

2 Conceitos Preliminares

Nesta tese, um ambiente virtual 3D interativo é um ambiente tridimensional dinâmico gerado em tempo real por técnicas de computação gráfica. Este ambiente está vinculado a um processo interativo contínuo onde as ações do usuário recebem respostas (*feedbacks*) geradas pelo sistema que gerencia o ambiente virtual. Ao longo do trabalho, o termo ambiente virtual 3D interativo será referenciado por *VE - Virtual Environment*.

Nos *VEs* existem tarefas 3D (*3DT - 3D Tasks*) executadas através de técnicas de interação 3D (*3DIt - 3D Interaction Techniques*). Basicamente as tarefas estão de alguma forma relacionadas com operações de seleção, manipulação e navegação (ver seção 2.4). Diversas *3DIt* foram desenvolvidas para cada tipo de *3DT*. Uma *3DIt* a princípio é considerada um software vinculado a uma aplicação de natureza 3D e depende de um conjunto de informações de entrada para ser executada. Estas informações são capturadas por meio dos dispositivos de entrada. Apesar de uma *3DIt* aparentemente ficar fortemente vinculada aos dados provenientes dos dispositivos de entrada, também há uma influência dos dispositivos de saída (*displays*) em questões como usabilidade e desempenho (ver Capítulo 3 na seção 3.2).

Este capítulo inicia esclarecendo o conceito de espaços de imagem que servem para entender quais são os lugares onde o usuário interage durante a realização de uma tarefa 3D. Em seguida são apresentados os estilos de interação 3D comumente encontrados nas interfaces de usuário 3D (*3DUIs - 3D User Interfaces*). Algumas tecnologias de dispositivos de entrada e *displays* utilizadas em interação 3D são descritas, assim como uma descrição dos tipos de *3DTs* relevantes para este trabalho. Por fim, há um overview sobre interface de transição, continuidade e congruência dimensional.

2.1 Espaços de Imagem e Estilos de Interação 3D

Estilos de interação são identificados na realização das *3DIts* e estão fortemente vinculados aos *displays* utilizados. Mas antes de descrevê-los, uma descrição dos lugares de atuação é necessária, ou seja, algumas formas de

espaços de imagem (BlundelS06, Blundel07) encontrados nos *displays* que exibem os gráficos 3D. A seguir há uma descrição sobre os tipos relevantes para esta tese:

Planar Image Space (PIS): Compreende literalmente a superfície (a maioria sendo 2D) dos *displays*, onde as imagens são retratadas. Um exemplo é a tela do monitor convencional de um *desktop*. Apesar desta superfície possibilitar a transmissão de informações visuais de profundidade (*depth cues*) do tipo monocular, outras informações importantes como as dos tipos binoculares e paralaxe de movimento ⁶ não são retratadas. Exibindo somente *monocular depth cues*, os olhos convergem e acomodam (focam) somente na superfície do *display*, e não nos detalhes visuais (cor, textura, forma, iluminação, etc) dos elementos retratados na imagem que estariam em profundidades diferentes.

Apparent Image Space (AIS): Este está relacionado com os *displays* que utilizam técnicas de estereoscopia (ver Apêndice A). Apesar das imagens serem geradas para uma superfície 2D, elas são percebidas como um volume 3D. Ou seja, as imagens estão fisicamente no plano de uma superfície 2D mas com o uso dos óculos estéreo, permitindo que informações binoculares cheguem ao cérebro, o aspecto tridimensional da imagem é percebido. Este aspecto 3D pode ser percebido estando para dentro do plano da imagem ou saindo dele. *Displays* com recursos de estereoscopia conseguem apresentar informações como paralaxe binocular e paralaxe de movimento. Mas problemas com acomodação e convergência ocorrem também. Além disso, podem ocorrer problemas com obstrução por objetos reais (ver Seção 2.3.1).

Virtual Image Space (VIS): Refere-se ao uso de um sistema óptico (normalmente usando recursos de reflexão) para projetar uma imagem que parece retratar elementos visuais que estão atrás de um componente óptico. Por exemplo, os *displays* do tipo *reachin displays* (ver Seção 2.3.2) usam uma superfície semi-transparente como componente óptico, e possibilitam interações manuais atrás dela. Este fato traz a vantagem do aparente contato direto com objetos virtuais (atrás da superfície semi-transparente) sem a interferência na visualização deles.

Em (BlundelS06, Blundel07), outros espaços de imagens (*Physical 3-D Image Space* e *Free Image Space*) são descritos, mas são relacionados a *displays* do tipo volumétricos, os quais não serão abordados nesta tese.

Os estilos de interação identificados nas *3DUIs* são:

Interação Indireta: Também conhecida por “Interação Transferida”,

⁶*Paralaxe de movimento* : Refere-se a uma dica de percepção de profundidade resultante do movimento da cabeça obtendo diferentes pontos de vista de um mesmo alvo de interesse.

consiste no uso de um dispositivo de entrada responsável pela movimentação de um objeto virtual no *VE*. Existe uma correspondência entre a movimentação deste dispositivo e a movimentação do objeto virtual, assim como acontece nas *2DUIs* de *desktop* em relação à movimentação do mouse e à movimentação do cursor, sendo que na situação das *3DUIs* há o acréscimo de mais uma dimensão. O objeto no *VE* serve como um *feedback* visual que mantém uma correspondência espacial e temporal.

Interação Direta: Neste estilo considera-se o uso do dispositivos de entrada literalmente na região do *AIS* e *VIS*. A liberdade de movimento do dispositivo é limitada pelo volume físico de interação e pelo aparente contato com os objetos virtuais que aparentam ser sólidos. Porém, para aparentar esta solidez seria necessário que o dispositivo fosse ou estivesse acoplado a algum outro dispositivo do tipo *force feedback*. Mas este estilo de interação pode levar a conflitos visuais como o explicado na propriedade *Masking* na seção de propriedades de *displays* visuais (Seção 2.3.1).

Blundel menciona mais um estilo de interação, porém é apropriado somente para *displays* volumétricos, e por isto está fora do escopo da tese.

2.2

Dispositivos de Entrada

Os dispositivos de entrada são os responsáveis pela coleta de informações enviadas pelo usuário para posteriormente serem utilizadas nas *3DIts*. Antes de enumerar algumas tecnologias comumente utilizadas, propriedades relevantes precisam ser descritas.

2.2.1

Propriedades

Graus de Liberdade - Degree of Freedom

Degree of Freedom (DOF) refere-se a uma forma independente de movimento que um corpo tem no espaço. Este movimento pode ser uma rotação ou uma translação ao longo de algum eixo. Um dispositivo de entrada pode ter mais de um DOF, e rastreadores de movimentos (*trackers*) tipicamente possuem 6 DOFs. Normalmente, o número de DOFs indica implicitamente o quão complexo é um dispositivo, e as possibilidades de seu uso nas *3DIts*.

Tipo de Dado

Os tipos de dados indicam quais são os modos de informação que são gerados e transmitidos por um dispositivo de entrada. Eles podem ser discretos,

contínuos ou uma combinação dos dois. Os discretos normalmente consistem em dados singulares, por exemplo, os do tipo booleano (verdadeiro e falso). Em aplicações interativas, estes dados são normalmente utilizados para ativar mudanças de estados. Os contínuos referem-se a geração de múltiplos valores de dados em relação às ações do usuário, por exemplo, as coordenadas espaciais 3D (x,y,z) enviadas por um rastreador de movimento ou até dados matriciais.

Affordances

Affordances de um dispositivo de entrada referem-se às possibilidades de ações que são percebidas ao observá-lo. No caso das aplicações interativas 3D, uma boa usabilidade pode surgir como resultado de um bom casamento entre um dispositivo que “*affords*” um tipo de interação que faz parte de uma metáfora de interação implementada por uma *3DIt*. Por exemplo, em um jogo virtual de tênis, onde o usuário tem a possibilidade de jogar contra um adversário virtual através de um *display* grande, provavelmente uma raquete de tênis (adaptada com sensores de rastreamento de movimento) seria um ótimo dispositivo ao invés de um outro qualquer que fornecesse também as mesmas informações de movimento.

Ergonomia

Questões ergonômicas são importantes para usabilidade de um dispositivo, por exemplo, peso, presença de cabos, conforto, restrição de movimentos, entre outras.

2.2.2

Tecnologias

Joysticks

São dispositivos que já acompanham o *desktop* há um bom tempo e ainda são úteis para aplicações 3D. Contêm uma combinação de eventos contínuos (posicionamentos 2D) mais um conjunto de componentes com eventos discretos, como botões, entre outros.

Trackers

Em muitas aplicações 3D é importante ter informações sobre a localização espacial 3D de objetos reais. Por exemplo, para se conseguir a percepção de profundidade através do movimento de paralaxe (ver Apêndice A) é necessária a localização da cabeça do usuário durante a interação em um *VE*. Para isto, existem os sistemas de rastreamento 3D ou simplesmente *Trackers*, que contêm



Figura 2.1: *CYCLOPE*⁷ optical tracker.

soluções compostas por um ou mais pontos de rastreamento, os quais fornecem as posições 3D. Estes pontos de rastreamento podem ser fixados em objetos de interesse, como por exemplo, ferramentas de interação e óculos. Atualmente existem diferentes tecnologias de rastreamento, as mais comuns são os *trackers* eletromagnéticos, acústicos, óticos (Figura 2.1) e inerciais.

Wands



Figura 2.2: *Wanda*⁸.

As *Wands* são considerados dispositivos que contêm eventos contínuos com informações de posicionamento 3D e eventos discretos resultantes de botões. São utilizados para interações espaciais utilizando movimentos no espaço, por exemplo, em tarefas de navegação e manipulação. Estes dispositivos também são às vezes classificados como 3D mice, ou seja, uma forma de mouse 3D. Um dos primeiros foi a *Wanda* (Figura 2.2) desenvolvido pela *Ascension*

⁷www.inrialpes.fr/sed/6doftracker

⁸www.inition.co.uk

Technology e utiliza internamente um sensor de rastreamento de um tracker eletromagnético.

Luvas



Figura 2.3: *Data Glove 5 Ultra*⁹.

Às vezes uma informação mais detalhada sobre a mão é importante para uma 3DUI, por exemplo, em técnicas de interação que utilizam o movimento da mão e dos dedos para movimentar uma mão virtual. Para isto, existem dispositivos na forma de luva que são sensíveis à dobra dos dedos e ao contato entre as pontas dos dedos (Figura 2.3).

2.3 Displays

A percepção de um *VE* está diretamente ligada ao que é apresentado por um sistema computacional. O sistema de percepção humano compreende cinco sentidos que enviam informações para o cérebro. Normalmente três (visão, audição e tato) desses cinco sentidos são estimulados pelo sistema computacional por meio da geração de estímulos sintéticos. O escopo desta tese se restringe à utilização de dispositivos que enviam estímulos visuais, logo não serão abordadas informações relacionadas aos demais sentidos.

No contexto dos *VEs*, os *displays* visuais são responsáveis por estimular o canal visual criando efeitos visuais que façam os usuários compreender um *VE* tridimensionalmente, ou seja, entender a relação de profundidade entre os elementos contidos neste ambiente. Esta relação de profundidade é inferida através de dicas visuais (*depth cues*) como: perspectiva, iluminação, sombra, oclusão, paralaxe de movimento, entre outros (BowmanKLP05, ShermanC02, BlundelS06) (ver Apêndice A).

Os *displays* para *VE* estão de alguma forma ligados com o termo imersão. É comum encontrar este termo vinculado com outro chamado presença. A

⁹www.5dt.com

origem do termo presença está relacionada com o conceito de tele-presença que foi inicialmente apresentado por Marvin Minsky (Minsky80) para descrever a sensação que alguém teria operando uma máquina remotamente, ou seja, a sensação de estar em um lugar diferente através de um sistema de tele-operação. Este conceito é aplicável no contexto de *VE*, mas para isto assume-se como mediadores desta sensação os equipamentos computacionais responsáveis por envolver o usuário tecnologicamente criando uma sensação de estar em um *VE*.

Sobre a relação entre imersão e presença, existem várias definições e até hoje não há um consenso sobre o significado dos dois. Desta forma, ao longo do texto será adotada a definição dada por Slater (Slater03) que caracteriza imersão de forma objetiva, descrevendo-a como o aparato tecnológico que dá suporte à sensação de presença, ou seja, algo que pode ser mensurado independente dos resultados da experiência virtual gerada. Para o termo presença são atribuídos os aspectos subjetivos resultantes na experiência virtual.

Em (Kjeldskov01) foi proposta uma classificação de *displays* utilizando imersão como critério. Duas categorias foram criadas: os Totalmente imersivos e os Semi-imersivos. Os primeiros são aqueles que têm *Field of View (FOV)* (ver Seção 2.3.1) disponibilizado em qualquer direção (CAVE de 6 lados, *HMDs*, *Booms*), enquanto os do segundo tipo não disponibilizam em todas as direções (Monitor, *WorkBenches*, *Panoramic Screens*, CAVE 3-5 lados). No estudo conduzido em (SteedP05) os *displays* foram divididos em *Head Mounted Displays (HMDs)* e *Immersive Projection Technologies (IPT)*. Neste trabalho será adotada a classificação encontrada em (ShermanC02), onde os *displays* estão divididos em três tipos: *Stationary Displays*, *Head-based Displays* e os *Hand-based Displays*. Os primeiro tipo compreende aqueles que estão fixos em um lugar. Os *displays* do segundo tipo são móveis e referem-se àqueles que de alguma forma estão fixos à cabeça do usuário, conseqüentemente permanecem numa posição fixa em relação aos olhos; normalmente são dispositivos na forma de capacetes. No último tipo estão *displays* móveis também, mas neste caso acompanham as mãos do usuário, por exemplo, o BOOM (ver Seção 2.3.4). O motivo da escolha por esta classificação foi a abrangência de *displays* que ela agrega, mas mesmo assim ainda podem surgir algumas confusões por haver *displays* que misturam características de mais de um tipo.

2.3.1

Propriedades

O conhecimento das propriedades dos *displays* é importante e pode influenciar na escolha de uma *3DIt*, assim como no grau de imersão. A seguir são listadas algumas propriedades relevantes para interação 3D (ShermanC02, BowmanKLP05, Kjeldskov01, BlundelS06):

Cor, Brilho e Contraste (*Display Quality*): Basicamente as opções sobre cores variam entre dois tipos: Monocromáticos (Preto e Branco) e os Tricromáticos (Coloridos). Estes últimos possuem três canais de cores e a combinação delas resulta no total de cores de disponibilizadas. Este total é influenciado pela profundidade de cor de cada canal: 4, 8, 16, 24 e 32 *bits*. O Contraste é a medida da diferença relativa entre claro e escuro. Quanto maior a faixa de contraste de um *display* mais fácil será distinguir diferentes elementos exibidos. O Brilho corresponde a uma medida da quantidade geral de luz emitida. É sempre desejável um alto brilho, pois existem tipos de *display* onde a quantidade de brilho é fundamental, como os projetores e alguns do tipo *AR See-through*. De forma geral todas essas propriedades influenciam a qualidade das imagens geradas pelas *engines* gráficas, e são importantíssimas para atingir imagens foto-realistas com convincentes *depth cues*. Alguns estudos relatam a influência da qualidade visual na sensação de imersão durante experiências virtuais (*Virtual Experience - VExp*).

Resolução Espacial: Compreende a quantidade de *pixels* contidos na direção vertical e horizontal de um *display*. Quanto maior a resolução mais detalhes são exibidos, em compensação, mais processamento computacional é necessário.

Field of View - FOV: É definido como a abertura máxima (número máximo de graus) de ângulo de visão que pode ser visualizado instantaneamente em um *display*. O *FOV* horizontal de um ser humano é aproximadamente 200 graus, com 120 graus na área de sobreposição binocular, ou seja, na área em que o *FOV* horizontal dos olhos se interceptam.

Field of Regard - FOR: É a quantidade de espaço físico (medida em graus de ângulos de visão) envolvendo o usuário. Esta quantidade de espaço físico corresponde à área onde estão sendo exibidas as imagens do *VE*. Em outras palavras, o quanto espacialmente o usuário é envolvido por imagens. Este não deve ser confundido com *FOV*, pois o *FOV* está relacionado com a influência visual em um instante de tempo, enquanto o *FOR* está vinculado ao espaço físico potencial que influencia visualmente o usuário. Por exemplo, pode acontecer que um dispositivo propicie um *FOV* de 30 graus mas tenha um *FOR* de 360, ou seja, a cada instante de tempo o usuário só pode visualizar

30 graus de um *VE* mas ele tem a possibilidade de visualizar os 360 graus em volta utilizando este *FOV* de 30. Um conceito muito parecido com o *FOR* é o *Available Field Of View - AFV*, descrito em (Kjeldskov01).

Geometria: Refere-se ao formato da superfície do *display* em que as imagens estão sendo exibidas. Na maior parte são geometrias planas representadas por superfícies no formato retangular mas há também aquelas com superfícies curvas como os *HemiSpherical Displays* ou *Panoramical Screens*. Às vezes, a geometria pode influenciar na qualidade da imagem exibida causando, por exemplo, o aparecimento de falhas visuais como buracos e distorções.

Taxa de Quadros: Corresponde à taxa em que as imagens são exibidas, e também é chamada *frame rate*. A unidade de medida correspondente é o números de quadros por segundo (*FPS - Frames Per Second*). Esta mesma medida pode ser expressa como *Hertz*, que é uma unidade de frequência para representar ocorrências por segundo. Um *FPS* acima de 30 é muito bom, uma taxa de 15 *FPS* é pouco aceito, e abaixo disso o cérebro começa a não mais perceber o efeito de movimento contínuo, e uma sequência de imagens torna-se evidente. Na literatura, esta propriedade também é encontrada com o nome Resolução Temporal (*Temporal Resolution*).

Opacidade: Também nomeada como *Real World Occlusion*, esta propriedade refere-se à capacidade do *display* em esconder, obstruir ou incluir a visão do mundo real. Os *HMDs* para *VR* escondem a visualização do mundo real. Os *displays* do tipo *Stationary*, como os monitores e a *CAVE*, obstruem parcialmente a visualização do mundo real. No caso de uma *CAVE* de 6 lados, ainda sim há uma obstrução da visão do mundo virtual porque o usuário pode ver o próprio corpo. Os *HMDs* para *AR* do tipo *See-through* fornecem uma visão combinada do mundo real com elementos virtuais, desta forma, eles incluem o mundo real.

Masking: Esta propriedade relaciona a obstrução de objetos reais por virtuais e vice-versa durante as interações. Normalmente estes objetos reais referidos são ferramentas de interação como *Wands* e/ou próprio corpo do usuário (frequentemente as mãos e braços).

No caso dos *Stationary Displays*, quando um objeto real, na frente da superfície, obstrui a visualização de um objeto virtual, duas situações importantes devem ser consideradas: a primeira quando o objeto virtual está realmente posicionado “virtualmente” atrás do objeto real, desta forma, a obstrução é válida; na segunda situação ocorre o inverso, o objeto virtualmente estaria posicionado na frente do real (entre os olhos do usuário e o objeto real), logo não poderia ocorrer a obstrução mas ela ocorre. Neste caso, a interação se dá com objetos virtuais que estão em paralaxe negativo (ver Apêndice A),

causando a impressão que estão saltando da superfície do *display*, e quando um objeto causa obstrução na visualização da imagem, ocorre um cancelamento do efeito da formação do objeto virtual que estaria na frente do objeto real. O objeto real cria uma região “vazia” chamada de “penumbra estereoscópica” que conseqüentemente destrói o efeito estereoscópico (BlundelS06).

Este problema é parcialmente contornado com tecnologias de *display* do tipo *See-through*, por exemplo um *HMD* para *AR* ou alguns tipos de *reachin displays*. Nestes casos é possível sobrepor o objeto real com um objeto virtual. Porém, deve-se ter o cuidado na situação inversa, ou seja, quando o objeto virtual estiver atrás do objeto real. Uma medida para contornar isto é não desenhar o objeto virtual; contudo é necessário o rastreamento da posição e forma do objeto real durante a interação.

Light Transfer: Descreve a forma como a luz é transferida para a superfície do *display*. Algumas formas são: projeção-traseira (*rear-projection*), projeção-frontal (*front-projection*), direto na retina através de *lasers*, e o uso de óptica através de jogos de lentes. A forma como a luz é transferida pode influenciar a escolha das tarefas de interação limitando a escolha de técnicas, por exemplo, para superfícies utilizando *front-projection*, o uso de *3DIts* para manipulação direta não funcionaria bem pois a proximidade das superfícies provocaria a geração de sombras e, conseqüentemente, problemas com a geração de efeitos de estereoscopia.

Ergonomia: Há várias características ergonômicas de um *display* que podem influenciar a interação 3D, causando o desvio da atenção do usuário resultando em problemas como restrição de movimentos e cansaço. Entre elas estão peso, acomodação física, presença de cabos, se é necessário carregar o *display*, presença de superfícies de apoio, entre outras.

Latência Gráfica: Às vezes é visível um certo atraso entre os movimentos do usuário e o *feedback* visual em resposta a estes movimentos. Esta propriedade não está relacionada com o *Taxa de Quadros do display*, pois as imagens podem estar sendo atualizadas corretamente. Por exemplo, um *HMD* com *head-tracking* permite que imagens sejam geradas de acordo com os movimentos da cabeça, entretanto, a latência pode ocorrer caso o usuário mova a cabeça e a atualização do ponto de vista ocorra alguns segundos ou milissegundos depois. Esta propriedade não está diretamente relacionada com o *display* em si, mas com o conjunto de elementos (dispositivo de entrada, poder de processamento da CPU, engine gráfica, etc) que compõem o *workspace*; de qualquer forma está sendo mantida aqui como informação adicional.

2.3.2 Stationary Displays

Monitores e FishTank

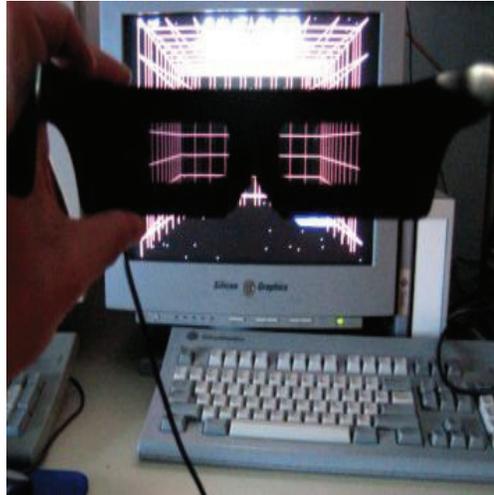


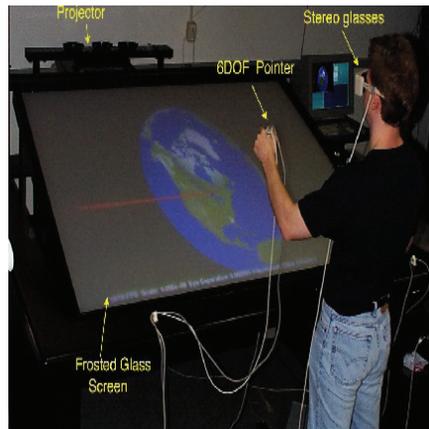
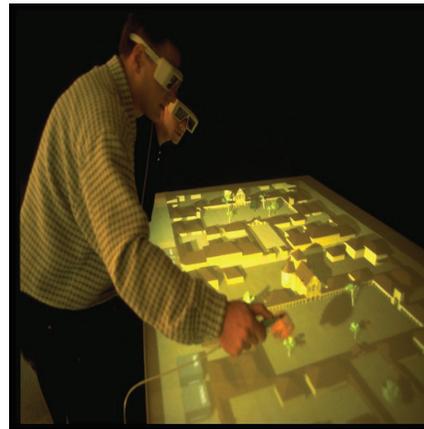
Figura 2.4: Monitor CRT com estereoscopia através do uso de óculos *estéreo*¹⁰.

Os monitores convencionais do tipo CRT (*Cathode Ray Tube*) são soluções baratas normalmente utilizadas para *3DIt*. São obtidos a custos relativamente baixos e oferecem estereoscopia em uma superfície plana através de óculos estereoscópicos (Figura 2.4). O efeito estéreo é conseguido graças às altas taxas de *refresh* (em torno de 100 a 120 Hz). O uso conjunto desses óculos com o *head-tracking* permite caracterizar um *setup* chamado “*FishTank*”. Este nome é inspirado na ideia de olhar um aquário através de um vidro, ou seja, o usuário pode mover a cabeça pra cima, para baixo e para os lados para ver o que tem dentro do aquário mas não pode realmente “entrar” nele. A vantagem principal nesses monitores é o seu baixo custo, entretanto há a desvantagem por não ser totalmente imersivo e por ter um reduzido FOR. A resolução e qualidade deste *display* são boas. Como é possível obstruir a superfície destes *displays* com objetos reais, problemas relativos ao *masking* podem ocorrer. Para os monitores do tipo *LCD* a estereoscopia ainda não está bem disponibilizada devido à baixa taxa de atualização das imagens (*refresh rate*) ainda presente nesta tecnologia.

Ao longo do trabalho o termo monitor será utilizado como sinônimo para o monitor convencional sem o uso de estereoscopia, enquanto *FishTank* significará o monitor acrescido de estereoscopia e *head-tracking*.

¹⁰<http://geektechnique.org/projectlab/851/making-3d-glasses-for-a-silicon-graphics>

Workbench Displays

2.5(a): *Perceptive Workbench*¹¹.2.5(b): *Responsive Workbench*¹².Figura 2.5: *Workbenches*.

As *Workbenches* são tecnologias baseadas em projeções e proporcionam um ambiente parcialmente imersivo de *VR*. Estereoscopia e *head-tracking* são normalmente utilizados em conjunto. As projeções podem ser posicionadas horizontalmente, verticalmente, uma composição das duas, ou em um ângulo específico para suprir a necessidade de uma aplicação particular (Figura 2.5). A qualidade visual é similar à obtida nos monitores mas com a vantagem de uma resolução e área física maior. Para tarefas de navegação envolvendo *3DIts* baseadas em locomoção física, elas não são apropriadas, uma vez que os usuários têm pouca mobilidade. Assim como os monitores, também têm um FOR pequeno e estão sujeitas ao uso de *masking* dada a possibilidade de obstrução da visualização por objetos reais.

Hemispherical Displays

Esta classe de *display* utiliza uma lente grande angular na frente de um projetor para permitir cobrir uma superfície na forma de uma semi-esfera. O usuário permanece sentado na frente desta semi-esfera interagindo com mouse, teclado e *3D pointers* (Figura 2.6). A área física de projeção é bem maior que um monitor convencional, proporcionando uma imersão parcial. Além disso, o brilho proporcionado por uma *front-projection* é alto. Alguns *displays* deste tipo também usam projeções do tipo *back-projection*. Entretanto, podem ocorrer problemas como a geração de sombras e quebra da ilusão de

¹¹http://mrl.nyu.edu/~perlin/courses/spring2006/shadowwalls/The_Perceptive_Workbench.htm

¹²<http://graphics.stanford.edu/projects/RWB/>



Figura 2.6: *Hemispherical Displays - VisionStation*¹³.

estereoscopia quando o usuário aproxima-se muito do *display*, devido ao uso da *front-projection*. Este problema de proximidade dificulta a execução de algumas *3DIt* envolvendo movimentação de braços e mãos, por exemplo, na tentativa de uma seleção mais direta de um objeto virtual próximo à superfície do *display*. Por causa da geometria ser curvilínea, às vezes, distorções radiais são percebidas.

Surround-Screen Displays - CAVE

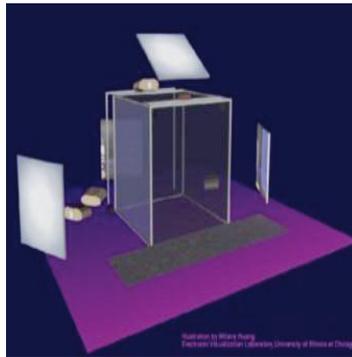


Figura 2.7: *Cave*¹⁴.

Consistem na ideia de envolver o usuário com superfícies grandes de projeção. Normalmente são caracterizados pela existência de três ou mais superfícies, cada uma contendo em geral *rear-projections* para eliminar problemas com sombras durante as interações. Em 1992, foi desenvolvido o primeiro *display* deste tipo que recebeu o nome *CAVE* e consistiu em 4 superfícies:

¹³<http://www.est-kl.com/projection/elumens/vs.html>

¹⁴<http://www.evl.uic.edu/pape/CAVE/>

uma frontal, duas laterais e uma no chão (Figura 2.7). Efeitos de estereoscopia, *head-tracking* e interação com dispositivos como *wands* são normalmente encontrados no *setup* de uma CAVE.

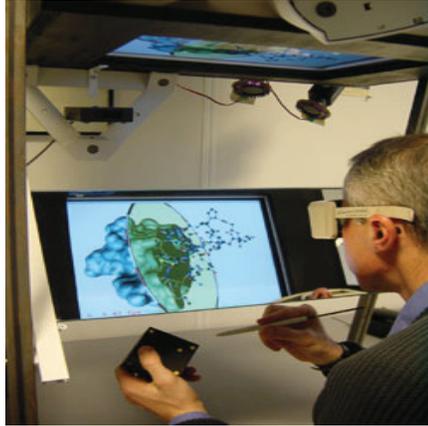
A vantagem principal da CAVE está na imersão causada pelo elevado estímulo visual proveniente das grandes superfícies de projeção. O arranjo físico das superfícies proporciona um elevado *FOV* que permite o uso da visão periférica. Um FOR de 270 graus normalmente é atribuído para estes setups mais comuns envolvendo 3 superfícies. Uma CAVE ideal com 6 lados (projeções frontal, traseira, as duas laterais, a de cima e a de baixo) apresentaria um FOR de 360 graus na vertical e lateral. Em ambos os casos, a CAVE não isola o usuário totalmente do mundo real (opacidade), pois mesmo envolvido totalmente ou parcialmente pelas superfícies, o usuário ainda consegue ver o próprio corpo. Conseqüentemente, isto leva à possibilidade do uso de *masking* durante a execução de algumas *3DIt*.

Entre as desvantagens estão: o custo elevado devido à quantidade de projetores e superfícies; necessidade de um espaço físico grande para ajustar os projetores (em boa parte das soluções são utilizados espelhos para reduzir o consumo de espaço). Como as imagens nas superfícies são geradas somente para um único ponto de vista, somente um usuário tem os movimentos da cabeça rastreado; quando há mais usuários dentro do ambiente, os demais enxergarão somente o ponto de vista do usuário rastreado.



Figura 2.8: *Panoramic display* (Kjeldskov01).

Há um tipo de *Surround-Screen Display* chamado *Panoramic display* que compreende telas curvas semelhantes a um semi-cilindro (Figura 2.8). Propiciam um *FOV* entre 160 e 180 graus. Basicamente tem algumas similaridades com as CAVES de 3 lados (Frontal e duas laterais) nos quesitos de imersão relacionados a *FOV* e *FOR*, assim como as suas desvantagens.



2.9(a): Setup utilizando espelho - *Personal Space Station*¹⁵.



2.9(b): Setup utilizando superfície semi-transparente.

Figura 2.9: *Reachin Displays* (MulderL02).

Reachin Displays

Partem da ideia de criar a ilusão de que o conteúdo do *VE* está ao alcance das mãos. Basicamente o usuário enxerga o reflexo de um *display* através de um espelho ou de alguma superfície semi-transparente, e interage diretamente com o conteúdo virtual posicionando as mãos atrás da superfície refletora (MulderL02)(Figura 2.9). Algum tipo de *Tracker* é frequentemente utilizado para tarefas de manipulação local. No caso da superfície refletora ser um espelho semi-transparente, este *display* assume a característica de não obstrução do mundo real, comum aos *displays* de *AR* do tipo *See-through*. Neste caso, as mãos e as ferramentas de interação são visualizadas em conjunto com os objetos virtuais, e o uso de *masking* é parcialmente resolvido, necessitando apenas o tratamento quando objetos reais estiverem na frente dos virtuais. No caso da superfície ser apenas um espelho, ocorre a oclusão total da visualização das mãos e ferramentas de interação, desta forma, o *masking* não é usado. Pode-se dizer que os *Reachin Displays* são uma mistura de características dos *HMD* e dos monitores convencionais.

2.3.3

Head-based Displays

Head-Mounted Displays

Os *HMDs* basicamente são compostos por um capacete contendo internamente dois *displays* pequenos do tipo *CRT* ou *LCD* com resoluções razoáveis,

¹⁵<http://homepages.cwi.nl/~robert1/pss/pss.html>



Figura 2.10: *HMD - Head-Mounted Display.*

e um sistema ajustável de lentes (Figura 2.10). Este par de *displays* internos permite o uso de estereoscopia, uma vez que garante uma imagem diferente para cada olho. Os dispositivos de interação utilizados com um *HMD* frequentemente são mouses 3D, *wands* e luvas. Tecnologias de rastreamento normalmente são acopladas aos *HMDs* para permitir *head-tracking*. O *FOR* de 360 é sem dúvida a grande vantagem de um *HMD* pois permite uma imersão completa do usuário. Para tarefas colaborativas imersivas, os *HMDs* apresentam a vantagem de permitir que vários usuários estejam em um mesmo ambiente virtual, sendo que cada um tem o seu ponto vista corretamente calculado e rastreado, o que não acontece em alguns sistemas de projeção onde o ponto de vista exibido é de apenas um usuário.

As desvantagens mais pertinentes nos *HMDs* são: *FOV* limitado (variando entre 30 e 60 graus normalmente) podendo causar distorções na percepção do tamanho e distância de objetos; fadiga causada pelo peso do equipamento durante períodos longos de uso (ergonomia); o isolamento visual do mundo real (Opacidade), sendo necessário ter o cuidado com o problema da colisão com objetos físicos durante tarefas envolvendo muita mobilidade física. Não existe o problema de *Masking* pois todo conteúdo visualizado é virtual. Neste caso, representações das ferramentas de interação são geradas virtualmente.

Todas as características acima mencionadas sobre os *HMDs* estão particularmente direcionadas para aplicações de *VR*. Os *HMDs* para este tipo de aplicação serão referenciados no texto como *HMDs-VR*. Porém os *HMDs* também são utilizados para aplicações *AR*, *HMDs-AR*, uma das formas é por meio do uso de câmeras. Este tipo *AR* é conhecida como *AR* do tipo *Video See-Through*, ou seja, a visualização do mundo real é intermediada por uma câmera. Alguns tipos de *HMD* já vêm de fábrica com câmeras acopladas (ver Figura 2.11(b)), entretanto é comum adaptar *HMDs* sem câmeras fixando-se uma ou duas câmeras na sua parte frontal.



2.11(a):
SAAB AddVisor 150¹⁶.



2.11(b): Trivision AR-vision HMD¹⁷.

Figura 2.11: HMDs para Realidade Aumentada.

Outra forma de *AR* usando *HMDs* se faz por meio de óptica (*Optical See-Through*) através de reflexões dos *displays* internos em superfícies transparentes (ver Figura 2.11(a)). Neste caso, o usuário enxerga o mundo real sobreposto por elementos virtuais projetados na superfície transparente.

2.3.4

Hand-based Displays

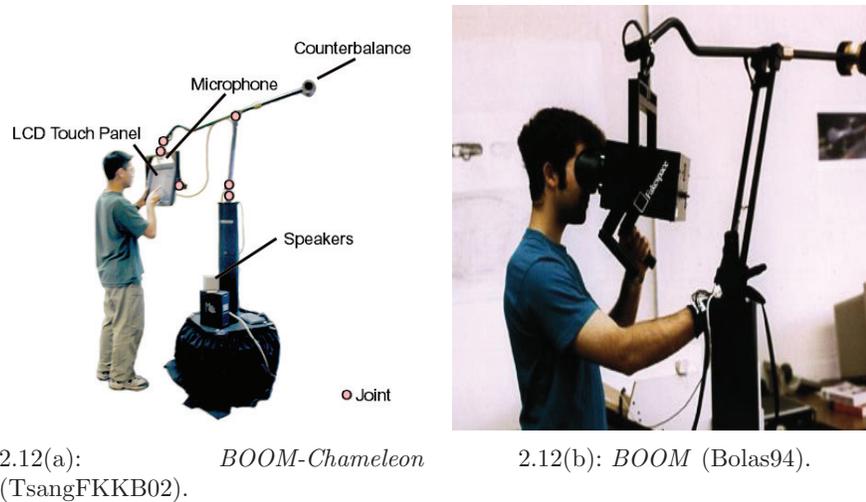
Arm-Mounted Displays

Consistem no uso de um *display* fixo a uma armação. Esta armação permite uma certa mobilidade em uma área ao redor dela, e contém um contra-peso para o usuário não precisar realmente “carregar” o *display*. Este contra-peso facilita a fixação de *displays* binoculares semelhantes aos *HMDs*, monitores com alta definição, entre outros portáteis. Normalmente, a própria armação é um *tracker* articulado e permite determinar a posição e orientação do *display* fixado.

O exemplo mais comum de um *Arm-Mounted Display* é o *BOOM - Binocular Omni-Orientation Monitor* (Bolas94) (Figura 2.12(b)). O *BOOM* tem qualidades visuais similares aos *HMDs* e os mesmos problemas com *FOV* mas tem uma vantagem, do ponto de vista ergonômico, porque o usuário não precisa vestir o capacete. Uma variação chamada *BOOM-Chameleon* (TsangFKKB02) (Figura 2.12(a)) foi desenvolvida utilizando um *display LCD*. Esta abordagem permitiu o uso da metáfora de uma janela para o mundo 3D. De um modo geral o *FOV* dos *Arm-Mounted Displays* fica dependente do tipo de *display* anexado, mas pode-se dizer que nos casos citados acima há uma variação entre 30 e 60 graus.

¹⁶www.est-kl.com/hardware/hmd/saab/addvisor150.html

¹⁷<http://www.inition.co.uk/inition/products.php>

Figura 2.12: *BOOMs*.

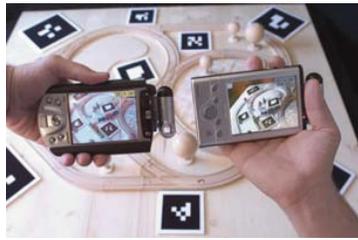
Algumas desvantagens são notadas para tarefas de navegação envolvendo locomoção física devido à restrição de movimento em volta da armação. Outra desvantagem acontece para as tarefas com manipulação bimanual porque o usuário precisa utilizar pelo menos uma das mãos para segurar o *display*. No caso do *BOOM-Chameleon*, por utilizar um *LCD*, a estereoscopia fica inviável.

Hand-Held Display

Estes compreendem dispositivos que são realmente carregados com as mãos. A grande parte são dispositivos pequenos (celulares, palmtops, tablets, notebooks, entre outros). Alguns possuem alguma câmera e possibilitam uma forma de *AR* do tipo *Video See-Through* (Figura 2.13(a)). Entretanto, aplicações de *VR* também são utilizadas de forma similar ao *BOOM-Chameleon* (TsangFKKB02), ou seja, usando a metáfora de uma janela 3D para o *VE*. Em ambos os casos, os pontos de vista das câmeras virtuais são gerados seguindo a orientação e posição dos dispositivos, os quais frequentemente estão nas mãos dos usuários.

Porém, existem também aqueles que possuem armações que permitem o posicionamento do display junto aos olhos do usuário (Figura 2.13(b)). Estes lembram os *HMD* porém não são vestidos e sim carregados.

De forma geral, os *Hand-Held Displays* compartilham as vantagens dos *Arm-Mounted Displays* com o acréscimo de mais mobilidade física, o que propicia uso em tarefas de locomoção. Entretanto, também têm a desvantagem de deixar ocupada pelo menos uma das mãos, o que impossibilita interações bimanuais. Questões referentes a fadiga podem surgir dependendo do peso e



2.13(a): *PalmTop Video See-Through* (WagnerS06).



2.13(b): *HandHeld* semelhante a um HMD (BillinghurstKP01).

Figura 2.13: *HandHelds Displays*.

tamanho do display.

2.4 Tarefas e Técnicas de Interação 3D

Neste trabalho, a classificação para as *3DT* encontrada em (BowmanKLP05) será adotada, e compreende:

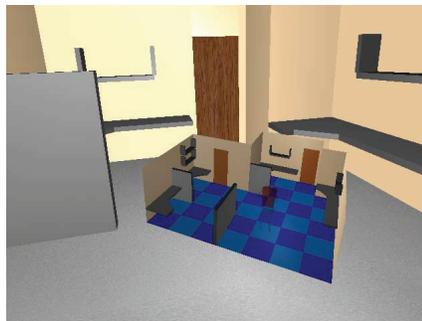


Figura 2.14: *WIM - World-in-Miniature* (PauschBBW95).

Navegação (*Travel*): São aquelas que permitem o deslocamento de um ponto do *VE* para outro. A navegação não deixa de ser uma forma de manipulação, pois consiste em manipular uma câmera virtual ou o ponto de vista do usuário dentro do *VE*. Um exemplo claro disso está na *3DIt World-in-Miniature (WIM)* (PauschBBW95) (Figura 2.14). Na *WIM* o usuário movimenta-se através de um *avatar* dentro de uma representação reduzida do *VE*. Este *avatar* é uma representação do próprio usuário dentro do *VE*. Outras *3DIts* consistem na indicação de um objeto no *VE*, geralmente através de um apontamento. Em seguida, a *3DIt* fica responsável por posicionar a câmera virtual a uma certa distância do objeto apontado. Por exemplo, na *3DIt*

Zoomback (ZelevnikLAK02) esta distância são dois pés (*two feet*), enquanto na *3DIt LaserGrab* (ZelevnikLAK02) esta distância corresponde à distância relativa entre a cabeça e a mão do usuário. Além disso, algumas *3DIt* assumem a manipulação da câmera virtual baseando-se em metáforas, por exemplo, a “*eyeball-in-hand*”, “*flying*” e “*scene-in-hand*” (WareJ88, WareO90).

Wayfinding: Relacionado com a tarefa de navegação está um processo cognitivo responsável por criar uma representação mental do ambiente virtual, também conhecido como mapa mental. Esta representação é criada através da exploração do *VE* e observação dos elementos contidos nele. O *wayfinding* compreende este aspecto cognitivo da tarefa da navegação, e existem meios para estimular este processo através da inclusão de ferramentas informativas do tipo: localização de pontos chaves em um cenários, geração de caminhos e distâncias entre pontos do cenário.

Seleção: Permite escolher um ou mais objetos num *VE*. Em (ShermanC02) são apontadas duas categorias de seleção, aquelas baseadas na escolha de uma direção e as baseadas na escolha de itens. Os autores mencionam também uma terceira categoria via entrada de dados alfanuméricos. Sobre as técnicas baseadas em direção, as *3DIts* mais comum são as baseadas em raios (*Ray-based techniques*), e consistem na seleção de um objeto pela intersecção do mesmo com um raio vindo (“metaforicamente”) do dedo ou mão do usuário. Esta forma de seleção pode ser considerada como uma analogia à escolha de um arquivo usando o ponteiro do *mouse*. A *raycasting* (Mine95) é um exemplo clássico desse tipo de *3DIt*. Variações dela são encontradas nas *3DIts*: *flashlight* (LiangG94) e *aperture* (ForsbergHZ96). Ambas usam a representação de um volume de um cone ao invés de um raio, e permitem escolher múltiplos objetos. A abertura deste cone é modificada interativamente para controlar a escolha dos objetos. O direcionamento desse raio pode ser resultante da direção somente de uma *Wand*, de duas *Wands* formando um vetor 3D, ou do *head-tracking* e uma *Wand*, entre outras. Nas técnicas baseadas na seleção de item, a maioria são consideradas como de seleção direta, por exemplo, a *Contact-Select* onde literalmente o usuário precisar “tocar” no objeto virtual para selecioná-lo. Isto pode ser feito utilizando uma luva rastreada, por exemplo. A escolha de item também pode ser feita, utilizando a *Contact-Select*, através de menus virtuais escolhendo itens em uma lista. As *3DIts* de seleção estão muito relacionadas com as *3DIts* de manipulação, pois são o pré-requisito delas. Desta forma, em seguida serão mencionadas implicitamente outras técnicas de seleção.

Manipulação: Modificam as propriedades, frequentemente a posição e orientação, de um objeto dentro do *VE*. Algumas *3DIts* foram naturalmente

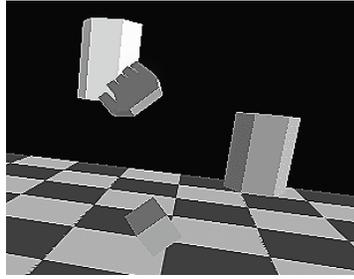


Figura 2.15: *Go-go technique* (PoupyrevBWI96).

inspiradas na intuitiva manipulação usando as mãos, ou seja, a manipulação direta ao alcance das mãos. Em (ShermanC02) estas técnicas estão classificadas como manipulações do tipo *Direct user control*. Estas *3DIt* também são denominadas como *reaching techniques* e a versão mais simplificada é a utilização de uma mão virtual (*virtual hand*). Neste caso, há o direto mapeamento do movimento da mão do usuário na mão virtual dentro do *VE*. O problema nesta *3DIt* é a impossibilidade de interação com objetos que estão fora do alcance dos braços. Esta limitação foi contornada na técnica *Go-go* (PoupyrevBWI96). Nela, o alcance do braço virtual é interativamente modificado quando a mão real do usuário ultrapassa uma certa distância limiar, logo, o acesso, a seleção e a manipulação dos objetos distantes ficam permitidos (Figura 2.15). Variações desta técnica também foram desenvolvidas, como a *fast Go-go* e a *stretch Go-go*. Outras técnicas são baseadas em raios virtuais e dão continuidade às técnicas de seleção desta natureza. Por exemplo, após o uso da técnica de seleção *ray-casting*, o objeto escolhido é fixado no raio e acompanha os movimentos dele. Alguns testes nesta abordagem foram relatados em (BowmanH97) e foram identificadas algumas desvantagens como dificuldade na realização de rotações e movimentos de aproximação e recuo; apesar das desvantagens foi indicada como uma boa vantagem na realização de arrastos (*grabbing*). Bowman e Hodges desenvolveram uma técnica híbrida chamada *HOMER - Hand-centered Object Manipulation Extending Ray-casting* (BowmanH97) juntando as características das técnicas baseadas em raios e as baseadas na mão virtual. Basicamente, a técnica consistia na seleção por *ray-casting* e em seguida uma mão virtual era posicionada no objeto escolhido para iniciar a manipulação. Em (ShermanC02) são relatadas outras categorias de manipulação: *Physical Control*, *Virtual Control* e *Agent Control*. A primeira refere-se à utilização de dispositivos físicos e à obtenção de *feedbacks* deles, por exemplo, feedback tátil ao pressionar botões e quando um controle vibra. As

outras duas categorias referem-se a formas de manipulação indireta utilizando controles virtuais (por exemplo, menus e sliders virtuais) e controle de agentes na forma de avatares. Em (Bolt80) foi relatada uma forma de interatividade *multi-modal* para manipulação de objetos. Através de gestos e comandos de voz, uma posição alvo era especificada para posicionar um determinado objeto. O nome desta técnica era “*put-that-there*”.

Controle de Sistema: Compreende as técnicas responsáveis por modificar o estado do sistema que gerencia o *VE*. Nas interfaces 2D do tipo *WIMP*, *toggles*, *radio buttons*, *checkboxes*, *menus* do tipo *pull-down* e *pop-up* são exemplos de técnicas de controle de sistema. Para *VE* foram criados *menus* flutuantes inspirados em interfaces 2D (TeylingenRM97), *menus* do tipo *TULIP - Three-Up*, *Labels in Palm* (BowmanWCL01), o *VEWL - Virtual Environment Windowing Library* (LarimerB03), *rings menus* (LiangG94), *spin menus* (GerberB05), entre outros. Além disso, comandos por voz e baseados em gestos já foram também utilizados. Outros esforços no desenvolvimento de *menus* também foram feitos para aplicações de *VR* voltadas para *desktop* e *AR*, mais detalhes sobre essas variações, assim como propostas sobre classificações estão em (RaimundH07).



2.16(a): *KITTY Glove*

2.16(b): Combinação de posições dos dedos na mão direita para letra K.

2.16(c): Combinação de posições dos dedos na mão esquerda para letra D.

Figura 2.16: Digitação sem o uso de teclados - *KITTY Project* (KuesterMPC05).

Symbolic Input: Esta é a tarefa em que o usuário comunica informações simbólicas (textos, números e outros símbolos) dentro do *VE*. Tarefas deste tipo são comumente encontradas nas *GUIs* através da utilização de aplicações para edição de texto, email, planilhas, entre outras. No contexto dos *VEs* esta tarefa ainda é muito pouco explorada. Um dispositivo chamado *KITTY* (ver Figura 2.16(a)) (KuesterMPC05), similar a uma luva, foi desenvolvido para digitação de texto independente do uso de teclados. A ideia básica é a associação de letras a uma combinação de posição dos dedos (ver Figura 2.16(b) 2.16(c)).

A maior parte dos trabalhos na literatura compartilham similaridades com essa classificação. Em alguns casos são usados nomes diferentes, como em (Kjeldskov01) que classificou as tarefas em: *Orienting*, *Moving* e *Acting*. A primeira referia-se a capacidade do usuário de olhar ao seu redor, a segunda é similar a tarefa de Navegação e a terceira consiste nas tarefas de Manipulação e Seleção juntas. Um aspecto adicional, relacionado com uma certa “sociabilidade virtual”, foi incluído na classificação de (ShermanC02) que além da Navegação e Manipulação inclui a Comunicação. Esta modalidade compreendia a comunicação com outros usuários e agentes contidos no VE. Para Smith et. al (SmithD99) as técnicas foram agrupadas em: *Navigation*, *Selection*, *Manipulation* e *Environment Commands*. Esta última corresponde à categoria “Controle de Sistema” descrita anteriormente, enquanto as demais são idênticas em nome e definição.

2.5

Interfaces Híbridas de Transição e Continuum da Realidade Mista

Em relação à localização das interfaces dentro dos espectros de interfaces propostos por Milgran e posteriormente por Molina (ver Seção 1.1.2), pode-se dizer que existem dois tipos: as estáticas e as dinâmicas. As interfaces com a característica estática são aquelas que estão estacionadas em algum ponto ou região dentro do espectro, por exemplo, a VR, a AR, e as WIMPs. Porém as com características dinâmicas têm a capacidade de “caminhar” ao longo deste espectro e são consideradas mais como um veículo de locomoção. Estas interfaces estão começando a ser exploradas e vêm sendo nomeadas como interfaces de transição (*transitional interfaces*). Ainda são pouco exploradas e o trabalho principal neste contexto é o projeto MagicBook (BillinghurstKP01).

O *MagicBook* apresentou o conceito das interfaces de transição vinculando-as ao *Continuum* da Realidade Mista definido por Milgram (MilgramK94). Na aplicação exemplo havia uma interface que caminhava da realidade comum (simplesmente o livro com as informações textuais e figuras bidimensionais), passava pela realidade aumentada (visualização do livro acrescidos de cenas 3D por meio de um hand-held display) e por fim parava na realidade virtual (imersão nas cenas 3D do livro). Esta transição usou como dispositivos um *hand-held display* e eventos de botões para interações simples de navegação.

Pode-se notar que uma vantagem clara nas interfaces de transição é a interligação de interfaces contidas no continuum da realidade mista com a finalidade de realizar uma tarefa, no caso do *MagicBook*, o entendimento das estórias de um livro. Para esta tarefa foi permitido experimentar tanto a

realidade aumentada e a virtual utilizando algumas formas de interação através de uma quantidade pequena de dispositivos. Entretanto, poderia-se pensar em outras tarefas utilizando mais recursos de hardware e interligando interfaces do continuum da MR, tais interfaces poderiam ser consideradas por um lado interfaces híbridas e por outro lado interfaces de transição.

2.6 Continuidade

O estudo e desenvolvimento das *HUI* está dando início ao surgimento de ambientes de trabalho cada vez mais complexos, contendo uma diversidade de dispositivos de entrada e *displays*. A composição destes dispositivos pode resultar em vários modos de operação caracterizando espaços de interação distintos. Todos estes fatores estão de certa forma de acordo com o que Ishii et al. (IshiiKA94) relatou sobre o desenvolvimento de sistemas de colaboração em tempo real. Duas metas importantes foram apontadas: continuidade com práticas de trabalho e transições suaves entre espaços funcionais. A primeira meta está relacionada ao aproveitamento das habilidades e práticas de trabalho que as pessoas estão acostumadas a fazer, ou seja, sistemas que forcem o abandono de habilidades adquiridas (conhecimento prévio) pelo aprendizado de novos métodos estão sujeitos a encontrar grande resistência. A segunda propriedade consiste no desenvolvimento de transições suaves entre modos de operação (espaços de interação), uma vez que, para a realização de algumas tarefas pode haver a necessidade da execução de outras “subtarefas” em ambientes distintos. Alcançar transições contínuas e suaves entre esses modos sem dúvida é uma meta importante, pois tenta evitar frustrações durante a realização de tarefas complexas.

Congruência Dimensional

O conceito da Congruência Dimensional apontado por (DarkenD05) está ligado à problemática da falta de um mapeamento direto entre as ações dos usuários e os efeitos dessas ações no ambiente virtual. Os autores definiram formalmente como:

...a condição aonde a demanda espacial de uma tarefa está diretamente de acordo com a técnica de interação que é usada para executá-la.

Este fato é bem percebido quando em um ambiente interativo o usuário é forçado a fazer tarefas de natureza 2D ou entrada de dados simbólica usando técnicas de natureza espacial 3D ou vice-versa. Um exemplo bem ilustrativo seria forçar um usuário a utilizar um editor de texto usando luvas através de um teclado virtual 3D flutuando dentro de um ambiente imersivo. A tarefa de

editar texto está presente no dia-a-dia de várias pessoas e está frequentemente associada ao uso de um teclado físico apoiado em uma mesa. Este teclado físico provê *feedback* tátil, ao contrário do teclado virtual em que a luva não iria tocar em nada. A edição de texto é uma tarefa de entrada de dados simbólica associada a um dispositivo de entrada de dados com um layout bidimensional.