UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Detecção de padrões de leitura com baixa amostragem

Monografia para a disiciplina MAC0499

Aluno: Carlos Eduardo Leão Elmadjian

Orientador: Prof. Dr. Carlos Hitoshi Morimoto

1 Resumo

Aplicações para computação vestível estão, potencialmente, sempre em funcionamento para auxiliar o usuário em suas tarefas cotidianas. Por outro lado, esse estado de alerta constante pode ser prejudicial ao usuário em alguns momentos. Assim, faz-se necessário o uso de contexto para melhorar a qualidade das interações, de modo que as aplicações possam fornecer informações úteis e relevantes na hora apropriada. Além disso, o consumo de energia está rapidamente se tornando um gargalo para a viabilidade da tecnologia vestível. Neste trabalho, iremos explorar como o reconhecimento de padrões de leitura pode prover contexto para aplicações e quais as abordagens algorítmicas necessárias para preservar consumo de processamento e, consequentemente, energia. Nossos resultados mostram que o algoritmo aqui proposto tem um desempenho significativamente melhor do que outras soluções em condições de baixa amostragem.

2 Introdução

2.1 Motivação

Computadores cada vez menores e dedicados às mais diferentes tarefas, presos às nossas roupas, cabeça, olhos, ouvidos e pulsos. Esse cenário, que há décadas atrás não passaria de ficção científica, hoje se encontra cada vez mais próximo da realidade.

Contrário ao que intui o senso comum, os primeiros protótipos em computação vestível começaram a circular no meio acadêmico já desde a década de 1960 [26]. No entanto, o interesse na pesquisa por *wearables* como uma nova maneira de interagir com computadores só ganhou verdadeiramente momento com os avanços tecnológicos no final da década de 1990 [16].

Idealmente, equipamentos vestíveis devem ser leves, resistentes, flexíveis, funcionar por um longo período de tempo e não devem monopolizar a atenção do usuário [16]. Além disso, muitos desses aparelhos não são de propósito geral, como os computadores tradicionais: cada um deles pode ser pensado para um fim que, grosso modo, está voltado para o auxílio do usuário em tarefas específicas do cotidiano. A somatória de todas essas características impõe novos paradigmas para como interagimos com a realidade [25]. Em particular, o uso de rastreadores de olhar como um acessório vestível já é aventado em pesquisas acadêmicas há alguns anos [4] e representaria também uma nova maneira de interagir com computadores.

Embora esses dispositivos façam parte de pesquisas científicas há quase um século, a maior parte dos estudos realizados com esses equipamentos esteve restrita à compreensão do comportamento ocular, algo de extremo valor para áreas como psicologia e neurociência. Já na ciência da computação, os rastreadores se tornaram objeto de estudo

apenas mais recentemente, com os avanços tanto em poder computacional quanto em algoritmos para processamento de imagens [7].

As aplicações mais recorrentes de rastreadores dentro da ciência computação estão tradicionalmente ligadas à pesquisa de tecnologias assistivas, uma vez que certas condições médicas como Esclerose Lateral Amiotrófica (ELA), Síndrome do Encarceramento ou alguns tipos de tetraplegia subjugam seus portadores a uma limitação física extrema, de modo que só lhes restam os movimentos oculares como maneira exclusiva de interagirem com a realidade.

Todavia, como tecnologia vestível, rastreadores ainda são pouco explorados. Apenas observando os principais tipos de movimentos dos olhos, como fixações (manutenção do olhar em uma região), sacadas (movimentos rápidos que duram entre 20 ms e 100 ms) e perseguições, pode-se inferir sobre os interesses, preferências, dificuldades e outros comportamentos individuais [23]. Embora já existam alguns trabalhos contemplando esses aspectos, não se encontram registros de como aplicar tais conhecimentos de modo contínuo no cotidiano das pessoas.

Por fim, sendo a leitura hoje uma prática essencial na atividade humana, isto é, uma das habilidades mais relevantes para a qual as pessoas não foram biologicamente programadas [23], pouco se observa também na literatura a respeito de como a interpretação dos movimentos oculares podem prover pertinência aos usuários, como *feedback* de relevância, assistência para interação sem as mãos, estatísticas para geração de conteúdo, entre outras possibilidades.

2.2 Justificativa

Na literatura, alguns autores classificam a interação do olhar com interfaces de janelas em cinco diferentes tipos: inspeção, busca, exploração, monitoramento e leitura. Cada um desses tipos está associado a um padrão de comportamento, sendo a leitura, possivelmente, o mais relevante entre todos, isto é, aquele para o qual o usuário despende mais tempo e em que seu interesse se manifesta de forma mais clara [6].

Por meio da análise da leitura em tempo real é possível depreender contexto: fixações prolongadas em uma região de um texto podem indicar dificuldade de compreensão de um termo, regressões constantes (e.g. releitura de uma linha) podem salientar uma dificuldade de compreensão geral, sacadas grandes sugerem falta de interesse sobre o conteúdo, enquanto que sacadas curtas alternadas com fixações rápidas indicam compenetração [21]. Kunze et al. [14] demonstraram que é possível, inclusive, inferir sobre o tipo de documento lido (e.g. revistas, jornais, quadrinhos) baseando-se apenas na análise dos padrões de leitura do usuário.

Contudo, embora o estudo de algoritmos para detecção de leitura já tenha quase duas décadas [6], a investigação sobre como fazê-lo com baixa amostragem ainda é inédita e fundamenta-se, sobretudo, pelo fato de que o consumo de energia está rapidamente se tornando o fator limitante para dispositivos vestíveis [19, 25].

Essa condição é ainda mais crítica para rastreadores de olhar móveis, que tipicamente apresentam ao menos duas câmeras — uma para capturar o que o usuário enxerga (cena), e outra utilizada para capturar os movimentos oculares. Somando-se a isso o fato de os rastreadores de desempenho mais limitados hoje trabalharem a uma taxa de 30 Hz, o grau de processamento envolvido nessa quantidade de imagens por segundo está longe de ser desprezível, o que implica, consequentemente, um consumo de energia maior.

Câmeras USB, quando em regime, podem ir de 100 mW (*stand-by*) de consumo para até mais de 1000 mW [2], rapidamente se tornando um gargalo para um uso prolongado, sobretudo em comparação com processadores desenvolvidos especificamente para equipamentos vestíveis, como a linha OMAP, da Texas Instruments, que tipicamente apresenta consumo inferior a 1000 mW [1].

O regime em potência baixa dos processadores também implica uma quantidade de cálculos por segundo inferior ao convencional para aplicações com grande demanda. Faz-se, portanto, necessário o emprego de técnicas que reduzam drasticamente a demanda sobre processamento das informações oriundas do olhar.

2.3 Objetivos

O presente trabalho tem como objetivo central demonstrar uma solução para o problema da detecção de padrões de leitura com baixa amostragem, ensejando, dessa forma, aplicações de interesse público com rastreadores de olhar em computação vestível.

Ainda dentro dessa meta, almejamos mostrar que, mesmo com tais condições restritivas, podemos estabelecer critérios de classificação do tipo de leitura (convencional ou *skimming*), permitindo que futuras aplicações possam basear-se em contextos, ainda que rudimentares.

Como objetivos secundários, pretendemos exibir também um panorama das soluções existentes para reconhecimento da leitura por máquina e analisar problemas pertinentes à baixa amostragem de sinais para este propósito e, por fim, exibir uma prova de conceito do reconhecimento da leitura como mecanismo de contexto, utilizando o algoritmo desenvolvido ao longo desta investigação.

2.4 Desafios

Os desafios para que o uso de contexto em computação vestível se expanda estão em pelo menos três frentes. Em primeiro lugar, há a questão da aceitação social no nível do *hardware*: aparelhos invasivos, indiscretos, inseguros ou intrinsecamente sem propósito dificilmente terão algum impacto em termos de consumo de massa [17]. Sem o fomento apropriado, é possível que os investimentos e o interesse pela pesquisa na área entrem em declínio.

Uma segunda frente a ser tratada está no nível das aplicações, e tal problema está intimamente ligado ao anterior, dado que uma massa crítica de desenvolvedores está associada a um panorama de oportunidades no setor [25]. Equipamentos potencialmente relevantes ao cotidiano do usuário são virtualmente inúteis sem aplicações que proporcionam ou facilitam alguma necessidade.

O último e mais relevante desafio reside nos algoritmos. Desenvolver aplicações apropriadas, com um comportamento atento, passivo, inoportuno e, ao mesmo tempo, com *consciência* de contexto, requer ainda uma pesquisa ampla na área de aprendizagem computacional das linguagens corporal e contextual humanas.

Embora o padrão de leitura simples seja algo claramente discernível para um especialista, ainda assim é um sinal bastante ruidoso, e uma mera filtragem sobre o sinal não é suficiente para eliminar — ou somente mitigar — o ruído, uma vez que ele está associado não só ao indivíduo *per se* como também ao comportamento individual em relação ao conteúdo lido.

Além disso, a leitura é tipicamente uma forma de interação não seletiva, ou seja, na qual o olhar não é empregado para o controle de uma interface. Portanto, não se deve presumir que o depreendimento das intenções do usuário nesse cenário seja equiparável

às tomadas de decisões explicitamente mensuráveis em interações seletivas.

Finalmente, poucos trabalhos têm tratado do uso desses padrões para novas formas de interação, tais como *retorno de relevância*, como sugerido por Buscher et al. [5] ou *leitura aumetada*, como proposta por Biedert et al. [3]. Interações mais sofisticadas (e possivelmente mais interessantes), considerando contextos com longos períodos de *feedback* dos usuários e com o emprego de inteligência artificial sobre esses comportamentos, ainda carecem de investigação.

3 Características da leitura

3.1 O movimento do olhar na leitura

Para que se possa dissertar sobre como o olhar se comporta na leitura, é preciso antes definir o que de fato denotamos como leitura. Neste trabalho, esse termo tem um significado muito mais preciso do que se pode inicialmente supor.

Em primeiro lugar, este estudo trata da leitura de textos em línguas indo-europeias em que a ordem com que se percorre os caracteres da escrita se dá da esquerda para a direita (primariamente) e de cima para baixo (secundariamente). Isso pode soar um tanto trivial para leitores ocidentais, mas essa ressalva é fundamental: em algumas línguas asiáticas, como chinês, japonês ou coreano, a convenção do sentido de leitura é da direita para esquerda. O mesmo vale também para línguas do Oriente Médio, com o intrigante fato de que em línguas arábicas é praxe escrever termos estrangeiros da esquerda para direita, o que pode tornar o sentido da leitura um tanto confuso em alguns casos.

Ademais, quando nos referimos à leitura, estamos tratando do comportamento dito *convencional*, segundo a literatura [18], isto é, excluímos aqui todas as outras manifestações de leitura que representam uma anomalia frente a esse padrão (Figura 1), como a de indivíduos portadores de dislexia, portadores do Transtorno do Déficit de Atenção e Hiperatividade, crianças em estágio inicial de alfabetização ou mesmo adultos semialfabetizados.

Feitas as devidas ressalvas, podemos caracterizar a leitura, em termos do movimentos oculares, como uma alternância entre sacadas e pequenas fixações com 200 a 300 ms de duração, sendo essas sacadas tipicamente curtas e feitas predominantemente da esquerda para a direita ao longo de uma linha de texto. Quando tal linha chega ao fim, o olho

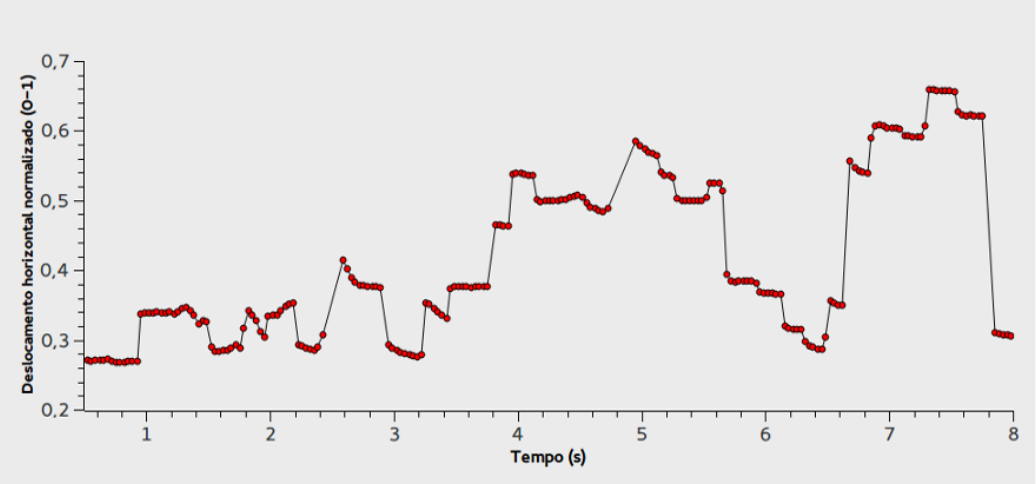

Figura 1: O padrão de um indivíduo com déficit de leitura. Nesse caso, as perturbações causadas por regressões são frequentes, o que dificulta a identificação do comportamento de leitura.

faz uma nova sacada, muito maior, em geral, para o início da próxima linha, um deslocamento comumente denominado de *regressão*, termo que também pode ser empregado para classificar movimentos de releitura [22].

Durante as sacadas, os olhos se movem a mais de 500^o por segundo, quando então ocorre a *supressão sacádica*, isto é, não é possível obter informação visual distinta da cena [21]. Desse modo, o cérebro só consegue extrair informações de um texto durante fixações.

Todavia, embora o campo de visão humano seja de aproximadamente 180^o, só temos acuidade visual em uma pequena região da imagem que capturamos da cena (Figura 2). Isso porque a percepção dos detalhes é uma propriedade dos cones, células fotossensíveis que se concentram apenas numa região específica da retina, denominada fóvea. Assim, o conjunto de fixações realizadas durante a leitura se distribui ao longo das linhas de um texto de forma a criar uma cobertura das informações com o máximo possível de acuidade.

Outros movimentos oculares, como vergência, vestibular e perseguição, contribuem

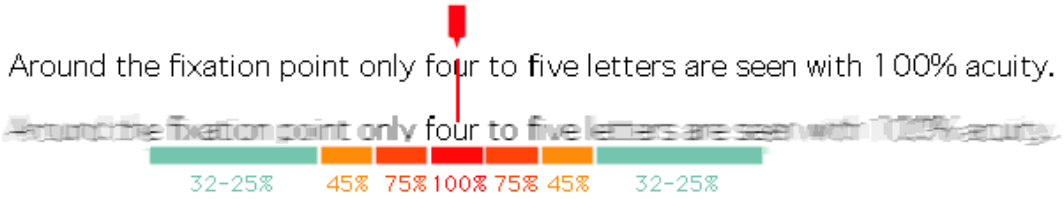

Figura 2: Representação artística da acuidade visual humana quando uma imagem é capturada durante uma fixação. O ponteiro vermelho indica o centro da fóvea. Fonte: *Hans-Werner34*

muito pouco para caracterizar a leitura, sobretudo dentro do modelo clássico de análise, em que se supõe textos a uma distância fixa, com o leitor relativamente inerte [15]. Se construirmos um gráfico dos movimentos horizontais do olho durante a leitura em função do tempo, percebemos uma padrão muito claro (Figura 3), dependente basicamente de fixações e sacadas, que por sua vez é conhecido na literatura como *staircase pattern* [15].

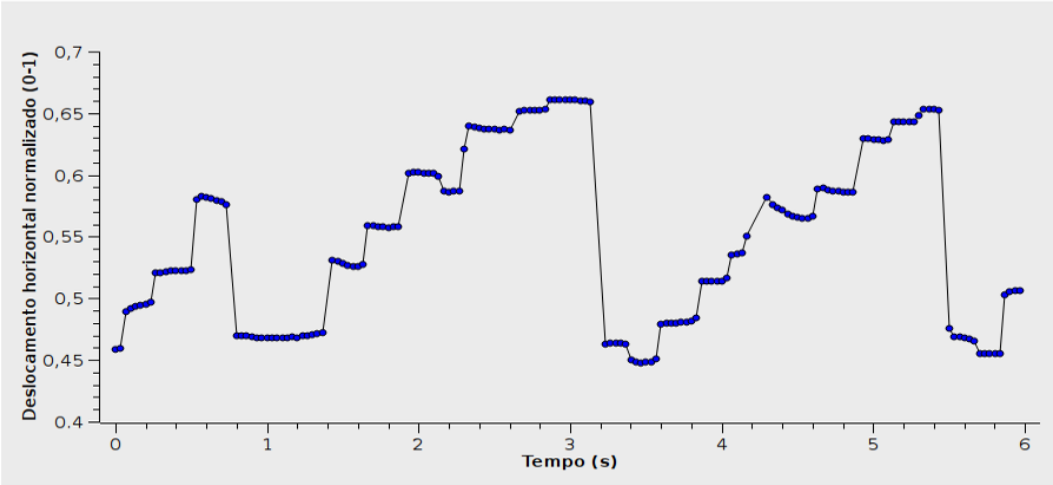

Figura 3: O *staircase pattern* de um leitor normal. Note que ocorrem regressões, mas elas são raras.

3.2 Depreendimentos do comportamento de leitura

Durante a chamada "segunda era" das pesquisas com rastreadores de olhar, marcadas pelas investigações behavioristas, surgiram os primeiros trabalhos que registraram os

movimentos característicos do olhar na leitura [7]. No entanto, somente em meados de 1970 é que se passou a estudar mais intensamente as implicações cognitivas do processo de leitura [21].

Os primeiros trabalhos nesse sentido pressupunham que leitores proficientes faziam um significativo uso do contexto textual para conseguir processar informações rapidamente. Embora essa concepção tenha sido validada empiricamente, estudos posteriores mostraram que o uso do contexto é ainda mais relevante para leitores iniciantes [22].

De todo modo, a familiaridade de indivíduos com um texto é algo sobre o que se pode inferir a partir dos movimentos oculares. Leitores proficientes, em geral, dependem menos da região de grande acuidade visual do que principiantes para capturar informações. Acredita-se que a experiência pessoal gradualmente forneça um arcabouço de inferências baseadas em contexto. Dessa forma, leitores experientes podem, por exemplo, ler apenas os três primeiros caracteres de uma palavra para processá-la e utilizar informações periféricas à região de acuidade, como o tamanho do próximo termo, para realizar suposições baseadas em contexto semântico [21, 22].

Outro dado relevante para inferir sobre processos cognitivos a partir dos movimentos do olhar na leitura é o tempo de fixação. Já se constatou em experimentos que passagens de texto consideradas difíceis geram, em média, tempos de fixação maiores nos leitores. Em geral, termos raros, desconhecidos ou contextualmente inapropriados recebem, por exemplo, um tempo maior de atenção do leitor em contraste com a média [23].

Finalmente, podemos inferir características do leitor baseando-nos no conjunto de sacadas acumuladas ao longo de um determinado tempo. Indivíduos com alfabetização precária ou com déficit de atenção costumam apresentar um padrão de regressões constantes, ou seja, termos, linhas ou mesmo parágrafos são relidos com uma frequência atípica

(Figura 1) [18, 21]. Por outro lado, poucas fixações por linha, intercaladas por sacadas espaçadas, podem insinuar pressa, busca por informação ou desinteresse do leitor [8].

3.3 Diferenças entre leitura e *skimming*

Se por um lado há um consenso no meio acadêmico sobre o que caracteriza a leitura convencional, por outro, não há um entendimento claro quanto à classificação da leitura rápida. Alguns autores subdividem essa atividade em dois grupos (*skimming* e *scanning*) [6], enquanto outros julgam que a partir de um determinado limiar de palavras por minuto, tudo que se assemelhe à leitura deve ser entendido como *skimming* [21].

O limiar mais aceito, nesse caso, está entre 600 - 700 palavras por minuto e foi estabelecido empiricamente [21]. Dado que um leitor proficiente lê, em média, a uma taxa de 250 palavras por minuto, intui-se que com mais do que o dobro da velocidade haja uma queda significativa da compreensão do texto.

Alguns estudos, porém, revelaram que embora haja uma correlação forte entre o nível de compreensão do indivíduo e a taxa de leitura, ainda assim é possível capturar as principais informações de um texto lendo rapidamente, de modo a se preservar um bom entendimento do conteúdo, em detrimento, é claro, dos detalhes. Curiosamente, observou-se ainda que a taxa de compreensão é ótima quando o indivíduo se sente levemente pressionado e lê a uma velocidade um pouco acima do que considera normal [8].

Portanto, mesmo sem uma definição precisa, existe um entendimento de que a atividade de *skimming* certamente não é algo excêntrico à leitura. Embora seja mecanicamente impossível para um ser humano ler a mais de 600 palavras por minuto (dadas as latências de preparação de sacadas, fixações e movimentos de estabilização do olho [23]), a literatura mostra que o *skimming* é claramente distinto, por exemplo, de uma busca por

termos em um texto.

Por outro lado, o *skimming* não é exatamente uma leitura acelerada. A uma velocidade entre 600 - 700 palavras por minutos, um indivíduo deixa de capturar várias informações que estão fora do seu campo de acuidade visual, o número de fixações por linha é menor, a latência das fixações é menor e o contexto desempenha um papel ainda mais significativo no entendimento, isto é, o cérebro acaba preenchendo com suposições semânticas o vazio deixado pela falta de dados visuais [21].

3.4 Reconhecimento de leitura por máquina

Ainda que a leitura tenha um padrão característico, tal padrão está longe de ter um aspecto determinístico. A leitura está intimamente ligada à cognição humana [21], um processo complexo e alimentado por diversos componentes como influências do ambiente, experiências pessoais, formação educacional, contexto textual, saúde, entre outros. Portanto, embora emerja uma ordem desse aparente caos, trata-se de uma ordem relativamente ruidosa [15].

Para que uma máquina possa reconhecer se uma pessoa está lendo a partir dos seus movimentos oculares, precisamos eliminar — ou pelo menos mitigar — esse ruído. Entretanto, essa não é uma tarefa simples.

Consideremos o modelo do movimento de leitura idealizado em comparação com um conjunto de dados coletados na prática (Figura 4). Não é difícil perceber a semelhança entre ambos, mas também não se deve supor que há um padrão no ruído. Um indivíduo pode reler o mesmo texto de diferentes maneiras, fazendo fixações com latências distintas para as mesmas regiões, sacadas de diferentes tamanhos, releituras ou saltos imprevisivelmente. Tudo depende do processo cognitivo, algo essencialmente subjetivo.

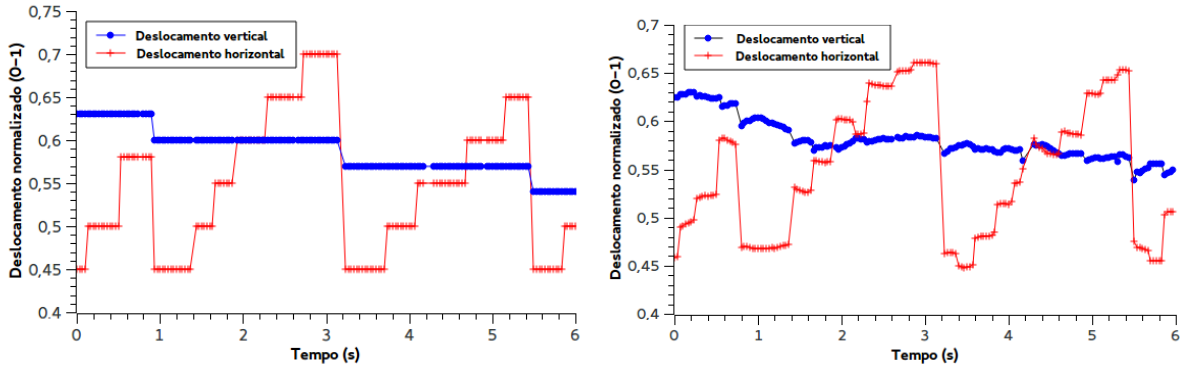

Figura 4: À esquerda, o modelo idealizado do padrão de leitura: as fixações ocorrem dentro de um intervalo limitado e os deslocamentos horizontal e vertical são uniformes. À direita, o padrão de leitura real de um indivíduo.

Mesmo entre leitores proficientes, o modelo idealizado está longe de ser perfeito. Estima-se que 10% do tempo seja gasto com regressões [23] e vários estudos exibem evidências de que esses leitores conseguem deduzir palavras subsequentes em algumas circunstâncias, mesmo quando tais termos não se encontram na sua região de alta acuidade visual, o que permite que façam menos fixações por linha [22], distorcendo o poder de previsão do modelo.

Como se isso não bastasse, deve-se considerar que a cabeça dificilmente permanece rígida enquanto lemos. Logo, os movimentos de tronco e cabeça durante a leitura fazem com que as fixações apresentem um caráter de instabilidade [15], sendo impregnadas por movimentos vestibulares e de estabilização do globo ocular.

Dessa forma, parece razoável supor que uma abordagem estatística para detectar o evento de leitura seja o caminho mais adequado para o reconhecimento de máquina. De fato, essa percepção é corroborada pela literatura, uma vez que todos os algoritmos para reconhecimento de padrões de leitura publicados até então utilizam estratégias probabilísticas, como veremos no próximo capítulo.

4 Algoritmos

Nesta seção, iremos tratar dos principais algoritmos presentes na literatura para reconhecimento da leitura a partir de rastreadores de olhar. Vamos fazer uma descrição geral sobre seu funcionamento e discutir brevemente os resultados alcançados pelos autores.

4.1 Algoritmo de Campbell e Maglio (2001)

Campbell e Maglio foram os primeiros autores na literatura a proporem um algoritmo para detecção de leitura por máquina. Sua solução fundamenta-se no reconhecimento de eventos do olhar e na subsequente classificação desses eventos de acordo com um modelo idealizado de leitura [6].

Para viabilizar o processo em tempo real, é construída uma representação simplificada dos movimentos do olhar, de maneira que os únicos eventos a serem detectados são fixações e sacadas. Para o bom funcionamento do algoritmo, são necessárias ainda algumas condições de controle, como uma postura rígida do leitor em relação ao texto, uma distância fixa pré-determinada e o conhecimento prévio do tamanho dos caracteres a serem lidos.

O método consiste em usar as coordenadas do olhar transformadas para um plano em intervalos de 100 ms, coletando-se 60 pontos por segundo. Para suavizar erros inerentes à função de calibração e minimizar perturbações involuntárias do usuário, os autores sugerem o emprego da média de três pontos adjacentes, totalizando, então, 20 pontos médios por segundo.

A ideia central é acumular "evidências de leitura" em uma variável sempre que o olho faz um movimento favorável ao modelo de leitura e decrementar o valor dessa variável quando o olho realiza um movimento contrário. No total, os autores definem 13 tipos

de eventos vindos do rastreador a serem interpretados (Tabela 4.1), todos em função de eixo, distância e direção do olhar. Os eventos são então "tokenizados" e recebem uma pontuação de acordo com o *token*.

eixo, distância e direção	**Token**	**Pontuação**
X curto à direita	leitura	10
X médio à direita	skimming	5
X longo à direita	scanning	Reset
X curto à esquerda	regressão	-10
X médio à esquerda	skimming	-5
X longo à esquerda	scanning	Reset
Y curto para cima	skimming	-5
Y médio para cima	scanning	Reset
Y longo para cima	scanning	Reset
Y curto para baixo	sacada antecipatória	0
Y médio para baixo	skimming	-5
Y longo para baixo	scanning	Reset
X longo ou médio para esquerda + Y curto para baixo	reinício de linha	5

Tabela 1: Esquema de pontuação de Campbell e Maglio por "tokenização" de movimentos do olhar. A pontuação positiva indica evidências de suporte à leitura, enquanto a negativa sinaliza o contrário.

Para que a leitura seja detectada, a pontuação acumulada (com múltiplos de cinco) deve ultrapassar um limiar de valor 30, determinado heuristicamente por Campbell e Maglio. Esse limiar também é fundamental para tratar dos problemas atrelados ao ruído do sinal, uma vez o sistema só deve reconhecer o comportamento de leitura com um conjunto significativo de indícios.

Em termos de desempenho, o algoritmo, de acordo com os autores, apresentou uma

acurácia superior a 90%, algo que não conseguimos reproduzir em nossos experimentos. Contudo, o índice de falsos positivos pôde ser verificado empiricamente. Deve pesar ainda o fato de que não tivemos acesso aos mesmos equipamentos, tampouco pudemos trabalhar com o mesmo patamar de amostragem da publicação.

4.2 Algoritmo de Buscher et al. (2008)

O algoritmo de Buscher et al.[5] representa, de certo modo, uma sofisticação do algoritmo de Campbell e Maglio (2001). Trata-se também de um dos poucos trabalhos na literatura que procura reconhecer não apenas a leitura como também o padrão de *skimming*.

A metodologia adotada é semelhante à de Campbell e Maglio, mas também há novos mecanismos de robustez. O primeiro deles é a detecção de fixações: no lugar de usar uma mera suavização de pontos do olhar, o algoritmo procura coletar amostras suficientes para determinar se o usuário está realizando uma fixação. Isso é feito delimitando-se um quadrado de tolerância sobre a superfície observada, de modo que todos os pontos próximos o bastante que recaiam sobre o quadrado são considerados partes de uma mesma fixação. Se mais de três pontos consecutivos não pertencerem a essa região, então a fixação terminou.

Uma outra melhoria está no uso de uma ferramenta de reconhecimento óptico de caracteres (OCR) do texto lido para parametrizar o algoritmo em função do espaço ocupado por letras. Como a caracterização das transições do olhar é bastante dependente do entrelinhamento e do tamanho da fonte do texto, obtém-se assim um esquema mais robusto de medição. Ademais, tais transições são medidas agora entre baricentros de fixações.

Embora aparente ser mais complexo que o de Campbell e Maglio, o algoritmo traz

uma simplificação importante: os eventos a serem detectados agora são sete em vez de 13 (Tabela 4.2). Novamente, o roteiro é semelhante à solução anterior: os eventos reconhecidos são "tokenizados" e cada *token* recebe uma pontuação. A diferença está no fato de que os mesmos *tokens* também são usados para contabilizar evidências para o *skimming* — com esquema de pontuação distinto, é claro. Assim, se o algoritmo coleta um conjunto de evidências que ultrapasse o valor 30, a leitura passa a ser detectada. Para o caso do *skimming*, o valor é 20.

Distância e direção em espaço de letras	**Token**	**Pontuação leitura**	**Pontuação *skimming***
$0 < X <= 11$	leitura	10	5
$11 < X <= 21$	skimming	5	10
$21 < X <= 30$	long skimming	-5	8
$-6 <= X < 0$	regressão curta	-8	-8
$-16 <= X < -6$	regressão longa	-5	-3
$X < -16$ e Y curto-baixo	reinício de linha	5	5
Outros movimentos	não relacionado	0	0

Tabela 2: Esquema de pontuação de Buscher et al. As transições entre uma fixação e a seguinte são classificadas em *tokens* e uma pontuação correspondente é atribuída, podendo sinalizar dois estados: leitura ou *skimming*.

Pelo fato de o algoritmo não ser o foco principal do artigo, Buscher et al. não exibem resultados de desempenho, como precisão e revocação. De todo modo, nossos experimentos mostraram que o algoritmo tem um desempenho compatível com o relatado por Campbell e Maglio [6] a uma taxa de 30 Hz.

4.3 Algoritmo de Kollmorgen e Holmqvist (2007)

Este algoritmo [13], assim como os demais apresentados nesta seção, utiliza um modelo do comportamento do olhar na leitura para detectá-la. A diferença, porém, reside no fato de que as alternâncias entre fixações e sacadas são representadas por um Modelo Oculto de Markov (HMM).

A primeira fase do algoritmo é de pré-processamento dos dados: deve-se analisar todos os pontos do olhar coletados e eliminar todos os movimentos que não sejam fixações ou sacadas. Piscadas também são computadas nesse processo e cada fixação é armazenada numa estrutura de dados contendo sua posição espacial, tempo de início e duração. Já as sacadas são contabilizadas como segmentos entre duas fixações.

Concluída essa fase, inicia-se a construção de uma máquina de estados finita, em que as transições e a saída são probabilísticas. Kollmorgen e Holmqvist acreditam que a leitura pode ser bem descrita por um HMM com seis estados e duas partições: leitura e não leitura (Figura 5). Definido o modelo e os parâmetros iniciais, os estados mais prováveis de transição são calculados a partir do algoritmo de Viterbi [20].

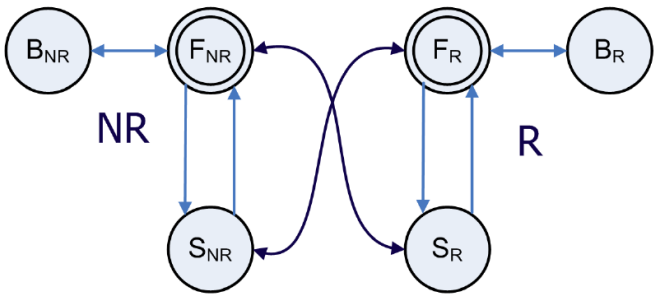

Figura 5: Modelo markoviano de Kollmorgen e Holmqvist. **R** indica leitura e **NR**, não leitura. Os estados **B**, **F** e **S** representam piscadas, fixações e sacadas, respectivamente.

Os autores argumentam que a tarefa mais complexa no algoritmo é determinar os parâmetros para o modelo. Um método proposto é analisar previamente o conjunto de dados e tentar realizar o *fitting*, isto é, utilizando supervisão. Eles também demonstram que é possível procurar os parâmetros de maneira não supervisionada, mas com resultados significativamente piores (<80 %).

Embora os autores relatem que o algoritmo possa ser executado em tempo real, o artigo condiciona sua execução a um pré-processamento intenso de dados e treinamento, o que de certa maneira contraria essa alegação. Somando-se a isso o fato de que a solução não apresenta um desempenho equiparável ao dos demais, tampouco uma complexidade computacional baixa, optamos por não implementá-la, pois não poderia posteriormente ser adaptada como uma alternativa eficiente e de baixo consumo para reconhecimento de padrões de leitura.

5 Estudo da baixa amostragem

5.1 A frequência de Nyquist

O Teorema da Amostragem de Shannon-Nyquist é a ferramenta fundamental que nos permite fazer processamento digital de sinais (DSP). A sua definição mais recorrente, atribuída a Shannon, é dada da seguinte forma: "se uma função $f(t)$ não contém frequências maiores do que W Hz, então ela pode ser completamente determinada por suas ordenadas em uma série de pontos espaçados $\frac{1}{2}W$ segundos" [24].

Em outras palavras, o teorema estabelece que se quisermos representar um sinal adequadamente, precisamos registrar pelo menos metade da sua frequência máxima. Caso contrário, a reconstrução da função se torna ambígua, ocorrendo o fenômeno denominado *aliasing* (Figuras 6 e 7).

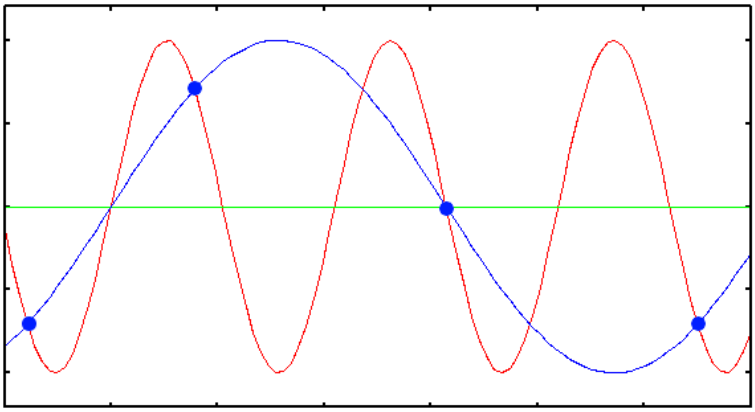

Figura 6: Exemplo de *aliasing* de uma senoide em função do tempo: a amostragem dada pelos pontos azuis é incapaz de diferenciar o sinal original, em vermelho, do sinal em azul.

O teorema se torna particularmente relevante em nosso caso, uma vez que computadores são máquinas de aritmética finita e, portanto, não possuem memória ilimitada para representar de modo contínuo um sinal capturado. Utilizando o teorema, intuímos que talvez seja possível reconstruir esse sinal de maneira discreta, conhecendo a quantidade

Figura 7: O fenômeno de *aliasing* também pode ocorrer em duas dimensões, como no caso de imagens. Nesse caso, temos uma função de distância com duas variáveis, mas os efeitos da subamostragem (imagem à esquerda) estão relacionados ao caso de uma variável.

mínima de amostras para isso [11]. A resposta para essa intuição está na *frequência de Nyquist*.

De certo modo, a frequência de Nyquist é uma concessão à impossibilidade de se representar funções contínuas com DSP: dado que não é possível adquirir infinitas amostras de um sinal contínuo, estabelece-se que para representar uma quantidade de S amostras por segundo devemos capturar pelo menos $\frac{1}{2}S$ amostras igualmente espaçadas no mesmo intervalo [10].

Na prática, essa limitação muitas vezes é imperceptível, já que as sensações humanas são processadas pelo cérebro também de maneira discreta. Assim, com uma representação suficientemente fina de um sinal (seja ele uma imagem ou um som, por exemplo), podemos provocar a ilusão da percepção de um sinal contínuo, desde que a frequência de *Nyquist* seja respeitada.

5.2 Padrões do olhar e o número de amostras

A maioria dos rastreadores de olhar empregados em trabalhos na literatura funciona a uma taxa de 50 a 60 imagens da pupila por segundo, uma frequência considerada adequada o suficiente para capturar interações do olhar com uma tela de computador [9].

Considerando que uma fixação típica para aquisição de informação dura entre 200 e 300 ms [21, 9], então um sistema de rastreamento trabalhando a 60 Hz poderá coletar em torno de 12 a 18 amostras desse evento. No entanto, é possível que ocorram fixações rápidas durante a interação, mas, em geral, com duração não inferior a 100 ms, o que nos leva a um mínimo de seis amostras para a detecção de uma fixação com esse equipamento.

Já as sacadas são movimentos muito rápidos e, portanto, de baixa duração (entre 20 e 100 ms [21, 23]). Detectá-las a 60 Hz ainda é possível (teríamos de uma a cinco amostras), mas identificá-las como tal seria algo um pouco mais complexo. Isso porque para os casos em que tivermos menos de três amostras, não há como garantir se os pontos coletados são, por exemplo, *outliers*, frutos de um erro de calibração.

Os movimentos de perseguição, por outro lado, não possuem uma duração definida, mas por apresentarem velocidade e amplitude menores que as sacadas, podem ser percebidos por um sistema de rastreamento com base na direção e deslocamentos realizados pela pupila (*scanpath*).

Outros movimentos mais sutis (de estabilização, por exemplo) também podem ser detectados, porém costumam estar associados a outros, como fixações. Além disso, a amostragem pode sofrer influência de ruídos inerentes ao processo de rastreamento, como movimentos involuntários de tronco e cabeça, no caso de rastreadores remotos, erros da função de calibração e o desalinhamento do centro da pupila em relação à fóvea.

Neste trabalho, iremos realizar nossas análises com um rastreador de 30 Hz vestível

[12]. Sendo assim, teremos acesso a apenas metade das amostras mínimas aqui contabilizadas e a identificação de sacadas poderá ficar prejudicada. Para efeito de investigação, todavia, essa restrição é irrelevante, dado que nenhum algoritmo para reconhecimento de leitura baseia-se em detecção de sacadas.

5.3 Resolução mínima para detecção da pupila

O reconhecimento da pupila, embora não seja o foco deste trabalho, desempenha um papel fundamental no escopo do problema: se quisermos aplicações de longa duração para dispositivos vestíveis, essas aplicações devem demandar o menor consumo de processamento possível, mantendo um nível de usabilidade satisfatório.

Quanto melhor a resolução da imagem da pupila, maiores são as chances de um sistema reconhecedor mapear corretamente o centro da pupila para um ponto em uma tela, por exemplo. Contudo, o custo do processamento de imagens cresce, no mínimo, quadraticamente em função do tamanho, dado que elas são representadas computacionalmente como uma matriz de *pixels*, em que que cada coordenada armazena uma certa quantidade de valores — a depender do modelo de cor adotado, como RGB, RGBA, HSV, escala de cinza, entre outros.

Não há ainda na literatura trabalhos que mostrem quais as condições mínimas de resolução para que a pupila seja detectada. Especula-se que uma redução severa da qualidade da imagem resulte ou em perda sensível de acurácia ou falhas de detecção, o que, consequentemente, leva-nos ou ao descarte de amostras ou a lacunas de amostragem, respectivamente.

Em testes com resultados ainda não publicados, Aluani et al. mostraram que o melhor custo-benefício entre consumo de energia e acurácia com o sistema de rastreamento

empregado nesta pesquisa [12] é de 240 linhas de resolução (Figura 8). Abaixo disso, a acurácia no mapeamento da imagem da pupila para a superfície observada se torna um problema significativo, a ponto de inviabilizar a maior parte das aplicações. Os resultados desse estudo podem ser verificados no Anexo I.

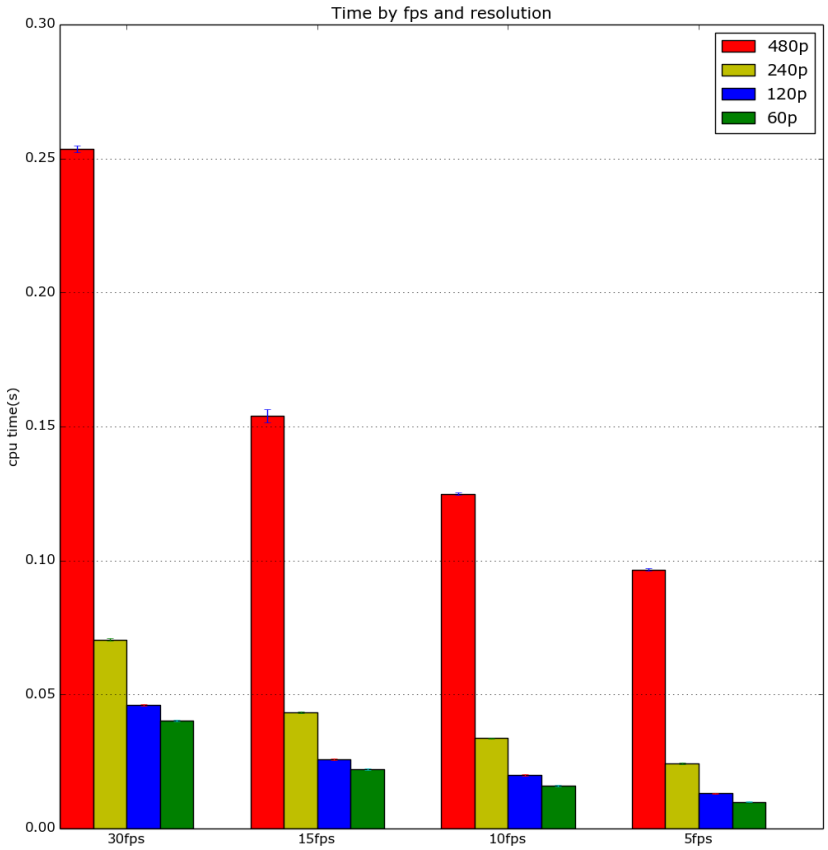

Figura 8: Relação entre consumo de CPU (em segundos) e o número de linhas de resolução. Com 240 linhas, há um ótimo custo-benefício entre redução de processamento e acurácia.

5.4 Estudo dos algoritmos e a amostragem

Quando analisamos os três algoritmos da literatura para detecção de leitura [6, 5, 13], logo notamos uma abordagem comum para reconhecimento de eventos do olhar — em

particular, das fixações.

Essa metodologia parece bastante razoável quando a limitação de amostras não é um problema. Mas se pretendemos utilizar tais algoritmos em condições críticas, isto é, em que o consumo de energia é o mínimo aceitável, não podemos supor que teremos uma quantidade suficiente de amostras para representar esses eventos.

Considere, por exemplo, um rastreador que funcione a 5 Hz (em contraste com os comumente empregados, de 50 - 60 Hz). Isso significa que teremos uma amostra a cada 200 ms, o que nos garante, aproximadamente, apenas um ponto para cada fixação ocorrida durante a leitura de um texto. Nesse caso, como seria possível, por exemplo, distinguir uma fixação de um movimento de perseguição (Figura 9) e evitar o *aliasing*?

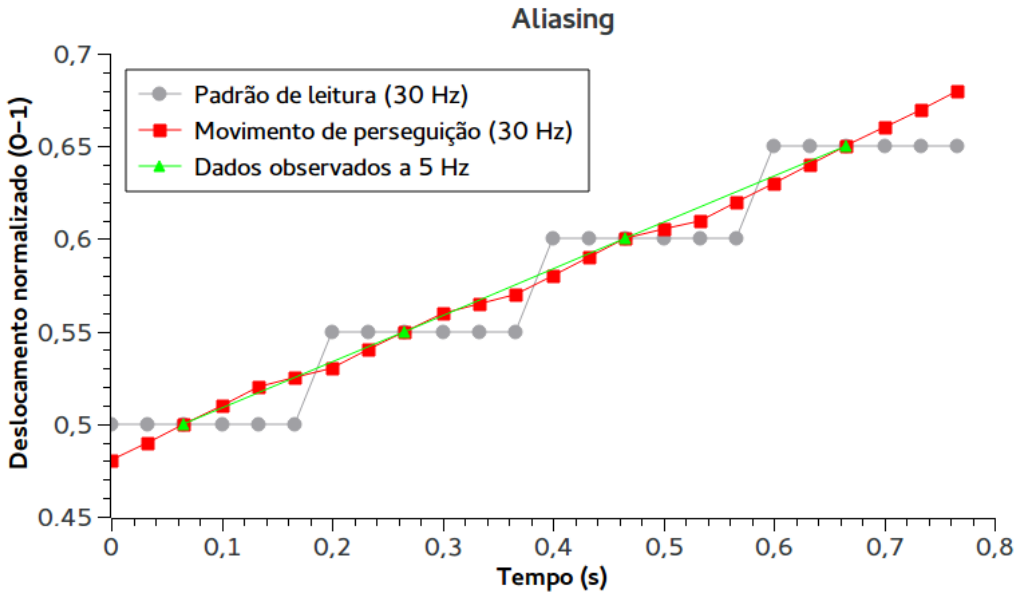

Figura 9: Exemplo de efeito de *aliasing* a 5 Hz: o conjunto de fixações que define um padrão de leitura passa a ser indistinguível de um movimento de perseguição com a mesma direção e duração.

Desde já, podemos intuir por que os algoritmos aqui abordados falham. Campbell e Maglio propõem uma filtragem dos dados por suavização [6], enquanto Buscher et al. e Kollmorgen e Holmqvist, cada qual à sua maneira, propõem uma interpretação de fixações

em função de *clusters* de pontos[5, 13]. Em todos os casos, simplesmente não há pontos suficientes para a tarefa a 5 Hz.

No próximo capítulo, iremos exibir uma solução para esse problema e um comparativo entre alguns dos algoritmos e a nossa proposta. A chave para resolver o dilema, dado que a frequência de Nyquist não pode ser ignorada, está no fato de que não precisamos reconstruir um sinal digital em sua integridade para reconhecer o processo de leitura.

6 Novo algoritmo

6.1 Concepção

Em breve.

6.2 Implementação

Em breve.

6.3 Testes e metodologia

Em breve.

6.4 Resultados

Em breve.

6.5 Discussão

Em breve.

7 Prova de conceito

7.1 Idealização e tentativas iniciais

Em breve.

7.2 Descrição do software

Em breve.

7.3 Possibilidades e futuras aplicações

Em breve.

8 Conclusões

8.1 Eliminando o *aliasing* na detecção da leitura

Em breve.

8.2 Comunicação passiva como forma de interação

Em breve.

8.3 Dificuldades a serem superadas

Em breve.

Referências

[1] http://processors.wiki.ti.com/index.php/omap-l138_power_consumption_summary. Acessado em 08-10-2015.

[2] http://www.tomsguide.com/us/squeezing-more-life-out-of-your-notebook,review-583-26.html. Acessado em 15-08-2015.

[3] Ralf Biedert, Georg Buscher, Sven Schwarz, Jörn Hees, and Andreas Dengel. Text 2.0. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 4003–4008. ACM, 2010.

[4] Andreas Bulling, Daniel Roggen, and Gerhard Tröster. *Wearable EOG goggles: eye-based interaction in everyday environments.* ACM, 2009.

[5] Georg Buscher, Andreas Dengel, and Ludger van Elst. Eye movements as implicit relevance feedback. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 2991–2996, New York, NY, USA, 2008. ACM.

[6] Christopher S. Campbell and Paul P. Maglio. A robust algorithm for reading detection. In *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, PUI '01, pages 1–7, New York, NY, USA, 2001. ACM.

[7] Andrew T. Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments and Computers*, 34(4):455–470, 2002.

[8] Mary C Dyson and Mark Haselgrove. The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies*, 54(4):585–612, 2001.

[9] Joseph H. Goldberg and Xerxes P. Kotval. Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24(6):631–645, 1999.

[10] U. Grenander. *Probability and Statistics: The Harald Cramér Volume*. Wiley Publications in Statistics. Almqvist & Wiksell, 1959.

[11] Abdul J Jerri. The shannon sampling theorem—its various extensions and applications: A tutorial review. *Proceedings of the IEEE*, 65(11):1565–1596, 1977.

[12] Moritz Kassner, William Patera, and Andreas Bulling. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 1151–1160. ACM, 2014.

[13] Sepp Kollmorgen and Kenneth Holmqvist. Automatically detecting reading in eye tracking data. *Lund University Cognitive Studies*, 2007.

[14] Kai Kunze, Yuzuko Utsumi, Yuki Shiga, Koichi Kise, and Andreas Bulling. I know what you are reading: Recognition of document types using mobile eye tracking. In *Proceedings of the 2013 International Symposium on Wearable Computers*, ISWC '13, pages 113–116, New York, NY, USA, 2013. ACM.

[15] Choongkil Lee. Eye and head coordination in reading: roles of head movement and cognitive control. *Vision Research*, 39(22):3761–3768, 1999.

[16] Steve Mann. Wearable computing: A first step toward personal imaging. *Computer*, 30(2):25–32, 1997.

[17] Steve Mann. Wearable computing as means for personal empowerment. In *Proc. 3rd Int. Conf. on Wearable Computing (ICWC)*, pages 51–59, 1998.

[18] George T. Pavlidis. Do eye movements hold the key to dyslexia? *Neuropsychologia*, 19(1):57–64, 1981.

[19] Johan Pouwelse, Koen Langendoen, and Henk Sips. Dynamic voltage scaling on a low-power microprocessor. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 251–259. ACM, 2001.

[20] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[21] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372, 1998.

[22] Keith Rayner, Barbara R Foorman, Charles A Perfetti, David Pesetsky, and Mark S Seidenberg. How psychological science informs the teaching of reading. *Psychological science in the public interest*, 2(2):31–74, 2001.

[23] Erik D. Reichle, Alexander Pollatsek, Donald L. Fisher, and Keith Rayner. Toward a model of eye movement control in reading. *Psychological review*, 105(1):125, 1998.

[24] Claude E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[25] Thad Starner. The challenges of wearable computing: Part 1. *IEEE Micro*, (4):44–52, 2001.

[26] Ivan E Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764. ACM, 1968.