Relatório de estudo

MAC5701 - Tópicos em Ciência da Computação

Thiago Schumacher Barcelos
Orientador: Carlos Hitoshi Morimoto

Sumário

1 Introdução

Este trabalho tem como objetivo consolidar uma série de estudos e levantamentos bibliográficos feitos no âmbito da disciplina MAC5701 - *Tópicos em Ciência da Computação.* O objetivo deste estudo, na área de Interação Humano-Computador, é obter uma consolidação do estado-da-arte na aplicação do rastreamento de olhar a interfaces.

Um conjunto de diferentes técnicas desenvolvidas ao longo dos últimos 20 anos podem ser classsificadas sob a denominação de técnicas de rastreamento de olhar. De maneira geral, podemos dizer que o objetivo de tais técnicas, apesar de suas diferenças de abordagem, é a estimação do ponto observado por um indivíduo em uma superfície 2D, ou a estimação da linha de visão do indivíduo, quando não é considerada a observação de uma superfície em particular.

Inicialmente, o uso de técnicas de rastremento de olhar foi razoavelmente empregado no campo da psicologia experimental, especialmente como ferramenta para a compreensão da busca visual como componente de processos cognitivos. Na última década, o contínuo aumento da capacidade de processamento e diminuição de custo dos microprocessadores, aliado à ampla disponibilidade de dispositivos que permitem a captura de dados para o processo de rastreamento de forma não intrusiva (p.ex. câmeras de vídeo), permitiu o desenvolvimento de técnicas de rastreamento de olhar robustas e que operam em tempo real.

Essas técnicas levantam a possibilidade do uso dos dados gerados pelo proceso de rastreamento de olhar como uma entrada no processo de interação humano-computador. Neste trabalho vamos expor como essa possibilidade foi explorada em trabalhos de pesquisa na área, e suas vantagens e desvantagens.

1.1 Estrutura deste documento

Na seção *Contexto: visão, rastreamento e aplicações* é feita uma introdução à estrutura do processo fisiológico da visão humana, e sua importância no processo de interação com dispositivos computacionais; então, é feita uma descrição das características das técnicas de rastreamento de olhar desenvolvidas até a presente data.

A seção *Técnicas de interação* explora em maior profundidade o desenvolvimento de técnicas de interação baseadas e/ou complementadas por técnicas de rastreamento de olhar. Características do processo de rastreamento que influenciam decisões de projeto dessas técnicas de interação serão comentadas; também descreveremos e comentaremos alguns sistemas já implementados, bem como características da metodologia de testes que tem sido adotada.

2 Contexto: visão, rastreamento e aplicações

2.1 Fisiologia do olho humano

A visão é um dos sentidos mais importantes na percepção do ambiente à nossa volta; a todo momento, a usamos para obter informações, aprender e guiar a maior parte dos nossos movimentos, dentre outras aplicações.

A retina do olho é a região que contem células fotossensíveis, as principais responsáveis pelo processo de detecção da luz. Dessas células, a proporção de 94% é dos chamados *bastonetes*, estruturas com alta sensibilidade à luz, permitindo a visão mesmo em ambientes com pouca luminosidade, mas pouca sensibilidade à cor; o restante compôe-se dos *cones*, pouco sensíveis à luz mas muito sensíveis às cores (Glenstrup e Engell-Nielsen [6], citando Bruce e Green [3]).

Os cones estão concentrados no centro da retina – de outra forma, a distância entre eles seria muito grande, prejudicando a acuidade visual em condições normais de iluminação do dia. Nessa área na retina há uma depressão, chamada *fóvea*, constituída completamente por cones. Essa área é responsável pela visão detalhada, e cobre um ângulo visual de aproximadamente 1 grau.

A cada momento, o objeto que recebe atenção visual deve ser focado pela fóvea; este evento é chamado de *fixação*, e dura entre 200 e 600 milissegundos [8]; na maior parte das vezes o olho muda de posição entre uma fixação e outra através de um *movimento sacádico*. Este é um movimento voluntário, que dura entre 30 e 120 milissegundos; porém uma vez iniciado, não é possível mudar sua direção ou a posição final [6]. Mesmo durante as fixações, o globo ocular não permanece estático o tempo todo. Pequenos movimentos de ajuste (movimentos microssacádicos) acontecem com função, aparentemente, de manter o objeto focado na fóvea. Em referências (p. ex. [3]) encontra-se a descrição de outros movimentos dos olhos, que acontecem para o ajuste do foco visual a objetos em movimento, ou a largos movimentos da cabeça. Tais movimentos trazem pouca ou nenhuma informação quando se considera a forma com que se opera os dispositivos computacionais amplamente disponíveis atualmente e em geral não são levados em consideração pelas técnicas de rastreamento de olhar existentes.

2.2 Técnicas de rastreamento de olhar

Existe uma grande variedade de técnicas de rastreamento de olhar, que podemos agrupar em duas categorias:

intrusivas: requerem o contato físico de algum tipo de instrumento de medição com o indivíduo. Um exemplo é a técnica baseada no potencial elétrico da pele ao redor dos olhos, que exige a colocação de eletrodos ao redor dos olhos para medição. Existe também a possibilidade do uso de uma lente de contato especial que induz campos eletromagnéticos de alta freqüência gerados ao redor do indivíduo, permitindo assim o rastreamento. Claramente, o desconforto causado por essas técnicas tornam sua aplicação no controle de interfaces praticamente impeditivo ([8], [6]), restringindo sua aplicação à execução de medições experimentais.

não-intrusivas: se utilizam da obtenção de imagens de vídeo do olho, da cabeça ou de ambos e posterior aplicação de algoritmos de visão computacional para estimar a posição observada pelo usuário. A maioria das técnicas existentes são baseadas na identificação da pupila que é menos sujeita a variabilidades individuais do que, por exemplo, a identificação da fronteira entre a esclerótica e a íris. Alguns sistemas ([5], [9]) utilizam propriedades de reflexão de luz infra-vermelha pela pupila e pela córnea, obtendo assim maior robustez na detecção.

Uma limitação das técnicas não intrusivas de rastreamento desenvolvidas é a necessidade da excução de algum procedimento de calibragem. O objetivo destes procedimentos é mapear a posição do componente do olho (p. ex. a pupila) detectado na imagem de vídeo em uma posição observada do ambiente. Os dados gerados pelo processo de rastreamento serão então ajustados [1] a partir dos dados obtidos na calibragem para o cálculo da estimativa da linha de visão ou ponto observado pelo indivíduo. Além da necessidade de cada usuário do sistema passar por esse processo, outra limitação é que a cabeça do indivíduo deve permanecer estática, em geral com o uso de um apoio para o queixo. Porém, realizando conjuntamente o rastreamento da cabeça, é possível reorientar a captura de imagens do olho e a calibragem do sistema, permitindo assim algum grau de mobilidade e um uso mais confortável do sistema.

3 Técnicas de Interação

Uma *técnica de interação* é uma forma de usar um dispositivo físico de entrada para executar uma tarefa em um diálogo entre um ser humano e um dispositivo computacional ([8], citando [1]). Após um período onde se surgiram técnicas de rastreamento de olhar razoavelmente robustas, não-intrusivas e capazes de fornecer dados em tempo real, surgiram as questões:

[1] por exemplo, por interpolação numérica [5], ou pela utilização de modelos geométricos do globo ocular associados a propriedades de refração da luz [9]

o rastreamento de olhar *pode* ser uma entrada no processo de interação? E se pode, *como* mapeá-lo à execução de tarefas? Vamos nesta seção comentar as aplicações já existentes do rastreamento de olhar a interfaces bem como os problemas e resultados identificados.

3.1 Decisões de projeto

O projeto de técnicas de interação que utilizam sistemas de rastreamento de olhar deve levar em consideração uma série de fatores inerentes tanto à estrutura fisiológica do processo de visão humana quanto à tecnologia empregada no rastreamento de olhar. Nesta seção vamos exibir os fatores que mais influenciaram o projeto das técnicas de interação existentes atualmente. Note que a partir deste momento consideramos que o objetivo do processo de rastreamento de olhar é obter as coordenadas (x, y) de um ponto sobre um plano observado pelo usuário no processo de interação, tais como um monitor, peça presente na grande maioria dos sistemas de mesa atuais, ou mesmo um *videowall*.

3.1.1 O olho

O processo de rastreamento de olhar encontra na estrutura do olho humano um fator fundamental de limitação de sua precisão. Conforme citado na seção 2.1, a fóvea é responsável pelo foco de nossa visão, e tem um área correspondente a 1 grau do ângulo visual. Não é possível saber qual o ponto exato focado pelo olho dentro desse intervalo de 1 grau, que corresponde aproximadamente ao tamanho do polegar de um indivíduo visto por ele com seu braço estendido. Além disso, os movimentos microssacádicos realizados inconscientemente durante as fixações são um fator de introdução de ruído nos dados gerados pelo rastreador.

3.1.2 Identificação de fixações

A entrada proveniente em tempo real de um rastreador de olhar traz grandes quantidades de dados que devem passar por um processo de filtragem para obter informações semanticamente relevantes sobre o processo visual. Em particular, os pontos (x, y) sobre o plano provenientes do rastreador devem ser agrupados em unidades que identifiquem as fixações e os movimentos sacádicos. Este é o primeiro passo para obter informação sobre o ponto de interesse do usuário.

A estratégia mais simples para implementar essa filtragem é o uso de um limiar espacial e outro temporal – somente ocorrências de pontos cuja distância entre si seja menor que um dado valor, e dentro de um intervalo mínmo de tempo são agrupados e considerados como componentes de uma mesma fixação ([8], [14] e [16]). Estes valores são definidos por uma combinação do prévio conhecimento da duração temporal média dos eventos do processo visual com testes empíricos. Uma estratégia mais sofisticada é apresentada por Salvucci e Goldberg em [10]; essa estratégia é baseada no uso de uma cadeia de Markov oculta com dois estados: *fixação* e *movimento sacádico*. Essa modelagem permite a reestimação dos parâmetros correspondentes aos limiares temporal e espacial. Em tese, essa reestimação permite o ajuste fino dos parâmetros, de forma a obter uma melhor detecção das fixações. Uma desvantagem seria a complexidade da implementação desse processo [13].

3.1.3 Seleção e o problema do "Toque de Midas"

A idéia mais divulgada de aplicação do rastreamento de olhar às interfaces já existentes é usar o olhar para executar tarefas de apontamento. Essa idéia está baseada em algumas evidências. Glenstrup e Engell-Nielsen [6] citam numerosos estudos que mostram evidências que há na maior parte das vezes relação entre o interesse dos indivíduos e o que eles estão olhando. Outra fonte de evidência é o modelo construído por Card, Moran e Newell [4], que prevê o tempo T de execução de tarefas de apontamento com dispositivos tradicionais (p. ex. o *mouse)*; a equação seguinte mostra uma versão simplificada do modelo:

$$T = T_M + T_V + T_P + T_R \tag{1}$$

Nele, a subtarefa motora (isto é, efetivamente mover o dispositivo para a posição correta), que leva tempo T_P, é precedida por uma tarefa cognitiva que leva tempo T_M e por uma tarefa visual (isto é, a busca visual do alvo em questão) que leva tempo T_V. T_R é o tempo de resposta do sistema. Hoje, apesar das evidências de que esse modelo não é preciso (p. ex. [7]), ele ainda traz uma estimativa razoável e justifica os esforços no sentido de aplicar o rastreamento de olhar para a execução de tarefas de apontamento.

Porém, o mapeamento direto do olhar (mais especificamente, das fixações) a um comando de seleção do sistema cria o problema identificado por Jacob [8] e apelidado de *"Problema do Toque de Midas"* – uma seleção pode ser ativada em qualquer lugar da tela observado pelo usuário, tenha ele a intenção de fazê-la ou não. Após a filtragem dos dados do rastreador de olhar, o desafio no projeto das técnicas de interação é evitar o problema do Toque de Midas, implementando mecanismos para que o usuário indique quando deseja executar um comando de seleção. A primeira abordagem a esse problema foi a implementação de um tempo de latência – a seleção de um objeto é efetuada somente após um intervalo de tempo. Alternativamente, a seleção pode ser ativada manualmente pelo pressionamento de um botão pelo usuário. Uma abordagem alternativa que incorpora o rastreamento de olhar ao apontamento tradicional foi desenvolvida por Zhai, Morimoto e Ihde [16]; denominada MAGIC Pointing, será descrita em maiores detalhes mais adiante.

3.1.4 Identificação da intenção do usuário

O estudo de registros gerados por rastreadores de olhar tem sido importante por vários anos para o campo da psicologia cognitiva, pois eles fornecem pistas importantes sobre a forma com que indivíduos buscam informação visual para resolver problemas, quano tempo gastam para analisar essa informação e quando precisam voltar à informação previamente acessada [11]. Recentemente, sistemas computacionais de rastreamento de olhar de forma não-intrusiva tem sido empregados no desenvolvimento de algoritmos de *interpretação* [2] de movimentos do olhar. A tarefa dos algoritmos de interpretação é mapear uma seqüência de ações definidas num certo protocolo [3]. Em outras palavras, um algoritmo de rastreamento toma com entrada uma seqüência de ações do usuário e devolve uma estimativa, baseada em um modelo previamente construído, de qual é a intenção do usuário ao tomar essas ações.

O projeto de técnicas de interação baseadas em rastreamento de olhar pode se beneficiar do uso de algoritmos de interpretação ([10], [12]). Uma vez que seja possível definir um modelo cognitivo da interface, as estimativas devolvidas por um algoritmo de interpretação servem como parâmetros de ajuste para as operações a serem executadas. No caso da aplicação do rastreamento de olhar a tarefas de apontamento, um algoritmo de interpretação pode ajudar a definir qual objeto o usuário tem a real intenção de selecionar, compensando as imprecisões inerentes ao olho humano (veja seção 3.1.1) e ruído gerado pelo processo de rastreamento.

[2]Nossa tradução do inglês *tracing algorithms*, usada aqui para evitar a confusão com a expressão *rastreamento* de olhar (*eye tracking*), usada para se referir ao processo de obter o ponto observado pelo usuário ou sua linha de visão

[3]A idéia de protocolo implica a filtragem semântica dos dados do dispositivo de entrada rastreador das ações; no caso do rastreamento de olhar aplicado a interfaces teríamos como componentes desse protocolo as fixações e os movimentos sacádicos

3.2 Técnicas de interação implementadas

Vamos agora exibir algumas técnicas de interação descritas na literatura. O objetivo dessa seção não é enumerar exaustivamente as aplicações do rastreamento de olhar a interfaces, mas ilustrar através de alguns dos trabalhos mais relevantes na área os conceitos de projeto expostos na seção anterior.

3.2.1 *Mundo das janelas*

Uma das aplicações pioneiras do rastreamento de olhar é o *mundo das janelas* descrito por Richard Bolt ([2], citado em [8]). O mundo das janelas constitui de um *videowall* que exibe diferentes programas em vídeo e, de acordo com a atenção visual recebida, as janelas correspondentes aos programas aumentam ou diminuem; o som do programa que recebe maior atenção visual também tem seu volume aumentado. É uma das primeiras tentativas de inferir a intenção do usuário através do olhar em um sistema interativo que se tem notícia.

3.2.2 *Pequeno príncipe*

Glenstrup e Engell-Nielsen [6] descrevem em seu trabalho um sistema desenvolvido por Starker e Bolt (1990) que conta, em uma versão animada, a história "O Pequeno Príncipe", de Antoine de Saint Exupery. A história é contada em um ambiente (o pequeno planeta habitado pelo príncipe) que contém vários objetos, tais como vulcões, escadas e flores.

O espectador é monitorado por um rastreador de olhar baseado no princípio de reflexão da pupila e da córnea; quando é detectada a atenção do espectador sobre algum objeto do ambiente, o sistema exibe maiores informações sobre o objeto. Essa detecção é feita através da contagem do número de fixações feitas sobre o objeto associada com o tempo de duração das fixações. É feita também uma tentativa rudimentar de infência do contexto: os objetos são agrupados em categorias, e quando se detecta que a atenção está distribuída de forma aproximadamente igual por vários objetos de uma mesma categoria, o sistema exibe informações sobre a categoria de objetos.

3.2.3 Seleção por tempo de latência

Em [8], Jacob descreve os protótipos de algumas técnicas de interação usando rastreadores de olhar. Seu trabalho levanta a questão do "Toque de Midas" (veja seção 3.1.3) e implementa duas alternativas para a seleção: o usuário pode pressionar um botão indicando a intenção de selecionar o objeto observado ou pode aguardar um tempo de latência razoavelmente curto (definido entre 150 e 250 milissegundos). Ambas as alternativas são oferecidas em paralelo, e é reportado que em experiências preliminares o tempo de latência se mostrou mais conveniente. (Uma análise mais formal do desempenho desse protótipo é feita em [14], que comentaremos mais adiante).

O protótipo baseia-se em uma tela onde são apresentados vários navios do lado direito, e uma área para exibição de texto do lado esquerdo (Figura 3.2.3). Quando um navio é selecionado da maneira descrita acima, informações sobre o mesmo são exibidas. Foi implementada também a operação de "arrastar-e-soltar" navios: uma vez selecionado um navio, o usuário pressiona um botão e olha para o lugar onde deseja posicionar o navio, soltando então o botão. A seleção usando o olhar também foi empregada no protótipo para a seleção de menus do tipo "pull-down"; nesse caso a estratégia do tempo de latência curto não se mostrou eficiente – o autor especula que isso acontece porque o tempo necessário para a leitura de itens do menu que não são familiares ao usuário é falsamente interpretado como o tempo de latência para a seleção. O aumento do tempo de latência também parece não ser razoável, pois gera a necessidade de manter a atenção visual sobre um item de menu durante um intervalo de tempo pouco natural.

O trabalho de Jacob, como uma das primeiras abordagens à implementação de tarefas de seleção usando o olhar, mostra que a estratégia do tempo de latência não é bem sucedida em todos os casos de uso que surgem em um protótipo mais próximo do funcionamento de interfaces reais.

3.2.4 MAGIC Pointing

Zhai, Morimoto e Ihde descrevem em [16] uma nova abordagem para o apontamento, denominada MAGIC (*Manual and Gaze Input Cascaded*) Pointing. Levando em consideração as limitações na precisão causadas pela estrutura do olho, o processo de rastreamento e a calibragem do sistema, propõem uma técnica de interação que combina o rastreamento de olhar com um dispositivo de apontamento convencional, como um *mouse* ou um *trackball*. A posição da tela indicada pelo rastreamento de olhar é considerada uma estimativa da posição observada pelo usuário, sujeita a erro em algum grau. Nesse momento o usuário ajusta a posição do cursor através do controle motor do dispositivo tradicional, obtendo maior precisão.

O cursor acompanha o olhar quando a atenção visual se volta para um alvo distante da posição atual do cursor. Duas estratégias diferentes de reposicionamento são implementadas: na primeira (chamada *estratégia conservadora)* o reposicionamento é ativado a partir de um movimento motor iniciado pelo usuário no dispositivo tradicional; na segunda (chamada *estratégia liberal*) o

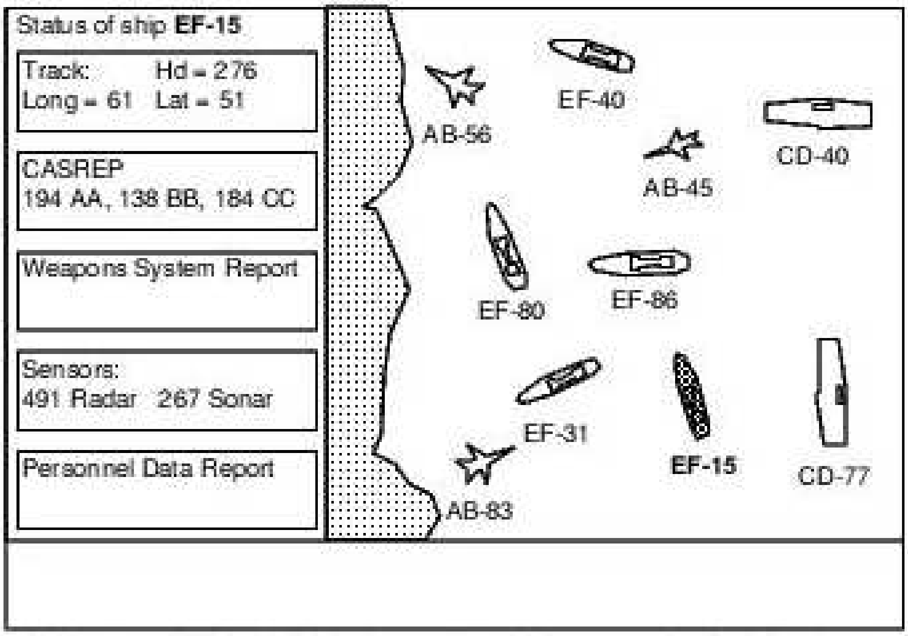

Figura 1: Sistema de seleção de barcos através do rastreamento de olhar usando tempo de latência (conforme exibido em [14])

movimento motor não é necessário.

O MAGIC Pointing avança no sentido de encarar os dados provenientes do rastreamento de olhar como um canal de entrada mais passivo, numa tentativa de diminuir o esforço cognitivo de explicitamente atuar sobre o ambiente com o olhar – a atuação continua sendo, em sua maior parte, motora.

Posteriormente, Wang, Zhai e Su [15] desenvolveram uma aplicação do MAGIC Pointing a um domínio específico: a digitação de ideogramas do idioma chinês através do *pinyin*, uma representação fonética usando caracteres do alfabeto ocidental. O problema surge com a grande quantidade de ideogramas que correspondem a uma mesma representação fonética, exigindo a eliminação de ambiguidades de alguma forma. Nesse sistema isso é feito através da botões que exibem os caracteres correspondentes à representação fonética que acabou de ser digitada pelo usuário. A seleção dos botões é feita usando o MAGIC Pointing, e é empregada uma estratégia de *design* para atenuar o efeito da precisão limitada do rastreamento de olhar: os botões não são dispostos em filas ou colunas, mas em um formato de 'W' (Figura 3.2.4).

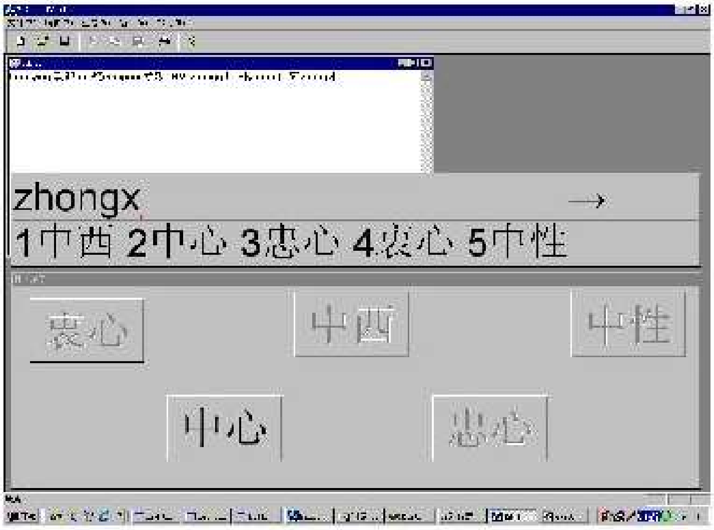

Figura 2: Digitação de ideogramas do idioma chinês através do *pinyin* [15]. Note a disposição dos botões em 'W' como uma forma de contornar a imprecisão do sistema de rastreamento de olhar

3.2.5 Modelagem do contexto

Como comentado na seção 3.1.4, um método de compensar a imprecisão do processo de rastreamento que tem sido empregado recentemente é incorporar mais informação sobre o *contexto* do diálogo em execução entre o usuário e o computador. Salvucci descreve em [12] uma interface, denominada IGO (*Intelligent Gaze-added Operated Interface*) que simula um desktop comum de um ambiente WIMP (Windows, Menus, Icons and Pointing) que pode ser operado através do olhar. A seleção é efetuada através do pressionamento de um botão no teclado, de maneira semelhante ao descrito em [8]; porém aqui a funcionalidade do apontamento tradicional é mantida, de forma que o apontamento pelo olhar seja uma forma adicional de interagir com o sistema. As fixações identificadas pelo sistema de rastreamento são mapeadas nos objetos através de um modelo probabilístico: dada uma fixação g, o item da interface i_{best} no qual a fixação será mapeada é aquele que maximiza a probabilidade $P(i|g)$, onde i representa o evento "selecionar o objeto i". Daí temos:

$$
\begin{aligned}
i_{best} &= \text{argmax}_{i \in I} \left(P(i|g) \right) \\
&= \text{argmax}_{i \in I} \left(\frac{P(g|i)P(i)}{P(g)} \right) \\
&= \text{argmax}_{i \in I} \left(P(g|i)P(i) \right)
\end{aligned}
$$

$P(g|i)$, a probabilidade de uma fixação (x, y) dada a intenção de se selecionar o item i é modelada através de uma distribuição gaussiana bidimensional centrada no item. No cálculo de $P(i)$ é onde especificamente é incorporada a informação de contexto da tarefa em execução atualmente. Cada um dos itens da interface recebe uma pontuação (posteriormente normalizada na escala de probalilidades) de acordo com a última tarefa executada, ou de acordo com algum estado da interface. Por exemplo, o botão "X" de fechamento de uma janela tem sua pontuação aumentada após a seleção de algum item na janela, considerando que o usuário tipicamente não fecha uma janela sem efetuar pelo menos uma seleção nela. O menu que contém o comando *Esvaziar lixeira* tem sua pontuação aumentada quando há itens na lixeira, e assim por diante.

Em [10] é desenvolvida uma modelagem mais complexa do contexto de uma aplicação que implementa a digitação de palavras olhando para um teclado exibido no vídeo. Para um dado conjunto de palavras pré-definido, é construída uma gramática (Tabela 3.2.5) que em suas regras de produção modela os passos cognitivos que podem ser tomados a partir de um dado estado.

Ha três possibilidades para a construção da gramática, enumeradas em ordem crescente do grau de informação seqüencial que possuem:

- **gramática de primeira ordem**: representa transições de letra para letra;

- **gramática de segunda ordem**: representa transições de pares de letras para letras;

- **gramática completa**: representa cada palavra separadamente.

A partir da gramática é contruída uma cadeia de Markov oculta que segue a estrutura da gramática dos passos cognitivos e tem como subunidades fundamentais de construção cadeias de Markov ocultas como as descritas na seção 3.1.4, para a identificação das fixações em cada alvo (no caso, em cada letra do teclado virtual). Uma vez montado esse modelo, um protocolo de fixações e movimentos sacádicos obtido a partir do rastreador de olhar

start	→	**word** rat
start	→	**word** trap
start	→	**word** part
rat	→	R A T end
trap	→	T R A P end
part	→	P A R T end
end	→	**out**

Tabela 1: Gramática completa que modela os passos cognitivos para a digitação das palavras *rat*, *trap* e *part* [10]. O usuário inicia a tarefa selecionando o botão **word**, que mostra a palavra a ser digitada, e encerra a digitação selecionando o botão **out**, representados por seus nomes na gramática.

é mapeado na seqüência mais provável de estados usando o Algoritmo de Viterbi.

Temos aí duas abordagens razoavelmente diferentes ao uso do contexto em interfaces que incorporam o rastreamento de olhar. A primeira tem uma estrutura e uma implementação mais simples, além de possibilitar de forma imediata a modelagem da mudança de estados da interface; ela exige, por outro lado, a codificação manual das regras de pontuação que geram as probabilidades de seleção de itens [4]. A segunda apresenta um modelo mais preciso dos passos cognitivos, porém o desempenho da interpretação em tempo real do modelo decresce à medida que se incorpora mais informação seqüencial na gramática e o número de palavras aumenta, o que parece limitar a incorporação de modelagens de contexto muito detalhadas em sistemas interativos. [5]

3.3 Projeto e execução de testes

Os testes formais na avaliação de técnicas de interação que incorporam o rastreamento de olhar (por exemplo, [14], [16], [12]) tem algumas características em comum, tanto no aspecto das decisões de projeto dos testes quanto dos resultados. Nesta seção vamos enumerar alguns pontos relevantes identificados na literatura.

[4]Note também que, para efeito de discussão, a interface exibida em [12] é uma simulação muito simplificada de uma interface WIMP real

[5]Essa informação carece de uma avaliação mais precisa, já que [10] não traz referências ao *hardware* usado na implementação do sistema.

Efeito da inexperiência dos usuários: obviamente, não existem usuários experientes no uso das novas técnicas de interação desenvolvidas. Assim, é uma técnica muito usual projetar testes que envolvem a execução de uma determinada tarefa em duas modalidades: usando o rastreamento de olhar e usando o controle manual tradicional (*mouse*). Em geral, se o tempo de execução da tarefa com o uso do rastreador é similar ao tempo de execução da tarefa usando o apontamento tradicional, esse fato costuma ser interpretado como uma indicação de que o desempenho na tarefa pode ser ainda melhor com o treinamento na nova técnica de interação.

Conformidade com a Lei de Fitts: implementações de algumas técnicas de interação para tarefas de seleção usando o olhar (tais como as descritas em [16], [14]) foram testadas para verificar a conformidade com um modelo do tempo de execução de movimentos manuais em direção a um alvo: a *Lei de Fiits*, que pode ser expressa na seguinte equação:

$$T = a_1 + a_2 \log_2(A/W) \tag{2}$$

Onde T é o tempo de execução do movimento, A é a amplitude do movimento realizado e W é o tamanho do alvo. a_1 e a_2 são coeficientes de regressão. [6] A Lei de Fitts é usualmente empregada na pesquisa em Interação Humano-Computador como ferramenta de avaliação de dispositivos de apontamento manual, daí o seu uso como parâmetro de comparação com as novas técnicas de interação desenvolvidas.

Os resultados obtidos até o momento são conflitantes. Sibert e Jacob [14] citam estudos anteriores que indicam que o apontamento auxiliado pelo olhar seguiria a Lei de Fitts, porém em seu experimento esse fato não se verificou. No teste piloto do MAGIC Pointing [16] a estratégia conservadora de posicionamento do cursor, que exige a ativação manual através do dispositivo de apontamento tradicional, trouxe evidencias de conformidade com o modelo, enquanto que a estratégia liberal, que não exige a ativação manual, não trouxe tais evidências.

Grau de similaridade da tarefa de teste com tarefas reais: nos trabalhos pesquisados, as conclusões em geral são validadas através de um procedimento formal de teste, feito da seguinte maneira: uma tarefa é executada por um grupo de usuários, tanto usando a técnica de

[6] A literatura mostra várias definições matemáticas desse modelo, que diferem especialmente no posicionamento das constantes aditivas e multiplicativas.

interação nova quanto da forma usual, e medidas relevantes à argumentação feita são tomadas (por exemplo, tempo de execução, número de erros). Nota-se uma marcante tendência à definição de tarefas altamente simplificadas. Podemos citar como exemplos a seleção de alvos representados por figuras geométricas simples ([16], [14]) ou digitação em teclados virtuais exibidos na tela ([15], [10]). Podemos inferir uma motivação para essa escolha: em tarefas mais simples e mecânicas, há menor possibilidade da medição ser atrapalhada por variabilidades pessoais no processamento cognitivo ou no efeito da fadiga. Uma análise experimental que destoa um pouco dessa tendência é a descrita em [12], onde os usuários executam tarefas de manipulação de objetos semelhantes àquelas que podem ser encontradas em interfaces WIMP usuais (por exemplo, mover e apagar arquivos, criar pastas). Mesmo nesse caso a interface é, compreensivelmente, uma simulação simplificada de um shell semelhante ao do MacOS.

4 Conclusão

Exibimos neste trabalho uma visão geral da aplicação de técnicas de rastreamento de olhar a interfaces; desde as aplicações pioneiras na área até o estado-da-arte, procuramos expor uma perspectiva da evolução no desenvolvimento de técnicas de interação.

Houve um desenvolvimento no sentido de cada vez mais considerar a visão como um canal de entrada passivo cuja atuação possa, na medida do possível, acontecer com um mínimo de controle consciente por parte do usuário. Para isso técnicas de ajuste dos dados provenientes do processo de rastreamento de olhar sempre foram empregadas, inicialmente de forma rudimentar e ingênua e atualmente evoluindo para modelagens mais sofisticadas do contexto da tarefa a ser executada.

Referências

[1] B.A.Myers. User-interface tools, introduction and survey. *IEEE Software 6(1)*, pages 15–23, 1989.

[2] Richard A. Bolt. Gaze-orchestrated dynamic windows. *Computer Graphics*, 15(3):109–119, august 1981.

[3] V. Bruce and P. R. Green. *Visual perception: physiology, psychology and ecology (2nd edition).* Lawrence Erlbaum Associates Ltd., 1990.

[4] S. K. Card, T. P. Moran, and A. Newell. The keystoke-level model for user performance time with interactive systems. *Communications of the ACM*, 23(7):396–410, 1980.

[5] C.H.Morimoto, D. Koons, A. Amir, and M. Flickner. Pupil detection and tracking using multiple light sources. *Image and Vision Computing*, v.18(no.4):pp. 331–336, 2000.

[6] A. Glenstrup and T. Engell-Nielsen. Eye controlled media: Present and future state. Master's thesis, Institute of Computer Science, University of Copenhagen, 1995.

[7] Anthony J. Hornof and David E. Kieras. Cognitive modeling demonstrates how people use anticipated location knowledge of menu items. In *Human Factors in Computing Systems: CHI 99 Conference Proceedings*, pages 410–417, 1999.

[8] Robert J. K. Jacob. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the CHI'90*, pages 11–18, 1990.

[9] Takehiko Ohno, Naoki Mukawa, and Atsushi Yoshikawa. FreeGaze: a gaze tracking system for everyday gaze interaction. In *Proceedings of the Eye Tracking Research & Application Symposium 2002 (ETRA-02)*, pages 125–132, 2002.

[10] Dario D. Salvucci. Inferring intent in eye-based interfaces. In ACM Press, editor, *Human Factors in Computing Systems: CHI 99 Conference Proceedings*, pages 254–261, 1999.

[11] Dario D. Salvucci. *Mapping Eye Movements to Cognitive Processes*. PhD thesis, Carnegie Mellon University, 1999.

[12] Dario D. Salvucci and John R. Anderson. Intelligent gaze-added interfaces. In *Human Factors in Computing Systems: CHI 2000 Conference Proceedings*, 2000.

[13] Dario D. Salvucci and Joseph H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research and Applications Symposium*, pages 71–78, 2000.

[14] Linda E. Sibert and Robert J. K. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the CHI'00*, pages 281–288, 2000.

[15] Jingtao Wang, Shumin Zhai, and Hui Su. Chinese input with keyboard and eye-tracking – an anatomical study. In *Proceedings of the CHI'01*, pages 349–356, 2001.

[16] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'99)*, pages 246–253, 1999.