

Detecção de padrões de leitura com baixa taxa de amostragem

Aluno: Carlos Eduardo Leão Elmadjian¹
Orientador: Prof. Dr. Carlos Hitoshi Morimoto

¹ elmad@ime.usp.br



USP

1. Objetivos

Aplicações para computação vestível estão, potencialmente, sempre em funcionamento para auxiliar o usuário em tarefas cotidianas, porém esse estado de alerta constante pode ocasionar um alto consumo de energia e, ao mesmo tempo, prejudicar o usuário com interrupções indesejáveis. Dessa forma, é necessário que essas aplicações demandem poucos recursos da máquina e façam uso de contexto para melhorar a qualidade das interações.

Neste trabalho, exploramos como o reconhecimento de padrões de leitura pode prover contexto para aplicações e quais abordagens algorítmicas podem ser empregadas para preservar energia por meio de soluções que trabalhem com baixa taxa de amostragem.

2. Características da leitura

A leitura pode ser caracterizada como uma alternância entre sacadas (movimentos rápidos do olho) e fixações de 200 a 300 ms, sendo as sacadas tipicamente curtas e feitas à direita (**Figura 1**). Ao final de uma linha de texto, os olhos se movem rapidamente para o começo da linha seguinte, executando um movimento denominado *regressão* [3].

Durante as sacadas, o processamento visual é interrompido. Assim, é durante as fixações que o cérebro obtém informações de um texto. Como nossa visão é foveada, temos uma acuidade visual limitada a apenas algumas letras (**Figura 2**), sendo necessário fazer diversas fixações ao longo de uma linha para capturar a maior quantidade possível de informações.

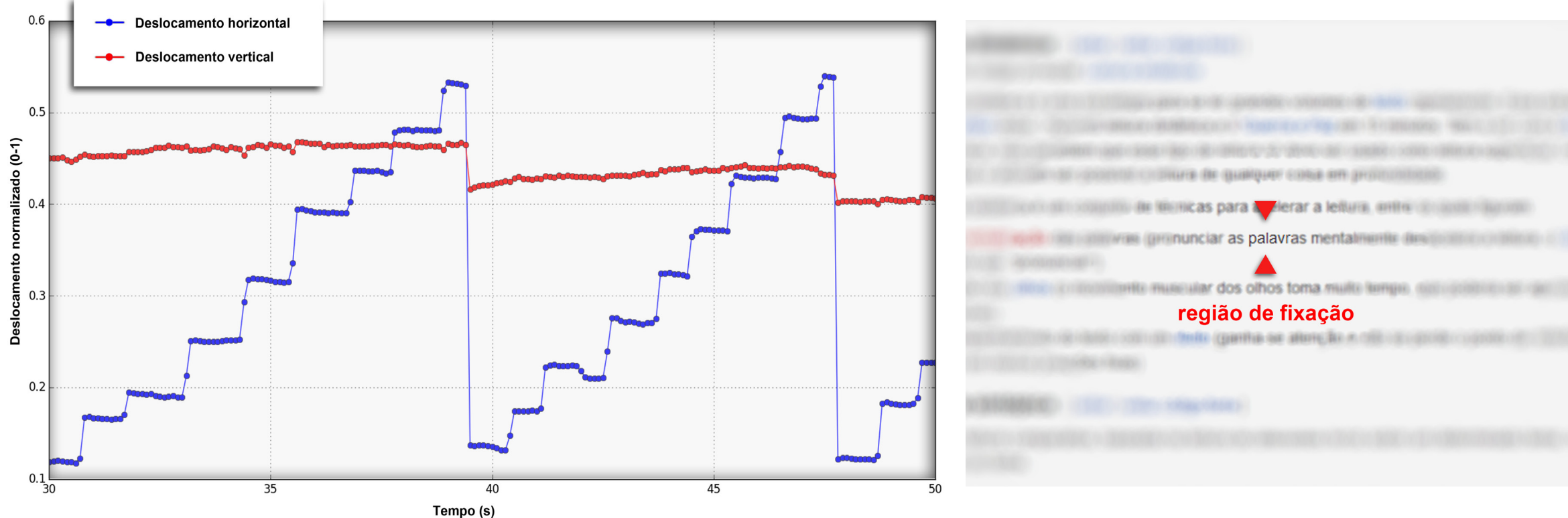


Figura 1. Padrão típico de leitura: os degraus formados pelo conjunto de pontos azuis indicam fixações ao longo de uma linha de texto, enquanto os saltos entre os degraus representam sacadas. Os pontos vermelhos, por outro lado, ressaltam a estabilidade vertical do olhar durante a leitura.

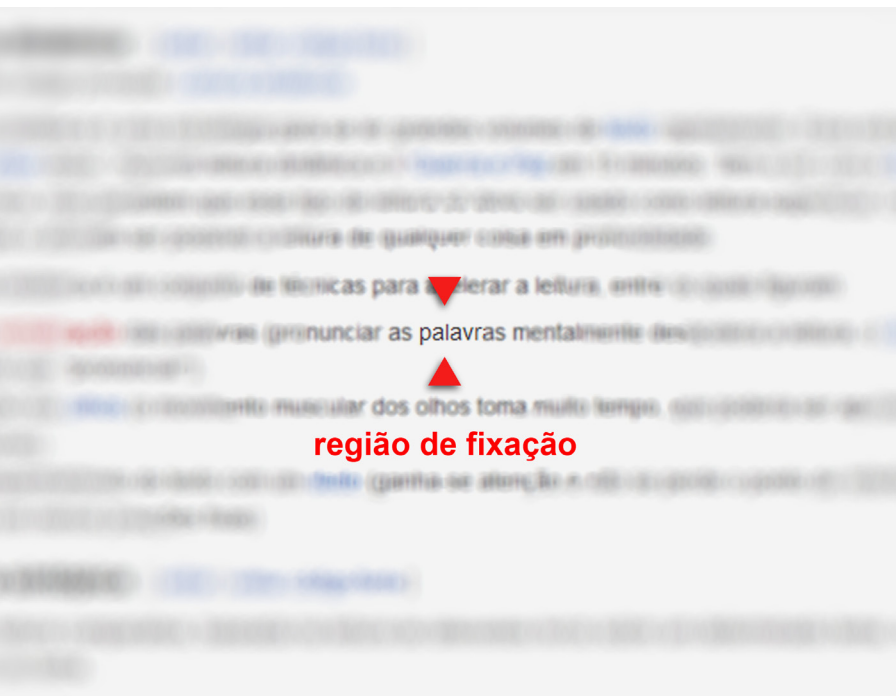


Figura 2. Representação artística da acuidade visual humana na leitura. Durante uma fixação, só é possível ter acuidade de 100% numa região muito limitada do texto (aproximadamente um grau).

3. O problema da baixa taxa de amostragem

Pelo Teorema de Nyquist-Shannon [4], todo sinal pode ser reconstruído a partir de um conjunto de amostras, desde que a taxa de amostragem seja de ao menos o dobro da sua frequência máxima. Caso contrário, sua reconstrução se torna ambígua, acarretando em um fenômeno conhecido como *aliasing*.

Para detectar a leitura, precisamos coletar dados sobre a posição da pupila de um indivíduo em relação à superfície observada. Isso é feito por meio de rastreadores de olhar, uma função de calibração e técnicas de visão computacional para processar imagens dos olhos. Contudo, com a diminuição da resolução dessas imagens e do número de amostras, a função de calibração perde acurácia (**Figura 3**) e ocorre o *aliasing* (**Figura 4**).

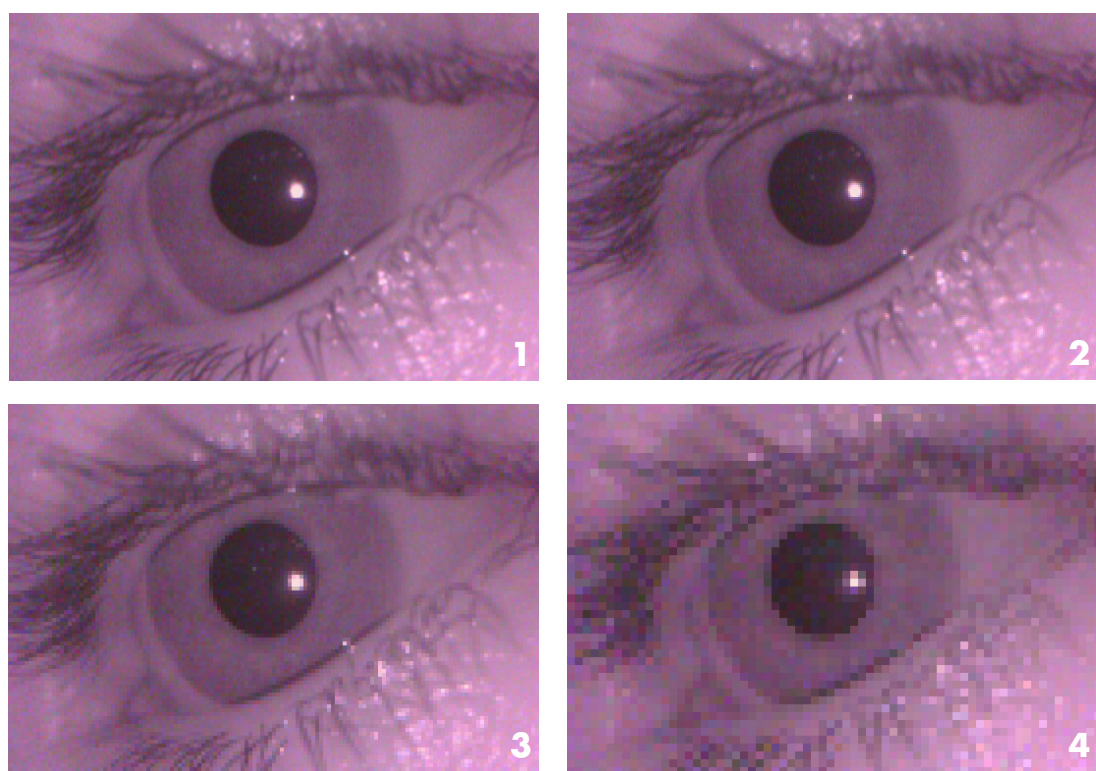


Figura 3. Imagem do olho capturada a 480 linhas (1) e subamostrada para 240 linhas (2), 120 linhas (3) e 60 linhas (4). Para os casos 3 e 4, o erro médio estimado do rastreador utilizado no projeto se mostrou superior a 10 graus (Aluani et al., comunicação pessoal).

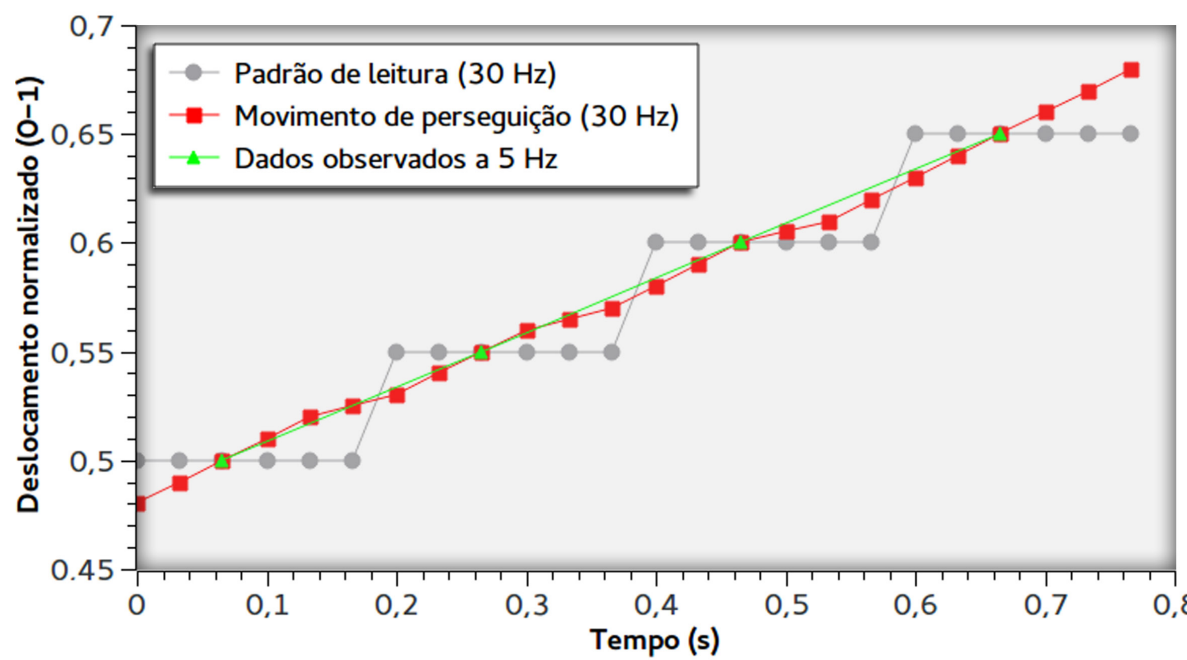


Figura 4. Exemplo de um possível caso de aliasing ao tentar reconstruir um sinal de leitura. O gráfico mostra que não é possível distinguir o padrão horizontal de leitura (cinza) de um movimento de perseguição lenta (vermelho) realizado na mesma direção a uma taxa de 5 Hz.

4. Algoritmo desenvolvido

Para tratar do problema da baixa taxa de amostragem, desenvolvemos um algoritmo que pudesse reconhecer padrões de leitura em qualquer faixa de frequência e sem reparametrizações. O algoritmo se baseou no princípio de *colecção de evidências*, proposto por outros trabalhos na literatura [1, 2], mas se distingue dos demais por não fazer uso da classificação dos movimentos do olhar em *tokens*. Para solucionar o problema, utilizamos o cálculo do quociente diferencial (**Figura 6**) entre duas amostras e atribuímos uma pontuação baseada em janelas de padrões válidos (**Figura 5**).

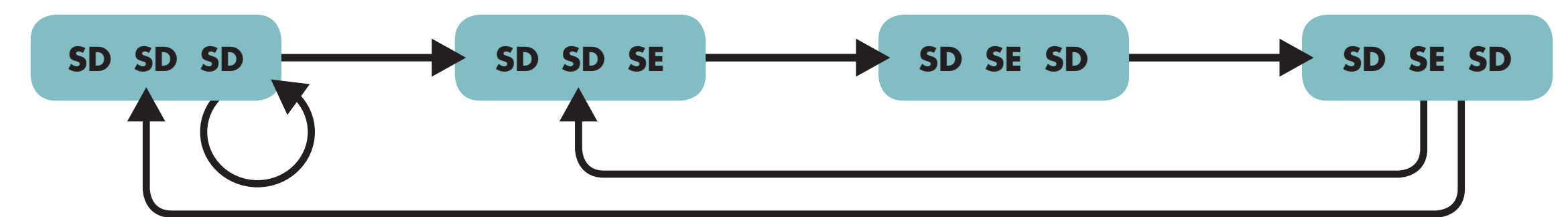


Figura 5. Diagrama de transição de estados do algoritmo proposto para uma janela de três eventos, que podem ser compostos por sacadas curtas à direita (SD) e sacadas de longa regressão à esquerda (SE).

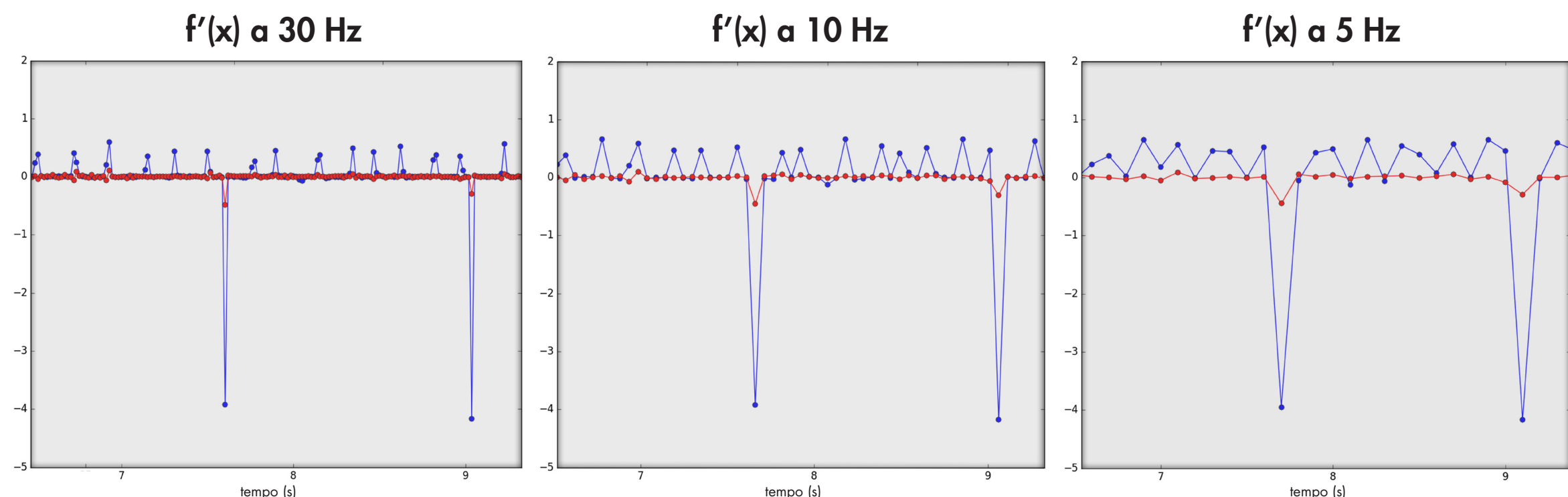


Figura 6. Três trechos de leitura de um usuário amostrados em diferentes frequências e transformados pelo quociente diferencial. Os pontos vermelhos indicam sacadas na vertical, enquanto os azuis, na horizontal. Observe que, empregando esta técnica, é possível limiarizar as sacadas correspondentes à leitura da mesma maneira para taxas de amostragem distintas, bastando que o denominador h seja independente do tempo no cálculo de $f'(x)$, dado por:

$$f'(x) = \frac{f(x+h) - f(x)}{h} \quad h > 0$$

5. Testes e metodologia

Para aferir o desempenho do algoritmo desenvolvido, coletamos dados de nove indivíduos. A cada participante, foi exibida uma sequência aleatória de cinco textos curtos (entre 120 e 170 palavras) e cinco imagens. No caso de um texto, o usuário deveria lê-lo por inteiro e depois responder um questionário para verificar seu grau de compreensão. No caso de uma figura, o usuário deveria contar os objetos presentes.

Os dados foram coletados a 30 Hz e depois subamostrados para 10 Hz e 5 Hz, totalizando 45 amostras de leitura e 45 de controle para cada faixa de frequência. Foram selecionados para um estudo comparativo os algoritmos de Buscher et al. [1] e Campbell e Maglio [2].

6. Resultados e discussão

Foram avaliadas a sensibilidade (*hit rate*) e a especificidade dos algoritmos (**Figura 7**). Os resultados mostram que o algoritmo proposto neste trabalho teve um desempenho estável e consistente em todas as frequências analisadas e especialmente superior aos demais a 5 Hz. Uma explicação para isso está no fato de os outros algoritmos dependerem da filtragem de uma grande quantidade de pontos.

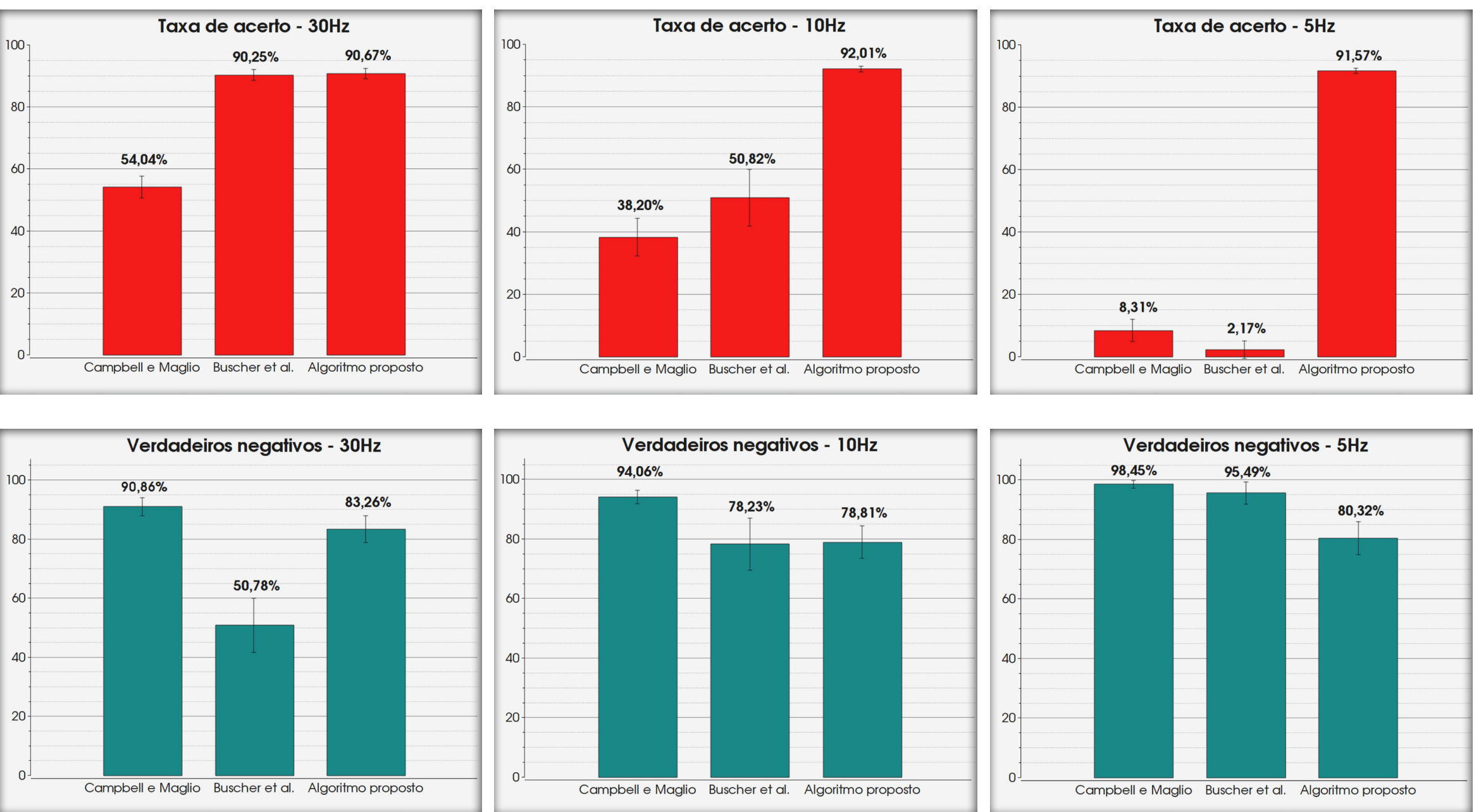


Figura 7. Os três primeiros gráficos (vermelho) mostram a taxa de acerto dos três algoritmos para dados reconhecidamente de leitura nas frequências estudadas. Os três últimos (verde) indicam a capacidade de reconhecer que uma entrada não é de leitura. O intervalo de confiança para todas as médias acima é de 95%.

7. Possíveis aplicações

A leitura é uma das principais formas de interação do usuário com sistemas gráficos e possivelmente aquela em que se dispense mais tempo. Mais do que isso, a leitura frequentemente é a tarefa primária do usuário e, portanto, reconhecê-la é um passo essencial para criar aplicações não invasivas e que façam uso de contexto.

Como proposta para o uso do algoritmo desenvolvido, criamos uma prova de conceito que reconhece o momento em que o usuário está lendo para prover ferramentas de auxílio (tradução de termos, *bookmarks* e autorrolagem) e, simultaneamente, viabilizar um canal de comunicação com o sistema para bloquear notificações indesejadas (**Figura 8**).

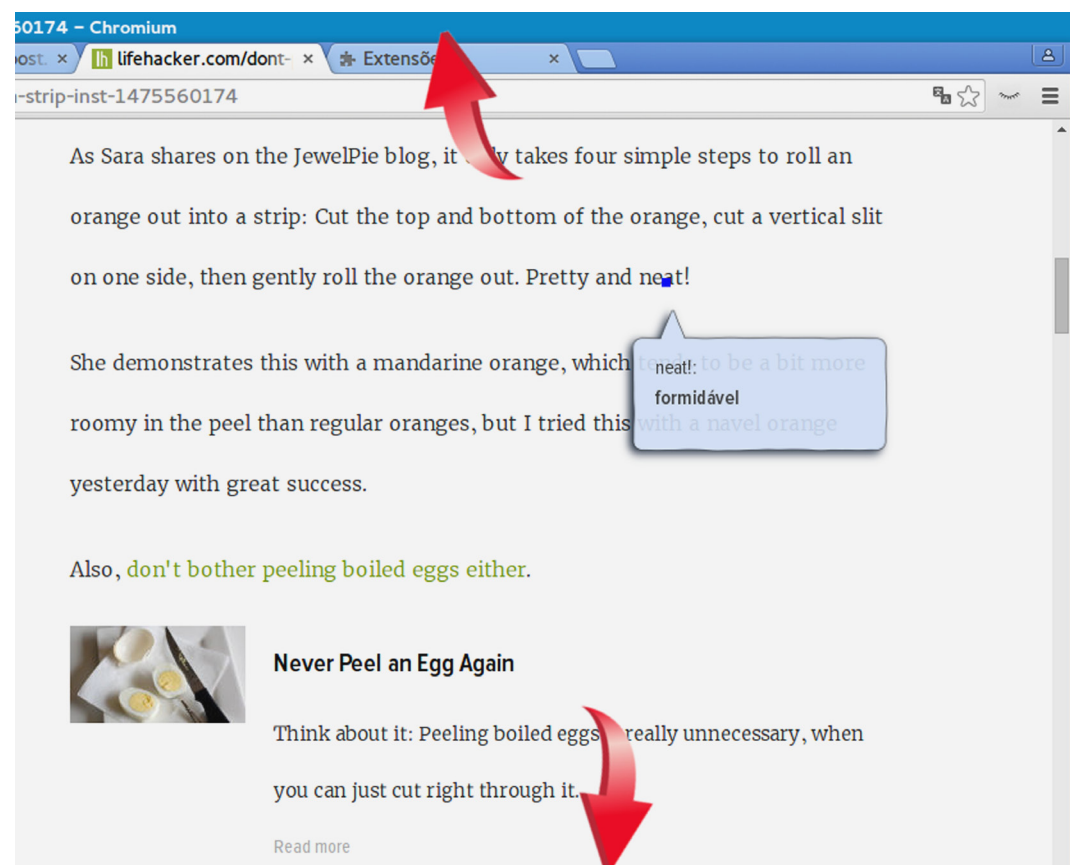


Figura 8. Exemplo de aplicação para detecção de leitura.

8. Referências

- [1] 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.
- [2] 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.
- [3] Keith Rayner. *Eye movements in reading and information processing: 20 years of research*. Psychological bulletin, 124(3):372, 1998.
- [4] Claude E. Shannon. *Communication in the presence of noise*. Proceedings of the IRE, 37(1):10-21, 1949.