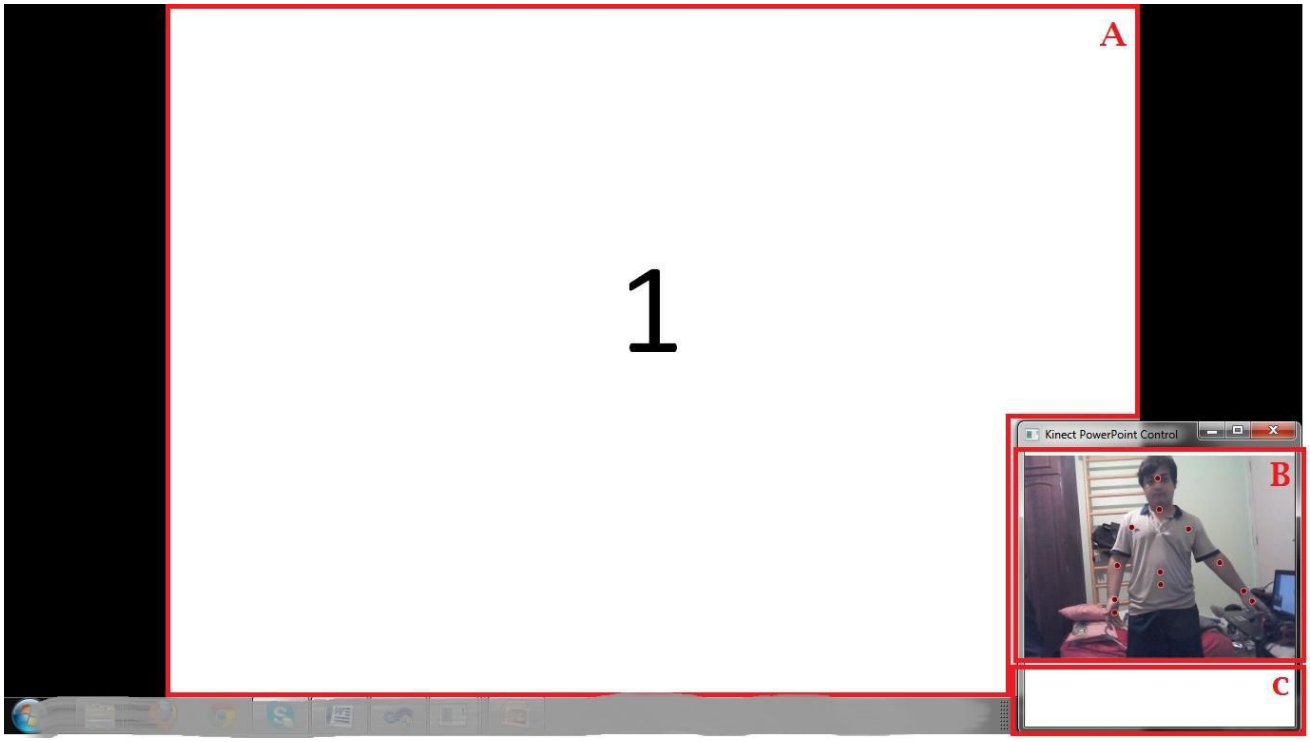


Figura 15 - Apresentação de Slides e Interface Kinect.

Figura 16 - Tela de Imagem e Interface Kinect.

5.2.1.2 API REST

O servidor inicializa um serviço WCF com comandos similares a uma arquitetura REST [44], e portanto possibilitam a comunicação externa através do protocolo HTTP. A lista de comandos disponíveis para serem utilizados, com objetivo de permitir a interação pelo *smartphone*, é explicada a seguir.

A Tabela 13 mostra os métodos disponíveis para controle de arquivos, e registro de notificações. Para que o *smartphone* receba atualizações quando mudanças ocorrem durante uma apresentação, por comandos disparados por gestos de corpo ou fala, ele deve requisitar o registro como um ouvinte da apresentação. O servidor registra o IP da requisição e supõe que o ouvinte registrado estará aguardando por requisições HTTP na porta 8080. Assim, para cada nova atualização necessária, o servidor enviará uma requisição HTTP para o IP do ouvinte cadastrado, indicando que uma atualização do estado da apresentação é necessária. Comandos de listagem de arquivos, envio de arquivo, ou recuperação de um arquivo específico estão disponíveis.

A Tabela 14 mostra os comandos relacionados à manipulação da apresentação de slides. Um comando é necessário para carregar o arquivo no servidor, para preparação inicial. Este comando leva em conta a abertura do mesmo pelo aplicativo Power Point, e pode demorar alguns segundos dependendo do tamanho do arquivo. Para gerar uma visualização do arquivo antes de inicializar uma apresentação, imagens de cada slide são geradas e acessíveis no servidor, podendo ser recuperadas conforme necessário. A sincronização do estado atual pode ser feita através da requisição de um dos comandos que retorna as informações do estado corrente da apresentação. A apresentação pode ser inicializada assim quando desejado, e comandos específicos podem ser enviados para controle da mesma.

A Tabela 15 apresenta os comandos disponíveis para manipulação da apresentação de imagem. Uma apresentação pode ser aberta utilizando o nome de algum arquivo presente no servidor. A imagem corrente que está sendo apresentada pode ser retornada, e o estado da mesma para atualização da aplicação remota. Um comando para mudanças de estado da imagem também é disponibilizado.

Detalhes específicos de como devem ser descritos os parâmetros das ações e dos comandos em geral não são apresentados aqui, mas todos utilizam um formato de representação JSON.

Tabela 13 - Comandos REST de Acesso aos Arquivos e Notificações.

Caminho (Path)	**Método**	**Parâmetros**	**Descrição**
/listeners	POST	Modo de apresentação: slides ou imagem.	Registra um novo endereço IP como ouvinte da apresentação.
/listeners	DELETE	-	Remove IP da lista de ouvintes.
/files	GET	Opcionalmente pode incluir o parâmetro *type* para filtrar por arquivos de imagem ou slides, incluído no *path*.	Retorna a lista de arquivos existentes no servidor.
/files/{fileName}	PUT	Nome do arquivo incluído no caminho, e *Stream* de dados.	Adiciona ou atualiza um novo arquivo no servidor, com determinado nome.
/files/{fileName}	GET	-	Recupera o arquivo desejado existente no servidor.

Tabela 14 - Comandos REST para Controle da Apresentação de Slides.

Caminho (Path)	**Método**	**Parâmetros**	**Descrição**
/presentation/prepare	PUT	Nome do arquivo a ser aberto.	Carrega o arquivo requisitado para apresentação (a apresentação não é inicializada por este comando).
/presentation	PUT	-	Inicializa a apresentação do arquivo carregado anteriormente.
/presentation/action	PUT	Ação a ser realizada na apresentação, e um argumento opcional.	Função que permite avançar slide, voltar slide, ir para determinado slide, ou fechar a apresentação.
/presentation/slides/{slideNumber}	GET	Número do slide a ser retornado,diretamente no *path*.	Retorna a imagem que representa o slide requisitado.
/presentation/info	GET	-	Retorna o estado atual da apresentação de slides, nome do arquivo, slide atual, e total de slides.

Tabela 15 - Comandos REST para Controle da Apresentação de Imagem.

Caminho (Path)	**Método**	**Parâmetros**	**Descrição**
/image	GET	-	Retorna a imagem corrente que está aberta no servidor.
/image	PUT	Nome do arquivo de imagem.	Abre a imagem requisitada.
/image/action	PUT	Comando e parâmetros a serem executados.	Executa uma determina ação sobre a imagem, como mover, rotacionar, ou aplicar zoom, utilizando os parâmetros enviados para os valores a serem executados.
/image/info	GET	-	Requisita informações do estado da imagem, como rotação, e porção da imagem que está sendo mostrada no servidor.

5.2.2 Cliente Android

As camadas da aplicação Android, que é executada no *smartphone,* são apresentadas na Figura 17. O Android define classes de interfaces do tipo *Activity* que são responsáveis por controlar os componentes gráficos da aplicação e os eventos que ocorrem sobre estes. Essa camada de interface da aplicação desenvolvida é responsável por registrar objetos para identificação dos gestos desejados, e ligar seus eventos de reconhecimento com requisições HTTP para o servidor, resultando no disparo dos comandos apropriados.

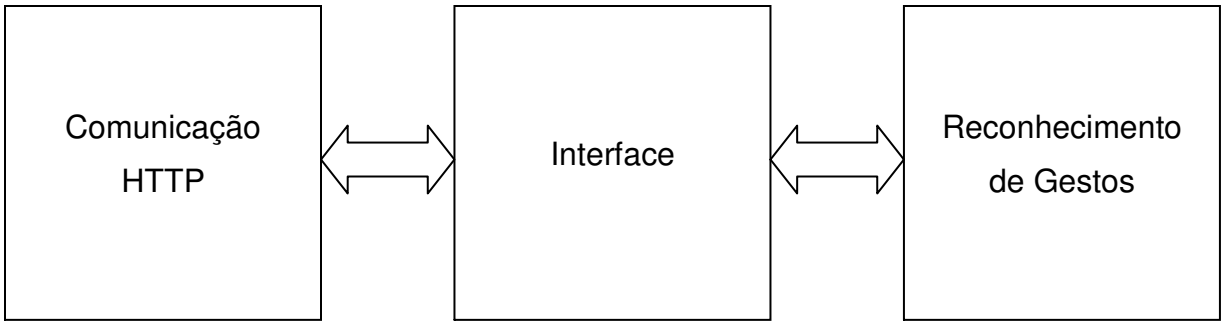
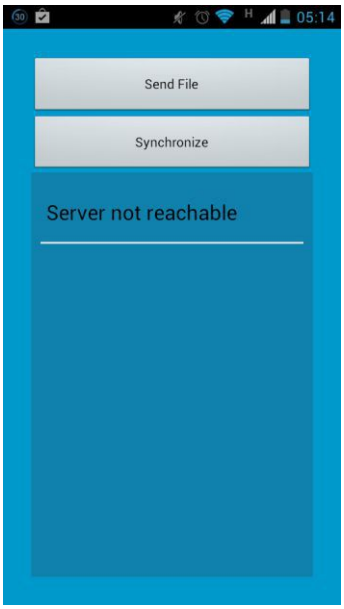
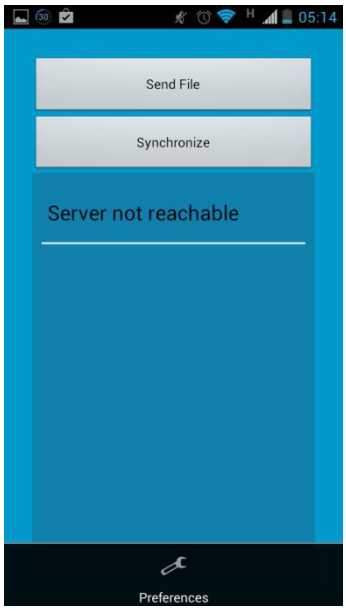


Figura 17 - Componentes do cliente Android.

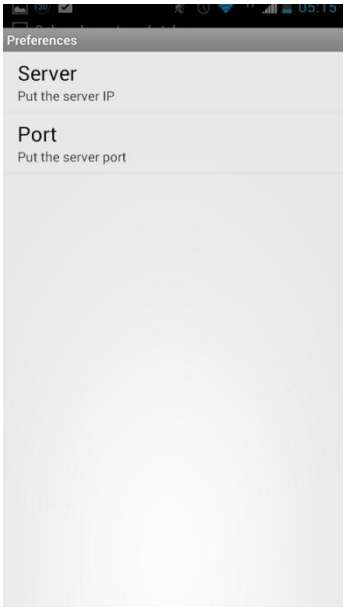
A tela principal da aplicação lista os arquivos disponíveis no servidor. É possível configurar o IP do servidor através do menu de opções (Figura 18). Existe também a possibilidade de enviar um novo arquivo ao servidor, que esteja contido ou seja acessível pelo *smartphone* (Figura 19). Ainda, uma opção de sincronizar o estado com o servidor é disponibilizada, de forma que, caso um arquivo já esteja em apresentação pelo mesmo, o *smartphone* atualize a tela com o estado atual do sistema.



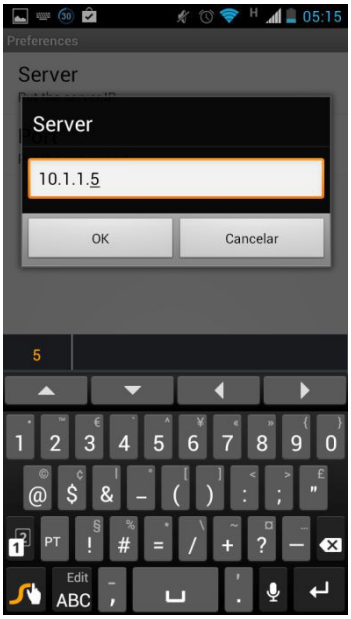
Tela inicial sem conexão com o servidor.



O menu de opções aberto.

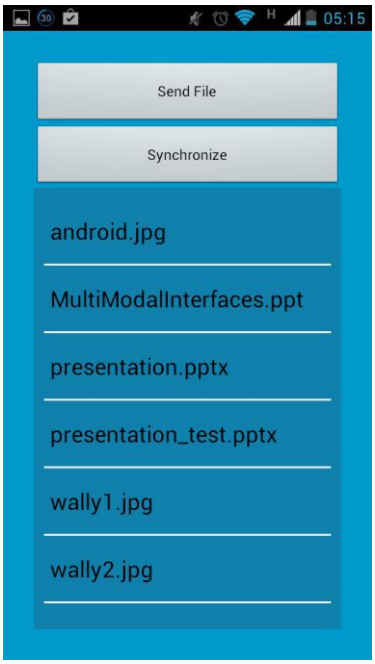


Tela de configurações com opção de IP e porta do servidor.

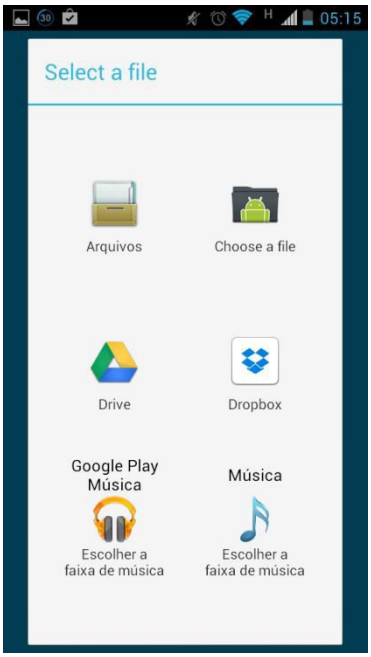


Configuração do IP do servidor.

Figura 18 - Tela inicial e Configuração do Endereço do Servidor.



Tela inicial com lista de arquivos do servidor.



Janela de opções para seleção de arquivo (mostrada a partir do botão “Send File”).

Figura 19 - Tela inicial Com Lista de Arquivos Existentes e Opção de Enviar um Novo.

Uma vez que um arquivo da lista é selecionado, uma requisição de abertura é enviada ao servidor. No caso de uma apresentação de slides, o *smartphone* requisita ao servidor a abertura e recupera o conjunto de imagens que representam os slides para serem mostrados em uma galeria na tela seguinte. A necessidade de transformar o arquivo em um conjunto de imagens teve que ser utilizada uma vez que não foi encontrada solução de abrir o arquivo diretamente no *smartphone*. Nessa tela o usuário pode navegar entre os slides (Figura 20); quando uma imagem é aberta, uma nova tela é apresentada com a imagem centralizada (Figura 21). O usuário pode manipular a mesma através dos gestos necessários. Foi optado por não permitir a rotação da imagem quando ela já estivesse ampliada em qualquer quantidade, pois o tratamento da imagem para a correta visualização em casos de rotação sobre um ponto que não fosse o central não foi implementado.

Imagem do primeiro slide da apresentação.

Imagem do segundo slide da apresentação.

Opção de *close* aberta.

Figura 20 - Apresentação da Imagem dos Slides na Tela.

Toda vez que uma apresentação é aberta ou quando o usuário requisita a sincronização com o servidor, durante uma apresentação ativa, a aplicação Android requisita o registro no sistema de notificações do servidor e inicializa um pequeno servidor HTTP local na porta 8080. Quando alguma requisição advinda do servidor é recebida, a aplicação requisita a atualização dos dados da apresentação ativa e atualiza a tela conforma necessário.

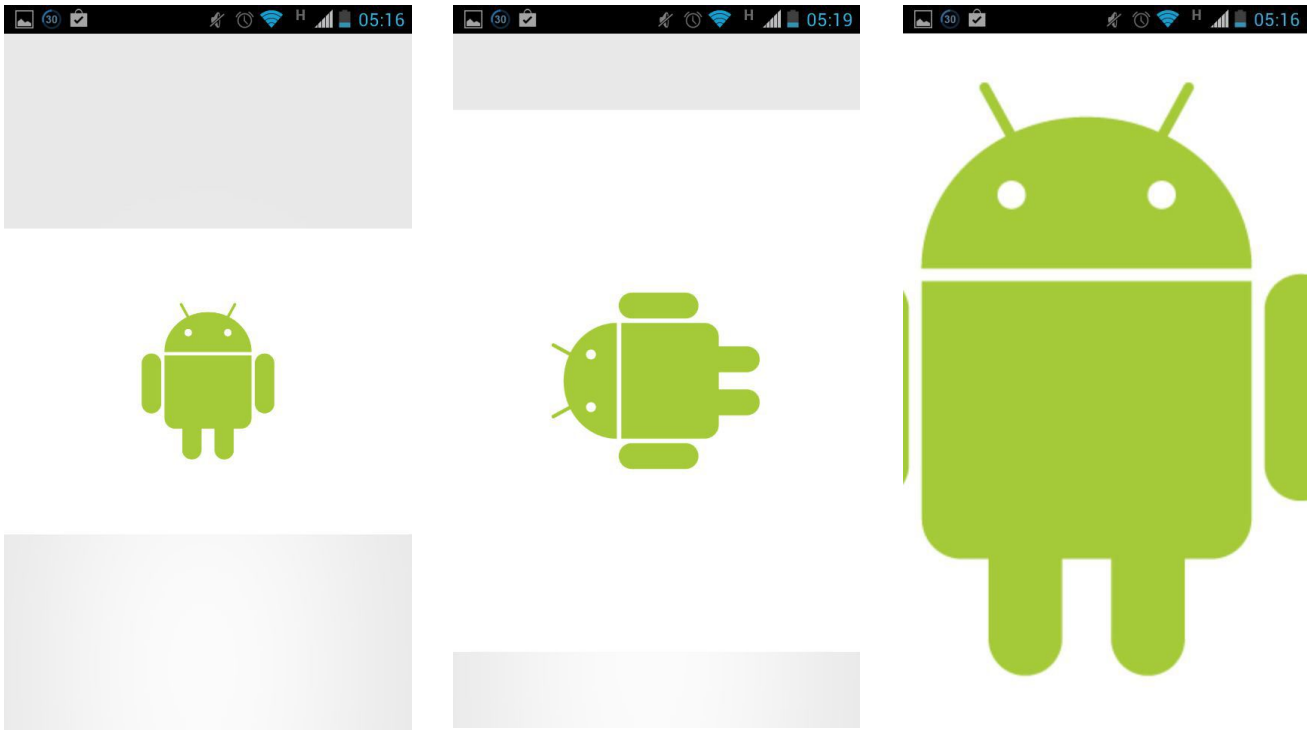


Imagem aberta. Imagem rotacionada. Imagem ampliada.

Figura 21 - Apresentação de Imagem em Andamento.

5.3 **Implementação de Gestos de Corpo**

Para identificação de gestos do Kinect, foi utilizada uma lógica simples de estados entre diferentes poses dadas pelo posicionamento dos pontos do corpo do usuário.

Como apresentado na Figura 22, o dispositivo Kinect é capaz de identificar 20 pontos do esqueleto do usuário. Além disso, introduzidos na versão 1.7 do SDK, é possível identificar novos estados das mãos dos usuários, como eventos de quando a mão é fechada, quando a mão é aberta, e a extensão de pressionamento (empurrar mão para frente, que identifica a mudança de distância entre a mão e o sensor) que está sendo feito pela mão.

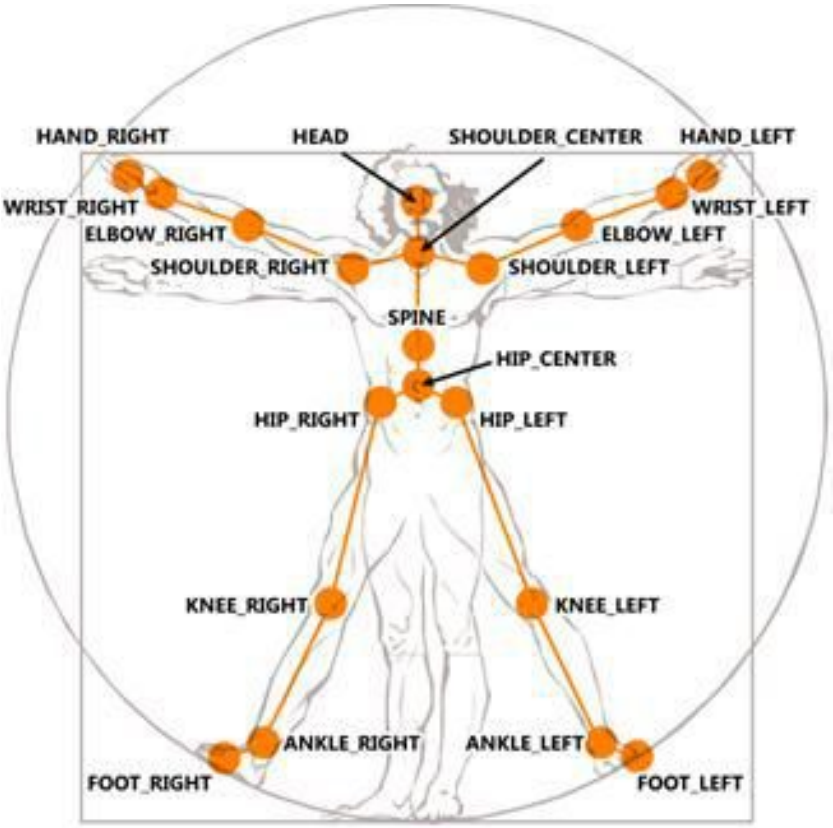


Figura 22 - Vinte Pontos do Esqueleto Reconhecidos pelo Kinect e Seus Identificadores. (Retirado de [56])

Para identificação de gestos, a cada *frame* as informações do estado do esqueleto do usuário são atualizadas e os objetos responsáveis pelo reconhecimento de cada possível gesto executam uma verificação para atualizar seu estado interno e determinar se o gesto foi executado.

Por exemplo, para identificar o gesto de avançar slide, foi implementado o código apresentado na Figura 23. No código apresentado, o estado de identificação inicia em zero. Quando o usuário está com a mão acima da altura do ponto da coluna (*spine*), e fechada, a posição inicial é capturada, avançando o estado para um. Caso a mão desloque-se pelo menos 15 centímetros para a esquerda, e esteja aberta, ou dez centímetros abaixo da posição inicial, a função retorna *true* identificando que o gesto foi reconhecido. O nome do gesto identificado fica em uma propriedade da classe, e outras variáveis servem para manter o estado interno entre diferentes frames.

```
Public bool IdentifyGesture(UserSkeletonState userState)
{
  ISkeleton skeleton = userState.Skeleton;
  var rightHand = skeleton.HandRight;
  var spine = skeleton.Spine;
  if (State == 1 && GestureUtils.HasMovedToLeft(initialPosition,
rightHand.Position, minimumDistanceToTrigger) &&
GestureUtils.IsHandBelow(rightHand, initialPosition, heightToReset))
            {
                State = 0;
                return true;
            }
  else if (GestureUtils.IsHandBelow(rightHand, spine))
            {
                State = 0;
                return false;
            }
  if (State == 0 && userState.IsRightHandGripped )
            {
                State = 1;
                initialPosition = rightHand.Position;
                return false;
            }
  if (State == 1)
            {
                SkeletonPoint nextPoint = rightHand.Position;
                if ((GestureUtils.HasMovedToLeft(initialPosition, nextPoint,
minimumDistanceToTrigger) && !userState.IsRightHandGripped))
                {
                    State = 0;
                    return true;
                }
  else  if (!userState.IsRightHandGripped)
                {
                    State = 0;
                    return false;
                }
                return false;
            }
  State = 0;
  return false;
}
```

Figura 23 - Exemplo de Código para Avançar Slide.

A identificação de outros gestos segue uma implementação similar sendo necessário definir diversas características de estado do gesto e manter mudanças no tempo. Métodos e propriedades com a mesma assinatura são mantidos entre diferentes classes

reconhecedoras de gestos, possibilitando adicionar e remover instâncias desses objetos facilmente em uma lista que recebe atualização do esqueleto pela aplicação.

Na etapa de codificação era necessário definir os limites de deslocamentos para ativação dos gestos. Os principais valores utilizados, e os gestos afetados, são descritos na Tabela 16.

Tabela 16 - Principais Valores de Distância Utilizados na Implementação dos Gestos.

Comando	**Valores Considerados**
Mover Imagem	Considera ativação quando a posição da mão muda pelo menos cinco centímetros. Mão esquerda deve estar atrás da direita, pelo menos metade da distância entre mão direita e coluna (*spine*).
Fechar	Mãos iniciam movimento quando estiverem em uma distância maior que a distância entre os ombros (*shoulder_left* e *shoulder_right*) do usuário mais 15 centímetros.
Avançar/Voltar Slides	É necessária a movimentação de pelo menos 15 centímetros para um dos lados. Gesto termina quando mão é aberta, ou quando ela é abaixada dez centímetros (mão ainda fechada).
Rotação	Inicia com as mãos afastadas pelo menos 12 centímetros de altura. Mãos devem estar próximas em Z pelo menos metade da distância entre a mão mais a frente e a coluna (*spine*). É ativado quando ângulo muda 60 graus, ou quando mãos trocam de altura (mão que iniciou acima vai para baixo).
Zoom	Mãos devem estar próximas pelo menos 12 centímetros de altura. Mãos devem estar próximas em Z pelo menos metade da distância entre a mão mais a frente e a coluna (*spine*). Ativa mudança de zoom quando mãos mudam pelo menos cinco centímetros de distância.

De forma geral, era estipulado também que os gestos deviam ser executados com as mãos acima da altura do ponto da coluna (*spine*), e não muito próximas ao corpo.

5.4 **Implementação de Gestos de Toque**

A página de desenvolvedores para Android apresenta informações sobre como implementar gestos [4]. O Android disponibiliza detectores para gestos como *scale*, *fling* e *scroll.* Entre os gestos necessários para implementação apenas foi necessário interpretar as mudanças de posição dos dedos para disparar eventos ao servidor nos intervalos e com parâmetros desejados, e implementar um detector de rotação entre dois dedos, não disponível diretamente pelo SDK.

Um exemplo de implementação do controle do gesto de arrastar para iniciar e fechar uma apresentação é explicada a seguir, com parte do código ilustrado na Figura 24. Um objeto do tipo *TouchListener* é vinculado ao objeto de interface desejado, para receber todos os eventos ocorridos sobre este. Este objeto recebe os eventos de uma forma mais crua e sem tratamento. Dentro dele é instanciado um objeto do tipo *GestureDetectorCompat*, utilizado para filtrar os eventos de gestos em eventos mais simplificados, avisando seu ouvinte, neste caso, o membro *flingListener*.

O método *onScroll* do objeto ouvinte é disparado sempre que o usuário executa um movimento com seu dedo enquanto este está em contato com a tela, para qualquer direção. Para o movimento de arrastar para cima e para baixo, estamos interessados apenas na distância Y percorrida, para atualizar a altura do objeto de imagem. Por fim, para disparar a real abertura ou encerramento da apresentação, um método especial é chamado pelo *TouchListener*, quando o dedo do usuário é levantado e perde o contato com a tela. Neste momento a altura resultante do movimento é comparada para verificar a necessidade de envio de algum comando ao servidor, e reiniciar a altura do objeto de imagem.

Dentro do Objeto de Interface Desejado

```
final GestureDetectorCompatsimpleGesturesDetector = new
GestureDetectorCompat(this, this.flingListener);
viewPager.setOnTouchListener(new OnTouchListener()
{
@Override
public boolean onTouch(View v, MotionEvent event)
      {
        simpleGesturesDetector.onTouchEvent(event);
        int action = MotionEventCompat.getActionMasked(event);
        switch(action) {
            caseMotionEvent.ACTION_UP:
            flingListener.onFingerUp();
            break;
        }

        if(viewPager.getScrollY() != 0)
        return true;

        return viewPager.onTouchEvent(event);
      }
});
```

Dentro do Objeto Listener (Classe do membro flingListener acima)

```
@Override
public boolean onScroll(MotionEvent e1, MotionEvent e2, float distanceX,
                  float distanceY) {
  if(Math.abs(distanceX) > Math.abs(distanceY)) return false;
  this.updatePositionY(viewPager.getScrollY(),
viewPager.getScrollY()+distanceY);
  return true;
}
```

Figura 24 - Código de Exemplo que Registra um Detector de Gestos para Controle do Evento de Arrastar.

5.5 Implementação de Comandos de Fala

A definição de comandos de fala foi feita utilizando-se arquivos XML com as sentenças, seguindo o formato da gramática SRGS 1.0 (*Speech Recognition Grammar Specification Version* 1.0) [52,92]. Uma vez compreendido o formato de especificação da gramática, foram criados os arquivos necessários para os diferentes cenários, e o código de disparo dos eventos para cada um destes. A Figura 25 demonstra um exemplo da gramática utilizada para controle da apresentação de slides. As sentenças definidas na gramática geram uma saída representando sua semântica, permitindo que mais de uma opção de fala gere a

mesma saída semântica. Essa saída é interpretada em código para disparar o comando desejado.

```
<?xmlversion="1.0"encoding="UTF-8" ?>
<grammarversion="1.0"xml:lang="en-US"mode="voice"root= "Expression"
xmlns="http://www.w3.org/2001/06/grammar"tag-format="semantics/1.0">
<ruleid="Expression"scope="public">
<one-of>
<item>
<rulerefuri ="#Forward"type="application/srgs+xml"/>
<tag>out.command=rules.latest();</tag>
</item>
<item>
<rulerefuri ="#Backward"type="application/srgs+xml"/>
<tag>out.command=rules.latest();</tag>
</item>
<item>
<rulerefuri ="#Close"type="application/srgs+xml"/>
<tag>out.command=rules.latest();</tag>
</item>
</one-of>
</rule>
<ruleid="Forward">
<one-of>
<item>
        next <tag>out = "next slide"; </tag>
</item>
<item>
        forward <tag>out = "next slide"; </tag>
</item>
</one-of>
</rule>
<ruleid="Backward">
<one-of>
<item>
        previous <tag>out = "previous slide"; </tag>
</item>
<item>
        back <tag>out = "previous slide"; </tag>
</item>
</one-of>
</rule>
<ruleid="Close">
<one-of>
<item>
        close file <tag>out = "close presentation"; </tag>
</item>
</one-of>
</rule>
</grammar>
```

Figura 25 - Exemplo de Gramática para Identificação dos Comandos de Controle de Apresentação.

Alguns testes iniciais demonstraram que algumas vezes em que havia apenas ruídos, sentenças não existentes eram proferidas, ou sentenças incompletas eram utilizadas, mesmo

assim o sistema acabava identificando algumas opções válidas com alto grau de confiança. Tal problema é também relatado em outras discussões do fórum oficial [57,58,59]. Recomendações gerais para reconhecimento de fala são: manter desabilitado o ganho automático; manter o cancelamento de ruído ativo (configurações padrões); e construir gramáticas com sentenças compostas de pelo menos duas palavras.

5.6 **Decisões de Design**

Nesta seção algumas considerações gerais sobre as decisões tomadas durante a implementação do sistema são apresentadas.

5.6.1 Feedback

O estado de implementação aqui apresentado possui algumas limitações. O objetivo principal neste trabalho era o foco nas modalidades de entrada, no entanto, a forma de apresentação dos dados na tela é uma característica fundamental para uso do sistema e pode influenciar a interação com o mesmo. Neste sentido, o *feedback* implementado aqui foi criado para cumprir o objetivo de avaliar os modos de entrada da aplicação, uma vez que a tela auxiliar com a imagem colorida da câmera poderia distrair ou atrapalhar o público da apresentação, em um contexto real. Foram pensadas em duas possíveis alternativas para melhorar o *feedback* do sistema, focado em um uso real, e portanto uma atualização da implementação atual, que não foram utilizadas devido ao tempo disponível, mas ficam aqui como referência para sugestões futuras.

A primeira alternativa é o uso constante de uma tela auxiliar (monitor), aonde o *feedback* dos gestos poderiam ser mostrados somente ao apresentador, ou pelo menos com este intuito, por exemplo, com tal *feedback* estando presente no computador ou notebook, e a tela de apresentação estar no projetor.

Uma segunda alternativa para melhorar o *feedback* atual, é a modificação do código para permitir que o reconhecimento de gestos esteja diretamente ligado ao *feedback* sobre os objetos manipulados, similar ao conceito da Tela de Arquivos. Esta tela principal de arquivos foi extraída dos exemplos do SDK do Kinect, e demonstra o uso da tecnologia WPF utilizando objetos de interação criados mais especificamente para o Kinect. Eles permitem

que pequenas alterações do movimento da mão do usuário, sejam refletidos rapidamente em *feedback* na tela, e a posição da mão influencie os objetos contidos nessa área de interação.

Durante o desenvolvimento do sistema, a segunda alternativa foi analisada, e poderia permitir que apenas ícones representando as mãos dos usuários pudessem ser desenhados sobre os slides da apresentação, por exemplo, de forma similar a Tela de Arquivos. Para isto, era necessário o uso de uma tela WPF contendo a apresentação Power Point. Visto que o programa Power Point é externo à aplicação, a melhor solução parece ser converter os slides em imagens, e assim formar uma galeria de imagens como apresentação, de forma similar ao utilizado no *smartphone*, eliminando, portanto, a interação com o Power Point para o controle da apresentação. Uma solução parecida poderia ser utilizada para o modo de apresentação de imagem. No entanto, devido ao tempo disponível foi optado por não fazer tal modificação que poderia demandar muito tempo para alteração.

5.6.2 Implementação dos Gestos de Corpo

A etapa de definição da interação surgiu pela análise de vídeo e identificação das posições gerais dos gestos. Muitos deles envolvem movimentação das mãos, e na codificação é necessário definir os limites de deslocamento que ativam determinado comando. A implementação atual utilizou de valores resultantes da decisão do pesquisador baseada em testes subjetivos e considerações das etapas anteriores. Idealmente, um refinamento anterior e posterior desses valores deveria ser feito, e um *feedback* condizente para avisar ao usuário dos detalhes da forma de execução dos gestos

Uma limitação presente no estado de desenvolvimento atual foi o uso exclusivo da mão direita para os gestos de mover imagem, e avançar/voltar slides.

5.6.3 Limitações de Funcionalidades

Não foi considerado, desde o início, que estariam presentes animações nas apresentações. Embora elas possam existir na apresentação e, ainda assim, o sistema funcione, a interação, neste caso, para avançar cada etapa de animação, provavelmente seria considerada muito mais enfadonha. Uma análise específica para este caso deveria ser estudada anteriormente.

Na etapa de teste com usuários foi apontado por um dos mesmos que não havia comando para movimentar a tela de arquivos através da fala. O motivo foi por não ter sido percebida a

necessidade nas etapas iniciais. Embora ele seja fundamental para permitir uma interação com o sistema completa para a fala, sem necessitar de uso de outra modalidade, para o cenário de teste este comando acabou não sendo necessário, uma vez que a quantidade de arquivos apresentados era pequena e visível na tela.

6 AVALIAÇÃO DO SISTEMA

Neste capítulo são descritas as etapas executadas e os resultados obtidos da avaliação do sistema desenvolvido, para compreender e comparar a satisfação dos usuários no uso das diferentes modalidades para a tarefa definida.

6.1 **Procedimento**

Os procedimentos executados para os testes com usuários para compreender e comparar a satisfação de uso dos modos do sistema foram os seguintes:

1. Introdução ao objetivo do trabalho ao participante e apresentação e assinatura do termo de consentimento livre e esclarecido (Anexo A);
2. Questões abertas sobre o perfil do participante quanto à sua familiaridade com as tarefas possíveis do sistema e uso de sistemas computacionais com tecnologias similares as utilizadas pelo sistema:
    a. Qual sua experiência e opinião sobre dispositivos de telas de toque? Possui *smartphone*, *tablet*, ou outros dispositivos deste tipo?
    b. Qual sua experiência e opinião sobre a interação por gestos de corpo? Quais sistemas já utilizou deste tipo? Exemplos: *Wii*, *Kinect*, *PS3 Move*, *Smart TVs*. Quais aplicações utilizou?
    c. Qual sua experiência e opinião sobre sistemas que possibilitam o uso de comandos de fala? Quais aplicações deste tipo utiliza ou já utilizou?
    d. Com que frequência você realiza aulas, ou faz apresentações em eventos? Descreva brevemente a forma que são organizadas e executadas suas apresentações. Exemplos: (1) Utiliza um notebook com Power Point e avança e retrocede slides utilizando os botões direcionais do notebook. (2) utiliza um programa de apresentação de slides no notebook e controla a apresentação através de um controle Bluetooth.
3. Demonstração dos comandos do sistema e uso de cada um dos modos com objetivo de ensinar os participantes a como utilizá-los. Uma folha com o resumo da interação com o sistema foi disponibilizada e o entrevistador auxiliou ativamente na etapa de

aprendizado. A ordem de apresentação dos comandos foi dividida pelos modos e pelas partes do sistema. Primeiro os comandos do *smartphone*, para abrir e fechar arquivos, comandos da imagem, e comandos de slides, foram apresentados. Em seguida o mesmo foi feito para o modo de gestos de corpo, e por último para o modo de fala;

4. Execução, por parte do participante, de cada um dos possíveis comandos do sistema para cada um dos modos, com objetivo de familiarizá-lo com o uso do sistema.Foram executados todos os comandos pelo menos uma vez com uso de um arquivo de imagem e slide para treinamento, até o participante ter sido capaz de executá-los ou julgado ter entendimento de seu uso.
5. Execução, por parte do usuário, para cada modo, de uma determinada tarefa de apresentação de slides e de uma tarefa de apresentação de imagem, com objetivo de medir tempo e permitir ao usuário ter experiência com o uso de cada modo. As tarefas foram duas:
    a. Abrir uma imagem que está de cabeça para baixo, de “onde está Wally?” (Figura 26). A informação da posição do personagem era dada previamente. O objetivo do usuário era destacar com o uso do zoom o personagem na tela e ao final fechar o arquivo.
    b. Abrir uma apresentação de slides pequena (cinco slides), navegar entre os slides até o final, voltar para início, e fechar.
6. Responder um pequeno formulário quanto à satisfação do uso de cada modo para as tarefas;
7. Aplicação de um pequeno questionário semiestruturado de perguntas abertas com objetivo de compreender as impressões dos participantes sobre o sistema e sua preferência de uso para cada um dos modos propostos:
    a. Qual destes modos você acredita que utilizaria para uma apresentação, e porque?
    b. Fale um pouco da impressão de cada um dos modos e como se sentiu utilizando-os para a tarefa executada.

c. Como você compararia essas formas de interação disponíveis com o uso direto do teclado e mouse?

Todas as etapas do procedimento foram acompanhadas de gravação de áudio, vídeo e anotações. Através da execução deste procedimento, foi possível extrair as seguintes informações:

- Tempo para executar tarefas;
- Taxa de erros / precisão na execução de comandos;
- Pontuação de satisfação do modo para a tarefa alvo;
- Opinião geral do usuário quanto ao sistema e os modos de interação disponíveis.

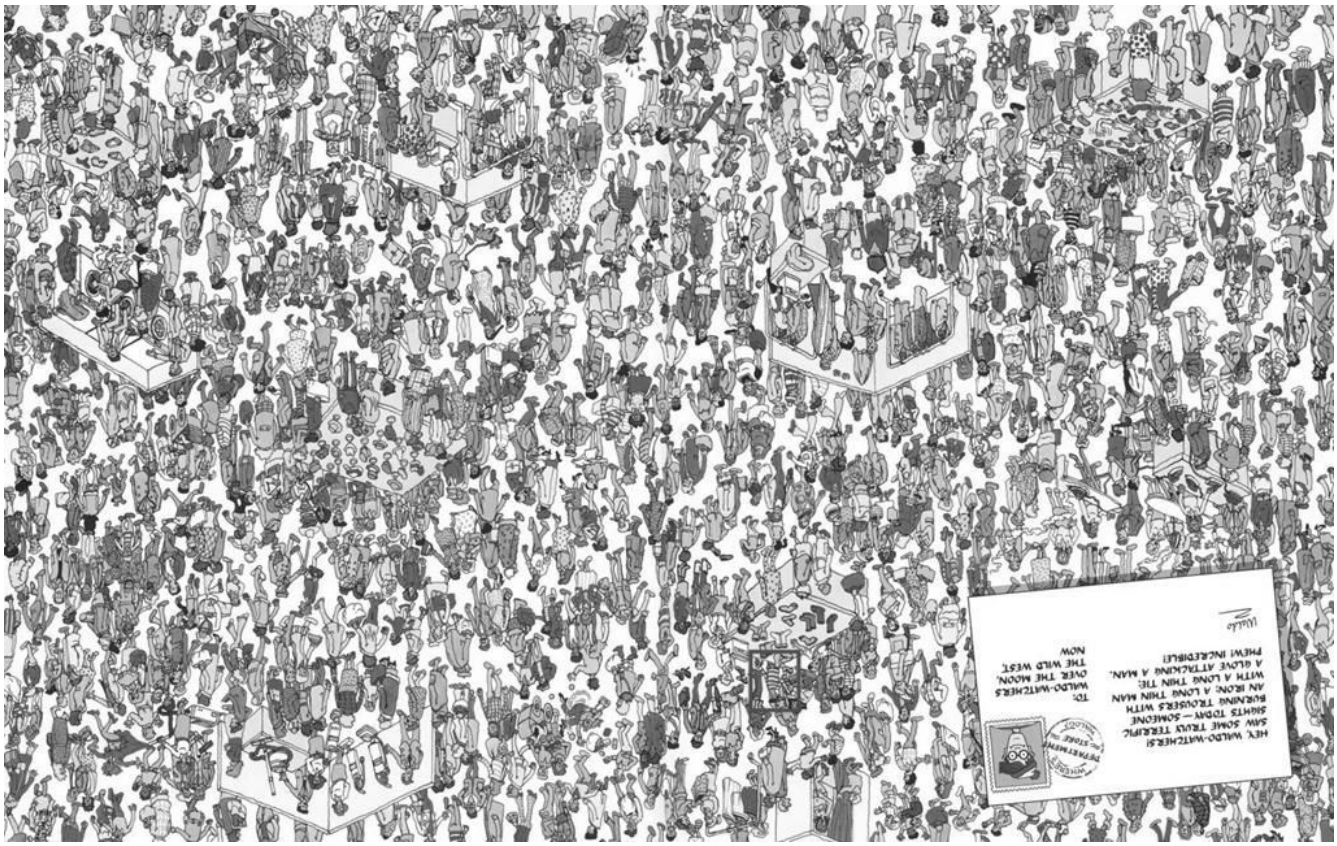


Figura 26 - Imagem de Wally utilizada para uma das tarefas

O foco da avaliação era qualitativo, com perguntas abertas para que os participantes tivessem maior liberdade em descrever suas opiniões sobre o sistema, mas também foram colhidas informações quantitativas, como tempo de execução, para melhor compreender as condições do sistema e analisar seus impactos na preferência dos usuários.

Durante a execução dos testes foram utilizados um notebook e um *smartphone* Android, ligados em rede através do programa Connectify [21] executado do notebook. Tal solução foi necessária uma vez que a rede sem fio disponibilizada pela universidade não é diretamente conectada com a rede a cabo, e era mais prático não precisar utilizar um roteador adicional. Neste modo, foi percebido que o sistema, em algumas poucas situações, demorava para responder, indicando algum atraso de rede, não encontrado em testes em um ambiente com um roteador exclusivo. Nas poucas situações que o atraso ocorreu durante as tarefas, esse atraso foi desconsiderado do tempo total para comparação.

Uma imagem da configuração do sistema no ambiente testado pode ser visualizada no início da seção 5.2.

6.2 **Teste Piloto**

Um teste piloto, com um participante, foi realizado para identificar possíveis problemas no procedimento de teste e verificar se o sistema estava pronto para tal. A análise deste teste levou a algumas modificações na implementação da interação do sistema de alguns comandos, pelos motivos explicitados a seguir.

6.2.1 Gesto de fechar a mão

O gesto que o usuário utilizaria para fechar a apresentação, levantando a mão, deixando-a parada por um tempo, e fechando-a, foi escolhido a partir das entrevistas e grupos focais para ser implementado. No teste piloto, no entanto, ele acabou sendo ativado diversas vezes sem intenção, enquanto o participante posicionava a mão para algum outro comando. Devido a estas ativações desnecessárias que poderiam comprometer o uso do sistema foi optado por remover tal gesto.

6.2.2 Gesto de rotação e ampliação

O gesto de rotação foi inicialmente definido e implementado como a mudança de ângulo entre as mãos. Devido a similaridade com o gesto para ampliação, um determinado limiar de ângulo ou distância das mãos que primeiro for disparado determinaria o gesto desejado. No entanto, durante o teste piloto algumas vezes o gesto de rotação era confundido com o de

ampliar, por vezes devido à baixa precisão de identificação dos pontos, que acabavam sofrendo mudanças de posições bruscas (ruídos/falhas de identificação).

Com objetivo de deixar os gestos para estes dois comandos distintos de forma a não serem confundidos, optou-se por considerar a posição inicial das mãos para a análise de intenção do gesto. No caso do zoom, as mãos deveriam estar afastadas pelo menos 12cm de altura, e alternarem posições de qual está acima ou abaixo da outra, ou 60º de mudança. Para o gesto de zoom, as mãos estariam entre os 12cm de diferença de altura, e a mudança de distância entre elas foi considerada para ativação.

6.2.3 Comando de fala para fechar apresentação

A definição inicial para o sistema era o uso da palavra “Close” com objetivo de fechar a apresentação corrente. O teste piloto identificou a baixa precisão na identificação de tal palavra, que era frequentemente ativada de forma indesejada durante a conversação para explicar o uso do sistema, ou quando algum outro comando de voz era requisitado. Tal problema é referenciado também em uma discussão do fórum oficial, e a recomendação é a criação de sentenças compostas de pelo menos duas palavras [59]. Optou-se por adicionar uma palavra extra, de forma a que a identificação indesejada ficasse mais difícil de ocorrer. O comando foi redefinido de “Close”para “Close File”.

6.3 **Resumo de Interação Implementada**

Nessa seção é demonstrada a representação da interação implementada no sistema, que serviu de referência para os participantes na forma como foi apresentada pelo entrevistador e para referência durante o aprendizado. A Tabela 17 apresenta a interação para os comandos de Iniciar e Fechar arquivos, para cada um dos modos do sistema. A Tabela 18 apresenta a interação para os comandos referentes à apresentação de slides. Por fim, a Tabela 19 apresenta os comandos para manipulação de imagem. A manipulação de imagem possuía comandos de fala que usavam valores de multiplicação para aumentar a quantidade de zoom ou movimento dado ao comando, sendo opcionais para uso, caso contrário um valor unitário era utilizado (adicionado ao valor atual).

Tabela 17 - Execução dos Comandos de Iniciar e Fechar Arquivo para Cada Modo.

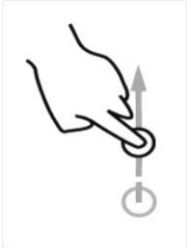
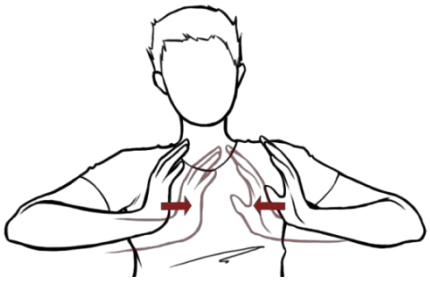
	Iniciar	**Fechar**
Dispositivo Móvel	Um toque na lista para visualizar arquivo. Para abrir, uma das seguintes opções: 1. Arrastar visualização para cima  adaptado de [31]. 2. Um toque na visualização	1. Arrastar para baixo  adaptado de [31]. 2. Clicar no botão de voltar do dispositivo 3. Um toque na tela + um toque na opção fechar
Gestos de Corpo	Empurrar com a mão em cima do arquivo selecionado. Mão fechada para rolar pagina. 	Afastaras mãos um pouco, e as fechar novamente até se encontrarem. 
Comando de Fala	“Open File” + número	“Close File”

Tabela 18 - Comandos de Avançar e Voltar Slides para Cada Modo.

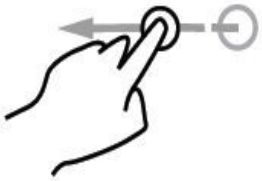
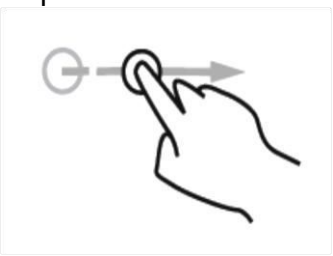
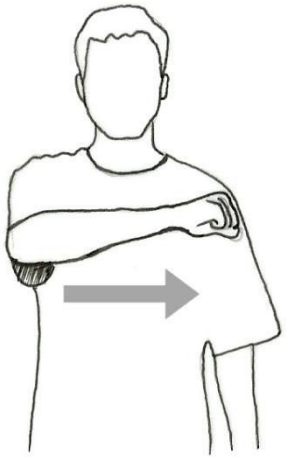
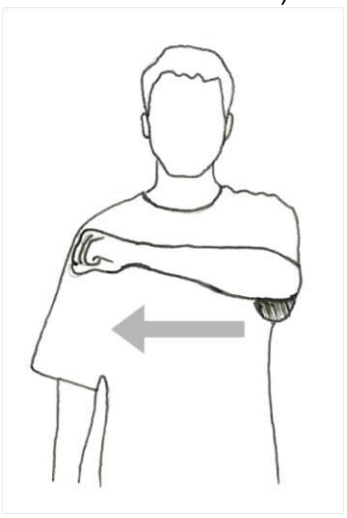
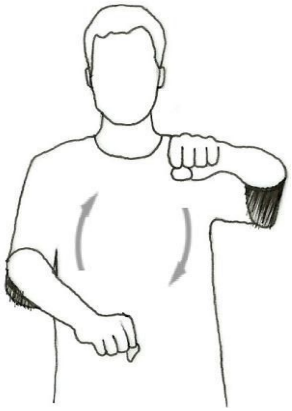
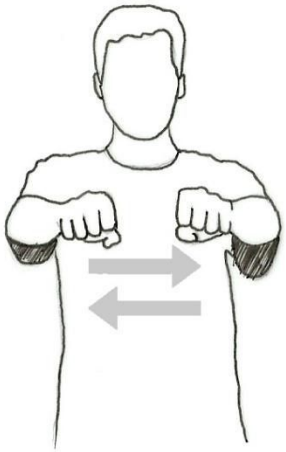
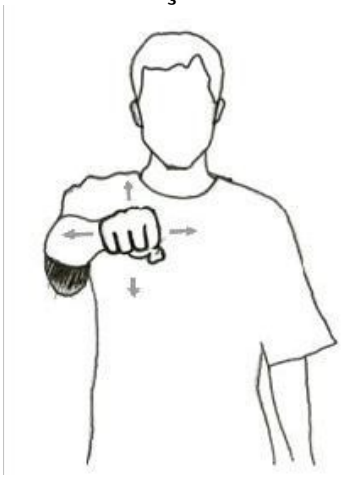
	Avançar Slide	**Voltar Slide**
Dispositivo Móvel	Arrastar para esquerda  adaptado de [31].	Arrastar para direita  adaptado de [31].
Gestos de Corpo	Empurrar para a esquerda (mão inicia fechada e termina aberta). 	Empurrar pra a direita (mão inicia fechada e termina aberta). 
Comando de Fala	“Next”, “Forward”	“Previous”, “Back”

Tabela 19 - Comandos de Manipulação da Imagem para Cada Modo.

	Rotacionar Imagem	**Aumentar/Diminuir Zoom**	**Mover Área de Visualização**
Dispositivo Móvel	Rotacionar com dois dedos  adaptado de [31].	Afastar/Aproximar dedos em linha  adaptado de [31].	Deslizar com um dedo como se estivesse arrastando a imagem.
Gestos de Corpo	1. Uma mão em cima e outra embaixo, trocam alturas circularmente. Mão que está em cima determina lado de rotação. (mãos fechadas)  2. Uma mão parada, a outra rotaciona em volta da primeira (mãos fechadas).	Afastar/Aproximar mãos horizontalmente ou diagonalmente 	Mover mão fechada para uma das direções. 
Comando de Fala	“Rotate Right/Left”	“Enlarge/Zoom in + 2-20 times (opcional)” “Reduce/Zoom out + 2-20 times (opcional)” Exemplos de comandos: 1. “Enlarge 2 times” 2. “Zoom out” 3. “Zoom in 5 times”	“Move Right/Left/Down/Up“ + 2-20 times (opcional) Exemplos de comandos: 1. “Move Right 2 times” 2. “Move Up” 3. “Move Left 10 times”

6.4 **Perfil dos Participantes**

Dez participantes foram recrutados para participar da análise do sistema, compondo uma amostra reunida por conveniência. Este grupo continha oito alunos e um professor da universidade aonde este trabalho está sendo desenvolvido, e um participante que não é relacionado à Instituição. Dos participantes, seis eram homens e quatro mulheres, com idades variando de 21 a 42 anos, com idade média de 26,4 anos. Apenas dois participantes não eram da área de Informática, sendo um músico, não cursando ensino superior, e um estudante de Farmácia. Dos participantes que são relacionados à área de Informática, quatro eram alunos de graduação, dois alunos de mestrado, um aluno de doutorado, e um professor.

Quanto à experiência com dispositivos computacionais, apenas um dos participantes não possuía um *smartphone* com tela de toque, embora tivesse acesso a um *tablet* (família). Além deste, apenas outros dois participantes possuíam *tablets*, embora todos já tivessem utilizado um *tablet* alguma vez. Dos *smartphones* que os participantes possuíam, todos com exceção de um eram dispositivos Android, sendo este ultimo um Apple. Dos *tablets*, dois eram Apple, e um Android.

Dos dispositivos com tecnologia de gestos de corpo, apenas um dos participantes possuía um dispositivo Kinect, e um possuía um dispositivo Ps3 Move. Apenas um participante disse não ter nunca utilizado nenhum dispositivo destes ou similar, enquanto os outros apenas tiveram pequeno tempo de uso em jogos.

Todos os participantes disseram já ter utilizado comandos de fala para interação com algum dispositivo computacional, mas apenas um deles disse achar útil, utilizando as vezes para fazer buscas na internet ao invés de precisar digitar no *smartphone*, enquanto os outros em sua maioria disseram ter tido muitos problemas de reconhecimento em seu uso e não costumarem utilizar.

Todos os estudantes disseram ter experiência com apresentação de trabalhos com o uso do Power Point pelo menos uma vez por semestre, e alguns fazem apresentações em sua bolsa de pesquisa, ou na empresa aonde trabalham.

6.5 **Resultado dos Testes**

Foram feitos testes individuais com os dez participantes para avaliar a aceitação dos modos para o sistema proposto. Como executado em outros trabalhos [5,13,40], medidas de usabilidade foram extraídas para comparar a eficiência entre os diferentes modos, como tempo e satisfação dos participantes. Foram feitas ainda perguntas abertas para melhor entender as preferências de cada participante. Os testes foram realizados em uma sala de aula da Faculdade onde este trabalho está sendo desenvolvido, e duraram entre 45 a 60 minutos. Os resultados dos testes são apresentados a seguir.

6.5.1 Tempo de Execução

Nesta seção são apresentados os resultados dos testes apresentando o tempo gasto para realização de cada tarefa em cada um dos modos. Ainda, valores referentes a erros de reconhecimento do sistema e erros de execução pelos usuários são apresentados.

O tempo aqui apresentado leva em consideração a execução de toda tarefa desde seu inicio até o fim, desde abrir o arquivo até fechá-lo, incluindo o tempo do sistema para processar cada comando. As tarefas se referem àquelas apresentadas na seção 6.1, e são referenciadas da seguinte forma nas tabelas:

- Tarefa1 – Imagem. Ampliar *Wally* na tela.
- Tarefa2 – PowerPoint. Avançar e voltar slides.

Tabela 20 - Tempo de Execução para as Tarefas Utilizando *Smartphone*.

	Tarefa 1	**Tarefa 2**
P1	21s	17s
P2	20s	23s
P3	28s	15s
P4	34s	16s
P5	19s	18s
P6	38s	22s
P7	24s	15s
P8	57s	17s
P9	26s	12s
P10	14s	10s
Média	28,1s	16,5s
Desvio Padrão	12,41s	3,98s

Na Tabela 20 são mostrados os tempos gastos para execução das duas tarefas, para cada um dos dez participantes. Alguns pontos foram observados ao longo da execução das tarefas com o uso do *smartphone:*

- A forma preferencial de abrir e fechar os arquivos foi o arrastar, visto que estes dois comandos possuem múltiplas opções. Um dos participantes esqueceu-se de tal gesto de arrastar para fechar o arquivo, e depois lembrou-se quando questionado se preferia a opção que havia utilizado (botão de fechar do menu), enfatizando que havia esquecido e depois demonstrando preferência por esse (arrastar). Apenas um

participante usou um toque para abrir, e apenas um utilizou o botão de voltar do *smartphone*;

- Quatro dos participantes na tarefa da imagem limparam a ampliação da mesma (retornando para a ampliação inicial) com o uso dos dois dedos, diminuindo a ampliação, e, após isso, arrastando a imagem para baixo para fechá-la. Nesta tarefa, tal escolha de ação não é a mais eficiente, uma vez que era possível simplesmente apertar o botão de voltar do *smartphone*, ou então remover totalmente a ampliação com o uso de dois toques na imagem;
- Três participantes tiveram dificuldade inicialmente no uso da rotação, de forma que executaram uma rotação incompleta na imagem, ou então acabaram ativando o comando de ampliar ao invés de *zoom*.

Tais pontos são consequência, provavelmente, do pequeno tempo de treinamento oferecido com o sistema, alguma regulagem necessária ao mesmo, e às variadas opções de gestos para alguns dos comandos (abrir e fechar). Como possível modificação poder-se-ia reduzir o número de opções para ativar um comando, ou de alguma forma orientar o usuário em um uso mais efetivo.

Na Tabela 21 são apresentados os tempos gastos para completar as tarefas, para cada um dos participantes, no modo de gestos de corpo. A taxa de precisão média para a execução de gestos de corpo na Tarefa 1 foi 78,92%, e para a Tarefa 2 85,92%. Essas médias foram calculadas através do percentual de cada indivíduo na execução dos gestos que foram identificados corretamente sobre o número total de tentativas, e posterior média entre todos os participantes. Uma pequena taxa de erro de 2,81% foi identificada, apenas na Tarefa 1, representando a execução errada dos comandos pelos usuários, sobre o número total de tentativas.

Erros de reconhecimento identificados neste modo foram, por exemplo, falha em identificar corretamente se a mão do usuário estava aberta ou fechada, no tempo adequado, ou se o usuário executou um gesto aparentemente correto que o sistema tenha rejeitado, por motivos de distância de execução ou proximidade com o equipamento (valores ajustados na etapa de desenvolvimento). Um exemplo de execução errada dos comandos pelos usuários foi o movimento mal executado do gesto de rotação com as mãos.

Embora, no caso de reconhecimento errado do estado da mão (aberta ou fechada), seja possível o usuário eliminar, ou pelo menos mitigar, através de uma execução mais nítida para a câmera, a resposta natural na execução dos gestos pelos participantes não foi assim. Os gestos foram demonstrados pelo entrevistador utilizando a palma da mão virada diretamente para a câmera, e os estados de mão fechada e aberta eram bem distintos (a mão totalmente aberta ou totalmente fechada). Uma posição mais lateral ou uma abertura parcial da mão por parte dos participantes foi o que ocasionou a maioria dos erros de identificação neste caso.

Para o gesto de fechar, a execução foi problemática para alguns participantes, pela forma como eles executavam o gesto. O gesto exigia que as mãos estivessem afastadas lateralmente entre si e a certa distancia à frente do corpo do usuário, mas alguns participantes deixavam as mãos afastadas lateralmente na mesma distância do corpo, e as aproximavam perto do mesmo ao uni-las à frente do corpo. Ainda, existia algum ruído advindo do reconhecimento do dispositivo quando as mãos se juntavam próximas da altura do peito, na qual a identificação dos pontos da mão parecia se perder.

Um maior tempo de treinamento e de uso do sistema provavelmente iria diminuir de forma significativa o número de erros, e consequentemente diminuir o tempo de execução dos comandos. Também, o algoritmo de reconhecimento poderia ser melhorado para permitir uma execução menos rígida da parte do usuário.

Um ponto adicional interessante foi observado para o modo dos gestos. Dois participantes mencionaram, durante o treinamento e uso, um pouco de dor nas mãos, pelo fato de eles estarem apertando a mão muito forte para manter o gesto da mesma fechada. Tal gesto não necessitava de força, mas eles pareceram exercer a mesma sem ter consciência disto, Ainda, no gesto de fechar, três participantes o executaram muito rápido, forçando as mãos a se baterem. Novamente, tal velocidade e força não era necessária para executar tal gesto, mas pareceram respostas naturais dos mesmos. Tal execução involuntária e exagerada poderia minimizar o potencial uso do sistema para longos períodos, exigindo uma educação do usuário para uso do mesmo com mais eficiência.

Tabela 21 - Tempo de Execução para as Tarefas Utilizando Gestos de Corpo.

	Tarefa 1	**Tarefa 2**
P1	40s	24s
P2	34s	31s
P3	26s	29s
P4	62s	65s
P5	27s	23s
P6	42s	38s
P7	81s	36s
P8	49s	22s
P9	69s	34s
P10	35s	27s
Média	46,5s	32,9s
Desvio Padrão	18,54s	12,55s

A Tabela 22 apresenta o tempo gasto para realização das tarefas no modo de fala. A taxa de precisão neste modo foi de 60,88% para a Tarefa 1, e 71,11% para a Tarefa 2, com uma taxa pequena de 2,73% de erros de execução dos comandos pelos participantes, apenas na Tarefa 1.

Erros de reconhecimento de fala incluíram em sua maioria a baixa precisão na identificação do comando falado, mas também existiram casos em que o comando foi interpretado como outro (‘move right’ foi entendido como ‘move left’, ‘next’ foi confundido com ‘back’). Erros do

usuário incluíram a chamada de comandos da forma errada (‘move ten times right’ ao invés de ‘move right ten times’).

Tabela 22 - Tempo de Execução para as Tarefas Utilizando Comandos de Fala.

	Tarefa 1	**Tarefa 2**
P1	82s	39s
P2	166s	38s
P3	44s	28s
P4	-	76s
P5	97s	57s
P6	114s	66s
P7	-	38s
P8	183s	40s
P9	-	80s
P10	-	-
Média	113,33s (*)	51,33s (*)
Desvio Padrão	52,30s (*)	18,90s (*)

(*) calculo baseado nos tempos dos participantes que terminaram a tarefa

É possível observar na Tabela 22 que quatro participantes não foram capazes de terminar a Tarefa 1 no modo de fala, e um não foi capaz de terminar a Tarefa 2. Erros e dificuldades são esperados visto que a língua escolhida (inglês) não é a nativa dos participantes. Na gravação é possível identificar que estes que não foram capazes de terminar sua tarefa pareciam pronunciar fonemas de maneira errônea. Não era esperada a correção da pronuncia dos participantes durante a fase de treinamento, embora alguns ficassem testando

e compreendendo a forma com a qual o sistema melhor identificava a pronuncia de algumas palavras.

Entre os erros de reconhecimento presentes no modo de fala, é importante destacar que considerando apenas os comandos de abrir e fechar arquivos, estes obtiveram uma média de 81,76% de precisão, entre todos participantes que terminaram as tarefas, sendo portando significativamente mais precisos se comparado ao percentual de precisão geral das tarefas 1 e 2 (que como apresentados, incluíam também tais comandos). No entanto, entre os participantes que não terminaram as tarefas, P4 não cumpriu a Tarefa 1 por dificuldades de abrir (após várias tentativas seguidas não conseguiu e desistiu), e P10 não conseguiu fechar os arquivos em nenhuma das duas tarefas.

A definição da língua não nativa e consequentemente os problemas de reconhecimento foram uma limitação claramente visível nos resultados. Mesmo testes anteriores com o sistema, na etapa de desenvolvimento, já apontavam para dificuldades e erros no sistema de reconhecimento. As configurações de ganho automático desligado, e o cancelamento de ruído ativado foram utilizados no sistema, pois são recomendados no fórum oficial do dispositivo [57,58,59]. O ambiente na qual os testes foram realizados, uma sala de aula da universidade, pode não ser o melhor lugar para o uso da fala, uma vez que existia um pequeno ruído, perceptível nas gravações, advindo do ar condicionado do ambiente, e o que parecia também um pequeno eco. Uma análise mais aprofundada dessas características do ambiente e de possíveis outras configurações do dispositivo não foram realizadas, mas é provável que o dispositivo e o algoritmo de reconhecimento utilizado estejam ainda em um estágio pouco desenvolvido para resultados melhores, também considerando que os participantes não eram nativos no idioma definido.

A Figura 27 mostra o resumo da comparação do tempo médio de execução das tarefas para cada modo.

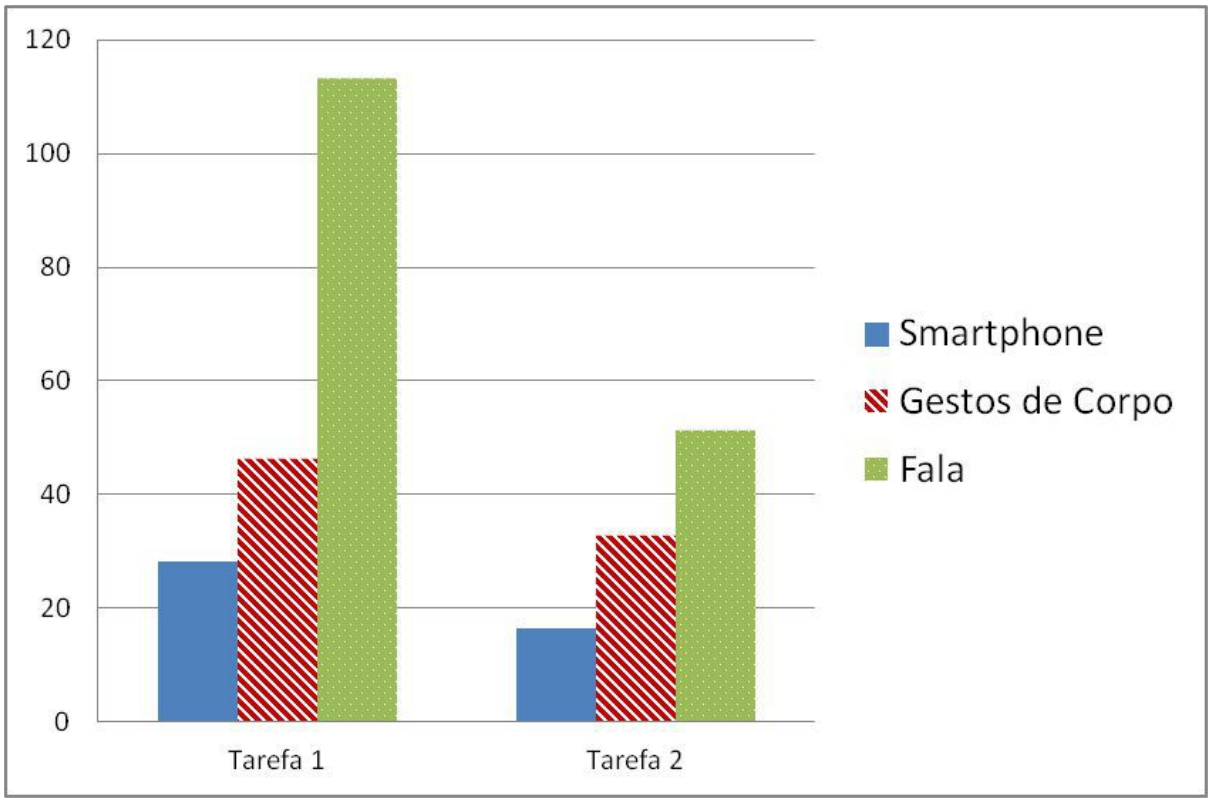


Figura 27 - Comparação de Tempo Entre Modos.

6.5.2 Satisfação

Foi requisitado que os participantes avaliassem com uma nota de um a cinco, a sua satisfação com cada um dos modos dentro de três diferentes categorias do sistema, sendo um muito ruim, e cinco muito bom. Durante o preenchimento e após o mesmo, os participantes tiveram a oportunidade de comentar sobre as impressões dos modos e os problemas presenciados.

A fala apresentou a menor pontuação nas tarefas de slides e imagem frente aos outros modos, enquanto o modo de *smartphone* obteve a maior pontuação para todas as situações. Entre os maiores problemas de baixa pontuação da fala foram destacados as falhas de reconhecimento, que no caso da fala ocorreram em menor número nas funções de abrir e fechar arquivos, e em maior número nas funções que envolviam a manipulação da imagem. Alguns participantes, no entanto, mesmo presenciando fortes problemas de reconhecimento atribuíram pontuações relativamente altas, se comparado ao desempenho obtido, justificando que o maior impedimento era o a tecnologia, mas que o modo de fala era fácil de utilizar, e seria uma forte preferência caso fosse melhorado.

A pontuação média para cada modo pode ser visualizada na Figura 28. Maiores detalhes das opiniões dos participantes são apresentados na subseção seguinte.

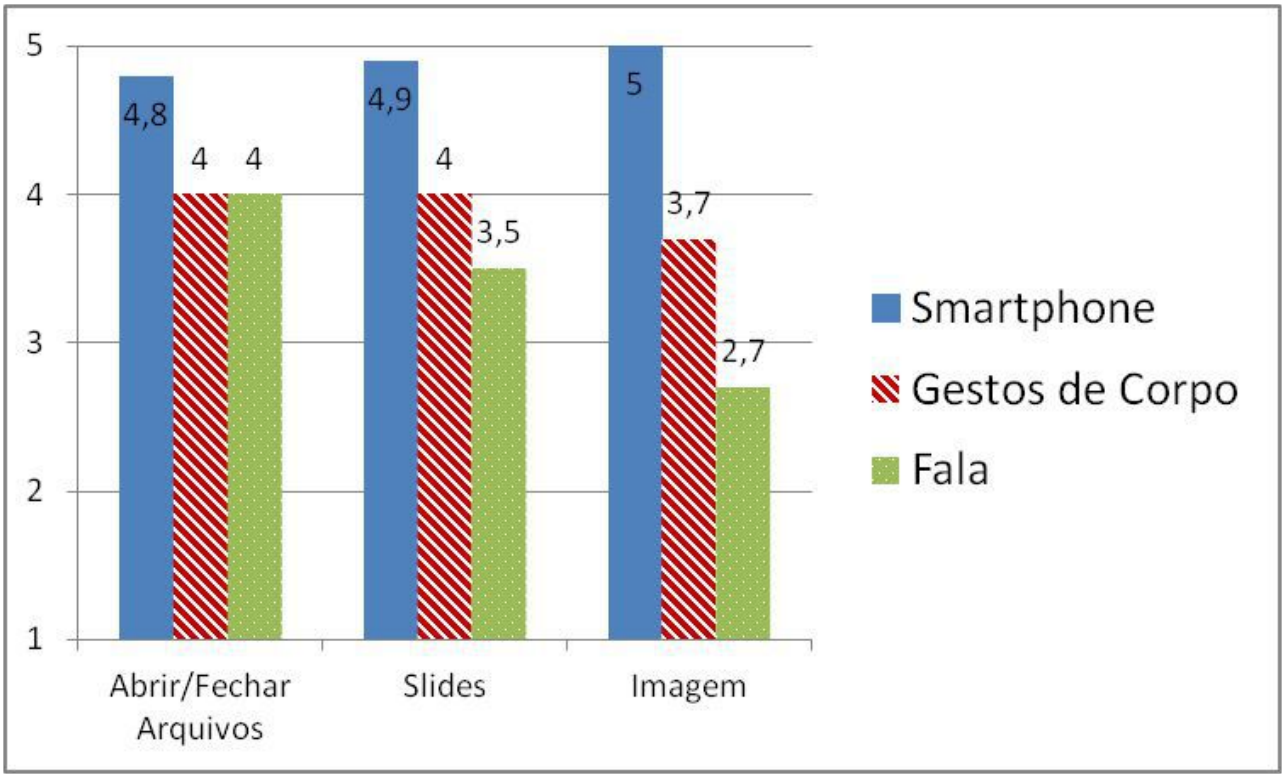


Figura 28 - Pontuação Média de Satisfação Entre os Diferentes Modos e Comandos do Sistema.

6.5.3 Opinião dos Participantes

Com objetivo de melhor entender as preferências dos participantes foram realizadas algumas perguntas abertas aonde os participantes tinham a oportunidade de explicar e detalhar melhor sua posição.

As considerações do participante P1 com relação os modos disponíveis foi de que "*Eu só não usaria o celular. Precisa só da voz e do corpo ali.*". Quanto a como usaria as duas em uma apresentação ele disse "*As duas ao mesmo tempo, a hora que eu quisesse fazer assim (gesto) eu passava ou falava 'next'. Depende da minha vontade*.".Ele disse preferir usar os gestos para a manipulação da imagem, e que a fala era bem útil caso estivesse longe do equipamento.

Para P2, a fala ainda está distante do ideal, e precisaria de melhorias. Usaria tanto gestos como o *smartphone*, e ficou em dúvida em qual dos dois teria maior preferência "*Aí é que*

ta... o smartphone eu também achei muito bom. Também são as mesmas vantagens... Fiquei muito na duvida.".

P3 disse que não usaria fala para uma apresentação, "*Eu não usaria voz... mesmo que seja perfeito, fica estranho uma pessoa gritando no meio da sala, uma questão de cultura*.". Quando questionado sobre os gestos ele teve uma opinião diferente, "*Eu acho que seria mais facilmente aceito do que voz. Porque voz tu incomoda as pessoas. Gesto tu pode ficar no teu canto fazendo teus gestos... não vai incomodar. Agora a voz mesmo tu olhando pra outro lado fica incomodando, incomodando. Acho que é mais fácil gestos serem aceitos do que a voz*.". Quanto a isso ele disse achar mais prático o uso dos gestos do que do *smartphone* para passar slides, sendo bom manter as mãos livres. Ele teve algumas reclamações quanto ao *feedback* de gestos na tarefa da imagem, aonde a rotação era "abrupta", contrário a implementação utilizada no smartphone. Também mencionou desgosto quanto à restrição do uso de uma única mão nas operações de mover imagem e passar slides.

P4 disse considerar a fala como melhor modo para apresentação de slides, "*para apresentações, 'next, 'previous', seria tudo de bom*.". Quanto a gestos de corpo P4 não achou intuitivo o gesto utilizado "*Assim (gesto de pegar), não é intuitivo. Se eu estivesse numa banca, e a pessoa fizesse isso eu ficaria nervosa*.". Sugeriu o uso de um gesto com a mão aberta, e mais discreto (movimentos pequenos, apenas com a mão, como folhar um livro). Ainda demonstrou o desejo de realizar os comandos de forma discreta, embora acredite que a fala acabe sendo algo que não atrapalhe "*'next slide', é uma coisa que até a gente acaba se acostumando, e acaba esquecendo quando a pessoa falar. Então seria mais ou menos assim... Tu ter a possibilidade de fala, e fazer a passagem na mão. Não havendo essa possibilidade, ai eu ia utilizar o smartphone. Sem a pessoa percebeu que eu to falando e to trocando slides*.". Também mencionou que o uso da fala para a manipulação de imagem era um ponto fraco *"... as questões do 'enlarge', 'zoom in', pode ser uma coisa meio desgastante... 'open', 'close', 'next', 'forward', 'back', perfeito*.".

P5 mencionou a dificuldade de trabalhar com os gestos de corpo, quando questionado da facilidade de usar o sistema, "*Olha, difícil não é. O problema é lembrar de todos os gestos*". Sua preferência de uso é do *smartphone*, "*Eu acho que é muito limitada a questão da voz. Voz é mais complicado. Pelo smartphone super tranquilo, fácil de trabalhar. Os movimentos*

(gestos de corpo) eu achei também interessante. Só que ainda por preferência prefiro o smartphone.", mas mencionou que seria interessante o uso dos gestos de corpo para manter as mãos livres "*Possivelmente usaria o gestual também... Possivelmente eu abriria o arquivo com o smartphone e tentaria levar a apresentação com o gestual pra não ficar com as mãos ocupadas... Eu preferia não ter que ficar com algo na minha mão*". Quanto a fala, mesmo que o reconhecimento fosse perfeito ele disse que não seria bom para uma apresentação *"(usaria fala) Só pra abrir e fechar arquivo. Porque... se eu to usando pra fazer uma apresentação... poderia ficar estranho para a apresentação, mais por isso*".

P6 mencionou que o uso dos gestos de corpo seriam mais fáceis: "*Eu achei os gestos, no geral, mais fáceis*", e quando questionado em porque eles seriam melhores do que, por exemplo, o uso do *smartphone*, ele mencionou a 'novidade', "*Acho que mais por ser novidade. Fiquei impressionado por ter usado o Kinect*.". Quando questionado se usaria os gestos em uma apresentação, P6 disse "*Acho que é bem possível. porque é muito mais fácil do que ficar fazendo assim (gesto com a mão) pra o colega passar*.". Ele mencionou que não usaria a fala "*Fala... Ela funciona bem, mas ainda acho estranho para uma apresentação*.".

P7 também teve preferência apenas pelo *smartphone*, disse não ter muita paciência com essas tecnologias, que não funcionam direito "*É Legal... eu não tenho muita paciência com essas coisas, sempre não pega direito uma coisa aqui, uma coisa aqui. Tem que ser exatamente como quer... não tenho muita paciência com isso*". Quando questionado se não usaria nada dos gestos ou da fala, disse "*Gestos acho que não. Falar se ele gostasse de mim eu usaria. Não me sinto muito a vontade com essa coisa de ficar lá assim (gestos)... Eu gosto mais do smartphone porque tem mais precisão... Em uma aula até pode ser, tu vai lá explica e 'next'. Agora se tu ta em uma apresentação só falando e 'next'... não ia dar certo. Eu prefiro ainda a precisão do touch*.".

P8 considerou os gestos interessantes, e preferencialmente iria usar eles com mais treinamento "*Em uma apresentação de verdade eu ia usar os gestos. Aprendendo os macetes de como ele eventualmente não capta algumas coisas. Eu acho que é melhor porque tu ta olhando a apresentação e ao mesmo tempo tu não precisa desviar a atenção dela. Tu vai olhando ali e vai fazendo os gestos todos. Aqui não (smartphone), aqui eu tenho que desviar os olhos da apresentação pra o celular. Não é que isso seja uma combinação grande, mas acaba perdendo um pouco o foco... Pra abrir talvez eu usasse o smartphone.*

Porque aqui é tranquilo. To iniciando uma apresentação... Dai depois disso não quero ser mais incomodada.". O maior problema da fala para P8 foram os erros de reconhecimento "*Eu usaria tanto a fala quanto os gestos... alternadamente, conforme o humor, se estivesse boa a fala.*".

P9 demonstrou maior preferência em utilizar a fala e o *smartphone*, caso a fala funcionasse bem, "*Provavelmente eu oscilaria entre os dois (fala e smartphone). Acho que depende do meu humor. Fala é muito mais útil se tu não quer ter nada em mãos. Fica com as mãos livres até para apontar coisas na apresentação, muito mais útil, mas se tu não te incomoda de ter o dispositivo junto contigo eu acho muito mais pratico o smartphone.*". Quanto aos gestos ele mencionou que poderiam ser embaraçosos devido a serem movimentos amplos, "*Em algo mais formal, uma apresentação de uma banca, eu acho que poderia ficar meio estranho, dependendo dos gestos. Por serem gestos mais amplos, se no futuro conseguisse colocar em uma escala menor, só com a mão, (gestos menores), dai seria mais conveniente, mas por enquanto poderia se passar até como embaraçoso, até mais se não funcionasse, e tu tivesse na frente de uma banca. Por isso que eu acho que a fala é mais fácil*". Quando P9 foi questionado quanto a se a fala atrapalharia em uma apresentação disse "*Não, eu acho que não, porque, na minha experiência de apresentação, já é isso que se faz, porque como tu tem alguém passando pra ti, tu tem que pedir pra passar, ou tu mesmo tens que te deslocar um pouco, pra poder passar. Então eu não vejo que isso atrapalharia.*".

P10 disse preferir o modo de gestos para slides "*Esse aqui (gestos de corpo) 'handsdown' o melhor... Mas esse aqui (smartphone) também achei muito bom, porque afinal é só passar pro lado.*", e disse que fala poderia ser estranho dependendo do caso "*Acho que em uma apresentação de verdade eu usaria mais os gestos. A fala ela pode atrapalhar dependendo da língua da apresentação. Se eu fizesse uma apresentação em inglês, se eu falasse 'close' toda hora e não fechasse ia ficar meio chato. Então eu ia usar o gesto de corpo, ou o smartphone.*". No entanto, ele disse que seria estranho utilizar dependendo do público "*Claro, se tu vai fazer uma apresentação que não estejam sabendo que tu vai fazer os gestos fica meio estranho. Tu ali pegando e jogando do nada... Eu usaria em determinada situação? Depende. Se eu estivesse em uma bancada discursando, eu usaria gestos. Usaria gestos com certeza. Se eu pudesse falar, depende sei la, no smartphone, falar baixinho 'close', que*

não interferisse na minha apresentação, eu usaria também. Eu faria comandos locais, digamos assim, acho q fica mais interessante. Mas o smartphone é o mais simples de usar.".

Em outras perguntas, foi possível observar dos comentários dos participantes que todos preferiram pelo menos um dos modos para uso em apresentação mais do que a forma que realizam hoje em dia (em sua maioria, *notebook* ou *desktop*, utilizando o teclado). Também todos foram questionados em comparação do uso do *smartphone* e um *pointer*, e todos demonstraram preferência pelo uso do *smartphone* sobre o outro, com exceção de um participante que não mencionou achar diferente para o caso de slides, incluindo razões de ser mais intuitivo, feedback, maior opções de comandos, e etc.

A presença de erros de reconhecimento nos modos de fala e gestos de corpo é impactante na sua preferência e real uso, mas os participantes foram questionados para responder o que achariam caso estes fossem mitigados. Neste caso, entre os participantes que demonstraram preferência nos modos de fala ou gestos houveram divergências de opiniões e pontos de vista. Alguns acreditam que o uso de fala pode incomodar ou seria estranho seu uso durante uma apresentação, e outros acreditam que o uso de gestos seria estranho ou embaraçoso. Mesmo assim, a grande vantagem citada no uso de um desses modos, sobre o *smartphone*, seria a liberdade nas mãos.

6.5.4 Mitigação de Erros

A fala apresentou poucos erros por parte do usuário, mas aqueles que ocorreram foram pelo uso errado na ordem das palavras na sentença disponível, ou omissão de alguma delas. Isto apresenta certa rigidez no sistema. Idealmente as sentenças deveriam ser interpretadas pela semântica e permitir flexibilidade na ordem das palavras e no uso de sinônimos, e/ou apresentar ao usuário detalhadamente aquelas disponíveis no sistema.

Quanto aos erros de reconhecimento, devido ao ambiente proposto para o sistema deste trabalho, seria fundamental um reconhecimento robusto a ruídos que podem advir do ambiente e de conversas paralelas. É importante considerar que o modo de fala poderia utilizar algum mecanismo de correção, por exemplo, pela apresentação ao usuário da lista das sentenças com maior grau de confiança para escolha e execução. Além disso, estudos [86] indicam que em casos de erro de reconhecimento a forma mais eficiente para correção é a troca de modalidade, ao invés de utilizar a mesma que resultou no erro. Com o tempo,

também, os usuários aprendem a utilizar a modalidade que melhor funciona para a dada tarefa. Portanto, o usuário poderia utilizar uma das outras formas de execução disponíveis para aquele comando para não ficar em um ciclo de erro contínuo.

Quanto aos erros encontrados na utilização de gestos de corpo, estes poderiam ser mitigados com um maior treinamento por parte do usuário, uma vez que é exigida uma certa rigidez na execução dos mesmos para um reconhecimento mais preciso, ou para se adequar as distâncias certas de execução. Neste sentido um *feedback* adequado parece ser necessário de forma a permitir ao usuário melhor compreender a execução dos gestos, e informar ao mesmo o estado em que ele se encontra na execução de um comando, por exemplo, fazendo com que a imagem seja rotacionada aos poucos juntamente com a execução do gesto, ou arrastar a imagem de slide junto com o movimento da mão. Tal *feedback* seria similar ao que já é apresentado no *smartphone*, na qual as mudanças dos gestos influenciam diretamente os objetos visuais da interface. Adicionalmente algum tutorial ou *feedback* com a correta execução do gesto poderia ser disponibilizado.

6.5.5 Discussão Geral

Era esperado que o uso de fala para a tarefa de manipulação de imagens apresentaria maior dificuldades, visto a natureza de tal modalidade [14,69], que não é apropriada em tarefas de manipulações espacial. Tal problema não pode ser analisado apenas por meio da pontuação de satisfação dada pelos usuários, pois tal pontuação sofreu influência da taxa de erros apresentada pelo sistema, e que foi maior na tarefa relacionada à imagem; assim, foram também consideradas as respostas dos usuários. P3 mencionou sua insatisfação pelo uso da fala na manipulação da imagem "*O modo de fala é complicado* [na tarefa de imagem].*não tenho muita ideia de como melhorar, mas não gostei muito dele*". P4 destacou problema com a fala na tarefa de imagem, "...*as questões do 'enlarge', 'zoom in', pode ser uma coisa meio desgastante para a pessoa*..." e também P5 ao mencionar que "*Agora, apresentação de imagem, a fala.... foi complicado porque parece que tem que raciocinar mais que o gesto.*".

Dois participantes citaram a palavra "novidade" junto à justificativa de porque usariam gestos ou fala para o sistema. Esse tipo de motivação poderia ser um forte catalisador para impulsionar usuários em experimentar a tecnologia. Um grande problema, no entanto, seria mitigar as confusões de uso do mesmo, e os erros de reconhecimento, que provocam uma

baixa eficiência dos modos. Um tempo mais elevado de execução não necessariamente resulta em um menor nível de satisfação do sistema, uma vez que existem outras características que são importantes para os usuários, como a facilidade de uso e de aprendizagem, e a eficiência percebida pode ser diferente do desempenho resultante [40]. No entanto, como referenciado nas perguntas iniciais para conhecer os participantes, muitos não utilizavam tecnologia de fala, mesmo que disponíveis, pois não funcionavam direito, resultando em frustração de uso.

As pontuações de satisfação geral de cada modo tiveram que ser analisadas em conjunto com as respostas dos participantes. Ao compararmos os tempos para execução das tarefas, há uma grande diferença entre os modos de gestos de toque e fala, e também gestos de corpo e fala. Primeiramente a fala resultou em quatro participantes não conseguirem terminar a tarefa 2, e desistir antes do seu final. Os erros foram altos e poder-se-ia pensar que a pontuação de satisfação seria ainda pior do que foram para a fala. A razão para uma pontuação mais razoável parece ser porque, embora os participantes tenham penalizado este modo pela dificuldade de uso apresentada, três deles demonstraram forte interesse em sua utilização (se funcionasse bem). No entanto, outros três enfatizaram que seria inapropriado seu uso em uma situação real da tarefa, mesmo que funcionasse perfeitamente.

O modo de gesto também foi destacado por três participantes, assim como a fala, como inapropriado em um uso real, mas foi considerado uma boa ou interessante opção para os restantes. Para minimizar tal problema, um sistema com gestos mais discretos, e em menor escala, considerando, por exemplo, a movimentação dos dedos do usuário, parece ser uma potencial alternativa, como sugestão mencionada por dois dos participantes.

Por fim o *smartphone* obteve uma alta pontuação. Os únicos pontos negativos gerais apontados foram o de ocupar o uso da mão do apresentador, e, como destacado por um dos participantes, a perda de foco exigida pela necessidade de ficar olhando para o mesmo.

7 CONCLUSÃO

Com o crescimento de uso e disponibilidade de diversos dispositivos computacionais é importante que o *design* de interfaces que melhor auxilie o usuário na execução das tarefas propostas. Alguns trabalhos [1,80,84] indicam que o uso de interfaces multimodais, com formas de interação mais naturais, estará amplamente presente em sistemas computacionais utilizados para as mais diversas tarefas do nosso dia-a-dia.

Em vista dessa preocupação, neste trabalho foi apresentado um estudo com as etapas de definição, implementação, e avaliação de um sistema multimodal de apresentações com interação através da fala, gestos de corpo, e gestos de toque com um *smartphone*. O objetivo era a avaliação da interação com o sistema, para comparar e compreender a satisfação de uso dos usuários com as modalidades disponíveis no sistema.

O sistema disponibiliza aos usuários a execução de tarefas de apresentação de slides e imagens para uma plateia. A interação com o sistema foi derivada através de um estudo com usuários, utilizando uma etapa de entrevistas individuais, e posteriormente uma etapa de execução de grupos focais, para definir a interação para cada um dos comandos nos três diferentes modos do sistema (gestos de corpo, fala, e gestos de toque com um *smartphone*). O sistema foi desenvolvido utilizando-se dispositivos amplamente disponíveis atualmente, sendo o dispositivo Kinect para captura de movimentação do corpo do usuário e fala, e um *smartphone* Android. Ao final, o sistema foi avaliado através de testes com participantes a partir da demonstração, uso e execução de duas tarefas com o sistema, para coleta de dados de uso e questões sobre a opinião dos participantes. Os testes foram executados em uma sala de aula da universidade, sem a presença de plateia.

O contexto de uso do sistema apresenta dificuldades para gestos de corpo e fala no ambiente proposto, resultando em que alguns participantes do teste classificaram essas formas como inapropriadas para o uso em um ambiente real. A influência do ambiente e audiência existente durante a interação com os sistemas através de gestos apresenta níveis de aceitação mais baixos devido à preocupação com o que os outros irão achar do usuário, devido a estes gestos chamarem atenção do público ao redor, como apresentado por Rico & Brewster [78]. No entanto, esse problema pareceu inexistente para outros participantes, e houve uma apreciação e interesse pelo uso dessas formas de interação.

No geral, o *smartphone* foi a forma de interação mais aceita, por sua precisão e discrição. Os gestos de corpo apresentaram a segunda melhor aceitação, sendo uma boa opção por permitir maior liberdade ao apresentador sem necessidade de este estar com as mãos ocupadas, embora alguns participantes tenham sugerido gestos menores e mais discretos, e alguns participantes tenham executado os gestos com um gasto de energia desnecessário, o que poderia prejudicar o uso para longos períodos. Por fim, a fala neste ambiente apresentou grandes dificuldades, devido tanto aos erros de reconhecimento, quanto menor aceitação por não ser apropriada para o ambiente, mas com alguns dos participantes demonstrando grande interesse em utilizá-la.

Quanto à fase de definição das formas de interação, esta foi muito útil, adaptando a interação para as limitações do dispositivo disponível. Sugestões extras existentes da etapa de entrevistas ainda podem ser utilizadas em trabalhos futuros. O procedimento utilizado nesta fase não seria a melhor solução caso fosse desejado a geração de uma interação multimodal com características mais complexas, como fusão, exigindo uma adaptação para propostas de uso de modalidades sendo utilizadas ao mesmo tempo. Ainda, apenas na etapa de avaliação do sistema houve a requisição mais explícita do uso de gestos especificamente mais discretos para uso no contexto real, e muitas das sugestões geradas parecem ter sido fortemente influenciadas por sistemas já utilizados pelos usuários, sendo um procedimento limitado para gerar propostas mais criativas e potencialmente melhores. Nesta fase também já era possível a discussão do uso do sistema e a opinião dos usuários a ele relacionada, possibilitando a identificação de pontos similares aos achados na fase de avaliação do mesmo, por exemplo, quanto ao desconforto no uso de gestos e fala para a tarefa, e preferência no uso do *smartphone* para uso real. O processo seguido nesta fase e sua aplicação neste e em outro projeto de Mestrado do grupo, foi aceito para publicação no HCII2014 [22].

No momento atual, para projetos de interfaces, o *smartphone* para auxílio na manipulação dos sistemas parece a melhor opção frente aos outros dois modos comparados. A utilização de gestos mais finos, e adaptados para o contexto do sistema, parecem ser uma opção forte mesmo para uso com plateia, enquanto tecnologias de fala ainda precisam ser muito mais robustas a erros e essa modalidade apresenta resistência para uso em situações de plateia, mas com um potencial de aceitação que depende da preferência particular de cada usuário.

7.1 **Limitações do Trabalho**

Testes em um ambiente real são importantes para melhor compreender e avaliar determinadas decisões de projeto de formas de interação. Devido aos testes executados neste trabalho serem em um ambiente isolado, sem plateia, não foi avaliado a forma com que o sistema iria diferenciar entre fala e gestos de corpo que estariam sendo realizados com propósito de execução de comandos, ou como forma de comunicação com a plateia. Da mesma forma, uma real avaliação do seu uso em um ambiente real poderia apresentar outros desafios e diferentes opiniões dos participantes (tanto quem o estivesse usando, quanto quem o estivesse assistindo) quanto ao sistema.

A coleta de dados de satisfação através de uma pontuação geral, realizada neste trabalho,foi afetada pelos erros de reconhecimento apresentados pelo sistema. Para melhor coletar tais informações de forma numérica dever-se-ia dividir os aspectos de satisfação em grupos, para melhor análise, contendo fatores, por exemplo, de facilidade de uso, esforço, facilidade de aprendizagem, etc. Neste sentido, a análise qualitativa das opiniões dos participantes extraídas através de perguntas abertas foi fundamental para compreender melhor a satisfação neste trabalho.

Por fim, uma limitação do trabalho foi a forma de amostragem utilizada ser não probabilística, limitando a generalização dos resultados e possível projeção para uma população.

7.2 **Recomendações**

A seguir são listadas recomendações para trabalhos futuros na área de interfaces naturais baseado nas conclusões e execução deste trabalho:

- Executar entrevistas e/ou grupos focais antes da criação do seu sistema irá ajudar a diminuir más decisões de design. Neste trabalho as opiniões dos usuários mesmo antes da implementação do sistema já refletiam grande parte dos resultados que foram colhidos na etapa final de avaliação, indicando que muito do trabalho poderia ter sido diminuído, por exemplo, com a exclusão do uso da fala devido as suas dificuldades de tecnologia e contexto de uso.
- Para esta etapa de coleta de sugestões de interação com o sistema deve-se tomar cuidado com a influência que o entrevistador pode causar na escolha de gestos ou

comandos com o sistema. A linguagem a ser utilizada na requisição de propostas de interação com o sistema deve ser neutra, e o entrevistador deve ter cuidado em não gesticular ao mesmo tempo em que descreve o comando a ser sugerido pelos participantes. Isto aconteceu em alguns momentos durante as entrevistas e grupo focal, e pode ter levado a influenciar algumas propostas. Uma sugestão para mitigar esse problema é o uso de vídeos durante estas etapas de sugestão, para garantir que estes problemas não aconteçam e que todos os participantes recebam exatamente o mesmo contexto do trabalho e conhecimento do que devem propor.

- As interações com gestos de corpo devem ser discretas e sutis de modo a diminuir a rejeição pelos usuários. Gestos muito amplos podem causar constrangimento por parte dos usuários em tarefas que possuem pessoas próximas, e também um uso prolongado do sistema pode causar cansaço.
- A baixa precisão de reconhecimento das tecnologias de fala e gestos influencia fortemente na rejeição de uso. Existe potencial de uso dessas modalidades caso a tecnologia evolua, mas o contexto de uso também deve ser levado em consideração.

7.3 **Trabalhos Futuros**

Próximo ao final do período que este trabalho foi realizado, a Microsoft lançou uma nova versão de seu vídeo game *Xbox*, e consequentemente do seu dispositivo *Kinect*, prometendo maior precisão e responsividade [61]. Outras tecnologias ainda começam a se tornarem amplamente disponíveis prometendo grande precisão e identificação de gestos finos, como o *leapmotion* [43], ou o *Myo* [88]. É interessante o uso e teste destes novos dispositivos de forma a diminuir os erros de reconhecimento que surgiram durante este trabalho e permitir uma mais livre implementação de gestos.

Revisões iniciais do novo Kinect existentes na internet [77,89,45] demonstram que a nova versão fornece uma grande integração com todas as funções do sistema, sendo uma forte opção para uso do mesmo, possibilitando o controle com o uso de fala e gestos. Também há a opção de controle das funções do vídeo game pelo *smartphone*, oferecendo, portanto, uma gama de modalidades similares as utilizadas neste trabalho. Alguns destes *reviews* apontam para a dificuldade existente no uso destes modos de fala e gestos, vivenciadas por tais revisores, sugerindo que ainda haverá desafios pela frente na evolução de tal tecnologia,

e/ou na adaptação dos usuários as mesmas. Um trabalho futuro interessante, sobre esse sistema, poderia ser voltado ao uso dessas modalidades em um contexto de visualização de mídias em uma sala de estar, com televisões que possuem sistemas integrados, para verificar o quanto eles realmente oferecem benefícios aos usuários, ou não.

Além das questões relacionadas à tecnologia, novos testes, agora em ambientes reais, são importantes para compreender com maior precisão os problemas a serem apresentados pelo ambiente de uso. Ainda, é interessante a análise da percepção do apresentador, e também da plateia, na aceitação do uso dessas tecnologias que recentemente começaram a ser utilizadas mais abertamente.

A forma de apresentação e *feedback* dos gestos de corpo e fala são outros pontos importantes a serem melhor estudados, de forma a permitir a interação com o sistema sem atrapalhar o usuário ou a plateia na troca de informações da tarefa.

Conforme destacado ao longo deste trabalho, interações naturais, como as proporcionadas por interfaces multimodais, estão, cada vez mais, sendo disponibilizadas a diferentes perfis de usuários. Estudos referentes às formas de interações e preferências nestas interfaces são essenciais para o sucesso de sua disseminação e para sua apropriação por parte dos usuários.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Abawajy, J. H., "Human-computer interaction in ubiquitous computing environments", *International journal of pervasive computing and communications*, vol. 5-1, 2009, pp. 61-77.

[2] Allen, J.; Byron, D.; Dzikovska, M.; Ferguson, G.; Galescu, L.; Stent, A., "Towards Conversational Human-Computer Interaction", *AI Magazine*, vol. 22-4, Outubro 2001, pp. 27-37.

[3] Android Developers. "Get the Android SDK". Capturado em: http://developer.android.com/sdk/index.html, Jul 2013.

[4] Android Developers. "Using Touch Gestures". Capturado em: http://developer.android.com/training/gestures/index.html, Jul 2013.

[5] Anthony, L.; Yang, J.; Koedinger, K. R., "Evaluation of multimodal input for entering mathematical equations on the computer". In *Proceedings of CHI '05 Extended Abstracts on Human Factors in Computing Systems*, 2005, pp. 1184-1187.

[6] Apache Software Foundation. "Apache HTTP Components". Capturado em: http://hc.apache.org/index.html, Jul 2013.

[7] Arroyo, E.; Selker, T.; Stouffs, A., "Interruptions as Multimodal Outputs: Which are the Less Disruptive?" In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, 2002, pp. 479-482.

[8] Asteriadis, S.; Tzouveli, P.; Karpouzis, K.; Kollias, S., "Estimation of behavioral user state based on eye gaze and head pose - application in an e-learning environment", *Multimedia Tools and Applications*, vol. 41-3, Fevereiro 2009, pp. 469-493.

[9] Avouac, P-A.; Nigay, L.; Lalanda, P., "Towards autonomic multimodal interaction". In *Proceedings of the 1st Workshop on Middleware and Architectures for Autonomic and*

Sustainable Computing, 2011, pp. 25-29.

[10] Baddeley, A., "Working memory", *Science*, vol. 255-5044, Janeiro 1992, pp. 556-559.

[11] Baillie, L.; Schatz, R., "Exploring multimodality in the laboratory and the field". In *Proceedings of the 7th International Conference on Multimodal Interfaces*, 2005, pp. 100-107.

[12] Balbo, S.; Coutaz, J.; Salber, D., "Towards automatic evaluation of multimodal user interfaces". In *Proceedings of the 1st International Conference on Intelligent User Interfaces*, 1993, pp. 201-208.

[13] Bernhaupt, R.; Navarre, D.; Palanque, P.; Winckler, M., "Model-Based Evaluation: A New Way to Support Usability Evaluation of Multimodal Interactive Applications". In Law, E. L. C.; Hvannberg, E. T.; Cockton, G. (Eds.), *Maturing Usability: Quality in Software, Interaction and Quality*. London: Springer-Verlag, 2007, pp. 96-119.

[14] Bernsen, N. O., "Multimodality theory". In Tzovaras, D. (Ed.), *Multimodal User Interfaces: From Signals to Interaction*. Berlin: Springer-Verlag, 2008, pp. 5-29.

[15] Bernsen, N. O., "Why are Analogue Graphics and Natural Language both Needed in HCI?" In Paterno, F. (Ed.), *Interactive Systems: Design, Specification, and Verification*. Berlin: Springer-Verlag , 1995, pp. 235-251.

[16] Blattner, M. M.; Dannenberg, R. B., "CHI'90 Workshop on Multimedia and Multimodal Interface Design", *ACM SIGCHI Bulletin*, vol. 22-2, Outubro 1990, pp. 54-58.

[17] Cherubini, M.; Anguera, X.; Oliver, N.; Oliveira, R., "Text versus speech: a comparison of tagging input modalities for camera phones". In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2009, pp. 1-10.

[18] Cohen, P. R.; Dalrymple, M.; Moran, D. B.; Pereira, F. C.; Sullivan, J. W., "Synergistic Use of Direct Manipulation and Natural Language". In *Proceedings of the SIGCHI*

Conference on Human factors in Computing Systems: Wings for the Mind, 1989, pp. 227-233.

[19] Cohen, P. R.; Oviatt, S. L., "The role of voice input for human-machine communication". In *Proceedings of the National Academy of Sciences*, 1995, pp. 9921-9927.

[20] Cohen, P.; Swindells, C.; Oviatt, S.; Arthur, A., "A high-performance dual-wizard infrastructure for designing speech, pen, and multimodal interfaces". In *Proceedings of the 10th International Conference on Multimodal Interfaces*, 2008, pp. 137-140.

[21] Connectify. "Connectify - Turn your PC into a Wi-Fi Hotspot and Get Faster Internet". Capturado em: http://www.connectify.me/, Jan 2014.

[22] Cossio, L.; Lammel, F.; Silveira, M., "Towards an Interactive and Iterative Process to Design Natural Interaction Techniques". In *Proceedings of HCI International 2014*, 2014, pp. 19-23.

[23] Coutaz, J.; Nigay, L.; Salber, D.; Blandford, A.; May, J.; Young, R. M., "Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The CARE Properties". In *Proceedings of INTERACT95*, 1995, pp. 115-120.

[24] D'Andrea, A.; D'Ulizia, A.; Ferri, F.; Grifoni, P., "Multimodal pervasive framework for ambient assisted living". In *Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments*, 2009, pp. 39:1-39:8.

[25] Dillon, R. F.; Edey, J. D.; Tombaugh, J. W., "Measuring the true cost of command selection: techniques and results". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People*, 1990, pp. 19-26.

[26] Drewes, H., "Eye Gaze Tracking for Human Computer Interaction", Dissertation, Faculty of Mathematics, Computer Science and Statistics, LMU München,2010, p. 164.

[27] Elting, C.; Möhler, G., "Modeling Output in the EMBASSI Multimodal Dialog System". In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, 2002,

pp. 111-116.

[28] Farah, M. J.; Hammond, K. H.; Levine, D. N.; Calvanio, R., "Visual and Spatial Mental Imagery: Dissociable Systems of Representation", *Cognitive Psychology*, vol. 20-4, Outubro 1988, pp. 439-462.

[29] Francese, R.; Passero, I.; Tortora, G., "Wiimote and Kinect: gestural user interfaces add a natural third dimension to HCI". In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2012, pp. 116-123.

[30] Frick, R., "Using both an auditory and a visual short-term store to increase digit span", *Memory & Cognition*, vol. 12-5, Setembro 1984, pp. 507-514.

[31] Gestureworks. "Gesture Markup Language". Capturado em: http://gestureworks.com/pages/core-features-gestures, Maio 2013.

[32] Gilroy, S. W.; Cavazza, M. O.; Vervondel, V., "Evaluating multimodal affective fusion using physiological signals". In *Proceedings of the 16th International Conference on Intelligent User Interfaces*, 2011, pp. 53-62.

[33] Grudin, J., "A Moving Target: The Evolution of HCI". In Sears, A.; Jacko, J. A. (Eds.), *The Human Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. Nova Iorque: Lawrence Erlbaum Associates, 2008, pp. 1-24.

[34] Hauptmann, A. G., "Speech and gestures for graphic image manipulation". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Wings for the Mind*, 1989, pp. 241-245.

[35] Heikkinen, J.; Olsson, T.; Väänänen-Vainio-Mattila, K., "Expectations for user experience in haptic communication with mobile devices". In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2009, Artigo 28, 10 páginas.

[36] Henze, N.; Löcken, A.; Boll, S.; Hesselmann, T.; Pielot, M., "Free-hand gestures for

music playback: deriving gestures with a user-centred process". In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*, 2010, Artigo 16, 10 páginas.

[37] Jaimes, A.; Sebe, N., "Multimodal human-computer interaction: A survey", *Computer Vision and Image Understanding*, vol. 108-1-2, Outubro 2007, pp. 116-134.

[38] Jetter, H-C.; Leifert, S.; Gerken, J.; Schubert, S.; Reiterer, H., "Does (multi-)touch aid users' spatial memory and navigation in 'panning' and in 'zooming & panning' UIs?" In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2012, pp. 83-90.

[39] Jöst, M.; Häußler, J.; Merdes, M.; Malaka, R., "Multimodal interaction for pedestrians: an evaluation study". In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, 2005, pp. 59-66.

[40] Käster, T.; Pfeiffer, M.; Bauckhage, C., "Combining speech and haptics for intuitive and efficient navigation through image databases". In *Proceedings of the 5th International Conference on Multimodal Interfaces*, 2003, pp. 180-187.

[41] Kaye, J. J., "Making Scents: aromatic output for HCI", *Interactions*, vol. 11-1, Janeiro 2004, pp. 48-61.

[42] Kjeldskov, J.; Stage, J., "New techniques for usability evaluation of mobile systems", *International Journal of Human-Computer Studies*, vol. 60-5-6, Maio 2004, pp. 599–620.

[43] LeapMotion. "Leap Motion". Capturado em: https://www.leapmotion.com/, Jan 2014.

[44] Liu, Y.; Connelly, K., "Realizing an Open Ubiquitous Environment in a RESTful Way". In *Proceedings of the 2008 IEEE International Conference on Web Services*, 2008, pp. 96-103.

[45] Machinima. "Youtube - Xbox One Review! - Inside Gaming Daily". Capturado em: http://www.youtube.com/watch?v=UdrqwavoDIw, Nov 2013.

[46] Mayer, R. E., "Systematic Thinking Fostered by Illustrations in Scientific Text", *Journal of Educational Psychology*, vol. 81-2, Junho 1989, pp. 240-246.

[47] Mayer, R. E.; Anderson, R. B., "Animations Need Narrations: An Experimental Test of a Dual-Coding Hypothesis", *Journal of Educational Psychology*, vol. 83-4, Dezembro 1991, pp. 484-490.

[48] Mayer, R. E.; Gallini, J. K., "When Is an Illustration Worth Ten Thousand Words?", *Journal of Educational Psychology*, vol. 82-4, Dezembro 1990, pp. 715-726.

[49] Mazza, R., "Evaluating information visualization applications with focus groups: the CourseVis experience". In *Proceedings of the 2006 AVI workshop on Beyond time and errors: novel evaluation methods for information visualization*, 2006, pp. 1-6.

[50] McGee-Lennon, M. R.; Wolters, M.; McBryan, T., "Audio Reminders in The Home Environment". In *Proceedings of the 13th International Conference on Auditory Display*, 2007, pp. 437-444.

[51] McGlaun, G.; Althoff, F.; Lang, M.; Rigoll, G., "Towards multi-modal error management: experimental evaluation of user strategies in event of faulty application behavior in automotive environments". In *Proceedings of the Seventh World Multiconference on Systemics, Cybernetics, and Informatics*, 2003, pp. 462-466.

[52] Microsoft. "Create Grammars Using SRGS XML (Microsoft.Speech)". Capturado em: http://msdn.microsoft.com/en-us/library/hh378349(v=office.14).aspx, Nov 2013.

[53] Microsoft. "Kinect for Windows - Developer Download". Capturado em: http://www.microsoft.com/en-us/kinectforwindows/develop/developer-downloads.aspx, Jul 2013.

[54] Microsoft. "Kinect for Windows Gallery". Capturado em: http://www.microsoft.com/en-us/kinectforwindows/discover/gallery.aspx, Maio 2013.

[55] Microsoft. "Kinect for Windows". Capturado em: http://www.microsoft.com/en-us/kinectforwindows/, Jul 2013.

[56] Microsoft. "Kinect Sensor". Capturado em: http://msdn.microsoft.com/en-us/library/hh438998.aspx, Jan 2014.

[57] Microsoft. "Msdn forums - Advanced audio capabilties of Kinect and Speech Platform". Capturado em: http://social.msdn.microsoft.com/Forums/en-US/f184a652-a63f-4c72-a807-f9770fdf57f8/advanced-audio-capabilties-of-kinect-and-speech-platform?forum=kinectsdkaudioapi, Jan 2014.

[58] Microsoft. "Msdn forums - Can I tell the Kinect to only recognise speech if it meets a certain volume threshold?" Capturado em: http://social.msdn.microsoft.com/Forums/en-US/a35ba7a0-6b7e-4d56-b3c8-3118798fa1dc/kinect-speech-recognition-not-working-properly, Jan 2014.

[59] Microsoft. "Msdn forums - Kinect Speech Recognition not working properly". Capturado em: http://social.msdn.microsoft.com/Forums/en-US/a35ba7a0-6b7e-4d56-b3c8-3118798fa1dc/kinect-speech-recognition-not-working-properly, Jan 2013.

[60] Microsoft. "Power Point". Capturado em: http://office.microsoft.com/pt-br/powerpoint/, Jul 2013.

[61] Microsoft. "Xbox One - O que ele faz". Capturado em: http://www.xbox.com/pt-BR/xboxone/what-it-does, Jan 2014.

[62] Morris, M. R.; Wobbrock, J. O.; D., Wilson A., "Understanding users' preferences for surface gestures". In *Proceedings of Graphics Interface 2010*, 2010, pp. 261-268.

[63] Mousavi, S. Y.; Low, R.; Sweller, J., "Reducing Cognitive Load by Mixing Auditory and Visual Presentation Modes", *Journal of Educational Psychology*, vol. 87-2, Junho 1995, pp. 319-334.

[64] Moustakas, K.; Tzovaras, D.; Dybkjaer, L.; Bernsen, N.; Aran, O., "Using Modality Replacement to Facilitate Communication between Visually and Hearing-Impaired People", *IEEE MultiMedia*, vol. 18-2, Abril 2011, pp. 26-37.

[65] Myers, B. A., "A Brief History of Human Computer Interaction Technology", *Interactions*, vol. 5-2, Março 1998, pp. 44-54.

[66] Nielsen, M.; Störring, M.; Moeslund, T. B.; Granum, E., "A procedure for developing intuitive and ergonomic gesture interfaces for HCI". In Camurri, A.; Volpe, G. (Eds.), *Gesture-Based Communication in Human-Computer Interaction*. Berlin: Springer, 2004, pp. 409-420.

[67] Nintendo. "Nintendo Wii". Capturado em: http://www.nintendo.com/wii, Jul 2013.

[68] Norman, D. A., "Natural user interfaces are not natural", *Interactions*, vol. 17-3, Maio 2010, pp. 6-10.

[69] Oviatt, S., "Multimodal interfaces for dynamic interactive maps". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground*, 1996, pp. 95-102.

[70] Oviatt, S., "Multimodal Interfaces". In Sears, A.; Jacko, J. A. (Eds.), *Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. Nova Iorque: Lawrence Erlbaum Associates, 2008, pp. 413-432.

[71] Oviatt, S.; Cohen, P., "Multimodal Interfaces that Process What Comes Naturally", *Communications of the ACM*, vol. 43-3, Março 2000, pp. 45-53.

[72] Oviatt, S. L.; Cohen, P. R.; Fong, M.; Frank, M., "A Rapid Semi-Automatic Simulation Technique for Investigating Interactive Speech and Handwriting". In *Proceedings of The Second International Conference on Spoken Language Processing*, 1992, pp. 1351-1354.

[73] Oviatt, S.; VanGent, R., "Error resolution during multimodal human-computer interaction". In *Proceedings of Fourth International Conference on Spoken Language Processing*, 1996, pp. 204-207.

[74] Perakakis, M.; Potamianos, A., "A Study in Efficiency and Modality Usage in Multimodal Form Filling Systems", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16-6, Agosto 2008, pp. 1194-1206.

[75] Playstation. "Playstation Move Motion Controller". Capturado em: http://us.playstation.com/ps3/playstation-move/, Jul 2013.

[76] Ren, X.; Zhang, G.; Dai, G., "The Efficiency of Various Multimodal Input Interfaces Evaluated in Two Empirical Studies", *IEICE Transactions on Information and Systems*, vol. E84-D-11, Outubro 2001, pp. 1421-1426.

[77] Rev3Games. "Youtube - Xbox One REVIEW! Adam Sessler Reviews". Capturado em: http://www.youtube.com/watch?v=3Y51zatx9qs, Nov 2013.

[78] Rico, J.; Brewster, S., "Usable gestures for mobile interfaces: evaluating social acceptability". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 887-896.

[79] Ruiz, J.; Li, Y.; Lank, E., "User-defined motion gestures for mobile interaction". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 197-206.

[80] Salber, D.; Dey, A.; Abowd, G., "Ubiquitous computing: Defining an HCI research agenda for an emerging interaction paradigm", GVU Technical Report, Georgia Institute of Technology, 1998.

[81] Samsung. "Smart TV". Capturado em: http://www.samsung.com/us/2013-smart-tv/, Maio 2013.

[82] Sarter, N. B., "Multimodal information presentation: Design guidance and research challenges", *International Journal of Industrial Ergonomics*, vol. 36-5, Maio 2006, pp. 439-445.

[83] Schapira, E.; Sharma, R., "Experimental evaluation of vision and speech based multimodal interfaces". In *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, 2001, pp. 1-9.

[84] Schmidt, A.; Kranz, M.; Holleis, P., "Interacting with the ubiquitous computer: towards embedding interaction". In *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies*, 2005, pp. 147-152.

[85] Starker, I.; Bolt, R. A., "A gaze-responsive self-disclosing display". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People*, 1990, pp. 3-10.

[86] Suhm, B.; Myers, B.; Waibel, A., "Multimodal error correction for speech user interfaces", *ACM Transactions on Computer-Human Interaction*, vol. 8-1, Março 2001, pp. 60-98.

[87] Sun, Q.; Lin, J.; Fu, C-W.; Kaijima, S.; He, Y., "A multi-touch interface for fast architectural sketching and massing". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 247-256.

[88] ThalmicLabs. "Myo - Gesture control armband by Thalmic Labs". Capturado em: https://www.thalmic.com/en/myo/, Jan 2014.

[89] The Verge. "Youtube - Xbox One review". Capturado em: http://www.youtube.com/watch?v=vRM77-7EsY8, Nov 2013.

[90] Vatavu, R-D., "User-defined gestures for free-hand TV control". In *Proceedings of the 10th European conference on Interactive tv and video*, 2012, pp. 45-48.

[91] Vernier, F.; Nigay, L., "A framework for the combination and characterization of output modalities". In *Proceedings of the 7th International Conference on Design, Specification, and Verification of Interactive Systems*, 2000, pp. 35-50.

[92] W3C. "Speech Recognition Grammar Specification Version 1.0". Capturado em: http://www.w3.org/TR/speech-grammar/, Nov 2013.

[93] Warnock, D., "A Subjective Evaluation of Multimodal Notifications". In *Proceedings of Pervasive Health*, 2011, pp. 461-468.

[94] Weiser, M., "The Computer for the 21st Century", *Scientific American*, vol. 265-3, Setembro 1991, pp. 94-104.

[95] Wobbrock, J. O.; Morris, M. R.; Wilson, A. D., "User-defined gestures for surface computing". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 1083-1092.

ANEXO A – TERMO DE CONSENTIMENTO

Apoio à avaliação de critérios de interação humano-computador de sistemas computacionais

Protocolo de Pesquisa Registro CEP 11/05667

Faculdade de Informática/PUCRS
Avenida Ipiranga, 6681 – Prédio 32 - 90619-900 – Porto Alegre – RS
Tel: (51) 3320-3558

Termo de Consentimento Livre e Esclarecido

A PUCRS, através das equipes do **Laboratório de Usabilidade e de Acessibilidade e de Realidade Virtual** da Faculdade de Informática**,** agradece a todos os participantes de testes realizados sob sua responsabilidade, a inestimável contribuição que prestam para o avanço da pesquisa sobre Interação Humano-Computador.

O objetivo desta pesquisa é investigar questões relacionadas à usabilidade e/ou acessibilidade de sistemas interativos. Para isto, os participantes dos testes são convidados a usar diferentes sistemas interativos, enquanto são observados por um ou mais pesquisadores. Esta observação será registrada em papel e, também, através de um software de captura de telas, que armazena tudo o que acontece na tela do computador, e, eventualmente, através de vídeo. Estas informações nos trarão dados importantíssimos para verificar a qualidade dos sistemas em questão.

Lembramos que o objetivo deste estudo **não é** avaliar o participante, **mas sim** avaliar o aplicativo computacional que o participante estará usando durante os testes. O uso que se faz dos registros efetuados durante o teste é **estritamente** limitado a atividades de pesquisa e desenvolvimento, garantindo-se para tanto que:

1. O anonimato dos participantes será preservado em todo e qualquer documento divulgado em foros científicos (tais como conferências, periódicos, livros e assemelhados) ou pedagógicos (tais como apostilas de cursos, *slides* de apresentações, e assemelhados).
2. Todo participante terá acesso a cópias destes documentos após a publicação dos mesmos.
3. Todo participante que se sentir constrangido ou incomodado durante uma situação de teste pode interromper o teste e estará fazendo um favor à equipe se registrar por escrito as razões ou sensações que o levaram a esta atitude. A equipe fica obrigada a descartar o teste para fins da avaliação a que se destinaria.
4. Os participantes que forem menores de idade terão, obrigatoriamente, que apresentar o consentimento de seu responsável, para participação no estudo, o qual será declarado ciente do estudo a ser realizado através de sua assinatura no presente Termo de Compromisso.
5. Todo participante tem direito de expressar por escrito, na data do teste, qualquer restrição ou condição adicional que lhe pareça aplicar-se aos itens acima enumerados (1, 2, 3 e 4). A equipe se compromete a observá-las com rigor e entende que, na ausência de tal manifestação, o participante concorda que rejam o comportamento ético da equipe somente as condições impressas no presente documento.
6. A equipe tem direito de utilizar os dados dos testes, mantidas as condições acima mencionadas, para quaisquer fins acadêmicos, pedagógicos e/ou de desenvolvimento contemplados por seus membros.

[a ser preenchido pelo observador]
Sistema:_____ Data: __ / __ / ____

Condições especiais (caso não haja condições especiais, escreva "nenhuma"):

_____
_____
_____
_____
_____
_____

☐ continua no verso

Por favor, indique sua posição em relação aos termos acima:

☐ Estou de pleno acordo com os termos acima.
☐ Em anexo registro condições adicionais para este teste.

_____
Assinatura do participante

_____
Assinatura do responsável
(caso o participante seja menor de idade)

_____
Assinatura do observador

Nome do Participante: _____

Nome do Responsável (se o participante for menor de idade): _____

Pesquisador Responsável: Prof. _____ – Faculdade de Informática - PUCRS