

UNIVERSIDADE DE LISBOA
Faculdade de Ciências
Departamento de Informática



**INTERACÇÃO GESTUAL SEM SUPERFÍCIES DE
APOIO**

Joana Raimundo Neca

MESTRADO EM ENGENHARIA INFORMÁTICA
Especialização em Sistemas de Informação

2011

UNIVERSIDADE DE LISBOA
Faculdade de Ciências
Departamento de Informática



**INTERACÇÃO GESTUAL SEM SUPERFÍCIES DE
APOIO**

Joana Raimundo Neca

DISSERTAÇÃO

Projecto orientado pelo Prof. Doutor Carlos Alberto Pacheco dos Anjos Duarte

MESTRADO EM ENGENHARIA INFORMÁTICA
Especialização em Sistemas de Informação

2011

Agradecimentos

Agradeço ao Professor Carlos Duarte pela sua orientação durante todo o projecto. Pelo seu apoio, motivação, paciência, compreensão, transmissão de ideias e boa disposição contagiante.

À Faculdade, ao LASIGE e em especial ao HCIM por todas as condições oferecidas durante a execução do projecto e pela troca de ideias que surgiu com os membros do grupo e em particular ao Tiago Reis.

Em especial ao meu pai, mana, avós e namorado pelo apoio de sempre.

Aos amigos e colegas da Faculdade por todos os bons momentos.

Agradeço ao Rafael Nunes pela ajuda no desenvolvimento do protótipo referente ao Google Earth.

Para o meu pai

Resumo

Os periféricos de entrada deixaram de ser a única forma de transmitir intenções à máquina, sendo agora possível fazê-lo com o próprio corpo. Dispositivos que permitem interacção gestual sem recurso a periféricos intermediários têm vindo a aumentar, principalmente na área dos jogos. Esta tendência levanta várias questões a serem investigadas na área da interacção pessoa-máquina.

A aproximação simplista de transferir conceitos de interacção do paradigma clássico WIMP, baseado nos dispositivos tradicionais de entrada, rato e teclado, rapidamente conduz a problemas inesperados. As características de uma interface concebida para uma interacção gestual em que não há contacto com nenhum dispositivo de entrada não se irão adequar ao paradigma utilizado nos últimos 40 anos. Estamos assim em condições de explorar como a interacção gestual com ou sem voz pode contribuir para minimizar os problemas com o paradigma clássico WIMP no tipo de interacção em que não há o contacto com nenhum periférico.

Neste trabalho irá ser explorado o campo da interacção gestual, com ou sem voz. Através de aplicações pretende-se conduzir vários estudos de manipulação de objectos virtuais baseada em visão computacional. A manipulação dos objectos é realizada com dois modos de interacção (gestos e voz) podendo estes surgir integrados ou não. Pretende-se analisar se a interacção gestual é apelativa para os utilizadores para alguns tipos de aplicações e acções, enquanto para outros tipos, os gestos poderão não ser a modalidade preferida de interacção.

Palavras-chave: Gestos com ou sem voz, Aplicações para manipulação de objectos, Estudos de Utilizadores, Tecnologia Perceptual

Abstract

The input peripherals aren't anymore the only way to transmit intentions to the machine, being now possible to do it with our own body. The number of devices that allow gestural interaction, without the need of intermediate peripherals, are increasing, mainly in the area of video games. This tendency raises several questions that need to be investigated in the area of person-machine interaction.

The simplistic approach of transferring interaction concepts from the classic paradigm WIMP, based on the traditional input devices, mouse and keyboard, quickly leads to unexpected problems. The characteristics of an interface conceived to a gestural interaction were there isn't any kind of contact with an input device won't suit with the paradigm of the last 40 years. So we're in conditions to exploit how the gestural interaction can contribute to minimize the classic paradigm issues.

In this work the field of gestural interaction, with and without voice, will be analyzed. Through the use of applications, it's intended to lead various studies of virtual objects manipulation based on computational vision. The objects manipulation is done with two kinds of interactions, gestural and voice, that may emerge integrated our not.

It's intended to analyze if the gestural interaction is appealing to the users for some kind of applications and actions, while for other types, gestural may not be the preferred interaction modality.

Keywords: Gestures, Objects Manipulation for Applications, Users Studies, Perceptual Technologies

Conteúdo

Lista de Figuras	xv
Lista de Tabelas	xvii
1 Introdução	1
1.1 Motivação	2
1.2 Objectivos	2
1.3 Contribuições	3
1.4 Estrutura do documento	4
2 Trabalho relacionado	5
2.1 Contexto e Conceitos	5
2.1.1 Paradigma WIMP	5
2.1.2 Reconhecimento Gestual	6
2.1.3 Interfaces Multimodais	7
2.2 Interação Gestual	10
2.2.1 Estilo de Gestos	10
2.2.2 Aplicabilidade	11
2.2.3 Aspectos de Interação	12
2.3 Discussão	17
3 Avaliação de protótipos de interacção gestual com e sem voz	19
3.1 Enquadramento e Preparação do estudo	19
3.2 Cenários e Acções	20
3.3 Participantes	21
3.4 Procedimento	21
3.5 Detalhes técnicos	21
3.6 Análise dos resultados	22
3.6.1 Primeiro cenário de interacção: interacção puramente gestual . . .	22
3.6.2 Segundo cenário de interacção: interacção gestual com voz: . . .	26
3.6.3 Comparação dos dois cenários de interacção	30
3.6.4 Discussão	34

4	Comparação de dois cenários de interacção	37
4.1	Enquadramento e Preparação do estudo	37
4.2	Cenários e acções	38
4.3	Gestos e comandos de voz padrão	39
4.4	Participantes	43
4.5	Procedimento	43
4.6	Detalhes técnicos	44
4.7	Análise dos resultados	45
4.7.1	Interacção gestual com voz:	45
4.7.2	Comparação dos dois cenários de interacção	49
4.7.3	Discussão	55
5	Conclusão e Trabalho Futuro	59
5.1	Conclusão	59
5.2	Trabalho futuro	61
	Bibliografia	74
	Índice	75

Lista de Figuras

2.1	Sistema Multimodal “Put that there” - Bolt, 1980 [7]	9
2.2	Protótipo QuickSet - Cohen, 1997 [7]	9
2.3	Pie-menu(esquerda) e a selecção equivalente usando marking-menu(direita) [51]	13
2.4	Gestos usados no sistema de making menu para navegação web [51]	13
2.5	Conjunto de dedos ou dedo e as suas categorias de menu [50]	13
2.6	Itens de um menu [50]	13
2.7	Feedback para a estratégia de selecção Apontar e Esperar [70]	14
2.8	Feedback para a estratégia de selecção Apontar e Agitar [70]	14
3.1	Aplicações de teste: (1)“Mesa” com imagens, (2)“Parede” de imagens	20
4.1	Aplicações de teste: (1)“Mesa” com imagens, (2)“Parede” de imagens, (3)Google Earth	39
4.2	Gesto de apontar (1), gesto de selecção (2)	47
4.3	Sequência de imagens que alguns participantes usaram para transmitir a acção apagar (1)mão fechada para seleccionar o objecto (2)mão aberta para a imagem ser fechada	49

Lista de Tabelas

3.1	Gestos padrões para cada acção na aplicação mesa de imagens	23
3.2	Gestos padrões para cada acção na aplicação parede de imagens	24
3.3	Generalização do gesto zoom in	25
3.4	Comandos padrão	27
3.5	Número de participantes que apontaram e gesticularam para cada acção .	29
3.6	Informação não repetida para cada acção, comandos de voz e nº de pessoas	31
3.7	Classificação média (1 a 5) para cada acção em cada cenário	32
3.8	Tempo médio (em segundos) para cada acção em cada cenário	33
4.1	Gestos e comandos de voz padrões para cada acção na aplicação mesa de imagens	40
4.2	Gestos e comandos de voz padrões para cada acção na aplicação parede de imagens	41
4.3	Gestos padrões para cada acção na aplicação Google Earth	42
4.4	Número de participantes que apontaram e gesticularam para cada acção .	46
4.5	Número de participantes que executaram o gesto de selecção ou apenas apontaram	48
4.6	Número de participantes que recorreu à folha auxiliar, para cada acção . .	50
4.7	Gestos alternativos para a acção de apagar, no cenário de interacção com gestos e voz	51
4.8	Tempo médio (em segundos) para cada acção em cada cenário	52
4.9	Classificação média (1 a 5) para cada acção em cada cenário	54
4.10	Mão dominante do participante consoante o cenário de interacção e a aplicação teste	55

Capítulo 1

Introdução

O termo computação é definido, cada vez mais, pela forma como ocorre a interacção do utilizador com a informação, dissociado do contexto tradicional a que nos habituámos em que a interacção pessoa-máquina era baseada no rato e no teclado.

Com o aumento da importância entre a interacção e o utilizador surge cada vez mais há um maior entusiasmo por novas interfaces, como as interfaces gestuais (tanto baseadas em multi-toque como em visão computacional). O reconhecimento de gestos nas interfaces é uma alternativa ao uso dos tradicionais menus, teclado e rato abrindo caminho para novas abordagens de manipulação de informação. Em consequência tem-se verificado um crescimento na investigação sobre formas alternativas de interagir com a tecnologia com o objectivo de aumentar a capacidade humana de expressar uma ideia e/ou melhorar a capacidade de absorver informação. Uma boa interpretação da linguagem corporal é fundamental para um sistema eficiente na recepção da ideia do utilizador.

Assim, os movimentos do corpo são usados para transmitir informação entre as pessoas e estão fortemente acoplados à fala ou a comandos co-verbais. Gesto e fala complementam-se muitas vezes na comunicação humana diária fornecendo um maior nível de fluidez se forem usados num sistema interactivo.

Temos como exemplo o filme, *Minority Report*, que apresentou dezenas de novas ideias e tecnologias que poderiam tornar-se realidade em 2054. Apenas com os movimentos das mãos (e.g. arrastar e soltar, aumentar) John Anderton acedia à informação. Felizmente, hoje em dia as interfaces gestuais estão cada vez mais em voga (e.g. consolas Microsoft Xbox com Kinect, Nintendo Wii, PlayStation 3 com o Move, touch-screens).

Neste trabalho iremos explorar a manipulação de objectos virtuais, em que o nosso principal objectivo é contribuir para uma melhoria na interacção do utilizador com o sistema, de forma a se obter uma interacção natural e intuitiva, assim como a alcançada no filme. Iremos explorar dois cenários de interacção, gestos com ou sem voz, em que o último será de uma forma complementar ao primeiro, em que não existe nenhuma ponte de comunicação física entre o utilizador e o sistema. Primeiramente, tentaremos encontrar quais os gestos e comandos de voz padrões para determinadas acções e posteriormente

confirmaremos o modo natural e intuitivo dos mesmo, tentando sempre perceber as vantagens de cada modalidade de interacção.

1.1 Motivação

A ideia de manipular directamente, através do toque e sem recurso a periféricos intermediários, dados digitais é bastante apelativa para os utilizadores. A apetência inata que os humanos têm para tocar e gesticular, aumenta a curiosidade para interagir com dispositivos tecnológicos e diminui a curva de aprendizagem.

Associando eliminação dos periféricos de entrada e interacção gestual, surgem variados contextos e cenários de utilização que várias consolas (e.g. consolas Microsoft Xbox com Kinect, Nintendo Wii, PlayStation 3 com o Move) conseguem proporcionar. Obviamente, que estes cenários e contextos aplicam-se aos jogos. Saindo do âmbito dos jogos mas continuando com a possibilidade de interagirmos directamente com o nosso corpo leva que as técnicas de interacção não estejam adequadas à nova realidade. É necessário então fornecer aos utilizadores mecanismos de interacção alternativos que permitam atingir um nível de usabilidade, pelo menos, semelhante ao atingido através de periféricos de entrada, e, sempre que possível, aumentar esse nível de usabilidade. A usabilidade de uma interface gestual está sempre dependente do contexto e cenário de utilização. Há uma necessidade de adaptação e inovação do paradigma WIMP. Este paradigma, com quem temos interagido nos últimos 40 anos, terá que mudar.

A mudança no paradigma de interacção requer o estudo de formas de melhorar a interacção através da adaptação de alguns conceitos subjacentes ao paradigma ou então criar um novo paradigma adaptado às novas necessidades, de forma a alcançar uma interacção mais simples e mais intuitiva nos sistemas perceptuais.

1.2 Objectivos

Neste trabalho pretende-se explorar as possibilidades oferecidas através da interacção gestual sem superfícies de apoio, complementada, ou não por comandos de voz. No seguimento do que foi exposto na secção anterior, a interacção gestual, principalmente sem superfície de apoio, não foi ainda caracterizada em detalhe suficiente. Como tal, o objectivo principal desta tese é contribuir para essa caracterização. Para isso, a primeira fase do trabalho irá focar-se no estudo da forma como o utilizador interage, utilizando gestos com ou sem comandos de voz. Vários estudos serão conduzidos com vista a atingir os dois grandes objectivos: primeiro, permitir uma caracterização da interacção sem restrições durante a realização de tarefas típicas para diversos cenários de estudo; segundo, validar o resultado da caracterização realizada, através da implementação de protótipos que incorporem o conhecimento adquirido. Numa primeira fase, o primeiro grande objectivo, um

estudo irá contribuir para compreender como é que os utilizadores abordam a interacção, tendo apenas uma modalidade de entrada, gestos, ou com a junção das duas modalidades de entrada, gesto e voz podendo-se analisar diversos factores:

- quais os gestos mais aptos para cada acção;
- quais os comandos de voz mais indicados para cada acção;
- preferência de complementar os gestos com voz para determinadas acções;
- vantagens e desvantagens da interacção com as duas modalidades de entrada ou com apenas uma modalidade de entrada.

Numa segunda fase, como já referido, serão realizadas as validações da primeira fase através da implementação dos gestos e comandos de voz nos protótipos para permitirem a interacção dos participantes.

1.3 Contribuições

Deste trabalho resultaram as seguintes contribuições:

1. Estudo de forma a obter um conjunto de gestos e comandos de voz padrão mais apropriados e tentar perceber quais as acções mais intuitivas, tentando colmatar quando não o são com o apoio da voz;
2. Definição de um mapeamento de acções/gestos e de acções/comandos de voz, para serem utilizados nos cenários mais comuns de interacção individual sem superfície de apoio, no contexto de uma interacção com uma superfície de projecção de grandes dimensões;
3. Estudo de comparação entre os dois cenários de interacção, interacção gestual com ou sem voz;
4. Três aplicações de teste que implementam os conjuntos de gestos e comandos de voz anteriormente definidos, permitindo interacção gestual com ou sem voz sem superfícies de apoio;
5. Conjunto de orientações para o desenho e aplicações que utilizem a interacção gestual com ou sem voz.

O trabalho realizado proporcionou contribuições para a comunidade científica na forma de artigos científicos, nomeadamente:

1. Neca, J., Duarte, C. (2011) **Evaluation of Gestural Interaction with and without Voice Commands**. Proceedings of IHCI 2011 – IADIS International Conference Interfaces and Human Computer Interaction 2011, Rome, Italy, 2011

2. Gomes, T., Duarte, C., Carriço, L., Neca, J., Reis, T. (2010) **Conjuntos de Gestos de Comando para Ferramentas de Desenho em Dispositivos sem Teclado**. Proceedings of 4th Conferência Nacional em Interação Pessoa-Máquina (Interação 2010), Aveiro, Portugal

1.4 Estrutura do documento

Este documento está organizado da seguinte forma:

- **Capítulo 2 – Trabalho relacionado:** É realizada uma contextualização do tema e dos conceitos em que o trabalho se baseia, fornecendo uma visão global sobre os temas relevantes neste trabalho. Começamos primeiramente por apresentar os contextos e conceitos mais relevantes (reconhecimento gestual, paradigma WIMP, visão computacional e interfaces multimodais); depois deparamo-nos com outra secção neste capítulo em que expomos o encontrado na literatura em relação à interação gestual, tema que merece ênfase no trabalho, fazendo uma abordagem ao seu significado, mostrando os estilos de gestos, a sua aplicabilidade e os seus aspectos de interação.
- **Capítulo 3 – Avaliação de protótipos de interação gestual com e sem voz:** Neste capítulo é apresentado um estudo sobre interação gestual com e sem o auxílio da voz, em superfícies de grande dimensão e sem qualquer tipo de contacto físico entre o utilizador e os periféricos de entrada. O objectivo deste estudo foi perceber as diferenças que os participantes mostram em dois cenários distintos (gestos com ou sem voz) na manipulação de objectos em duas aplicações de teste de forma a executarem acções.
- **Capítulo 4 – Comparação de dois cenários de interação:** Este capítulo apresenta um estudo que explora o mapeamento entre os gestos e os comandos de voz para as acções, resultantes do capítulo anterior. São medidos diversos valores (tempo, classificação, número de vezes que acedem à folha explicativa, etc.). É possível verificar se o mapeamento previamente adquirido foi apropriado ou não e qual a maneira mais adequada de transmitir a informação.
- **Conclusão e trabalho futuro:** Por último, neste capítulo são apresentadas as conclusões do trabalho e as perspectivas de trabalho futuro.

Capítulo 2

Trabalho relacionado

Este capítulo apresenta uma contextualização do tema e dos conceitos em que se baseia o trabalho e um resumo da pesquisa da investigação efectuada na literatura existente, quer sobre interacção gestual, quer sobre interacção gestual sem superfícies de apoio com ou sem comandos de voz.

2.1 Contexto e Conceitos

A interacção gestual proporciona aos seus utilizadores uma experiência natural e intuitiva de interacção com sistemas computacionais, em domínios aplicativos muito vastos. Este trabalho foca-se em particular na interacção gestual sem superfícies de apoio, procurando contribuir para melhorar a sua usabilidade. De seguida apresentam-se com maior detalhe alguns dos conceitos necessários para melhor compreender este trabalho:

2.1.1 Paradigma WIMP

Em 1980 surgiu o paradigma de interacção WIMP (*Windows, Icon, Menu, Pointing Device*) também classificado como paradigma clássico, concebido por Merzouga Wilberts [81].

De uma forma genérica, neste tipo de interacção, a informação encontra-se organizada em janelas e é representada por ícones. Os comandos disponíveis são organizados em conjuntos de menus acessíveis através do dispositivo apontador, normalmente um rato, reduzindo-se assim a carga cognitiva necessária para lembrar os comandos disponíveis e o tempo de aprendizagem. Outro benefício deste estilo de interacção é que se consegue alcançar uma abstracção dos espaços de trabalho, documentos e das suas acções, recorrendo-se a analogias (e.g. analogia dos documentos com folhas de papel ou pastas do mundo real), e consequentemente há uma maior facilidade de utilização para os utilizadores menos experientes.

As vantagens do uso do paradigma clássico explicam porque é que este se tornou prevalente até à actualidade, apesar de alguns investigadores reconhecerem a sua per-

manênciã como uma falta de inovaçã na procura de novos modelos de interacçã e um sinal de estagnaçã no design de interfaces com o utilizador[81].

No entanto, tem-se assistido a um crescimento de dispositivos que possibilitam a interacçã gestual com ou sem voz, pelo que os periféricos de entrada deixaram de ser o único meio de interacçã possível. O paradigma WIMP deixa de ser o mais adequado para uma interacçã direccionada a este tipo de dispositivos. Surge assim a necessidade de estudar formas de melhorar a interacçã gestual com ou sem voz, quer seja através da adaptaçã de alguns conceitos subjacentes ao paradigma ou criando um novo paradigma, adaptado às novas necessidades e mais actual.

2.1.2 Reconhecimento Gestual

O reconhecimento gestual preocupa-se em desenvolver algoritmos capazes de interpretar os gestos. Os gestos podem provir de qualquer movimento ou estado corporal, mas geralmente são originados na mão.

Com o reconhecimento de gestos podemos enriquecer a interacçã entre os humanos e os computadores, fazendo com que os últimos compreendam a linguagem corporal humana, obtendo-se assim, uma relaçã mais rica entre ambos, não limitando a entrada só ao teclado e ao rato.

Nas abordagens clássicas, o reconhecimento gestual é executado através de entradas não-perceptuais, isto é, através de dispositivos ou objectos que necessitam de contacto físico. Temos como exemplos: rato, caneta, toque, ecrãs com sensores de pressã e periféricos com sensores electrónicos (vestuário, luvas, objectos com sensores embebidos e interfaces tangíveis). A vantagem que o exemplo de toque permite é um estilo de interacçã mais natural do que a utilizaçã dos dispositivos intermédios (e.g. rato) [28] [89].

Outras abordagens [7] [40] [37] [43], que saem fora do âmbito das tradicionais, são conhecidas como entradas perceptuais, não dependendo de contacto físico com os dispositivos de entrada. Assim, de forma a interpretarem a linguagem de sinais executam o reconhecimento gestual através da utilizaçã de câmaras e algoritmos de visã computacional.

A última abordagem será mais aprofundada porque o nosso trabalho recai nesta opção. Nesse tipo de abordagem a câmara é usada para “sentir” a presençã de um utilizador, podendo ser usada para rastrear a mão e controlar o cursor ou executar comandos com gestos auxiliados por vezes pela voz [31]. A visã computacional é a tecnologia das máquinas que “vêm”, onde o verbo ver neste caso significa que a máquina é capaz de extrair informaçães de uma imagem.

A interacçã humano-computador baseada na visã computacional encontra-se suficientemente madura para substituir dispositivos de legado físico, como por exemplo, o rato e o joystick numa série de aplicaçães diferentes [32]. O facto da sua maturidade na

substituição de dispositivos físicos juntamente com o seu baixo custo [90], [33] [41] , poderá provir num maior crescimento de interesse por este tipo de interacção.

Interfaces baseadas em visão remontam ao início dos anos 80: o rastreamento e reconhecimento do gesto humano (gesto da mão, expressão facial ou gesto do corpo) têm sido estudados na comunidade da visão computacional desde os primeiros dos trabalhos [85] [5] . Para simplificar a detecção e o rastreamento baseado em visão computacional, uma solução adoptada no passado consiste na colocação de marcadores visuais ou, em condições de iluminação pobres, Infrared Light Emitting Diodes (IR-LED) nas mãos e/ou nos dedos [62] [18]. Outra abordagem consiste na detecção das mãos “nuas”, aplicando técnicas de segmentação baseada em cores [9] [77]: as regiões que apresentam uma distribuição de cor semelhante à da pele humana são extraídas para uma cena. A principal desvantagem na aplicação destas técnicas é como identificar essas regiões da imagem, dado que a cor da pele não é uniforme e varia de utilizador para utilizador e com as condições de iluminação do ambiente. Para ultrapassar estes problemas são usados artifícios ou elementos adicionais, como por exemplo, o utilizador ter de usar uma luva de tecido com uma cor uniforme. A mão pode ser detectada por meio de técnicas de subtração da imagem de fundo, onde o fundo é removido pelo processamento sucessivo de frames na sequência do vídeo [82], ou, como no caso de visão estéreo [35] [25], por frames adquiridas por várias câmaras a partir de diferentes pontos de vista. Quando o fundo desejado estiver removido a mão é procurada nas frames pré-processadas. Novamente deparamo-nos com limitações, especialmente se a cena tiver um fundo complexo e/ou a iluminação for variável. Outra abordagem é baseada no modelo 3D da mão [23] [64]. Estas técnicas dão uma estimativa da posição da mão do utilizador, combinando o modelo 3D com imagens 2D adquiridas por uma ou mais câmaras. Eventualmente, pode ser necessário que a cena tenha de ser adquirida a partir de vários pontos de observação. Esses métodos são geralmente afectados por problemas de oclusão e envolvem elevado custo computacional. No que diz respeito à detecção de dedos, há várias técnicas de detecção que fazem uso do conhecimento da geometria do objecto que está sendo procurado, como por exemplo o comprimento e a largura dos dedos. Muitos outros métodos de monitorização e análise de algoritmos de gestos da mão podem ser encontrados em [61] [87] .

O nosso trabalho vai basear-se nas abordagens perceptuais, devido à maturidade que estas abordagens já conseguiram e também pelo facto do reconhecimento gestual ser de baixo custo. Optámos assim pela utilização do Kinect por essa razão e pelas possibilidades que oferece, sendo assim possível explorar novos pontos de vista de interacção.

2.1.3 Interfaces Multimodais

Diversos estudos confirmam que as pessoas preferem utilizar múltiplas modalidades para a manipulação de objectos virtuais [29] [57]. Hauptmann [29], em 1993, conclui que 71%

dos indivíduos estudados preferem utilizar as mãos e a voz para controlar objectos do que uma modalidade isolada. Passados quatro anos, Oviatt [57] também demonstrou que 95% das pessoas tendem a utilizar gestos juntamente com a voz para a tarefa de manipulação de mapas. Cohen, em 1989 [Cohen89], mostrou que as modalidades também podem complementar-se, ou seja, gestos manuais são considerados ideais para manipulação directa de objectos e a linguagem natural tem melhor aplicação em tarefas descritivas.

A utilização simultânea de vários canais sensoriais, ou modos de comunicação, tais como visão e audição, aumentam a capacidade humana de absorção e troca de informação e evitam que apenas um canal seja sobrecarregado [17]. Temos assim um aumento do número de tarefas realizadas pelo utilizador, existindo assim uma redução no tempo necessário para completar uma tarefa e também uma redução no esforço do utilizador, e as aplicações proporcionam maior satisfação aos seus utilizadores [56].

As interfaces multimodais suportam de maneira flexível a comunicação humano-computador e podem permitir que os utilizadores, com diferentes preferências e níveis de habilidade, escolham o modo como interagem. Por exemplo, um utilizador que tenha uma deficiência visual ou motora pode sentir-se mais confortável em utilizar a voz para interagir. Por sua vez um utilizador que tenha problemas na fala ou na audição pode sentir-se mais confortável numa interacção gestual. Quando a interacção é feita de modo multimodal a linguagem do utilizador é muitas vezes simplificada [55]. Por exemplo, se um utilizador quer criar uma linha verde no mapa isso pode ser feito com um comando verbal como “criar uma linha verde com as coordenadas 9 e 20”, em contrapartida a mesma tarefa pode ser realizada com o comando verbal “linha verde” e simultaneamente, desenhando o gesto de uma linha na localização desejada. Quando se utiliza um comando multimodal que envolve reconhecimento de voz apenas com duas palavras, enquanto o equivalente comando unimodal envolve o reconhecimento de uma expressão complexa de dez palavras, podemos perceber que a linguagem é mais simplificada na presença de interacção multimodal e será mais simples para um reconhecedor de voz atingir melhores resultados.

Aplicabilidade

Nos parágrafos seguintes apresentaremos uma visão geral sobre algum dos mais relevantes sistemas que gozem da junção do gesto e voz, área relevante para o nosso trabalho, em que estas interfaces se inserem.

Na década de 80 surgiu uma das primeiras demonstrações do conceito de interfaces multimodais, o “Put-That-There” [7], que usa a fala e gestos baseados em apontar, permitindo criar e mover objectos num espaço 2D, apontando para eles e usando comandos de voz, sendo possível visualizar o sistema na figura 2.1.

Passados 9 anos surgiram outras grandes contribuições na área. O CUBRICON [52] utiliza a fala com gestos deiticos e expressões gráficas numa aplicação de mapas.

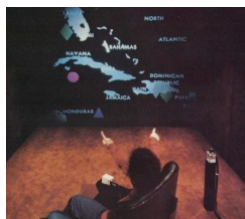


Figura 2.1: Sistema Multimodal “Put that there” - Bolt, 1980 [7]

Mais tarde, as maiores contribuições baseiam-se na integração da fala, olhar, e gestos da mão como nos trabalhos de Koons [42] e a plataforma Quickset [15].

No trabalho desenvolvido por Koons [42] foram estudadas três modalidades de entrada: olhar, fala e gestos da mão num sistema gráfico 3D intitulado “o mundo de blocos”.

Em 1994 surgiu o primeiro protótipo QuickSet [15]. Este é um sistema multimodal de colaboração que permite a vários utilizadores criarem e controlarem simulações militares, para interagir com aplicações distribuídas (figura 2.2). Utiliza a interacção baseada na caneta e na fala.

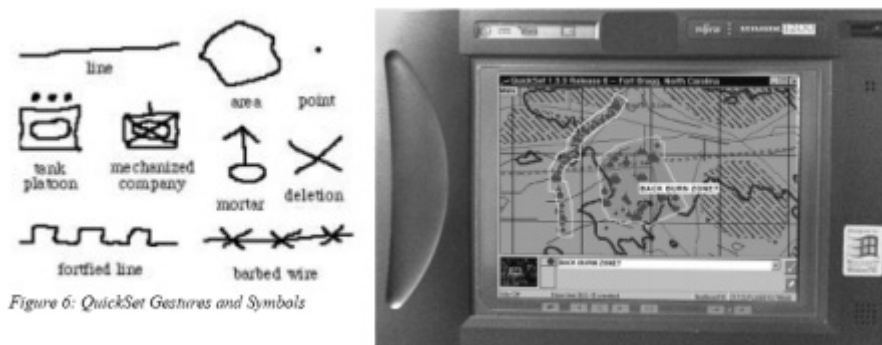


Figura 2.2: Protótipo QuickSet - Cohen, 1997 [7]

Em 2005 foi apresentado um sistema multimodal para um directório corporativo e para mensagens interactivas [34]. Estamos perante uma interacção por caneta e por fala.

Cada vez mais surgem aplicações multimodais como os sistemas de entretenimento, por exemplo: NintendoWii [71] ou Microsoft Kinect.

No entanto, em todos os sistemas referidos anteriormente apenas são focados aspectos tecnológicos na sua literatura. Não temos conhecimento de factores como: a adaptação do utilizador, se os gestos (quando existentes, sem ser de apontar) são adequados, assim como os comandos de voz. Deparamo-nos com uma grande carência de estudos em relação aos sistemas que têm como modalidades de entrada gestos e voz, desconhecendo se estes eram naturais e intuitivos. É esta carência de estudos que pretendemos colmatar e dar uma contribuição com os conhecimentos adquiridos nos estudos a realizar na área perceptual.

2.2 Interação Gestual

A interação gestual como meio de interagir com os computadores não é novidade, tendo sido o foco de bastante investigação ao longo das últimas décadas. Quase todas as formas possíveis de gestos humanos são encontrados na literatura, como meio de fornecer uma forma natural e intuitiva de interagir com computadores, na maioria, senão em todos os domínios da informática.

Nesta secção os gestos são discutidos pelo seu tipo e são considerados os diferentes domínios aplicativos onde os gestos têm sido utilizados assim como os estudos que se podem realizar nesses domínios.

2.2.1 Estilo de Gestos

A maioria dos sistemas que utilizam gestos depende de conjuntos de posições da cabeça, corpo ou mão que são pré-associados a acções que têm que ser aprendidos pelos utilizadores. No contexto da interação gestual há várias classificações que podem ser propostas [19]. Neste trabalho vamos distinguir quatro classes de gestos comunicativos: deícticos, manipulativos, semafóricos e gesticulados.

Os gestos deícticos servem para fornecer o contexto ou informações explicativas. Envolvem apontar para o objecto de forma a precisar a sua identidade (e.g. apontar para o objecto do tema de conversa) ou a sua localização espacial (e.g. apontar para a direcção de uma acção a ser tomada) dentro do contexto do domínio da aplicação. O “Put that there” [7] foi o primeiro sistema a recorrer aos gestos deícticos. Quando o utilizador aponta com o seu braço para um local distante de um ecrã de parede e emite comandos verbais como “move that ...” e em seguida aponta para um local diferente e continua o comando “... there”, o objecto que está a ser apontado é identificado e posteriormente movido.

Quando existe uma relação directa entre o movimento da mão ou do braço e a entidade que está a ser manipulada, estamos perante um gesto manipulativo. Os gestos manipulativos são utilizados em três áreas de interação: na interação com o ambiente de trabalho, no espaço 2D, usando um dispositivo de manipulação directa como o rato ou o estilete [68]; em interfaces de realidade virtual, espaço 3D, simulando a manipulação de objectos físicos com os movimentos de mãos vazias [86]; em interfaces tangíveis para manipular objectos físicos reais que mapeiam objectos virtuais [60].

Uma das formas mais naturais de comunicação é talvez a gesticulação que tem sido abordada em vários trabalhos de investigação [63] [39] [20] que tipicamente incluem combinações de interfaces de fala e de gestos. Deste modo, pretende-se criar um estilo de interação natural e intuitiva sem ser necessário recorrer a dispositivos físicos que iriam diminuir a forma inata com que as pessoas utilizam os gestos. Este estilo de gestos depende da análise computacional que interpreta os gestos mediante o contexto, e não existe um mapeamento de gestos pré-estabelecidos.

Por fim, temos os gestos semafóricos que são definidos [63] como um sistema gestual que emprega um dicionário estático ou dinâmico de gestos. É o tipo de interacção gestual mais aplicado, embora nas interacções humanas a sua percentagem de uso seja mínima. Os gestos semafóricos são uma forma prática de fornecer computação à distância em salas e ambientes inteligentes [14] [49] [84] e como forma de reduzir a distração com tarefas primárias quando se executam tarefas secundárias [38]. Este género de gestos pode ser realizado utilizando os dedos [24] [66], as mãos [65] [1], a cabeça [72] ou periféricos de entrada como uma varinha ou um rato [84] [51]. Estes gestos são muito utilizados como forma de interagir com aplicações através do mapeamento do movimento gestual em comandos. Como exemplos temos os movimentos do rato para controlar acções de retroceder e avançar em navegadores Web [51], controlar os movimentos dos avatares através da realização de gestos, letras do alfabeto, com uma caneta [2], para lançar comandos de aplicações estilo desktop [86] [53] [59] ou para navegação de ecrãs ou selecção em menus [75] [49] [91].

2.2.2 Aplicabilidade

Desde o seu aparecimento, a interacção gestual tem sido aplicada em quase todas as áreas, desde sistemas cooperativos, aplicações computacionais para “desktop”, nas salas inteligentes e na computação pervasiva. Em relação às aplicações para “desktop” os gestos surgem, geralmente, como uma substituição do rato e do teclado. Podemos realizar com os gestos diferentes tarefas que as aplicações de desktop requerem, como: manipulação de objectos gráficos [8], anotar e editar documentos [15] [67], fazer “scroll” em documentos [75], activar selecções em menus [49] e navegar em web browsers [51], ou interagir com caixas de diálogos abanando a cabeça para a frente [16]. Através dos gestos o dispositivo de entrada pode ser manipulado para desenhar linhas, círculos, ou fazer movimentos em diferentes direcções para desenhar, controlar a funcionalidade proposta pela aplicação, alternar modos e emitir comandos [10].

Na área dos sistemas cooperativos, há uma variedade de aplicações baseadas em gestos, desde em computadores “desktop”, ecrãs de mesa [65] e ecrãs de grandes dimensões [14].

Na computação pervasiva e nos dispositivos de computação móvel, os gestos foram aproveitados para permitir interacções eyes-free, permitindo que os utilizadores foquem-se na mobilidade em vez de terem que interagir visualmente com os dispositivos [59] [72], aceitando gestos como modo de entrada e utilizando áudio como modo de saída.

Nas salas inteligentes, a interacção baseada em gestos é utilizada para controlar dispositivos e ecrãs à distância. Pode-se interagir com as máquinas que pertencem ao ambiente [79], ou então com ecrãs de grande dimensão [58].

2.2.3 Aspectos de Interação

Em todas as áreas de aplicação existem vários estudos envolvendo vários temas de interação humano-computador, que saem do aspecto tecnológico da aplicação e do seu processo de reconhecimento gestual. Estes temas são enumerados seguidamente:

Paradigma de Selecção

Os gestos são uma alternativa de interação e isso faz com que o paradigma WIMP que tinha sido pensado para realizar interação através destes dispositivos deixe de ser o mais adequado para uma nova forma de interação. Por este facto surgiram estudos com o intuito de adaptar alguns conceitos do paradigma WIMP ou criar um novo paradigma mais actual e mais adaptado às novas necessidades. Os novos paradigmas que surgem adaptados à interação gestual pretendem que os utilizadores tenham ao seu dispor os mesmos comandos existentes no anterior paradigma e que a interação seja mais fluida e multifacetada.

Sendo a utilização dos menus um dos pontos fulcrais do paradigma WIMP, o recurso aos mesmos através de gestos tem sido amplamente discutido. Exemplos dessa discussão podem ser encontrados em trabalhos como [13] [46] [21] [27] [30] [3]. Quando a interação é gestual o sistema providencia mecanismos para aceder a comandos através de menus radiais, organizados de maneira a otimizar a performance do utilizador. Os pie-menus [13] são menus radiais que ao serem invocados, no caso de interação gestual, através de um toque mais demorado na superfície de interação aparecem precisamente nesse local e o utilizador escolhe a acção que pretende através da execução de um gesto na direcção do mesmo (figura 2.3, esquerda).

Os marking menus [45] são uma variante dos anteriormente apresentados, permitindo aos utilizadores escolher a opção do menu antes do mesmo estar visível (figura 2.3, direita). Assim. Os utilizadores experientes podem seleccionar a opção pretendida com um gesto rápido na direcção dessa opção. A pensar nos utilizadores menos experientes foi implementado um intervalo de tempo (cerca de meio segundo), tendo como objectivo verificar se existe hesitação por parte do utilizador. Caso exista essa hesitação, é então mostrado o pie-menu que permite ao utilizador fazer a sua escolha, como também ajuda na memorização do gesto correcto para determinada opção.

Moyle [51], baseando-se nos marking menus avaliou um sistema de gestos para navegação na Web (figura 2.4). O sistema de gestos permite aos utilizadores emitir comandos usados com frequência, como retroceder e avançar nas páginas Web, com um simples gesto em forma de seta (flick). Através de experiências que compararam o uso do botão padrão de retroceder e avançar com o sistema de gestos de Moyle concluiu-se que é mais rápido e eficiente realizar um gesto em forma de seta (flick) para retroceder ou avançar numa página, verificando-se uma redução significativa do tempo de realização das tarefas com o sistema de gestos. No estudo de Lepinski [50], há um aumento do número de itens

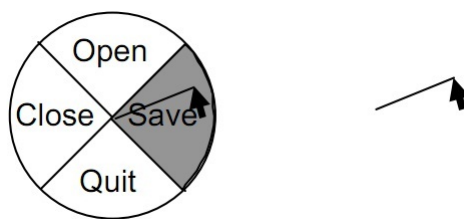


Figura 2.3: Pie-menu(esquerda) e a selecção equivalente usando marking-menu(direita) [51]

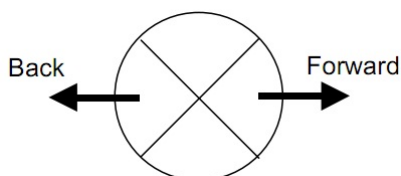


Figura 2.4: Gestos usados no sistema de making menu para navegação web [51]

por menu (figura 2.6), possibilitando um menor nível de profundidade em comparação com os menus hierárquicos tradicionais e o utilizador consoante o conjunto de dedos ou apenas dedo usado existirá uma correspondência a um menu diferente (figura 2.5). Encontrou-se novamente um melhor desempenho nos marking menus comparados com os menus hierárquicos tradicionais.

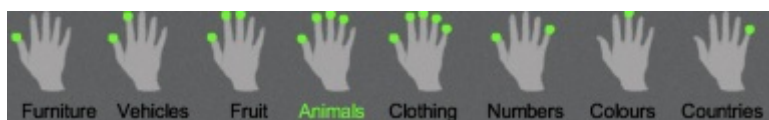


Figura 2.5: Conjunto de dedos ou dedo e as suas categorias de menu [50]



Figura 2.6: Itens de um menu [50]

Apontar e Seleccionar

Apontar é o modo mais primitivo de comunicação entre os humanos. Apontar é também bastante comum nas tarefas de interacção. Os dispositivos de apontar têm sido extensivamente estudados, e quando se fala de interacção gestual temos que tentar perceber se conseguimos obter as mesmas características que nos habituamos quando utilizamos dispositivos para apontar, como por exemplo o rato.

O estudo de Schapira [70] desenvolveu três estratégias para selecção baseada em visão:

- Apontar e Esperar: consiste em esperar um período de tempo apontando na direcção, o cursor providencia feedback no progresso da selecção com o crescimento do preenchimento do círculo (figura 2.7).



Figura 2.7: Feedback para a estratégia de selecção Apontar e Esperar [70]

- Apontar e Agitar: o utilizador tem que de mover a mão, num movimento rápido, repetitivo e pequeno, sobre o alvo e agitá-lo para fazer a selecção. O feedback desta estratégia é pequenos círculos 2.8 em que a quantidade de movimento define o seu número.



Figura 2.8: Feedback para a estratégia de selecção Apontar e Agitar [70]

- Apontar e Falar: o utilizador aponta para o objecto e com o auxílio de um comando voz ditará que o objecto está seleccionado. O feedback dado é o preenchimento do círculo vermelho, como na primeira estratégia descrita, mas não se aplica o crescimento do preenchimento do círculo. A frase que dita a selecção não é definida e como consequência disso não existe o reconhecimento de frases mas sim o cálculo da amplitude do som para um determinado intervalo de tempo. Se a amplitude superar o limite o comando de selecção é emitido.

Os participantes no estudo, de entre as três estratégias, preferiram “Apontar e Esperar”, seguida da “Apontar e Falar”.

Memorização dos Gestos

A memorização e a aprendizagem do conjunto de gestos de comando é um tema relevante na interacção gestual. A linguagem dos comandos baseada numa estrutura de menu tem a vantagem cognitiva dos comandos poderem ser reconhecidos em vez de ser necessário recordá-los. Num cenário de interacção gestual baseado em tecnologia perceptual a realização dos comandos com menus não é de todo atraente. No entanto Lenman [49] propõe o uso dos marking menus de modo a providenciarem uma base para o que

seria conseguido em conjuntos de comandos de gestos autónomos. Os utilizadores ao posicionarem a mão numa determinada posição, por exemplo dedo indicador e polegar estendidos, indicariam que se está no modo comando e depois só é necessário movimentar a mão na direcção do comando. A aprendizagem do conjunto de comandos de gestos é realizada através do aparecimento dos pie-menus quando o utilizador hesita na execução do gesto. Desta forma é dado feedback ao utilizador da direcção do gesto que tem que executar para realizar um comando, o que permite ao utilizador memorizar progressivamente os gestos.

Conforto

O conforto do utilizador quando interage com o sistema é importante, o que não é alcançado quando este tem que vestir tecnologia, por exemplo uma luva [4], limitando os seus movimentos. No entanto o desconforto é ultrapassado com a visão computacional. No caso de estudo de Charade [4] estamos perante um caso de “síndrome de imersão”, em que todos os gestos que são capturados podem ser interpretados pelo sistema, e pode ou não destinar-se a essa intenção, cortando a comunicação gestual que poderia ser desejada com outras pessoas ou dispositivos. A interacção é conseguida através de uma dataglove que diferencia os gestos através da postura da mão e da orientação dos dedos. Os gestos apenas são interpretados quando o utilizador aponta a mão para a zona activa. O cursor aparece no ecrã e segue a mão podendo depois receberem-se comandos.

No que respeita o uso de tecnologia perceptual deparamo-nos com o factor desconforto e fadiga [49] quando o utilizador tem como principal membro de interacção o braço e a mão sem nenhuma superfície de apoio.

Cooperação das Duas Mãos

Outro tema de estudo na interacção gestual é a cooperação das duas mãos para a realização de uma tarefa. Buxton [11] dividiu uma tarefa composta em duas subtarefas, podendo serem executadas em paralelo com ambas as mãos. O desempenho (tempo) de execução de uma tarefa composta é melhor comparado com a realização da mesma tarefa com apenas uma mão.

Adicionalmente Kabbash [36] chegou a uma conclusão semelhante. No entanto, mostrou-se neste estudo que o uso das duas mãos pode ser prejudicial quando as técnicas de interacção são empregadas inapropriadamente ou quando a carga cognitiva é aumentada.

Estudos demonstram que as pessoas naturalmente atribuem diferentes tarefas para cada mão, e que a mão não-dominante pode suportar as tarefas da mão dominante [26]. Interfaces de duas mãos são frequentemente usadas para especificar relações espaciais que seria difícil de descrever por voz, como por exemplo, demonstrar o tamanho relativo

dos objectos ou então especificar como um determinado objecto (mão dominante) deve ser movido em relação ao seu ambiente (mão não-dominante).

Quando nos deparamos com sistemas perceptuais, como o protótipo GWindows [83], apenas é possível a detecção de duas mãos e não é estudado a cooperação entre elas.

Interacção gestual baseada em tecnologia perceptual

A aplicabilidade dos gestos nas interfaces gestuais baseada em tecnologia perceptual é patente nas mais comuns tarefas realizadas num computador, como por exemplo: a manipulação de objectos gráficos, desde a manipulação de quadrados, esferas [7], escolha das opções de uma caixa de diálogo com os movimentos da cabeça em que o conjunto de gestos comunicativos inclui “sim”, “não”, “tem dúvidas” e “surpresa” [16]. Exemplos da aplicabilidade fora do âmbito do computador são o KidsRoom [6] que pretende que as crianças se sintam dentro de um conto de fadas interagindo com o espaço.

- **Interacção gestual em ambientes tridimensionais:**

Variados estudos em ambientes tridimensionais recaem nos jogos virtuais e os que mais se identificam com o nosso trabalho são os que utilizam os movimentos do corpo também como modo de entrada [74]. Uma abordagem existente para interacção 3D com jogos [48] usa um rastreamento espacial do movimento dos utilizadores e dos seus gestos, onde os utilizadores interagem e controlam os elementos dos mundos do jogo 3D com os seus corpos. Esta aproximação, chamada interacção baseada no corpo, usa movimentos do mundo real para controlar acções virtuais sendo aplicada tanto na realidade virtual imersiva (VR) como nas consolas de jogos. Os benefícios promovidos na realidade virtual resultantes da interacção baseada no corpo são uma melhor compreensão espacial [73] e um alto sentido de presença [80]. Em ambiente de desktop a interacção baseada no corpo [74] não tem o objectivo de substituir o rato e o teclado mas apenas reforçar a interacção podendo ser baseada nos movimentos corporais. Como exemplos temos uma interacção com jogos de tiros [44] em que os utilizadores rodam os seus corpos para navegarem e têm como input um dispositivo tipo arma, resultando disto uma experiência mais agradável e o jogo flui melhor; outro exemplo é quando o avatar do Second Life é controlado por gestos do utilizador [76]. Neste caso os utilizadores acharam mais fácil lembrarem-se dos gestos do que dos comandos de texto que o jogo providencia, verificando-se novamente uma maior facilidade em memorizar comandos com a interacção gestual. Outro exemplo é um protótipo que permite a navegação no jogo World of Warcraft através dos movimentos do corpo [74] com o intuito de dividir tarefas compostas por uma ou mais tarefas complexas. Por exemplo, se for tirada uma tarefa pertencente à tarefa complexa da interacção com o teclado e for passada para a interacção com o corpo será mais fácil para o jogador concentrar-se nas res-

tantes acções do jogo que necessitam de teclado. Concluiu-se também que é viável a combinação da interacção baseada no corpo com o teclado e o rato como input para jogos de desktop. Os participantes acharam esta interacção mais atraente porque as personagens imitavam os seus movimentos do corpo e foi fácil a memorização dos movimentos do corpo com as acções no jogo devido à sua parecença.

Investigação sobre a usabilidade em interfaces 3D para ambientes de realidade virtual foram realizados por Cabral [12] com a CAVE e a Powerwall. Verificou-se que os utilizadores geralmente têm mais ausência de stress físico sobre respostas do sistema quando estão na presença de manipulação de objectos no mundo virtual, preferindo liberdade dos movimentos para navegação. Por sua vez Kwon [47], que combinou o uso de sensores no corpo com câmaras de captura de movimento para uma aplicação de treino de movimento, descobriu através de testes de usabilidade que os utilizadores muitas vezes formam a opinião de que uma interface baseada em gestos faz com que a aplicação seja mais atraente e incentiva os utilizadores a concentrarem-se mais nas tarefas a serem alcançadas. Em relação aos tipos de gestos que são preferíveis os utilizadores realizarem para interagir com interfaces 3D baseados em gestos 3D [76] destacou-se como input a preferência de associar aos gestos que não são mapeados naturalmente para gestos com as mãos a utilização dos botões e do joystick da Wii. Concluiu-se também que com a interacção gestual 3D aplicada ao jogo Second Life se obtém uma interface mais intuitiva e agradável.

Galeria 3D de objectos de um museu [78], aplicações de ambientes 3D de manipulação directa com os objectos (p.e. mapas 3D [88] são alguns dos trabalhos sobre ambientes tridimensionais encontrados na literatura. Estes trabalhos focam-se nos aspectos tecnológicos de reconhecimento e não em estudos de usabilidade com os utilizadores.

Restringindo o mundo da interacção gestual em ambientes tridimensionais baseado apenas na tecnologia perceptual, excluindo a realidade virtual imersiva, os estudos que se focam na forma como os utilizadores interagem usando os gestos são escassos, como podemos verificar.

2.3 Discussão

Neste capítulo começamos por apresentar o contexto e alguns conceitos em que se baseia este trabalho. Posteriormente focámo-nos na área de interacção gestual, apresentado os principais tópicos para se contextualizar e conhecer o que tem vindo a ser feito nesta área, de forma a melhorá-la e adequá-la às necessidades dos utilizadores.

Tecnologicamente existem duas maneiras de perceber os gestos realizados pelos utilizadores. As tecnologias perceptuais não exigem que o utilizador tenha qualquer con-

tacto com um objecto físico, tendo assim a capacidade de interpretar a intenção do utilizador somente pelo seu movimento corporal. Por sua vez, as tecnologias não-perceptuais implicam que o utilizador esteja em contacto com um periférico de entrada ou com o próprio dispositivo com o qual pretende interagir.

Neste âmbito distinguem-se quatro estilos de gestos diferentes, nomeadamente: gesticulados, actividade inata de gesticular; deícticos, gesto de apontar para um objecto; manipulativos, há uma relação directa entre o movimento da mão, ou do braço, e o que está a ser manipulado; e semaforicos, o gesto do utilizador tem que coincidir com algum gesto que esteja presente no dicionário de gestos do dispositivo com o qual quer interagir.

A não adequação do paradigma WIMP a esta nova forma de interacção leva a que estejam constantemente a ser estudadas novas formas de melhorar a interacção gestual e, em alguns casos, aproximá-la dos hábitos já adquiridos pelos utilizadores.

O crescimento exponencial das tecnologias perceptuais faz com que o paradigma WIMP deixe de ser o mais adequado e surge a necessidade dos investigadores da HCI estudarem formas de melhorar a interacção através da adaptação de alguns conceitos subjacentes ao paradigma ou então criando um novo paradigma adaptado às novas necessidades. Mesmo com os estudos realizados existem contextos em que uma interacção puramente gestual apresenta alguns desafios que ainda não foram superados.

Existe escassez de investigação de modo a encontrar gestos padrões naturais para determinadas tarefas, assim como o estudo da selecção de objectos em ambientes perceptuais. Algumas das soluções propostas para a selecção incluem: apontar e esperar; apontar e agitar; ou apontar e falar. A preferência recai sobre a estratégia apontar e esperar seguida da estratégia apontar e falar. Outros aspectos como: conforto do utilizador e memorização dos gestos estão pouco explorados.

Percebe-se que a interacção gestual ainda se encontra em fase de constante desenvolvimento e maturação, pelo que é necessário procurar soluções para os desafios existentes. Neste sentido, e recorrendo a dispositivos com capacidade de captar perceptualmente os gestos dos utilizadores, o nosso trabalho passará por tentar dar um contributo nesta área em crescimento.

Capítulo 3

Avaliação de protótipos de interacção gestual com e sem voz

Neste capítulo é apresentado um estudo sobre interacção gestual com e sem o auxílio da voz, em superfícies de grande dimensão e com a particularidade do utilizador não ter contacto físico com nenhum periférico de entrada.

Neste estudo pretendeu-se perceber as diferenças que existem nos gestos que os utilizadores realizam em dois cenários que se distinguem pela possibilidade de utilização de comandos de voz como modalidade complementar aos gestos. Para isso foi pedido aos participantes no estudo para manipularem objectos em duas aplicações de teste com o intuito de executar acções sobre os objectos ou a área de trabalho. Desta forma pretendeu-se: 1. Compreender como podem os utilizadores beneficiar do uso de interacção gestual com e sem o suporte da voz; 2. Quais as acções dessas aplicações que estão aptas a serem realizadas através de gestos e/ou voz; 3. Quais os gestos e palavras mais apropriados, intuitivos e confortáveis para acções distintas.

3.1 Enquadramento e Preparação do estudo

O objectivo principal deste estudo foi comparar a utilização da interacção gestual com e sem voz quando não se estabelece nenhum contacto entre o utilizador e a superfície de interacção ou qualquer outro dispositivo de entrada. Como resultados pretende-se obter, para determinadas acções, um conjunto de orientações que possibilite determinar qual o melhor cenário de interacção, gestos com ou sem voz, assim como os conjuntos de gestos e palavras que são realmente utilizados.

Para este fim foram seleccionadas duas aplicações que pudessem beneficiar da utilização de interacção gestual. Para cada uma das aplicações identificou-se um conjunto de acções passíveis de serem executadas pelo utilizador nos dois cenários a estudar, e definiu-se um conjunto de tarefas que exercite essas acções. Durante a realização das tarefas os participantes tiveram total liberdade na criação dos gestos e dos comandos de voz para alcançar os objectivos pedidos.

3.2 Cenários e Acções

Seleccionaram-se duas aplicações (figura 3.1) que permitem a manipulação de imagens, existindo algumas acções comuns às duas aplicações e outras particulares a cada uma. Na primeira aplicação, temos as imagens como se estivessem espalhadas aleatoriamente numa mesa, sendo no entanto possível movimentar as imagens num espaço tridimensional, afastando-as ou aproximando-as da superfície da mesa. Por sua vez, na segunda aplicação temos uma exposição de fotos ao longo de uma “parede” curva. Qualquer uma destas aplicações, em virtude da sua natureza espacial, pode beneficiar da utilização da interacção gestual com ou sem auxílio de comandos de voz, principalmente em cenários em que as aplicações são projectadas em superfícies de grandes dimensões e o utilizador interage distante da superfície de projecção. Pode assim considerar-se que estas duas aplicações são adequadas para os objectivos que se pretende atingir com este estudo.



Figura 3.1: Aplicações de teste: (1)“Mesa” com imagens, (2)“Parede” de imagens

Posteriormente definiu-se um conjunto de acções a realizar em cada uma das aplicações. As acções a executar em cada uma das aplicações foram as seguintes:

“Mesa” com imagens: Acções aplicadas a uma imagem: zoom in, zoom out, rodar no sentido dos ponteiros do relógio, rodar no sentido contrário aos ponteiros do relógio, deslocar a imagem, sobrepor uma imagem a outra e apagar a imagem. Undo e redo.

“Parede” de imagens: Acções aplicadas a uma imagem: zoom in, zoom out, mostrar imagem seguinte e anterior. Acções aplicadas a todo o ambiente: zoom in, zoom out, movimentar para a esquerda, movimentar para a direita.

Para exercitar estas acções definiram-se tarefas que implicavam manipular imagens num workflow com significado para o utilizador. Por exemplo, em vez de pedir ao utilizador para rodar uma imagem, colocou-se uma determinada imagem invertida sobre a superfície e pediu-se aos participantes para consultar um determinado detalhe da imagem, o que na prática “obrigava-os” a rodar e aproximar a imagem, permitindo desta forma perceber como realizam as acções individuais no contexto de um workflow mais natural.

3.3 Participantes

Um total de dez indivíduos, com uma média de 22 anos, participou numa sessão experimental. Todos os participantes eram estudantes universitários e, segundo os mesmos, oito com mais de 10 anos de experiência de utilização de computadores e dois com menos de 10 anos. Todos eles afirmaram conhecer pelo menos um dispositivo de interacção gestual, dois dos quais com utilização regular, referindo-se aos dispositivos Apple iPhone, iPod, Nintendo Wii e outros telemóveis com reconhecimento gestual. Oito participantes afirmaram também conhecer pelo menos um dispositivo com reconhecimento de voz, variando desde um programa para cegos, telemóveis com reconhecimento de voz até a um kit de mãos livres. No entanto, apenas um dos participantes possui experiência regular com o dispositivo de mãos livres. Nenhum participante conhece um único dispositivo que disponibilize ambas as formas de interacção.

3.4 Procedimento

O estudo iniciou-se com uma explicação dos objectivos aos participantes, seguido pelo preenchimento de um questionário de perfil.

Durante a experiência os participantes tiveram toda a liberdade de interacção para executar os gestos e os comandos de voz, ou seja, não foram explicitadas quaisquer restrições na utilização de gestos e palavras. Por exemplo, não existia restrição no número de palavras nos comandos de voz.

Cada participante esteve envolvido numa sessão individual, manipulando as duas aplicações de teste. Foi pedido aos participantes que desempenhassem um conjunto de tarefas, pensadas de modo a englobarem todas as acções, de forma a que os participantes tivessem de criar gestos e palavras para as acções definidas. Não foi imposto nenhum limite de tempo durante as tarefas. No fim, foi pedido que preenchessem um questionário de satisfação em que os dois modos de interacção foram classificados numa escala de 1 a 5, onde 1 significa a interacção menos adequada e 5 a mais adequada e intuitiva. Foi-lhes também pedido que tentassem justificar a escolha dos gestos. Com a finalidade de registar todos os gestos e comandos de voz as sessões foram filmadas.

Quer a ordem de utilização das aplicações (mesa de imagens e parede de imagens), quer a ordem de cenários de interacção (sem e com auxílio de comandos de voz) foi decidida aleatoriamente para cada participante.

3.5 Detalhes técnicos

Realizou-se uma sessão experimental onde se recorreu ao Microsoft Kinect assim como a um projector, ambos ligados a um computador Windows 7, com uma resolução de ecrã

de 1366 por 768 pixéis, e projectou-se na parede ocupando uma área de 280 cm por 200 cm.

Os participantes da sessão experimental encontravam-se de pé e a interacção com os protótipos foi realizada sem nenhum meio intermédio. O avaliador encontrava-se afastado dos participantes e com excelente visibilidade para o participante e para o local onde as aplicações foram projectadas. Um sistema de seguimento permitia aos utilizadores controlarem a posição do cursor apontando para a superfície de projecção. Dado que se pretendia oferecer total liberdade para criação de gestos e comandos de voz, estes eram interpretados pelo avaliador que os mapeava em acções sobre o sistema através de atalhos de teclado, desempenhando assim o papel de Wizard of Oz. Os participantes não foram previamente informados desta “solução técnica”, e como o avaliador não estava no seu campo de visão, acreditaram sempre estar a interagir com uma solução completamente implementada.

3.6 Análise dos resultados

A análise dos resultados encontra-se dividida em três partes: primeiro cenário de interacção, apenas com interacção gestual; segundo cenário com interacção gestual auxiliada por comandos de voz; análise comparativa dos dois cenários.

3.6.1 Primeiro cenário de interacção: interacção puramente gestual

O primeiro resultado que foi possível observar é que em determinadas acções há uma grande convergência dos gestos resultados por diferentes participantes, enquanto que para outras acções o número de participantes que realiza o mesmo gesto é bastante menor. Para efeitos de análise definiu-se que existia concordância num gesto quando a maioria dos participantes realizava um gesto com as mesmas características. Nas tabelas 3.1 e 3.2 apresenta-se para cada acção, o número de participantes que realizou o gesto e ilustra-se esse gesto com uma sequência de imagens representativa e uma descrição textual. Nos casos em que não houve concordância apresentam-se as alternativas realizadas pelos participantes para a mesma acção.

Na descrição textual utiliza-se o conceito de ponto. Neste caso, o ponto corresponde a uma abstracção do que é usado pelos utilizadores para apontar para a superfície: alguns utilizadores apontavam com a mão esticada, outros com a mão fechada, outros com um dedo esticado. Não se encontrou nenhuma semântica associada à forma como a mão ou dedos eram utilizados, e assim sendo, optou-se por analisar os gestos da mesma forma sendo considerado importante apenas o ponto que a mão ou dedo representam. Como exemplo apresentam-se na tabela 3.3 diferentes formas como foi realizado o zoom in, mas sempre com as duas mãos a convergirem até se encontrarem juntas.


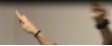




















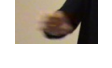


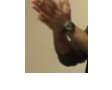

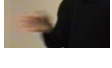

Acções	Nº	1ª	2ª	3ª	Descrição
Zoom In	6				Dois pontos a afastarem-se (tipicamente, paralelamente à superfície de interacção).
Zoom Out	6				Dois pontos a aproximarem-se (tipicamente, paralelamente à superfície de interacção).
Sobrepor	6				Um ponto a afastar-se da superfície de interacção.
Movimentar	9				Um ponto a deslocar-se paralelamente à superfície de interacção.
Parar Movimento	9				Um ponto imóvel.
Rotação SPR/SCPR	3				Um ponto centrado no objecto e outro que descreve um movimento de rotação com início no outro ponto.
Rotação SPR/SCPR	3				Um ponto a efectuar um movimento de rotação, mas com um raio bastante menor que na alternativa anterior.
Apagar	3				Construção de uma forma de X, recorrendo quer a um quer a dois pontos.
Apagar	2				Deslocar repetidamente um ponto para a esquerda e direita.
Apagar	2				Bater uma palma, ou seja, dois pontos a aproximarem-se paralelamente à superfície de projecção.
Undo/Redo	4				Um ponto num movimento para a esquerda ou para a direita paralelamente à superfície de interacção.

Tabela 3.1: Gestos padrões para cada acção na aplicação mesa de imagens

Acções	Nº	1ª	2ª	3ª	Descrição
Zoom In	7				Dois pontos a afastarem-se (tipicamente, paralelamente à superfície de interacção).
Zoom Out	9				Dois pontos a aproximarem-se (tipicamente, paralelamente à superfície de interacção).
Seguinte/ Anterior	9				Um ponto num movimento para a esquerda ou para a direita paralelamente à superfície de interacção.
Zoom in total	8				Dois pontos a afastarem-se (tipicamente, paralelamente à superfície de interacção).
Zoom out total	8				Dois pontos a aproximarem-se (tipicamente, paralelamente à superfície de interacção).
Movimentar tudo para a esquerda/direita	5				Dois pontos num movimento para a esquerda ou para a direita paralelamente à superfície de interacção.
Movimentar tudo para a esquerda/direita	4				Um ponto num movimento para a esquerda ou para a direita paralelamente à superfície de interacção.

Tabela 3.2: Gestos padrões para cada acção na aplicação parede de imagens

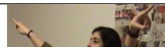


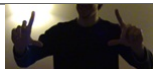

	Zoom In
Ponta dos dedos	
Dedos juntos	
Todos os dedos afastados	
Posição triangular dos dedos	
Palma da mão e dedos	

Tabela 3.3: Generalização do gesto zoom in

De seguida analisam-se alguns aspectos da forma como os participantes interagiram com as duas aplicações e as acções que realizaram. Adicionalmente analisa-se a evolução verificada nos participantes com o avançar da sessão.

-Gestos padrão

Para as tarefas apagar, rodar uma imagem e movimentar tudo para a direita ou esquerda não foi possível encontrar apenas um gesto padrão. Os conjuntos de gestos que representam a acção apagar não têm semelhanças entre si. Por sua vez os conjuntos de gestos de rodar uma imagem e movimentar tudo são mais similares, distinguindo-se essencialmente pelo número de pontos usados pelos participantes. Para as outras tarefas uma maioria dos participantes realizou um gesto semelhante.

-Gestos iguais para acções diferentes

Como se pode ver nas tabelas podemos concluir que existem gestos iguais para acções com finalidades diferentes. Os gestos de zoom in e zoom out de uma única imagem são idênticos aos gestos de zoom in e zoom out de todo o ambiente de trabalho. O zoom out, independentemente de que objecto sofre transformação, é similar a um dos gestos de apagar uma imagem. Também os gestos de undo e redo, seguinte e anterior, movimentar o ambiente todo para a esquerda ou todo para a direita em nada diferem.

- Mudança de gestos

Durante a sessão experimental existem várias acções de zoom in e zoom out que são solicitadas mais do que uma vez como resultado das tarefas a desempenhar. Das primeiras vezes essas acções são referente apenas a uma imagem enquanto que posteriormente dizem respeito a todo o ambiente de interacção. Verificou-se que, tanto relativamente a uma imagem como a todo o ambiente, o número de participantes que realizou o gesto padrão cresceu com o avançar da experiência. Em ambos os casos, da primeira vez que tiveram de realizar o gesto, 6 participantes realizaram o gesto padrão, tendo o número

crescido para 8 no final da experiência. De seguida apresenta-se uma análise mais fina desta evolução.

Da primeira vez que é executada a acção de zoom in 6 participantes adoptaram o gesto padrão indicado na tabela 3.1 . Da segunda vez houve mais um participante a realizar o gesto e da última vez ainda mais um. Por sua vez, o gesto padrão de zoom out foi adoptado inicialmente por 6 participantes, existindo um acréscimo de três pessoas da segunda vez e terminando com um decréscimo de um participante. O crescimento do número de pessoas que adoptam os gestos padrão, para as duas acções, pode estar relacionado com a adaptação do utilizador ao sistema e a sua habituação assim como a confirmação destes dois gestos padrão serem os mais naturais e intuitivos para os participantes.

Da segunda vez que se tem que recorrer às acções de zoom in e zoom out temos, respectivamente, 7 e 9 participantes a realizar o gesto padrão. Esta diferença poderá resultar do facto de entre as duas acções (zoom in e zoom out) o tamanho do objecto com que se está a interagir ser diferente (a imagem em que se faz zoom out é maior do que aquela em que se faz zoom in) o que pode justificar alguns utilizadores não utilizarem gestos “simétricos” para zoom in e zoom out. Por outras palavras, podemos dizer que quando a imagem é maior é mais natural para os participantes usarem as duas mãos (para fazer o zoom out). Também a suportar esta afirmação temos o facto de exactamente o mesmo número de participantes, oito, empregar os gestos padrão quando o zoom (in ou out) é realizado sobre a área de trabalho (um “objecto” de grandes dimensões). Verificou-se ainda que, durante a última actividade, dois dos participantes terem-se apercebido de que os gestos de zoom in e zoom out do ambiente de trabalho não poderem ser idênticos aos de zoom in e zoom out de uma imagem adoptando por isso um gesto diferente.

-Gestos em que a profundidade é relevante

Verificou-se que existe apenas uma acção para a qual é importante considerar a profundidade para o reconhecimento do gesto padrão. Essa acção é sobrepor, e o gesto realizado pela maioria dos participantes envolve afastar a mão da superfície de projecção.

3.6.2 Segundo cenário de interacção: interacção gestual com voz:

Numa das fases da sessão experimental os participantes tinham de realizar as tarefas pedidas complementando os gestos com comandos de voz. A tabela 3.4 apresenta os comandos padrão para cada acção no modo de interacção gesto e voz bem como o número de participantes que os utilizou.

Um factor que é interessante analisar é qual o uso que os participantes fizeram dos gestos quando também têm a possibilidade de comunicar por voz. Na tabela 3.5 verificamos a quantidade de participantes que realizou o gesto correspondente à acção e os que, por sua vez, apenas usufruíram da interacção gestual com a finalidade de apontar para seleccionar a imagem ou o ambiente de trabalho. Na última coluna são apresentados apenas os gestos quando estes são diferentes dos gestos padrão que foram encontrados quando

Mesa de imagens	Zoom in	“zoom in”	4
	Zoom out	“zoom out”	5
	Mover	“Arrastar”	3
		“Move”	3
	Parar Movimento	“Pára”	3
	Sobrepôr	“Frente”	3
		“Para a frente”	3
	Rodar SPR	“Roda”	5
	Rodar SCPR	“Roda”	5
	Apagar	“Apagar”	6
	Undo	“Undo”	5
Redo	“Redo”	5	
Parede de imagens	Zoom in	“Aumentar boneco neve”	2
		“Zoom in”	2
		“Abre”	2
		“Aumentar”	2
	Zoom out	“Fechar”	6
	Seguinte	“Próxima”	3
		“Seguinte”	3
		“Next”	3
	Anterior	“Anterior”	6
	Zoom out total	“Diminuir”	3
		“Zoom out”	3
		“Diminuir tudo”	2
	Movimentar para a direita	“Rodar”	4
		“Rodar direita”	3
	Movimentar para a esquerda	“Rodar”	4
“Rodar esquerda”		3	

Tabela 3.4: Comandos padrão

interagiram unicamente com gestos.

De seguida analisam-se alguns dos aspectos relevantes do cenário de interacção com gestos e comandos de voz combinados.

-Mudança de gestos com/sem voz

Quando se alterou o cenário de interacção, de gestos sem comandos de voz para gestos com comandos de voz, verificou-se uma mudança de alguns gestos. Quando é possível interagir através das duas modalidades os três gestos mais relevantes para apagar um objecto passam apenas a um gesto (dois movimentos laterais) e é utilizado unicamente por um participante. Quando a acção é a movimentação de todo o ambiente de trabalho verificou-se que cinco participantes passaram a realizar o gesto apenas com uma mão (mais um que anteriormente) e menos um participante realizou o gesto com as duas mãos

-Apontar ou gesticular

Quando o utilizador se depara com acções que correspondem a gestos pouco intuitivos e quando o cenário de interacção permite duas modalidades os participantes tipicamente enviam informação para executar a acção através de comandos de voz e apenas apontam para a imagem ou objecto que é alvo da acção. Exemplos destas acções são apagar, undo e redo, que são também acções em que não havia concordância nos gestos realizados pelos participantes quando não tinham a possibilidade de empregar comandos de voz.

O zoom in e zoom out de uma imagem ou da área de trabalho são abordados de forma diferente pelos participantes. Zoom in e zoom out de uma imagem são acções em que a quase totalidade dos participantes apontam enquanto que quase nenhum aponta quando o objecto da acção é a área de trabalho. Isto dever-se-á principalmente ao facto de ser necessário distinguir o alvo da acção quando se faz o zoom de uma imagem. Mesmo quando se faz o zoom out de uma imagem que está a ocupar a quase totalidade do ecrã, os participantes sentem a necessidade de seleccionar de alguma forma o objecto alvo da acção. Quando o alvo da acção é toda a área, não há essa necessidade de distinguir o alvo.

Analisando a tabela 3.5 podemos concluir que o número de participantes que opta pela realização do gesto representativo de cada acção quando está a interagir com duas modalidades acaba por repetir a informação que também indica através do comando de voz. A excepção são as acções de rotação e de mover em que apenas alguma informação é repetida, outra é complementar ao comando. Um exemplo disso é quando o participante diz “roda” e gesticula de modo a transmitir a informação da direcção e a amplitude da rotação, informação que não é disponibilizada com o comando de voz. Outro exemplo é a acção de movimentação do ambiente total para a esquerda ou para a direita quando dizem apenas “move” ou “rodar” e a informação da direcção é transmitida através do gesto.

A tabela 3.6 ilustra de que modo a informação é transmitida pelas duas modalidades sem se repetir. Todos os comandos de voz aí apresentados são acompanhados pelo gesto de apontar do participante. Alguns dos comandos também referem qual é o objecto a que a acção se aplica, replicando assim a informação transmitida ao apontar. Em todos


	Acções	Apontar (nº de pessoas)	Gesticular(nº de pessoas)	Gesto
Mesa de imagens	Zoom in	7	3	
	Zoom out	7	3	
	Mover	6	4	
	Parar Movimento	6	4	
	Sobrepor	8	2	
	Rodar SPR	2	8	
	Rodar SCPR	2	8	
	Apagar	9	1	 dois movimentos laterais
	Undo	9	1	
	Redo	9	1	
Parede de imagens	Zoom in	8	2	
	Zoom out	8	2	
	Seguinte	1	9	
	Anterior	1	9	
	Zoom in total	2	8	
	Zoom out total	2	8	
	Movimentar para a direita	0	10	
	Movimentar para a esquerda	0	10	

Tabela 3.5: Número de participantes que apontaram e gesticularam para cada acção

estes casos a especificação da acção a realizar é efectuada por uma única modalidade, o comando de voz, enquanto que a outra modalidade é usada para indicar um argumento da acção, o objecto sobre a qual ela é aplicada.

-Objecto da acção referido no comando de voz

Como referido anteriormente, nalguns casos, o objecto da acção foi incluído no comando de voz para determinadas acções. Na aplicação mesa de imagens houve um participante que referiu o objecto para as acções de zoom in, zoom out, movimentar e sobrepor uma imagem. Os correspondentes comandos de voz foram “aumentar gato”, “diminuir gato”, “flor amarela acompanhe” e “flor amarela para a frente”. No entanto, na aplicação parede de imagens houve um acréscimo de mais dois participantes a incluir essa informação no comando de voz mas um decréscimo nas acções em que isso aconteceu. Apenas na acção de zoom in de uma imagem foram proferidas as frases “aumentar boneco de neve”, duas vezes, e “papagaio cinzento”, uma vez. Neste último caso, a informação da acção a executar foi transmitida através do gesto realizado.

3.6.3 Comparação dos dois cenários de interacção

Dos questionários efectuados após a sessão experimental obteve-se a classificação dada pelos participantes à interacção em cada um dos modos (gestos com ou sem comandos de voz) para cada uma das acções. Na tabela 3.7 podemos visualizar a classificação média de cada uma das acções dependendo do seu modo de interacção. A classificação varia de 1 a 5 em que 5 significava que a interacção utilizada era natural e intuitiva e 1 o seu oposto:

No modo de interacção em que se tinha de utilizar as duas modalidades (gestos e voz) as acções apagar, undo e redo tiveram uma classificação mais elevada do que na interacção gestual. Isto deve-se à dificuldade sentida pelos participantes em encontrar um gesto relacionado com a acção. Essa dificuldade é imediatamente ultrapassada quando se pode transmitir a acção através de um comando de voz.

Há uma preferência mais acentuada na utilização de só uma modalidade (gestos) apenas na acção que envolve a movimentação de todo o ambiente para a esquerda ou direita. Isto deve-se ao facto da tarefa envolver a deslocação de todas as imagens, abrangendo quase toda a área de interacção mas também pode ser resultado de ser parte da tarefa final da sessão experimental, altura em que os participantes já estão mais à vontade e interagem de forma mais intuitiva.

Pretendeu-se também perceber quais as acções que causaram maior dificuldade aos participantes em encontrar um gesto representativo. Para isso procedeu-se a uma análise dos vídeos em que se gravou toda a interacção dos participantes durante a sessão experimental. Na tabela 3.8 visualizam-se a média de tempo, em segundos, de reflexão de cada acção até à sua execução nos dois modos de interacção.

Para as acções que foram realizadas posteriormente à sua acção similar, por exemplo, undo e redo, constatou-se que a posterior é sempre executada de forma mais rápida porque

	Acção	Número de vezes utilizado	Comandos de voz
Mesa de imagens	Zoom in	2	“Amplia”
		2	“Zoom in”
		1	“In”
		1	“Amplia”
		1	“Aumentar”
		1	“Aumentar gato”
	Zoom out	3	“Zoom out”
		2	“Encolhe”
		1	“Diminuir”
		1	“Diminuir gato”
	Mover	2	“Move”
		1	“Mexe”
		1	“Arrastar”
		1	“Segue”
		1	“Flor amarela acompanhe”
	Parar movimento	5	“Pára”
		1	“Stop”
	Sobrepor	3	“Frente”
		3	“Para a frente”
		1	“Avança”
		1	“Flor amarela avance”
	Rodar SPR	1	“Rodar direita X graus”
		1	“Rodar X graus direita”
	Rodar SCPR	1	“Rodar esquerda X graus”
		1	“Rodar X graus esquerda”
	Apagar	5	“Apagar”
		3	“Delete”
		1	“Del”
	Undo	5	“Undo”
		2	“Voltar atrás”
1		“Retroceder”	
1		“Previous”	
Redo	5	“Redo”	
	1	“Voltar à frente”	
	1	“Frente”	
	1	“Avança”	
	1	“Next”	
Parede de imagens	Zoom in	2	“Aumentar boneco de neve”
		2	“Abre”
		2	“Zoom in”
	Zoom out	1	“Aumentar”
		1	“Amplia”
		6	“Fechar”
	Seguinte	2	“Zoom out”
		1	“Next”
	Anterior	1	“Previous”
	Zoom in total	1	“Aumentar tudo”
1		“Aumentar imagens”	
Zoom out total	1	“Diminuir tudo”	
	1	“Diminuir imagens”	

Tabela 3.6: Informação não repetida para cada acção, comandos de voz e nº de pessoas

	Acções	Classificação média (1-5) Gestos	Classificação média (1-5) Gestos e Voz
Mesa de imagens	Zoom in	4.8	4.8
	Zoom out	4.8	4.8
	Mover	4.8	4.7
	Parar Movimento	4.4	4.2
	Sobrepor	3.8	3.8
	Rodar SPR	4.6	4.5
	Rodar SCPR	4.6	4.5
	Apagar	4.2	4.8
	Undo	3.1	4
Redo	3.1	4	
Parede de imagens	Zoom in	4.8	4.2
	Zoom out	4.7	4.2
	Seguinte	4.7	4.4
	Anterior	4.7	4.4
	Zoom in total	4.6	4.3
	Zoom out total	4.6	4.3
	Movimentar para a direita	4.8	3.9
	Movimentar para a esquerda	4.8	3.9

Tabela 3.7: Classificação média (1 a 5) para cada acção em cada cenário

são gestos idênticos em que apenas é alterado a direcção dos gestos.

A acção com maior fracção de tempo, independentemente das modalidades de interacção, é a undo: 7 segundos no modo de interacção gestual e 3,2 segundos quando se tem o auxílio da voz, o que representa uma redução para metade.

Em várias das acções há uma diferença significativa nos tempos quando se mudam os cenários de interacção. No modo de interacção puramente gestual, a tarefa apagar tem o valor médio de 7,1 segundos mas quando se interage com as duas modalidades a média de tempo de reflexão é bastante menor: 0,4 segundos. A causa principal é a mesma que justifica as diferenças na acção undo: é difícil encontrar um gesto relevante para a acção, mas é bastante intuitivo com o recurso à voz. A mesma explicação é válida para perceber porque é que 9 participantes, nas acções undo e apagar, quando tinham duas modalidades disponíveis utilizaram a interacção gestual apenas para apontar para a imagem que se pretendia apagar ou para o ambiente de trabalho e a informação da acção foi transmitida por comandos de voz.

- Comentários dos participantes

Os comentários dos participantes acerca do porquê dos gestos realizados para determinadas acções apontam diversas justificações. Para o gesto de zoom in temos comparações a gestos executados para abrir as coisas, por exemplo abrir uma porta. Para ambos os

	Acções	Média de tempo(segundos) Gestos (1-5) Gestos	Média de tempo(segundos) Gestos e Voz
Mesa de imagens	Zoom in	1.3	0.8
	Zoom out	0.2	0.2
	Mover	1.3	1.1
	Parar Movimento	0.5	0.4
	Sobrepor	3	3.1
	Rodar SPR	1.1	1.4
	Rodar SCPR	0.2	0.5
	Apagar	7.1	0.4
	Undo	7	3.2
Redo	0.6	1	
Parede de imagens	Zoom in	2	1
	Zoom out	0.3	0
	Seguinte	0.6	1.2
	Anterior	0.3	0
	Zoom in total	1.5	1.4
	Zoom out total	0.3	0
	Movimentar para a direita	0.7	2.1
	Movimentar para a esquerda	0.1	0

Tabela 3.8: Tempo médio (em segundos) para cada acção em cada cenário

zoom, in e out, inspiraram-se nos gestos de outras aplicações com interacção gestual para abrir e fechar uma imagem, como os zooms em filmes com interacção gestual e gestos que realizam no iPod. Surgem também explicações dos zooms e das movimentações de todo o ambiente comparadas ao que a personagem interpretada por Tom Cruise fazia no filme “Minority Report”. O gesto de rotação é equiparado com a rotação de um volante e da maçaneta de uma porta. Para mover uma imagem relacionam ao gesto de agarrar um objecto fisicamente e parar o movimento ao deixar o objecto para não se movimentar mais. Apagar foi baseado no gesto de expulsar uma imagem do ecrã, à borracha da aplicação Paint e ao rasgar uma folha. Os gestos de seguinte e anterior são comparados aos gestos de aplicações dos seus telemóveis touch e com o virar das páginas de um livro.

Alguns participantes não se conseguiram explicar e alguns salientaram que foi difícil arranjar gestos para apagar e sobrepor porque não estavam habituados a ter essas tarefas gestualmente nos dispositivos utilizados.

Todos estes comentários confirmam que com a habituação e anterior interacção ou realização dos gestos em outros ambientes ou outras aplicações obtém-se uma interacção mais natural, sendo mais rápido ao nosso cérebro ir buscar soluções para as tarefas que eram apresentadas.

3.6.4 Discussão

Através da sessão experimental obteve-se para algumas acções os gestos padrões, ou seja, os gestos que continham as mesmas características para uma acção. No entanto, para o gesto de apagar não se conseguiu alcançar um gesto padrão porque os diferentes conjuntos de gestos executados tem características bastantes diferentes. Os conjuntos de gestos para as acções de rodar uma imagem e movimentar o ambiente todo para a esquerda ou direita são divergentes mas apenas no número de pontos (uma mão ou duas), tendo as mesmas características.

Outro facto relevante que se obteve da análise dos resultados foram gestos iguais que mapeiam acções diferentes. Por exemplo, para as acções de zoom in, zoom out de uma imagem ou de todo o ambiente os gestos são idênticos, e o gesto de apagar é idêntico ao gesto de zoom out independentemente do sujeito. Um gesto idêntico para acções diferentes em que o argumento da acção é diferente, ou seja, podendo ser uma imagem ou um conjunto de imagens, não disputa problemas. Se quiséssemos usar o gesto apagar e zoom in em que o argumento da acção seria, em ambos os casos, uma imagem, era importante arranjar uma maneira de resolver esta ambiguidade. Os comandos por voz seriam uma solução para este problema.

Um dado curioso, no cenário de interacção gestual, é que com o passar do tempo os participantes tendem mais a realizar o gesto padrão (caso do gesto de zoom in e zoom out), isto poderá estar influenciado com o nível de à vontade ganho durante o passar do tempo da experiência. Relevante também é que quanto maior o tamanho do argumento da acção mais tendência os participantes têm para interagir com as duas mãos, diferença analisada, por exemplo, no gesto de zoom in e out com foco num objecto ou vários. É discutível se obteríamos os mesmos gestos padrões e preferências pela interacção gestual sem superfície de apoio numa tela de projecção de menor dimensão. Alterando o cenário de interacção, o tamanho do argumento continua a ser relevante, para a acção de zoom in e zoom out, a acção é transmitida por voz quando o argumento é de menor tamanho (uma imagem), e há repetição de informação quando o argumento é transmitido por voz e gesto quando a sua área é maior (conjunto de imagens). Como era de esperar, no que respeita as classificações, há uma preferência pela interacção gestual quando em que o argumento é uma área maior.

Mesmo estando presente em aplicações tridimensionais, com um nível de profundidade presente, o único gesto em que a profundidade foi relevante foi o sobrepor.

Na possibilidade de interacção com voz e gesto, verificamos que as acções que tinham sido menos intuitivas são aquelas em que um maior número de participantes transmite o objectivo da acção através de comandos de voz e apenas utiliza a interacção gestual para identificar o argumento da acção, referimo-nos neste caso às acções de apagar, undo e redo. No entanto, para acções pouco intuitivas quando os participantes repetem a informação, o gesto executado difere do conjunto dos gestos para essa acção, no caso

da acção apagar. Para a acção de movimentar o ambiente para a esquerda ou direita, o número de participantes que executou o gesto apenas com uma mão, poder-se-ia ter em conta que este gesto é mais intuitivo do que com o uso das duas mãos, só que como o aumento foi de apenas um participante e a diferença é diminuta continuámos sem conseguir definir apenas um conjunto de gestos para a acção de movimentar o ambiente todo para a esquerda ou direita como aconteceu no outro cenário de interacção.

Inserindo o factor tempo e classificação, é previsível, que as acções nas quais os participantes demoraram mais tempo a reflectir sobre elas e as quais tiveram uma menor pontuação são as menos intuitivas. Optando pelo método de comparação entre os dois cenários as acções sobrepor, apagar, undo e redo tiveram uma melhor pontuação e um menor tempo no modo de gesto e voz do que apenas com gesto, isto porque foi difícil encontrar um gesto, ultrapassando-se essa barreira com o uso da voz. Com a oportunidade de utilizador comandos de voz para as acções menos intuitivas, referidas anteriormente, a maioria dos participantes utilizada a interacção gestual apenas para apontar.

Podemos assim concluir que para acções mais intuitivas o modo gestual pode ser sempre utilizado, com ou sem voz. Contrariamente, para as acções menos intuitivas, em que não foi possível obter apenas um conjunto de gesto mais relevante com as mesmas características, é preferível transmitir a acção pelos comandos de voz.

Os resultados do estudo mostram também, que é um erro assumir que a interacção gestual é uma boa solução para todas as acções. Isto foi corroborado pelos resultados de algumas acções terem uma preferência na interacção gestual com voz, tendo menores tempos de reflexão na tarefa e consequentemente uma melhor classificação nesse modo, levando a concluir que são menos intuitivas e não são adequadas para realizar através de gestos.

Capítulo 4

Comparação de dois cenários de interacção

Nesta secção é apresentado um estudo que explora o mapeamento entre gestos e comandos de voz para acções, assim como apenas o mapeamento de gestos para acções, que resultou do estudo apresentado no capítulo anterior. Enquanto no estudo anterior era pedido aos utilizadores que criassem gestos e/ou palavras para diferentes acções consoante o cenário de interacção em causa (gestos com ou sem voz), neste estudo, utilizou-se o processo inverso, ou seja, os gestos e/ou comandos de voz estavam pré-definidos e foi pedido aos participantes que os executassem para cada acção em cada cenário (gestos com ou sem voz). Através dos vários valores medidos ao longo da sessão, como o tempo que demoravam a realizar o comando, assim como a quantidade de vezes que os participantes recorreram à folha, onde estava ilustrado o gesto ou descrito o comando de voz, foi possível analisar de que modo o mapeamento previamente adquirido era apropriado, intuitivo e confortável. Pretendeu-se também, compreender quais as acções dessas aplicações que estão aptas a serem realizadas através de gestos e/ou voz.

4.1 Enquadramento e Preparação do estudo

O objectivo principal do estudo foi validar as conclusões da sessão experimental anterior, ou seja, confirmar se os comandos de voz e gestos para cada acção são intuitivos e qual o cenário de interacção mais intuitivo para cada acção num ambiente em que não existe nenhum contacto entre o utilizador e a superfície de interacção, ou qualquer outro dispositivo de entrada.

De forma a alcançar esse propósito reunimos os gestos e comandos de voz padrões para cada acção e reutilizámos os dois cenários de interacção (gestos com ou sem voz). Seguidamente, às duas aplicações de teste anteriores (“parede” e “mesa” de imagens) adicionámos uma nova aplicação de teste, o Google Earth, de forma a reforçar que a interacção gestual com ou sem voz se aplica a aplicações que permitam o manuseamento de objectos que sejam similares à manipulação no mundo real. Por fim, enumerámos

tarefas, englobando todas as acções já definidas na sessão experimental anterior, e criámos novas acções para a manipulação do Google Earth. Durante a realização das tarefas os participantes poderiam recorrer às folhas de apoio que continham os gestos e os comandos de voz correspondentes para cada acção.

Como resultados pretendemos obter, para cada acção, qual o modo de interacção mais intuitivo. Para tal vários factores foram avaliados: tempo médio de reflexão para cada acção foi contabilizado; número de vezes que os participantes se socorreram da folha; de que modo utilizaram a interacção gestual (apontaram ou realizaram gestos) no cenário de interacção com ambas as modalidades; se realizaram gestos nesse cenário de interacção (voz e gestos), se estes eram diferentes dos estabelecidos no cenário de interacção gestual; se a informação da acção é transmitida duas vezes através dos gestos e da voz; se o gesto de selecção (fechar a mão) foi executado para seleccionar os objectos quando se usufruiu da interacção gestual para apontar. O último factor de estudo é novo, é analisado se os participantes quando tinham presentes as duas modalidades de interacção e não repetiam a informação pelas duas modalidades possíveis de que forma é que utilizam a mão, apenas para sinalizar a informação sem qualquer movimento específico (mão aberta) ou uniam os dedos, fechando a mão, e assim executando o gesto para seleccionar o objecto, quando a acção de seleccionar e aumentar é pedida.

4.2 Cenários e acções

Seleccionaram-se três aplicações de teste, duas delas (“mesa” e “parede” de imagens) migram da primeira sessão experimental e a terceira é uma nova aplicação de teste, o Google Earth. As três aplicações são representativas de configurações que podem beneficiar da utilização da interacção gestual com ou sem voz sem qualquer superfície de apoio. Durante a sessão experimental existem dois cenários de interacção, um apenas com gestos e o outro cenário é com ambas as modalidades, gestos e voz.

Duas aplicações possibilitam a manipulação de imagens, e a nova aplicação de teste, Google Earth, permite, como o próprio nome indica, a manipulação do globo. A terceira aplicação surgiu com o intuito de reforçar o modo intuitivo da interacção gestual com ou sem voz para algumas acções que são comuns às aplicações anteriores e demonstrar como outras aplicações podem beneficiar da interacção gestual com ou sem voz sem nenhuma superfície de apoio.

Qualquer uma das três aplicações pode beneficiar da utilização da interacção gestual com ou sem auxílio de comandos de voz, principalmente em cenários em que as aplicações são projectadas em superfícies de grandes dimensões, efectuando-se uma manipulação directa sobre a representação gráfica dos objectos, uma manipulação próxima da que ocorre no mundo real.

As acções das duas aplicações de teste, “mesa” e “parede” com imagens, foram reu-



Figura 4.1: Aplicações de teste: (1)“Mesa” com imagens, (2)“Parede” de imagens, (3)Google Earth

tilizáveis, assim como os gestos e os comandos de voz correspondentes, realizados pelo maior número de participantes na sessão experimental do primeiro estudo (tabelas 4.1 , 4.2 e 4.3). Posteriormente, para o Google Earth, definiu-se um conjunto de acções que permite manipular o objecto e em que algumas acções são comuns ou similares às acções das primeiras aplicações.

“Mesa” com imagens:

- Acções aplicadas a uma imagem: zoom in, zoom out, rodar no sentido dos ponteiros do relógio, rodar no sentido contrário aos ponteiros do relógio, deslocar a imagem, sobrepor uma imagem a outra e apagar a imagem.

“Parede” de imagens:

- Acções aplicadas a uma imagem: zoom in, zoom out, mostrar imagem seguinte e anterior, seleccionar e aumentar, fechar uma imagem que se encontrava aumentada
- Acções aplicadas a todo o ambiente: zoom in, zoom out, movimentar para a esquerda, movimentar para a direita.

Google Earth:

- Acções aplicadas a ao globo: zoom in, zoom out, deslocar para cima/baixo/esquerda/direita, marcar um local, deslocação que rastreia a direcção da mão do participante, parar a deslocação.

De maneira a que todas as tarefas enumeradas anteriormente fossem executadas durante a sessão experimental foram pensadas tarefas que englobassem as acções. Por exemplo, em vez de se pedir ao utilizador para deslocações no Google Earth, solicitava-se uma movimentação até à China.

4.3 Gestos e comandos de voz padrão

Do estudo anterior, por vezes, para uma acção obtivemos mais do que um comando de voz ou mais do que um gesto padrão. Nesta situação tivemos que optar apenas por um. Não optámos por testá-los todos, de forma a não sobrecarregar a memória dos participantes e provocar confusão para optar por um para executar uma tarefa. As escolhas que realizámos serão reavaliadas posteriormente.

Na aplicação “mesa” de imagens (tabela 4.1) a acção rotação tem dois gestos com

Acções	Comando de voz	1ª	2ª	3ª	Descrição
Zoom In	“Zoom in”				Dois pontos a afastarem-se (tipicamente, paralelamente à superfície de interacção).
Zoom Out	“Zoom out”				Dois pontos a aproximarem-se (tipicamente, paralelamente à superfície de interacção).
Sobrepôr	“Frente/Para a frente”				Um ponto a afastar-se da superfície de interacção.
Movimentar	“Arrastar”				Um ponto a deslocar-se paralelamente à superfície de interacção.
Parar Movimento	“Pára”				Um ponto imóvel.
Rotação SPR/SCPR	“Rodar esquerda/direita”				Um ponto centrado no objecto e outro que descreve um movimento de rotação com início no outro ponto.
Apagar	“Apagar”				Construção de uma forma de X, quer recorrendo a um ou dois pontos.

Tabela 4.1: Gestos e comandos de voz padrões para cada acção na aplicação mesa de imagens








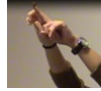
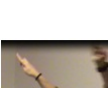


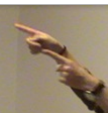
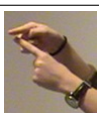
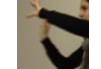


Acções	Comando de voz	1ª	2ª	3ª	Descrição
Seleccio- nar e aumentar	“Aumen- tar”				Um ponto que altera a sua composição (é importante o afastamento dos dedos).
Fechar	“Fecha”				Dois pontos a aproximarem-se (tipicamente, paralelamente à superfície de interacção).
Seguinte/ Anterior	“Seguinte/ Anterior”				Um ponto num movimento para a esquerda ou para a direita paralelamente à superfície de interacção.
Zoom in total	“Zoom in”				Dois pontos a afastarem-se (tipicamente, paralelamente à superfície de interacção).
Zoom out total	“Zoom out”				Dois pontos a aproximarem-se (tipicamente, paralelamente à superfície de interacção).
Movimen- tar tudo para a esquerda ou direita	“Movi- mentar esquerda” “Movi- mentar direita”				Dois pontos num movimento para a esquerda ou para a direita paralelamente à superfície de interacção.

Tabela 4.2: Gestos e comandos de voz padrões para cada acção na aplicação parede de imagens


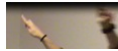


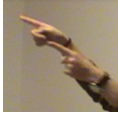


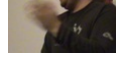
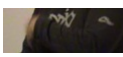






Acções	Comando de voz	1ª	2ª	3ª	Descrição
Zoom in	“Zoom in”				Dois pontos a afastarem-se (tipicamente, paralelamente à superfície de interacção).
Zoom out	“Zoom out”				Dois pontos a aproximarem-se (tipicamente, paralelamente à superfície de interacção).
Cima ou Baixo ou Esquerda ou Direita	“Cima” “Baixo” “Esquerda” “Direita”				Um ponto num movimento para cima ou baixo ou esquerda ou direita paralelamente à superfície de interacção.
Agarrar	“Arrastar”				Um ponto a deslocar-se paralelamente à superfície de interacção.
Largar	“Pára”				Um ponto imóvel.
Marcação	“Marcação”				Um ponto a aproximar-se da superfície de interacção e depois a afastar-se.

Tabela 4.3: Gestos padrões para cada acção na aplicação Google Earth

o mesmo número de participantes, e esses gestos diferem apenas do número de pontos (número de mãos). Optámos pelo gesto que inclui dois pontos, obtendo assim um raio maior, fulcral na rotação da imagem, do que na alternativa com um ponto (uma mão). Para a acção de apagar decidimos optar pelo gesto que teve maior número de participantes na primeira sessão experimental, 3 participantes.

Na aplicação Google Earth (tabela 4.3), para as acções de zoom in e out optámos pelos mesmos gestos e comandos de voz, já estudados. As acções cima/baixo/esquerda/direita são similares à seguinte/anterior e movimentar o ambiente para a esquerda/direita, diferenciando-se do gesto que corresponde à última acção pelo número de pontos e também por cada gesto ter uma direcção específica. Para a acção de agarrar e largar o globo seleccionámos os mesmos comandos padrão das acções de mover e parar o movimento de uma imagem, porque a tarefa consiste no mesmo, só se altera o objecto. Os comandos para marcar um local foram novos, optámos por arriscar num gesto que já tínhamos observado ser implementado em filmes na internet com o Google Earth. O comando de voz que pareceu mais intuitivo foi “marcação”, consoante os comandos das outras aplicações já estudadas, por exemplo, para apagar uma imagem o comando padrão foi “apagar”.

4.4 Participantes

Nesta sessão experimental participaram 15 voluntários, com idades compreendidas entre os 14 e os 55 anos, sendo a média de idades de 24 anos. Quase todos os participantes eram estudantes universitários, exceptuando três deles, dois trabalhadores e um estudante do ensino básico. Quase todos eles têm mais de 10 anos de experiência de utilização de computadores, excepto três com menos de 10 anos. Todos afirmaram conhecer pelo menos um dispositivo de interacção gestual, por exemplo, dispositivos Apple iPhone, iPad, iPod, Nintendo Wii, Xbox 360 e telemóveis com reconhecimento gestual. No entanto, apenas seis participantes usam com regularidade telemóveis com reconhecimento gestual, Nintendo Wii e iPhone. Nove participantes também conhecem pelo menos um dispositivo com reconhecimento de voz, desde telemóveis com reconhecimento de voz, Nintendo WiiU, kit de mãos livres a programas de cursos de línguas. Apenas quatro participantes usam com regularidade o reconhecimento de voz. Três participantes conhecem dispositivos que usem as duas modalidades de interacção, a Nintendo WiiU, mas nenhum utiliza esse dispositivo.

4.5 Procedimento

Cada participante realizou uma sessão individual, manipulando as três aplicações de teste através de duas formas distintas de interacção, gestos com ou sem voz. No início de cada sessão o propósito do estudo foi explicado aos participantes. Antes dos testes com

cada protótipo uma folha com as ilustrações dos gestos e/ou com os comandos de voz foi fornecida aos participantes (tabela 4.1 , 4.2 e 4.3) de forma a visualizarem os gestos e comandos de voz que correspondiam a cada acção. Quando os participantes se sentiram preparados, a folha era retirada e colocada num sítio estratégico, para que quando os mesmos recorressem a ela fosse possível o avaliador observar. Posteriormente, era pedido aos participantes que desempenhassem um conjunto de tarefas delineadas de forma a englobarem cada uma das acções pelo menos uma vez. Não existia nenhum limite de tempo para a execução das tarefas e era possível recorrer à folha de ilustrações sempre que fosse necessário. Este procedimento repetiu-se duas vezes, quando a interacção inclui apenas gestos e no de gestos e voz.

Por fim, foi pedido aos participantes que preenchessem um questionário de perfil e um questionário de satisfação no qual todas as acções dos três protótipos nos dois cenários de interacção foram classificadas numa escala de 1 a 5, onde 1 significa que o comando de gesto ou voz pré-definido é o menos adequado e intuitivo e 5 o mais adequado e intuitivo.

4.6 Detalhes técnicos

Os detalhes técnicos desta sessão experimental são quase todos semelhantes aos da primeira sessão experimental. Resume-se na utilização do Microsoft Kinect e de um projector ligados a um computador Windows 7, com uma resolução de ecrã de 1366 por 768 pixéis, e projectou-se na parede ocupando uma área, aproximadamente, de 280 cm por 200 cm. Nesta sessão os participantes encontravam-se, novamente, de pé e a interacção foi executada sem nenhum meio de intermediário.

A diferença reside no papel do avaliador que já não executava o papel de “Wizard of Oz”, ou seja, o reconhecimento de voz e gestual estava implementado, excepto para o gesto de apagar que se recorreu a uma tecla de atalho para mapear a acção. Isso sucedeu, devido ao gesto de apagar ser um gesto mais complexo do que os outros, porque precisava de ser indicado quando era iniciado a acção e quando esta terminava de modo a seguir todos os pontos da mão e ver se correspondia ao símbolo de 'x'. Por causa do restante tempo que dispunhamos optámos por não realizar o mapeamento desta acção e realizar os testes. Foi interessante notar que durante a sessão experimental os participantes não notavam que a acção apagar não estava mapeada.

De modo a se conseguir alcançar a implementação do reconhecimento de gesto e voz recorreu-se à biblioteca OpenNi e à biblioteca da Microsoft Speech Recognition para o reconhecimento de voz na nossa língua materna. Devido à biblioteca OpenNi limitar-se às linguagens C++ e C#, a escolha recaiu para C#. Para poder reutilizar os protótipos da primeira sessão experimental realizou-se a interoperabilidade entre C# e ActionScript.

A duração de cada sessão experimental com cada participante foi de, aproximadamente, 55 minutos. Cada aplicação teste levou cerca de 15/20 minutos nos dois cenários

de interacção a executar as tarefas pedidas. O tempo médio de utilização das aplicações ultrapassou o tempo indicado anteriormente se adicionarmos o tempo em que cada participante após terminar todas as tarefas pedidas decidiu “brincar” com as aplicações.

4.7 Análise dos resultados

Os resultados obtidos na sessão experimental, alguns captados pela câmara de filmar, e a sua correspondente análise encontram-se organizados por duas secções: primeiro os dados relativos ao cenário de interacção gestual com voz, seguindo-se, de dados de comparação para os dois cenários de interacção. Não existe uma sessão referente apenas aos dados da interacção apenas com gestos, porque os assuntos que foram analisados também foram no cenário com gestos e voz, então optámos pela realização de uma comparação entre as duas, abordando os seus pontos. Primeiramente temos para a interacção com gesto e voz:

- a repetição da informação da acção pelas duas modalidades;
- e a execução do gesto de selecção ou não.

Na análise dos dados que correspondem aos dois cenários segue-se:

- contabilização do tempo que cada participante demorou a executar cada uma das acções;
- auxílio da folha com gestos e comandos de voz;
- execução de outros gestos assim como outros comandos de voz sem serem os definidos.

4.7.1 Interacção gestual com voz:

Durante a sessão experimental as tarefas tinham que ser realizadas com comandos de voz complementando a interacção gestual.

É relevante analisar, como na sessão experimental anterior, como é que os participantes podem usufruir do modo de interacção gestual para diferentes finalidades, ou seja, apontar para o objecto ou ambiente de trabalho de modo a seleccioná-lo ou repetir a informação que é transmitida pelo comando de voz, a acção, pela execução do gesto. Na tabela 4.4 observamos de que forma a interacção gestual foi utilizada e o seu número de participantes. A última coluna ilustra quais os gestos que foram realizados quando estes foram diferentes dos padronizados para o outro cenário de interacção (interacção gestual).

-Mudança de gestos com voz

Quando os participantes interagiram com voz e gestos para transmitir a acção e optavam por utilizar a interacção gestual para realizar novamente a acção, por vezes ocorreu uma mudança dos gestos padrões, como se pode observar na última coluna da tabela anterior.


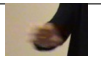


	Acção	Gesticular (nº de pessoas)	Apontar(nº de pessoas)	Gesto	Descrição
Mesa de imagens	Zoom in	2	13		
	Zoom out	2	13		
	Mover	10	5		
	Parar Movimento	10	5		
	Sobrepôr	4	11		
	Rodar SPR ou Rodar SCPR	4	11	 (2 pessoas)	Um ponto a efectuar um movimento de rotação.
	Apagar	3	12	 (2 pessoas)	Deslocar re- petidamente um ponto para a esquerda e direita.
 (1 pessoa)				Um ponto a afastar-se da superfície de interacção.	
Parede de imagens	Seleccionar e Au- mentar	5	10		
	Fechar	4	11		
	Seguinte	11	5		
	Anterior	11	5		
	Zoom in total	8	7		
	Zoom out total	8	7		
	Movimentar para a direita	13	2		Um ponto num movimento para a esquerda ou para a direita paralelamente à superfície de interacção.
Movimentar para a esquerda	13	2			
Google Earth	Zoom in	7	8		
	Zoom out	7	8		
	Cima	9	6		
	Baixo	9	6		
	Esquerda	9	6		
	Direita	9	6		
	Agarrar	3	12		
	Largar	3	12		
	Marcação	1	14		

Tabela 4.4: Número de participantes que apontaram e gesticularam para cada acção

O gesto padrão de apagar foi o que mais variação teve quando era possível interagir com as duas modalidades. Duas pessoas fizeram o gesto que corresponde a dois movimentos laterais e um participante realizou um gesto descrito como um ponto a afastar-se da superfície de interacção (gesto idêntico ao gesto padrão para a acção de sobrepor).

Por sua vez, para as tarefas de rodar uma imagem e movimentar o ambiente todo, para a esquerda ou direita, a única diferença em relação ao gesto padrão é o número de pontos que os participantes usaram. Dois participantes para a rotação usaram apenas um ponto (uma mão) e para movimentarem o ambiente para a esquerda ou direita os treze participantes que gesticularam usaram apenas um ponto (uma mão) também.

-Apontar ou gesticular

O facto dos participantes optarem por transmitir a informação da acção apenas pela voz, não repetindo a informação através das duas modalidades, pode resultar de dois factores:

-primeiro, porque estão na presença de acções para as quais os gestos são pouco intuitivos. Desta forma, os utilizadores apontam para o objecto que é o alvo e transmitem a acção através do comando de voz. Por exemplo, o gesto padrão para a acção de apagar é pouco intuitivo e até problemático, porque apenas três indivíduos repetiram informação e dos três gestos que executaram nenhum é idêntico ao gesto padrão.

-pode depender do alvo. Se for um único objecto ou se o alvo, sendo um objecto ou um conjunto, abrange quase toda a área. Como exemplo, temos a comparação da tarefa de zoom in e out de uma imagem com a de toda a área. O número de indivíduos aumenta de 2 para 8 ou 7 quando passamos de um zoom in e out de uma imagem para o zoom in e out do conjunto de imagens ou o zoom in e out no Google Earth, que ocupava quase a totalidade do ecrã. O que potencia também esta análise é a tarefa de movimentação de todo o ambiente para a esquerda ou para a direita em que a informação é repetida pela maioria dos participantes, treze.

Quando a interacção gestual tem apenas a finalidade de identificar e seleccionar o objecto, ou seja, apontar para o alvo, isso pode ser feito de duas maneiras. Como é visível na tabela 4.5 dentro de apontar temos duas subcategorias: apontar com o gesto de selecção (figura 2, imagem 1) ou apontar sem o gesto de selecção (figura 4.2, imagem 2).

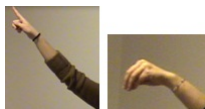


Figura 4.2: Gesto de apontar (1), gesto de selecção (2)

Através da análise da tabela podemos observar que há uma grande percentagem de participantes que realizam o gesto de selecção, para identificar e seleccionar o argumento da acção. Isto acontece quando o argumento da acção é singular e a área do objecto é pequeno, ou seja, uma imagem. Caso contrário, quando a área do alvo é maior, ocupando

	Acções	Gesto de selecção (nº de pessoas)	Sem gesto de selecção(nº de pessoas)
Mesa de imagens	Zoom in	6	9
	Zoom out	6	9
	Mover	10	5
	Parar Movimento	10	5
	Sobrepor	6	9
	Rodar SPR/SCPR	5	10
	Apagar	8	7
Parede de imagens	Seleccionar e Aumentar	8	7
	Fechar	0	15
	Seguinte	1	14
	Anterior	1	14
	Zoom in total	0	15
	Zoom out total	0	15
	Movimentar para a direita	0	15
	Movimentar para a esquerda	0	15
Google Earth	Zoom in	0	15
	Zoom out	0	15
	Cima	0	15
	Baixo	0	15
	Esquerda	0	15
	Direita	0	15
	Agarrar	3	12
	Largar	3	12
	Marcação	0	15

Tabela 4.5: Número de participantes que executaram o gesto de selecção ou apenas apontaram

quase todo o ambiente, independentemente de o alvo ser apenas um objecto ou vários, a preferência é apontar sem realizar o gesto de selecção.

4.7.2 Comparação dos dois cenários de interacção

-Folha explicativa:

Durante a sessão experimental quando necessário os participantes poderiam recorrer à folha ilustrativa dos gestos e dos comandos de voz. Na tabela 4.6 observamos o número de vezes para cada acção e cada cenário de interacção que o recurso foi usado.

O gesto que mapeava a acção fechar foi o que teve maior número de participantes, nove, que necessitou de ser lembrado- Como consequência tiveram que recorrer à folha. Destes nove, oito deles realizaram o gesto da figura 4.3 (fechar e abrir a mão), gesto idêntico ao de seleccionar e aumentar. A maioria dos participantes realizou gesto idêntico à acção seleccionar e aumentar de forma a acontecer o oposto, ou seja, fechar. Um deles não executou nenhum gesto imediatamente a seguir à tarefa ser pedida, reflectindo alguns segundos sobre o mesmo e só depois recorreu à folha.

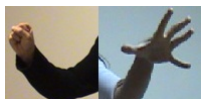


Figura 4.3: Sequência de imagens que alguns participantes usaram para transmitir a acção apagar (1)mão fechada para seleccionar o objecto (2)mão aberta para a imagem ser fechada

Segue-se o gesto de apagar uma imagem com 5 pessoas. Na tabela 4.7 podemos ver os diferentes gestos que foram realizados, assim como o número de pessoas que o executou em alternativa ao gesto padrão. A principal razão para os participantes terem recorrido a outros gestos, sem ser o padrão, reflecte que para a tarefa de apagar o gesto escolhido (o que tinha mais participantes no estudo anterior) não foi o mais indicado e nenhum será o mais intuitivo. O uso de outros gestos sem ser o padrão é por causa dos conjunto de gestos que surgiram não terem semelhanças entre si. Continua assim a ser complicado encontrar um gesto padrão para a tarefa apagar.

Por sua vez, para a acção de movimentar uma imagem, quando o cenário de interacção envolvia as duas modalidades, 3 participantes recorreram à folha para recordarem qual o comando de voz. A palavra que foi pronunciada pelas três pessoas foi “mover” e era “arrastar”. Seguem-se as acções sobrepor, rodar imagem e seleccionar e aumentar, em todas existiram sempre dois participantes que recorreram à folha. Os comandos enunciados erradamente foram: “sobrepor” em vez de “frente” ou “para a frente”; “abre” e “abrir” em vez de “aumenta”. Para a acção de rodar nenhum dos dois indivíduos referiu algum comando antes de recorrer ao auxiliar.

-Tempo médio de reflexão para cada acção:

	Ações	Nº de vezes: (Gestos)	Nº vezes: (Gestos e voz)
Mesa de imagens	Zoom in	0	0
	Zoom out	0	0
	Mover	0	3
	Parar Movimento	0	0
	Sobrepor	2	2
	Rodar SPR/SCPR	0	2
	Apagar	5	0
Parede de imagens	Seleccionar e Aumentar	0	2
	Fechar	9	0
	Seguinte	0	0
	Anterior	0	0
	Zoom in total	0	0
	Zoom out total	0	0
	Movimentar para a direita	1	0
	Movimentar para a esquerda	0	0
Google Earth	Zoom in	0	0
	Zoom out	0	0
	Cima	0	0
	Baixo	0	0
	Esquerda	0	0
	Direita	0	0
	Agarrar	0	0
	Largar	0	0
	Marcação	0	0

Tabela 4.6: Número de participantes que recorreu à folha auxiliar, para cada acção

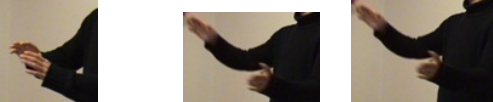
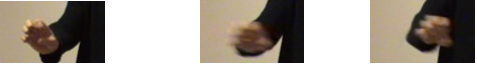
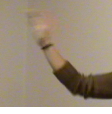
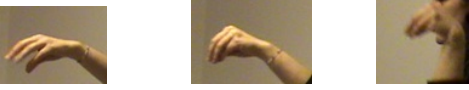
Gestos que não eram mapeados para a acção de apagar	Descrição	Nº de participantes
	Dois pontos a afastarem-se (tipicamente, paralelamente à superfície de interacção).	1
	Deslocar repetidamente um ponto para a esquerda e direita.	2
	Dois movimentos frontais. Um ponto a alternar a profundidade.	1
	Um ponto a afastar-se da superfície de interacção.	1

Tabela 4.7: Gestos alternativos para a acção de apagar, no cenário de interacção com gestos e voz

Na tabela 4.8 podemos visualizar, nos dois modos de interacção, a média de tempo, em segundos, de reflexão de cada acção até à sua execução.

	Ação	Média de tempo(segundos) Gestos (1-5) Gestos	Média de tempo(segundos) Gestos e Voz
Mesa de imagens	Zoom in	0	0
	Zoom out	0	0
	Mover	0.13	0.14
	Parar Movimento	0	0
	Sobrepor	0.46	0.68
	Rodar SPR/SCPR	0	0.03
	Apagar	3.1	0
Parede de imagens	Seleccionar e aumentar	0	1.95
	Fechar	4.6	0.13
	Seguinte	0	0.11
	Anterior	0	0
	Zoom in total	0	0
	Zoom out total	0	0
	Movimentar para a direita ou esquerda	1.2	0
Google Earth	Zoom in	0	0
	Zoom out	0	0
	Cima	0	0
	Baixo	0	0
	Esquerda	0	0
	Direita	0	0
	Agarrar	0	0
	Largar	0	0
	Marcação	0.46	0.33

Tabela 4.8: Tempo médio (em segundos) para cada acção em cada cenário

Através da tabela é possível observar que em três acções (apagar, seleccionar e aumentar, e fechar imagem) ocorre uma diferença significativa do tempo de execução consoante o cenário de interacção.

A tarefa apagar tem o valor médio de 3,1 segundos, apenas com interacção gestual, mas quando há a possibilidade de utilização da voz na interacção, a média de tempo de reflexão diminui drasticamente sendo de 0 segundos. O facto do tempo médio ser elevado unicamente com interacção gestual deveu-se a cinco participantes terem realizado gestos que não correspondiam ao padrão, logo não eram mapeados para a acção correspondente e, por fim, estes 5 recorreram à folha ilustrativa.

É difícil encontrar um gesto relevante para a acção apagar, mas é bastante intuitivo com o recurso à voz, superando assim o problema do gesto. A mesma explicação aplica-se para entender porque é que quando os dois modos de interacção são viáveis, doze

participantes utilizam a interacção gestual apenas para apontar para a imagem e a voz para comunicar a acção (tabela 4.4). Os restantes participantes, três, transmitiram a acção pela interacção gestual mas todos realizaram gestos diferentes do padrão (tabela 4.4).

Contrariamente à anterior, a acção seleccionar e aumentar uma imagem, no modo de interacção gestual, obteve o tempo médio de 0 segundos mas com os dois modos houve um aumento do tempo médio alcançando 1,95 segundos. O motivo desse incremento foi devido à invocação de outros comandos (“abre” e “abrir”) por parte dos utilizadores e não referiram o comando padrão (“aumentar”). Foi necessário recorrer à folha auxiliar.

Por fim, para a acção fechar uma imagem, o cenário de interacção com maior fracção de tempo foi o gestual com 4,6 segundos. Existiu uma redução do tempo quando estamos na presença de voz e gestos para menos 4,47 segundos (foi de 0,13s). No modo de interacção gestual 9 participantes recorreram à folha ilustrativa para realizar o gesto devido a executarem outro gesto, diferente do padrão.

No entanto no cenário com as duas modalidades de interacção o tempo médio obtido ser de 0,13 segundos foi consequência de nenhum participantes ter recorrido à folha explicativa e este tempo ser de uma reflexão mais demorada por parte de dois participantes antes de realizarem o gesto. Conclui-se assim que para a realização da tarefa fechar é muito mais intuitivo recorrer à voz e o gesto de apagar é problemático e não é natural e intuitivo. E de forma a reforçar a conclusão, quando os participantes tinham duas modalidades disponíveis a maioria, 11, utilizou a interacção gestual apenas para apontar para o alvo e transmitir a informação da acção por comandos de voz (tabela 4.4).

Questionários de satisfação:

Após o término da sessão experimental os participantes preencheram os questionários de satisfação, obtendo-se assim a classificação dada a cada acção consoante o cenário de interacção.

Na tabela 4.9 podemos visualizar a classificação média de cada uma das acções dependendo do modo de interacção (gestos com ou sem voz). A classificação varia de 1 a 5 em que 5 significa que a interacção utilizada era a mais natural e intuitiva e 1 o seu oposto:

Em termo de comparação, podemos verificar que para a acção apagar e fechar há uma maior preferência para interacção com ambas as modalidades (gestos e voz). Este episódio deve-se à dificuldade que os participantes tiveram em recordar os gestos padrões, executando outros gestos. A dificuldade é ultrapassada quando se transmite a acção através de comandos de voz.

Em relação a todas as outras acções, não há nenhuma grande discrepância na preferência de interacção. No entanto, verificamos preferência geral pela interacção com ambas as modalidades. Podemos justificar isso, devido ao reconhecimento gestual ter algumas falhas, por exemplo, quando o ângulo e/ou a posição dos braços, mãos ou dedos não se encontrava na posição estipulada não ocorria um mapeamento do gesto na acção

	Acção	Classificação média (1-5) Gestos	Classificação média (1-5) Gestos e Voz
Mesa de imagens	Zoom in	4.5	4.9
	Zoom out	4.5	4.9
	Mover	4.5	4.4
	Parar Movimento	4.4	4.7
	Sobrepôr	4.5	4.6
	Rodar SPR	4.5	4.9
	Rodar SCPR	4.5	4.9
	Apagar	3.2	4.8
	Seleccionar e aumentar	4.5	4.2
	Fechar	3.2	4.7
Parede de imagens	Seguinte	4.7	5
	Anterior	4.7	5
	Zoom in total	4.7	5
	Zoom out total	4.7	5
	Movimentar para a direita ou esquerda	4.1	4.6
Google Earth	Zoom in	4.7	5
	Zoom out	4.7	5
	Cima	4.3	5
	Baixo	4.3	5
	Esquerda	4.3	5
	Direita	4.3	5
	Agarrar	4.7	5
	Largar	4.7	5
	Marcação	4.5	4.9

Tabela 4.9: Classificação média (1 a 5) para cada acção em cada cenário

correspondente e isso pode ter influenciado as preferências.

Mão dominante:

Na tabela 4.10 observamos qual foi a mão dominante dos participantes durante a sessão experimental, consoante o cenário de interacção e a aplicação teste.

Seis participantes alteraram a mão dominante durante a interacção com a “mesa” de imagens independentemente do cenário. A alternância dependia da tarefa pedida e da mão que estivesse mais próxima do alvo. Após uma “pausa” na interacção para se informarem da tarefa seguinte, isto é, nenhuma mão a interagir nesse momento, a mão dominante ia depender da sua proximidade ao objecto alvo.

Nas outras duas aplicações, catorze indivíduos usaram sempre a mão direita para interagir e um deles utilizou a mão esquerda. Temos conhecimento que um participante não é esquerdino a escrever mas pratica as suas actividades diárias com a mão esquerda como dominante, e foi este participante que usufruiu da mão esquerda nas duas aplicações. Nenhum participante alternou a mão, iniciando a interacção, nas duas aplicações, com a mão

Cenário de interacção	Aplicações de teste	Mão Direita (nº pessoas)	Mão Esquerda (nº pessoas)	Alternado (nº pessoas)
Gestos	Mesa de imagens	8	1	6
	Parede de imagens	14	1	0
	Google Earth	14	1	0
Gestos e voz	Mesa de imagens	8	1	6
	Parede de imagens	14	1	0
	Google Earth	14	1	0

Tabela 4.10: Mão dominante do participante consoante o cenário de interacção e a aplicação teste

direita e mantiveram essa postura até terminar. Esse facto pode estar relacionado com os sujeitos da acção e as suas orientações e organizações. Apesar de algumas tarefas envolverem uma única imagem esta encontrava-se sempre com uma distância similar das duas mãos e existia uma harmonia entre todas as imagens, organizadas como um todo (num conjunto) e não se encontram casos mais extremos em que o alvo se situa, por exemplo, no canto superior direito como na aplicação “mesa” de imagens. Quando o sujeito da acção é o ambiente todo o participante continua a interacção com a mão dominante porque o alvo é o todo e não um objecto localizado mais à esquerda ou direita.

4.7.3 Discussão

Baseado no tempo médio de reflexão e no número de vezes que cada participante recorreu à folha de ilustrações de modo a recordar os gestos ou comandos de voz padrão para cada acção, conclui-se como esperado, que esses dois valores estão relacionados. As acções para as quais os participantes demoraram mais tempo, são as que foram mais vezes auxiliadas pela folha e/ou também houve um momento de reflexão antes da execução do gesto ou comando de voz (apagar, fechar, seleccionar e aumentar, mover).

As acções menos intuitivas estão directamente dependentes do cenário de interacção. Quando temos um cenário de interacção gestual as acções apagar e fechar são as mais demoradas, têm maior número de indivíduos que recorreram à folha, e por sua vez as menos intuitivas. Incrementando à interacção comandos de voz as acções mover e seleccionar e aumentar uma imagem são as que maiores dificuldades trouxeram aos participantes.

Temos sempre um cenário alternativo de interacção para as acções problemáticas. A alternativa surge porque quando estamos, por exemplo, apenas com interacção gestual, as acções que foram problemáticas deixam de o ser quando interagimos com voz e gestos. Neste modo alternativo de interacção os participantes preferem apenas apontar e quando

efectuam gestos esses não correspondem ao padrão, assim como o número de acessos à folha é nulo e o tempo de reflexão não existe. A mesma explicação sucede quando surge a alternativa de gesticular aos comandos de voz.

Todas as outras acções, exceptuando as referidas nos parágrafos anteriores, são intuitivas, independentemente do cenário de interacção. Nestas acções os participantes demoraram pouco tempo, houve pouco ou nenhum acesso à folha auxiliar e obtiveram uma boa classificação.

Analisando os resultados dos questionários, conclui-se que nas três aplicações de teste o cenário de interacção com voz e gestos é o que apresenta os melhores resultados, excepto para as acções de mover, seleccionar e aumentar. No entanto a diferença de classificação, entre os dois cenários, é reduzida. Esta preferência difere da classificação dada na primeira sessão experimental. O cenário de interacção com voz e gestos apresenta melhores resultados devido a dois factores distintos. Um dos factores é porque com a modalidade da voz os participantes conseguem ultrapassar as dificuldades quando não se recordavam dos gestos (por exemplo, acções apagar e fechar), o outro factor pode ser devido ao reconhecimento gestual ter alguns problemas.

No entanto verificamos que existe uma conformidade entre três factores medidos na sessão experimental: os resultados dos questionários (classificação), o tempo médio de reflexão, e o número de vezes de uso da folha explicativa e número de participantes que repetiram ou não a informação no cenário de interacção com gestos e voz. Em geral, as acções para as quais os participantes executaram o gesto e/ou comando de voz mais rapidamente e recorreram menos ou nenhuma vez à folha tiveram uma classificação boa e vice-versa. No cenário de interacção com as duas modalidades, por vezes os indivíduos repetiram a informação da acção com gestos. Quando a maioria dos participantes transmite a acção através de gesto e voz é comprovado que o gesto padrão para a acção é intuitivo (por exemplo as acções: mover, parar movimento, seguinte, anterior, movimentar esquerda/direita, cima, baixo, cima, etc.), excepto quando os participantes executam gestos diferentes do padrão (por exemplo a acção apagar).

Os dados recolhidos em relação à repetição de informação (tabela 4.3) permitem tirarmos outra conclusão, para acções em que o alvo abrange uma área menor (acções como, zoom in, zoom out, sobrepor, rodar, apagar, seleccionar e aumentar uma imagem ou marcação num local) a percentagem de indivíduos que repete a informação é diminuta. Caso contrário, se o argumento da acção abranger uma grande área ou for um conjunto de objectos os participantes têm preferência por repetir a informação da acção com gestos e voz.

A nova aplicação de teste, o Google Earth, serviu para salientar que o dicionário de gestos e comandos de voz definido pode servir para outras aplicações (por exemplo: zoom in, zoom out). Convém realçar que se forem adicionados novos gestos ou comandos de voz que mapeiem novas acções estes devem ser directos e o mais possível relacionados

com o mundo real. Por exemplo, na nova aplicação de teste, a acção de movimentar o globo numa direcção (esquerda, direita, cima, baixo) permite realizar tarefas que se tornam similares ao manusear um objecto não digital, como um globo ou um mapa. Os gestos e comandos de voz que mapeavam as acções de movimentação do globo são similares, por sua vez, aos gestos e comandos de voz para as acções de seguinte ou anterior de uma imagem, na aplicação “mesa” de imagens, e para as acções de movimentar o ambiente total para a esquerda ou direita, na aplicação “parede” de imagens. Todas estas acções têm em comum o manuseamento do objecto para uma certa direcção, verificando-se que as acções cima, baixo, esquerda, direita são intuitivas também na manipulação do Google Earth, independentemente do cenário de interacção. Desta forma, quando surgem novas acções em novas aplicações que são do mesmo “tipo” podemos definir novos comandos de voz e gestos que sigam o molde do seu “tipo”.

Capítulo 5

Conclusão e Trabalho Futuro

Neste capítulo apresentam-se as conclusões obtidas depois dos estudos realizados e abrem-se perspectivas para trabalho futuro.

5.1 Conclusão

Durante a realização deste trabalho foram exploradas maneiras de melhorar o nível de usabilidade em dois cenários (interacção gestual com ou sem voz), sem qualquer contacto físico entre os periféricos de entrada e o utilizador, em superfícies de grande dimensão. Com este fim, realizaram-se dois estudos de modo a contribuir na caracterização nos dois cenários possíveis e compreender melhor que factores afectam a experiência.

No primeiro estudo através de “Wizard of Oz”, de dois protótipos e de um conjunto de acções, foi pedido aos participantes que executassem tarefas nos dois cenários distintos de interacção com o principal intuito de encontrar gestos e comandos de voz padrão para cada acção. Os padrões para cada acção foram encontrados para a maioria das acções. No entanto, para algumas acções não foi possível obter apenas um gesto padrão mas um conjunto de gestos, como é o caso da acção apagar porque não haviam características comuns entre os conjuntos de gestos nem uma maioria de participantes a realizar um determinado conjunto de gestos. E quando os participantes dispunham das duas modalidades de interacção estes utilizavam a voz para transmitir as acções menos intuitivas, como o caso da acção apagar, e a interacção gestual apenas como apontador de forma a seleccionar o alvo da acção. Deste modo é possível concluir que para acções que não sejam muito intuitivas é bom termos presente nas aplicações comandos de voz, porque assim, podemos colmatar este problema.

A interacção por voz foi preferida em outras situações, nesta sessão experimental. Quando os participantes têm como argumento da acção apenas um objecto (por exemplo) uma imagem, estes preferem interagir com voz sem repetição de informação. Contrariamente, se o alvo da acção for um conjunto de objectos, como o exemplo de movimentar todas as imagens para a esquerda ou direita, os participantes têm uma preferência por

utilizar os gestos e na presença das duas modalidades de interacção usam as duas em simultâneo.

Os comandos por voz são importantes para solucionar a ambiguidade que existiu quando para diferentes acções obtivemos os mesmos gestos, se não fosse possível diferenciar pelo alvo em causa poderíamos usufruir da voz de forma a eliminar esta ambiguidade. Um exemplo disso, teria sido se tivéssemos optado na segunda sessão experimental pelo gesto padrão de zoom out para a acção de apagar também. Uma maneira possível de diferenciar essa acção seria por comandos de voz e a realização do gesto na mesma ou então o uso de interacção gestual apenas para seleccionar o alvo.

No cenário com voz e gestos, podemos definir acções em que a informação é transmitida pelas duas modalidades, sem se repetir, tivemos o exemplo dos participantes transmitirem o comando “rodar” e a direcção da acção com a interacção gestual. Pode-se optar por usufruir das duas modalidades sem sobrecarregar nenhuma modalidade com toda a informação que é necessária transmitir.

Através da primeira sessão experimental podemos também concluir que as acções que foram mais problemáticas na interacção gestual, nas quais os participantes demoraram mais tempo a reflectir sobre que gesto realizar, e por sua vez uma menor classificação quando comparadas no modo de interacção gestual com voz, são as que são menos intuitivas e as que os participantes não estão habituados a utilizar no seu dia-a-dia com dispositivos de toque, como por exemplo o iPad, ou como por exemplo a Nintendo Wii.

A segunda sessão experimental surgiu com o objectivo de avaliar os conjuntos de gestos e comandos de voz da sessão anterior. Verificou-se que os menos intuitivos na sessão anterior continuaram a ser um pouco problemáticos e os participantes tiveram que recorrer à folha auxiliar mais vezes de forma a lembrarem-se do gesto, ou então realizaram o gesto erradamente. Esses gestos são os menos intuitivos e aqueles que foram mais facilmente esquecidos pelos participantes. As acções que são mais problemáticas na interacção gestual, como a acção apagar, são exemplos de acções em que informação da acção deve ser transmitida através de comandos de voz, neste caso.

Através da inserção de outra aplicação teste, o Google Earth, foi possível concluir que para acções que são mais usuais em outros dispositivos ou aplicações não surge nenhum problema de usabilidade com os comandos de gesto ou voz definidos. Logo, é possível definir novas aplicações de teste desde que estas não saiam do contexto das anteriores e faça sentido a interacção sem qualquer dispositivo físico de entrada.

A segunda sessão experimental foi executada sem “Wizard of Oz” mas com o reconhecimento de gestos e voz. Foi patente que a falta de eficácia do reconhecimento de gestos poderia acabar por desmotivar os participantes, afectando a atribuição da classificação no cenário de interacção gestual. Num uso mais extenso poderia trazer consequências maiores, como uma maior desmotivação e provocar o abandono da interacção gestual. Concluimos deste modo que a interacção, apesar de estar cada vez mais em voga, conti-

nua a colocar alguns entraves numa interacção fluida e eficaz, devido a alguns problemas tecnológicos com que se depara. No entanto, no nosso estudo na presença do cenário de interacção com gesto e voz a barreira tecnológica facilmente era ultrapassada pelo reconhecimento do comando de voz que disputava a acção.

Durante os dois estudos foi perceptível que o paradigma WIMP não foi concebido para sistemas sem as modalidades de interacção clássicas, como o rato e o teclado e também não é adaptado a ecrãs de grandes dimensões. O uso destes dois tipos de cenários de interacção, gestos com ou sem voz, tem como intuito minimizar os problemas e as limitações introduzidas pela ausência das modalidades clássicas. Isto se os gestos e comandos de voz padrões forem utilizados para a mesma finalidade, ou então outros que sigam os padrões. Mas de forma pragmática a comparação nos dois cenários revelou vantagens e desvantagens de cada um e como se podem colmatar as desvantagens um do outro.

5.2 Trabalho futuro

A interacção gestual esta cada vez mais em voga, nas mais variadas plataformas. O foco deste trabalho foi na interacção numa superfície de grandes dimensões e sem qualquer contacto físico entre utilizador e dispositivo físico. Como tal seria interessante utilizar as três aplicações de testes durante sessões experimentais idênticas em dispositivos de menores dimensões e com outras características e a interacção gestual seria realizada por toque com o dispositivo. Por exemplo, no contexto de um Tablet PC se o dicionário gestual se manteria e os gestos menos intuitivos também. Desta modo, verificaríamos que diferenças existem neste tipo de interacção nas várias em diferentes plataformas e analisar de que modo as mesmas técnicas e mapeamentos poderiam ser utilizados, assim como é que isso iria influenciar o uso da voz e da preferência por o cenário de interacção com gesto e voz. Analisaríamos de que forma o mapeamento dos gestos e comandos de voz em acções é transversal às mais variadas dimensões dos dispositivos e se os problemas se mantinham ou se dependiam do tipo de plataforma e tipo de interacção gestual. As futuras conclusões iriam facilitar o desenho das aplicações baseadas em gestos com ou sem voz e quando seria preferível e se adequado a utilização dos comandos de voz.

Também seria interessante aprofundar o tema da mão dominante que surgiu na última sessão experimental de modo a perceber melhor e confirmar a influência da localização do objecto com a mão que os participantes utilizam para interagir.

Bibliografia

- [1] Micah Alpern and Katie Minardo. Developing a car gesture interface for use as a secondary task. In *CHI '03 extended abstracts on Human factors in computing systems*, CHI EA '03, pages 932–933, New York, NY, USA, 2003. ACM.
- [2] Francesca A. Barrientos and John F. Canny. Cursive:: controlling expressive avatar gesture using pen gesture. In *Proceedings of the 4th international conference on Collaborative virtual environments*, CVE '02, pages 113–119, New York, NY, USA, 2002. ACM.
- [3] Olivier Bau and Wendy E. Mackay. Octopocus: a dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, UIST '08, pages 37–46, New York, NY, USA, 2008. ACM.
- [4] Thomas Baudel and Michel Beaudouin-Lafon. Charade: remote control of objects using free-hand gestures. *Commun. ACM*, 36:28–35, July 1993.
- [5] Andrew Blake, Rupert Curwen, and Andrew Zisserman. A framework for spatiotemporal control in the tracking of visual contours. *Int. J. Comput. Vision*, 11:127–145, October 1993.
- [6] Aaron F. Bobick, Stephen S. Intille, James W. Davis, Freedom Baird, Claudio S. Pinhanez, Lee W. Campbell, Yuri A. Ivanov, Arjan Schütte, and Andrew Wilson. The kidsroom: A perceptually-based interactive and immersive story environment. *Presence: Teleoper. Virtual Environ.*, 8:369–393, August 1999.
- [7] Richard A. Bolt. “put-that-there”: Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '80, pages 262–270, New York, NY, USA, 1980. ACM.
- [8] Richard A. Bolt and Edward Herranz. Two-handed gesture in multi-modal natural dialog. In *ACM Symposium on User Interface Software and Technology*, pages 7–14, 1992.

- [9] Lars Bretzner, Ivan Laptev, Tony Lindeberg, and Yngve Sundblad. A prototype system for computer vision based human computer interaction, 2001.
- [10] W Buxton, E Fiume, R Hill, A Lee, and C Woo. *Continuous hand-gesture driven input*, volume 83, pages 191–195. 1983.
- [11] W. Buxton and B. Myers. A study in two-handed input. *SIGCHI Bull.*, 17:321–326, April 1986.
- [12] Marcio C. Cabral, Carlos H. Morimoto, and Marcelo K. Zuffo. On the usability of gesture interfaces in virtual reality environments. In *Proceedings of the 2005 Latin American conference on Human-computer interaction*, CLIHC '05, pages 100–108, New York, NY, USA, 2005. ACM.
- [13] J. Callahan, D. Hopkins, M. Weiser, and B. Shneiderman. An empirical comparison of pie vs. linear menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '88, pages 95–100, New York, NY, USA, 1988. ACM.
- [14] Xiang Cao and Ravin Balakrishnan. Visionwand: interaction techniques for large displays using a passive wand tracked in 3d. *ACM Trans. Graph.*, 23:729–729, August 2004.
- [15] Philip R. Cohen, Michael Johnston, David McGee, Sharon Oviatt, Jay Pittman, Ira Smith, Liang Chen, and Josh Clow. Quickset: multimodal interaction for simulation set-up and control. In *Proceedings of the fifth conference on Applied natural language processing*, ANLC '97, pages 20–24, Stroudsburg, PA, USA, 1997. Association for Computational Linguistics.
- [16] James W. Davis and Serge Vaks. A perceptual user interface for recognizing head gesture acknowledgements. In *Proceedings of the 2001 workshop on Perceptive user interfaces*, PUI '01, pages 1–7, New York, NY, USA, 2001. ACM.
- [17] Alan J. Dix, Janet E. Finlay, Gregory D. Abowd, and Russel Beale. *Human-Computer Interaction*. 2nd edition, 1998.
- [18] Klaus Dorfmüller and Hanno Wirth. Real-time hand and head tracking for virtual environments using infrared beacons. In *Proceedings of the International Workshop on Modelling and Motion Capture Techniques for Virtual Environments*, CAPTECH '98, pages 113–127, London, UK, 1998. Springer-Verlag.
- [19] Carlos Duarte and António Neto. Gesture interaction in cooperation scenarios. In *Proceedings of the 15th international conference on Groupware: design, implementation, and use*, CRIWG'09, pages 190–205, Berlin, Heidelberg, 2009. Springer-Verlag.

- [20] Jacob Eisenstein and Randall Davis. Visual and linguistic information in gesture classification. In *ACM SIGGRAPH 2007 courses*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.
- [21] George Fitzmaurice, Azam Khan, Robert Pieké, Bill Buxton, and Gordon Kurtenbach. Tracking menus. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, UIST '03, pages 71–79, New York, NY, USA, 2003. ACM.
- [22] Fulano, Cicrano, and Beltrano. A paper on something. In *The 7th Conference on Things and Stuff (CTS 2009)*, Lisbon, Portugal, May 2009. Accepted for publication.
- [23] D M Gavrilu and L S Davis. Towards 3-d model-based tracking and recognition of human movement: a multi-view approach. *on automatic faceand gesturerecognition*, pages 3–8, 1995.
- [24] Tovi Grossman, Daniel Wigdor, and Ravin Balakrishnan. Multi-finger gestural interaction with 3d volumetric displays. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, UIST '04, pages 61–70, New York, NY, USA, 2004. ACM.
- [25] Radek Grzeszczuk, Gary R. Bradski, Michael H. Chu, and Jean-Yves Bouguet. Stereo based gesture recognition invariant to 3d pose and lighting. In *CVPR*, pages 1826–1833. IEEE Computer Society, 2000.
- [26] Yves Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model, 1987.
- [27] François Guimbretiére and Terry Winograd. Flowmenu: combining command, text, and data entry. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, UIST '00, pages 213–216, New York, NY, USA, 2000. ACM.
- [28] Carl Gutwin and Reagan Penner. Improving interpretation of remote gestures with telepointer traces. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, CSCW '02, pages 49–57, New York, NY, USA, 2002. ACM.
- [29] Alexander G. Hauptmann and Paul McAvinney. Gestures with speech for graphic manipulation. *Int. J. Man-Mach. Stud.*, 38:231–249, February 1993.
- [30] Ken Hinckley, Patrick Baudisch, Gonzalo Ramos, and Francois Guimbretiére. Design and analysis of delimiters for selection-action pen gesture phrases in scriboli.

- In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '05, pages 451–460, New York, NY, USA, 2005. ACM.
- [31] Eric Horvitz and Tim Paek. A computational architecture for conversation. In *Proceedings of the seventh international conference on User modeling*, pages 201–210, Secaucus, NJ, USA, 1999. Springer-Verlag New York, Inc.
- [32] Giancarlo Iannizzotto, Carlo Costanzo, Francesco La Rosa, and Lanzafa Pietro. A multimodal perceptual user interface for video-surveillance environments. In *Proceedings of the 7th international conference on Multimodal interfaces*, ICMI '05, pages 45–52, New York, NY, USA, 2005. ACM.
- [33] Giancarlo Iannizzotto, Massimo Villari, and Lorenzo Vita. Hand tracking for human-computer interaction with graylevel visualglove: turning back to the simple way. In *Proceedings of the 2001 workshop on Perceptive user interfaces*, PUI '01, pages 1–7, New York, NY, USA, 2001. ACM.
- [34] Michael Johnston and Srinivas Bangalore. Finite-state multimodal integration and understanding. *Nat. Lang. Eng.*, 11:159–187, June 2005.
- [35] Nebojsa Jojic, Thomas Huang, Barry Brumitt, Brian Meyers, and Steve Harris. Detection and estimation of pointing gestures in dense disparity maps. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, FG '00, pages 468–, Washington, DC, USA, 2000. IEEE Computer Society.
- [36] Paul Kabbash, William Buxton, and Abigail Sellen. Two-handed input in a compound task. In *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, CHI '94, pages 417–423, New York, NY, USA, 1994. ACM.
- [37] Ed Kaiser, Alex Olwal, David McGee, Hrvoje Benko, Andrea Corradini, Xiaoguang Li, Phil Cohen, and Steven Feiner. Mutual disambiguation of 3d multimodal interaction in augmented and virtual reality. In *Proceedings of the 5th international conference on Multimodal interfaces*, ICMI '03, pages 12–19, New York, NY, USA, 2003. ACM.
- [38] Maria Karam and m. c. schraefel. A study on the use of semaphoric gestures to support secondary task interactions. In *CHI '05 extended abstracts on Human factors in computing systems*, CHI EA '05, pages 1961–1964, New York, NY, USA, 2005. ACM.

- [39] Sanshzar Kettebekov. Exploiting prosodic structuring of coverbal gesticulation. In *Proceedings of the 6th international conference on Multimodal interfaces, ICMI '04*, pages 105–112, New York, NY, USA, 2004. ACM.
- [40] Sanshzar Kettebekov and Rajeev Sharma. Toward natural gesture/speech control of a large display. In *Proceedings of the 8th IFIP International Conference on Engineering for Human-Computer Interaction, EHCI '01*, pages 221–234, London, UK, 2001. Springer-Verlag.
- [41] Rick Kjeldsen and Jacob Hartman. Design issues for vision-based computer interaction systems. In *Proceedings of the 2001 workshop on Perceptive user interfaces, PUI '01*, pages 1–8, New York, NY, USA, 2001. ACM.
- [42] David B. Koons, Carlton J. Sparrell, and Kristinn R. Thorisson. *Integrating simultaneous input from speech, gaze, and hand gestures*, pages 257–276. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1993.
- [43] David B. Koons, Carlton J. Sparrell, and Kristinn R. Thorisson. Readings in intelligent user interfaces. chapter Integrating simultaneous input from speech, gaze, and hand gestures, pages 53–64. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998.
- [44] Steffi Beckhaus Kristopher, Kristopher J. Blom, and Matthias Haringer. A new gaming device and interaction method for a first-person-shooter. In *IN COMPUTER SCIENCE AND MAGIC 2005, GC DEVELOPER SCIENCE TRACK*, 2005.
- [45] Gordon Kurtenbach and William Buxton. Issues in combining marking and direct manipulation techniques. In *Proceedings of the 4th annual ACM symposium on User interface software and technology, UIST '91*, pages 137–144, New York, NY, USA, 1991. ACM.
- [46] Gordon Kurtenbach and William Buxton. User learning and performance with marking menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence, CHI '94*, pages 258–264, New York, NY, USA, 1994. ACM.
- [47] Doo Young Kwon and Markus Gross. Combining body sensors and visual sensors for motion training. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology, ACE '05*, pages 94–101, New York, NY, USA, 2005. ACM.
- [48] Joseph J. LaViola Jr. Bringing vr and spatial 3d interaction to the masses through video games. *IEEE Comput. Graph. Appl.*, 28:10–15, September 2008.

- [49] Sören Lenman, Lars Bretzner, and Björn Thuresson. Using marking menus to develop command sets for computer vision based hand gesture interfaces. In *Proceedings of the second Nordic conference on Human-computer interaction*, NordiCHI '02, pages 239–242, New York, NY, USA, 2002. ACM.
- [50] G. Julian Lepinski, Tovi Grossman, and George Fitzmaurice. The design and evaluation of multitouch marking menus. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 2233–2242, New York, NY, USA, 2010. ACM.
- [51] Michael Moyle and Andy Cockburn. The design and evaluation of a flick gesture for 'back' and 'forward' in web browsers. In *Proceedings of the Fourth Australasian user interface conference on User interfaces 2003 - Volume 18*, AUIC '03, pages 39–46, Darlinghurst, Australia, Australia, 2003. Australian Computer Society, Inc.
- [52] Jeannette G. Neal and Stuart C. Shapiro. *Knowledge-based multimedia systems*, pages 403–438. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1994.
- [53] Jiazhi Ou, Susan R. Fussell, Xilin Chen, Leslie D. Setlock, and Jie Yang. Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks. In *Proceedings of the 5th international conference on Multimodal interfaces*, ICMI '03, pages 242–249, New York, NY, USA, 2003. ACM.
- [54] Sharon Oviatt. Mutual disambiguation of recognition errors in a multimodel architecture. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, CHI '99, pages 576–583, New York, NY, USA, 1999. ACM.
- [55] Sharon Oviatt. Ten myths of multimodal interaction. *Communications of the ACM*, 42(11):74–81, 1999.
- [56] Sharon Oviatt. Taming recognition errors with a multimodal interface. *Commun. ACM*, 43:45–51, September 2000.
- [57] Sharon Oviatt, Antonella DeAngeli, and Karen Kuhn. Integration and synchronization of input modes during multimodal human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '97, pages 415–422, New York, NY, USA, 1997. ACM.
- [58] Joseph A. Paradiso, Kai yuh Hsiao, Joshua Strickon, Joshua Lifton, and Ari Adler. Sensor systems for interactive surfaces. *IBM Systems Journal*, 39(3,4):892–, 2000.

- [59] Robert Pastel and Nathan Skalsky. Demonstrating information in simple gestures. In *Proceedings of the 9th international conference on Intelligent user interfaces, IUI '04*, pages 360–361, New York, NY, USA, 2004. ACM.
- [60] Shwetak N. Patel, Jeffrey S. Pierce, and Gregory D. Abowd. A gesture-based authentication scheme for untrusted public terminals. In *Proceedings of the 17th annual ACM symposium on User interface software and technology, UIST '04*, pages 157–160, New York, NY, USA, 2004. ACM.
- [61] Vladimir I. Pavlovic, Rajeev Sharma, and Thomas S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19:677–695, July 1997.
- [62] Wayne Piekarski, Ben Avery, Bruce H. Thomas, and Pierre Malbezin. Colorplate: Integrated head and hand tracking for indoor and outdoor augmented reality. *Virtual Reality Conference, IEEE*, 0:276, 2004.
- [63] Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCullough, and Rashid Ansari. Multimodal human discourse: gesture and speech. *ACM Trans. Comput.-Hum. Interact.*, 9:171–193, September 2002.
- [64] J M Rehg and T Kanade. Digiteyes: vision-based hand tracking for human-computer interaction. *Proceedings of 1994 IEEE Workshop on Motion of Nonrigid and Articulated Objects*, (November):16–22, 1994.
- [65] Jun Rekimoto. Smartskin: an infrastructure for freehand manipulation on interactive surfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves, CHI '02*, pages 113–120, New York, NY, USA, 2002. ACM.
- [66] Jun Rekimoto, Takaaki Ishizawa, Carsten Schwesig, and Haruo Oba. Presense: interaction techniques for finger sensing input devices. In *Proceedings of the 16th annual ACM symposium on User interface software and technology, UIST '03*, pages 203–212, New York, NY, USA, 2003. ACM.
- [67] Volker Roth and Thea Turner. Bezel swipe: conflict-free scrolling and multiple selection on mobile touch screen devices. In Dan R. Olsen Jr., Richard B. Arthur, Ken Hinckley, Meredith Ringel Morris, Scott E. Hudson, and Saul Greenberg, editors, *CHI*, pages 1523–1526. ACM, 2009.
- [68] Dean Rubine. Combining gestures and direct manipulation. In *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '92*, pages 659–660, New York, NY, USA, 1992. ACM.

- [69] Dan Saffer. *Designing Gestural Interfaces: Touchscreens and Interactive Devices*. O'Reilly, Beijing, 2008.
- [70] Emilio Schapira and Rajeev Sharma. Experimental evaluation of vision and speech based multimodal interfaces. In *Proceedings of the 2001 workshop on Perceptive user interfaces*, PUI '01, pages 1–9, New York, NY, USA, 2001. ACM.
- [71] Thomas Schlömer, Benjamin Poppinga, Niels Henze, and Susanne Boll. Gesture recognition with a wii controller. In Albrecht Schmidt, Hans Gellersen, Elise van den Hoven, Ali Mazalek, Paul Holleis, and Nicolas Villar, editors, *Tangible and Embedded Interaction*, pages 11–14. ACM, 2008.
- [72] Chris Schmandt, Jang Kim, Kwan Lee, Gerardo Vallejo, and Mark Ackerman. Mediated voice communication via mobile ip. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, UIST '02, pages 141–150, New York, NY, USA, 2002. ACM.
- [73] Philip Schuchardt and Doug A. Bowman. The benefits of immersion for spatial understanding of complex underground cave systems. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology*, VRST '07, pages 121–124, New York, NY, USA, 2007. ACM.
- [74] Mara G. Silva and Doug A. Bowman. Body-based interaction for desktop games. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, CHI EA '09, pages 4249–4254, New York, NY, USA, 2009. ACM.
- [75] G. M. Smith and m. c. schraefel. The radial scroll tool: scrolling support for stylus- or touch-based document navigation. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, UIST '04, pages 53–56, New York, NY, USA, 2004. ACM.
- [76] Sreeram Sreedharan, Edmund S. Zurita, and Beryl Plimmer. 3d input for 3d worlds. In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*, OZCHI '07, pages 227–230, New York, NY, USA, 2007. ACM.
- [77] Thad E. Starner and Alex Pentland. Visual recognition of american sign language using hidden markov models, 1995.
- [78] Björn Stenger, Thomas Woodley, Tae-Kyun Kim, and Roberto Cipolla. A vision-based system for display interaction. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, BCS-HCI '09, pages 163–168, Swinton, UK, UK, 2009. British Computer Society.

- [79] Norbert A. Streitz, Jorg Geissler, Torsten Holmer, Shin'ichi Konomi, Christian Müller-Tomfelde, Wolfgang Reischl, Petra Rexroth, Peter Seitz, and Ralf Steinmetz. i-land: an interactive landscape for creativity and innovation. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, CHI '99, pages 120–127, New York, NY, USA, 1999. ACM.
- [80] Martin Usoh, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P. Brooks, Jr. Walking: walking-in-place - flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '99, pages 359–364, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.
- [81] Andries van Dam. Post-wimp user interfaces. *Commun. ACM*, 40:63–67, February 1997.
- [82] Christian von Hardenberg and François Bérard. Bare-hand human-computer interaction. In *Proceedings of the 2001 workshop on Perceptive user interfaces*, PUI '01, pages 1–8, New York, NY, USA, 2001. ACM.
- [83] Andrew Wilson and Nuria Oliver. Gwindows: robust stereo vision for gesture-based control of windows. In *Proceedings of the 5th international conference on Multimodal interfaces*, ICMI '03, pages 211–218, New York, NY, USA, 2003. ACM.
- [84] Andrew Wilson and Steven Shafer. Xwand: Ui for intelligent spaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '03, pages 545–552, New York, NY, USA, 2003. ACM.
- [85] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19:780–785, July 1997.
- [86] Mike Wu and Ravin Balakrishnan. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, UIST '03, pages 193–202, New York, NY, USA, 2003. ACM.
- [87] Ying Wu and Thomas S. Huang. Vision-based gesture recognition: A review. In *Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction*, GW '99, pages 103–115, London, UK, 1999. Springer-Verlag.
- [88] Ying Yin and Randall Davis. Toward natural interaction in the real world: real-time gesture recognition. In *International Conference on Multimodal Interfaces and*

- the Workshop on Machine Learning for Multimodal Interaction, ICMI-MLMI '10*, pages 15:1–15:8, New York, NY, USA, 2010. ACM.
- [89] Robert Zeleznik and Andrew Forsberg. Unicam — 2d gestural camera controls for 3d environments. In *Proceedings of the 1999 symposium on Interactive 3D graphics, I3D '99*, pages 169–173, New York, NY, USA, 1999. ACM.
- [90] Zhengyou Zhang, Ying Wu, Ying Shan, and Steven Shafer. Visual panel: virtual mouse, keyboard and 3d controller with an ordinary piece of paper. In *Proceedings of the 2001 workshop on Perceptive user interfaces, PUI '01*, pages 1–8, New York, NY, USA, 2001. ACM.
- [91] Shengdong Zhao and Ravin Balakrishnan. Simple vs. compound mark hierarchical marking menus. In *Proceedings of the 17th annual ACM symposium on User interface software and technology, UIST '04*, pages 33–42, New York, NY, USA, 2004. ACM.

