

UMA INTERFACE HUMANO-MÁQUINA INTELIGENTE BASEADA NO RASTREAMENTO OCULAR PARA COMUNICAÇÃO ESCRITA DE PACIENTES COM SÍNDROME *LOCKED-IN*

AMANDA LEONEL, FERNANDO BUARQUE DE LIMA NETO, SÉRGIO CAMPELLO OLIVEIRA, HUGO SERRANO B. FILHO

*Departamento de Sistemas e Computação, Escola Politécnica de Pernambuco, Universidade de Pernambuco
Rua Benfca, 455 – Madalena, 50.720-001 – Recife, Pernambuco – Brasil
E-mails: {aln, fbln, scampello, hsbf}@dsc.upe.br*

Abstract— Patients with ‘Locked-in Syndrome’ (LIS) report a rare and severe clinical case as they become paralyzed and voiceless. Due to a brain damage, patients lose their movements of the voluntary muscles of the body and their ability to speak. Surprisingly, their cognitive functions remain in perfect order. In general, the only voluntary movement that still remains is the control of the eyes. This work presents a proof-of-concept and algorithm based on an Intelligent Human-Computer for writing Interface to LIS patients. The system uses techniques from Computer Vision and Artificial Intelligence in order to acquire and detect eye movements affording written communications of LIS patients. The design of the system is composed of modules which results were very satisfactory. The processed images show representative patterns at low spatial resolution. Artificial Neural Networks were trained in order to recognize the patterns analyzed. The high rates of success and short duration of the experiments demonstrate the feasibility of the proposed interface model.

Keywords— Human-Computer Interfaces, Paralysis, Locked-in Syndrome, Eye-tracking, Artificial Neural Networks.

Resumo— Pacientes portadores da Síndrome *Locked-in* apresentam um quadro clínico de paralisia rara e grave. Devido a um dano cerebral, eles perdem os movimentos dos músculos voluntários do corpo e a capacidade de falar. Surpreendentemente, suas funções cognitivas mantêm-se em perfeito estado, i.e. quando acordados, eles podem ver, ouvir e compreender tudo à sua volta. Em geral, o único movimento voluntário que ainda se mantém é o do controle dos olhos. Este trabalho apresenta uma prova de conceito baseada em uma Interface Humano-Máquina inteligente para comunicação escrita de pacientes com Síndrome *Locked-in*. Propomos a utilização de técnicas de Visão Computacional e Inteligência Artificial, a fim de mapear o rastreamento dos movimentos dos olhos em escrita na tela do computador. A modelagem do sistema é composta por módulos cujos resultados obtidos foram bastante satisfatórios. As imagens processadas apresentaram padrões representativos em baixa resolução espacial. Redes Neurais Artificiais foram treinadas para reconhecimento dos padrões analisados. As altas taxas de acerto e curta duração dos experimentos comprovaram a viabilidade do modelo de interface proposto.

Palavras-chave— Interfaces Humano-Máquina, Paralisia, Síndrome *Locked-in*, Rastreamento Ocular, Redes Neurais Artificiais.

1 Introdução

Uma interface para usuário é um grupo de recursos com o qual o usuário interage com um sistema. O termo "Interface Humano-Máquina" (IHM), do inglês "Human-Computer Interface" (HCI), é mais usado no contexto de Engenharia e Computação [1].

A fim de melhorar o desempenho de HCIs, técnicas de Computação como Reconhecimento de Padrões, Inteligência Artificial e Engenharia de Software são frequentemente aplicadas no estado da arte.

Novas tendências e potenciais grupos de usuários de sistemas de computacionais surgem constantemente. Em particular, HCIs podem ser utilizadas para beneficiar pacientes com capacidade de comunicação restrita ou comprometida. É o caso dos pacientes com Síndrome *Locked-in* [2] (do português, "Síndrome do Encarceramento").

A Síndrome *Locked-in* (LIS) é uma condição rara em que um paciente humano apresenta os músculos voluntários paralisados (devido a um acidente vascular cerebral grave, por exemplo). Porém, mantém todas as funções cognitivas em perfeito estado, i.e., quando acordado, pode ver, ouvir e compreender tudo à sua volta. Tais pacientes necessitam da utilização de uma HCI específica, de requisitos especiais e restritos, para se comunicar. Em geral, quando existentes, essas HCIs apresentam alto custo financeiro,

inviabilizando, assim, sua utilização na reabilitação de muitos casos.

Este trabalho tem como principal objetivo modelar e implementar uma HCI inteligente para comunicação escrita, baseada no rastreamento de movimentos oculares, utilizando Redes Neurais Artificiais e um sistema de baixo custo financeiro e computacional. A meta principal é construir uma prova de conceito a partir da modelagem e implementação de uma HCI que possibilite a comunicação de pacientes com Síndrome *Locked-in* (ou casos de paralisia semelhantes).

Este artigo é organizado conforme a seguir: a seção 2 reúne os principais conceitos necessários para a fundamentação teórica e compreensão deste trabalho; o modelo proposto é apresentado na seção 3; a seção 4 ilustra os experimentos e principais resultados; e, na seção 5, são apresentadas as conclusões e sugestões de trabalhos futuros.

2 Fundamentação Teórica

2.1 Síndrome *Locked-in*

Segundo a "Organização Mundial da Saúde" (OMS), síndrome é "um agrupamento de sinais e sintomas com base em sua frequente co-ocorrência, que pode sugerir uma patogênese básica, curso, padrão familiar ou tratamento comuns" [3]. Isto é, uma síndrome não caracteriza necessariamente uma única doença, mas

um grupo de doenças semelhantes em determinado aspecto.

A Síndrome *Locked-in* foi definida em 1966, mas desde 1986 passou a ser descrita como qualquer “quadriplegia e anartria com preservação da consciência” [2]. “Quadriplegia” porque o paciente apresenta paralisia de aproximadamente todos os músculos voluntários do corpo. Nos casos clássicos de LIS, o paciente possui apenas movimentos oculares. “Anartria” devido à impossibilidade de articular palavras. “Consciência preservada” uma vez que o paciente mantém todas as suas funções cognitivas, i.e., percepção (e.g. visão e audição), atenção, habilidades intelectuais, memória visual e verbal etc.

A principal causa de quadros de LIS são danos em uma região cerebral denominada de “ponte ventral”. A ponte é uma estrutura do tronco encefálico localizada entre o encéfalo e a medula espinhal no sistema nervoso central humano e responsável pelo envio de informação sensorial ao cérebro [4].

Em particular, o caso do jornalista francês Jean-Dominique Bauby tornou a síndrome conhecida ao escrever um livro [5] sobre suas memórias e seu estado psicológico durante o “encarceramento”. Diariamente, uma funcionária da sua editora ditava-lhe o alfabeto repetidas vezes, e apenas com o piscar da pálpebra esquerda, ele sinalizava a letra escolhida para compor cada palavra, frase e capítulo do seu livro.

2.2 O olho humano

Em pacientes portadores de LIS clássica, os únicos movimentos voluntários restantes são, em geral, os dos olhos. O olho humano é composto por diversas estruturas, entre elas: pupila, íris, córnea, cristalino, retina e nervo óptico [6].

O globo ocular apresenta movimentos em torno de seu eixo. Os quatro principais tipos de movimentos do olho humano são [7]:

- Sacádicos (*saccades*): movimentos bruscos e rápidos. Na leitura, ocorrem pequenos sacádicos na mudança de olhar de uma palavra para outra.
- Busca suave: para rastrear estímulos visuais de movimento, mantendo a linha da visão no objeto selecionado durante os intervalos entre os sacádicos;
- Fixação: para fixar estímulos visuais estacionários.
- Piscado: dura em média 0,3 a 0,4 segundos. Durante a leitura, um adulto pode piscar apenas 3 a 4 vezes por minuto.

2.3 Aspectos de Processamento Digital de Imagem

Os movimentos oculares podem ser captados por câmeras comuns, que geram vídeos (ou sucessão de imagens) em diversos formatos e taxas.

Uma imagem digital é a materialização e base de grande parte dos processos que envolvem áreas de Processamento Digital de Imagem (PDI), Visão

Computacional e Segmentação de Imagem [8], aplicadas neste trabalho.

Na representação matricial de uma imagem, cada ponto da matriz-imagem é chamado de elemento da imagem ou *pixel* e possui uma cor (valor) associada a ele. A resolução espacial, i.e. o número de linhas e colunas da matriz, representa as dimensões de uma imagem.

Operações com imagens desempenham um importante papel, pois podem proporcionar mudanças das características de uma imagem, para melhoria ou busca de informações. As operações no espaço de imagem utilizando filtros são chamadas de filtragem.

2.4 Redes Neurais Artificiais

Redes Neurais Artificiais (RNAs) [9] podem se relacionar com áreas como Visão Computacional, possibilitando o reconhecimento e classificação de padrões ilustrados em imagens.

RNAs são um conjunto de meta-heurísticas de Inteligência Artificial inspiradas e fundamentadas em diversas disciplinas. Em RNA, um conjunto de unidades simples de processamento – i.e. neurônios – são conectadas em rede, que passa a apresentar habilidades de processamento não-linear, geralmente com potencialidade de resolver problemas computacionais complexos.

Existem incontáveis modelos, arquiteturas e algoritmos de aprendizagem para RNAs. Os mais conhecidos são, de longe, as redes *Perceptron* e *MLP – Multilayer Perceptron* – treinadas com o algoritmo de treinamento *Backpropagation* [9].

As decisões de projeto de uma RNA ficam a cargo do projetista da rede e dependem bastante dos problemas a serem resolvidos. Contudo, algumas fases devem ser seguidas cuidadosamente: (i) coleta e seleção de dados; (ii) configuração da rede; (iii) treinamento; (iv) validação; e (v) teste.

2.5 Trabalhos relacionados

Existem incontáveis produtos comerciais de HCIs com características técnicas que dispensam o contato físico do usuário para interação. Contudo, apenas poucos deles podem ser usados por pacientes com LIS. Em geral, tais equipamentos são caros, e os requisitos de acessibilidade, usabilidade e segurança demandados pelos pacientes são bastante específicos, devido às suas limitações de interação.

O rastreamento dos movimentos dos olhos tem sido a abordagem mais promissora para casos de LIS clássica, uma vez que é possível ser desenvolvida a baixo custo [10]. Nessa linha, alguns trabalhos têm sido desenvolvidos com o uso de algoritmos convencionais e técnicas clássicas de Visão Computacional, ou de técnicas alternativas de Inteligência Artificial, (tais como RNAs), e alguns poucos apresentam um sistema para comunicação escrita [11].

3 O modelo proposto

O modelo proposto é baseado no rastreamento do movimento dos olhos, através de técnicas de Processamento Digital de Imagem e Inteligência Artificial, para possibilitar a comunicação escrita de pacientes portadores de LIS, via computadores pessoais. Uma visão geral do sistema e de seus módulos é apresentada na Figura 1 e descrita a seguir.

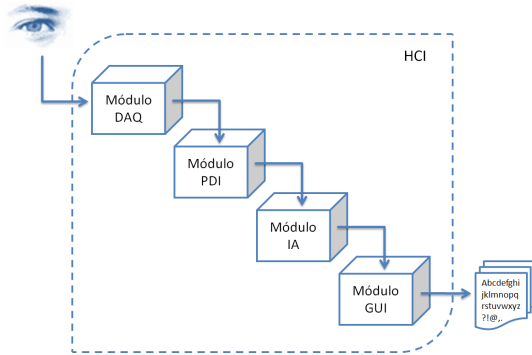


Figura 1. Estrutura geral e organização dos módulos da HCI proposta.

3.1 Módulo GUI: Interface Gráfica do Usuário

O módulo GUI apresenta uma interface gráfica do usuário para mapeamento visual dos padrões identificados pelos módulos anteriores em caracteres para escrita.

Genericamente, o módulo GUI pode ser descrito como um *grid* de dimensões 4×7 , em resolução de aproximadamente $1024 \times 768 \text{ pixels}$, contendo 28 quadrados e um campo superior retangular, conforme a Figura 2. Os quadrados definem as classes utilizadas para mapeamento dos movimentos oculares e devem conter caracteres do alfabeto português entre outros, que podem ser selecionados pelo usuário para escrita. O campo superior retangular é reservado para exibição dos caracteres selecionados, ao final de cada ciclo de processamento da HCI, conforme a Figura 1.

1x1	1x2	1x3	1x4	1x5	1x6	1x7	
2x1	2x2	2x3	2x4	2x5	2x6	2x7	
3x1	3x2	3x3	3x4	3x5	3x6	3x7	
4x1	4x2	4x3	4x4	4x5	4x6	4x7	

Figura 2. Disposição das classes na interface gráfica do usuário.

3.2 Módulo DAQ: Aquisição de Dados

Um módulo DAQ tem a finalidade de adquirir dados de entrada para processamento em um sistema. Em geral, possui sensores apropriados que monitoram os sinais a serem adquiridos.

Neste modelo, os sinais de entrada são representados por imagens de movimentos dos olhos. O "dispositivo de rastreamento de movimentos oculares

baseado em *webcam* e iluminação com LED infravermelho", desenvolvido por T. Vieira e E. Fontana [10] foi utilizado para a captação e geração da base de imagens necessária para a prova de conceito do sistema proposto.

3.3 Módulo PDI: Processamento Digital de Imagem

O módulo PDI aplica técnicas de Processamento Digital de Imagem, a fim de extrair das imagens os dados necessários para reconhecimento de padrões no módulo IA.

A ferramenta utilizada para processamento de imagens foi o *Matlab*® versão 7 (R2006a), em particular, seu complemento "*Image Processing Toolbox*" versão 5.2 (R2006a).

A sequência interna de processamentos do módulo PDI é ordenada a seguir:

- i. Extração e classificação dos quadros;
- ii. Pré-processamento das imagens: rotação, redução da resolução espacial e conversão para tons de cinza;
- iii. Segmentação da pupila: detecção de bordas, binarização e dilatação;
- iv. Redução da densidade de resolução;
- v. Conversão binária.

3.4 Módulo IA: Inteligência Artificial

O módulo IA define a arquitetura e os parâmetros utilizados na aplicação de Redes Neurais Artificiais para reconhecimento dos padrões direcionais analisados. O modelo de RNA escolhido foi a arquitetura *Multilayer Perceptron*, por ser um classificador neural bastante conhecido e utilizado.

Para realizar a configuração, treinamento, validação e teste da MLP, foi utilizado o módulo *Experimenter* da ferramenta *WEKA*® versão 3 – um pacote de *softwares* com algoritmos de Aprendizagem de Máquina em linguagem de programação Java.

4 Experimentos e Resultados

Os experimentos e resultados de cada módulo são descritos a seguir.

4.1 Resultados do módulo GUI

A interface gráfica do usuário foi desenvolvida em linguagem de programação Java, para (i) mapear visualmente as classes (quadrados) em caracteres e (ii) e exibir as saídas do módulo IA em formato de texto acima da tela.

O mapeamento é ilustrado no *grid* da Figura 3. O caractere '<' representa a operação de *backspace* em um teclado normal; e o quadrado em branco, um espaço em branco. O símbolo '@' foi adicionado após relato de um paciente, solicitando que incluíssem o sinal, para que pudesse escrever endereços de e-mail ao se comunicar.

Como resultado do módulo GUI, foi desenvolvido também um pequeno protocolo de comunicação

para utilização do sistema. O usuário deve confirmar o caractere desejado olhando fixamente para seu quadrado e piscando por aproximadamente 1 segundo. À taxa de captação de vídeo de 10 quadros/segundo, a detecção consecutiva de mais de 10 imagens classificadas como “piscado” efetua a seleção do último sinal de imagem não classificado como “piscado”, para escrita desta classe. Essa abordagem reduz os erros de escrita accidental, decorrentes do movimento natural de piscado do olho humano.

A	B	C	D	E	F	G	
H	I	J	L	M	N	O	
P	Q	R	S	T	U	V	
X	Z	@	?	<	.		

Figura 3. Disposição final dos caracteres na interface gráfica.

4.2 Resultados do módulo DAQ

A obtenção dos vídeos deu-se da seguinte forma:

- O dispositivo selecionado e em uso foi acoplado devidamente à cabeça do usuário, que se manteve fixa (assim como um paciente portador de LIS clássica), à distância de aproximadamente 30 cm da tela do computador.
- Solicitou-se ao usuário que observasse quadrado a quadrado do *grid*, pausada e sucessivamente, percorrendo linha a linha até o final do *grid*.
- Em cada quadrado, os olhos do usuário realizaram movimentos de busca suave na área interna do quadrado, permitindo maior variedade dos dados para a RNA.
- Entre quadrados, os olhos realizaram movimentos sacádicos ou de piscado – este último, sinalizando explicitamente a troca de quadrado.

Por fim, os vídeos obtidos apresentam as seguintes características técnicas exibidas na Tabela 1.

Tabela 1. Características técnicas dos vídeos gerados.

PARÂMETROS	VALORES
Formato do arquivo	.avi
Compressão de vídeo	nenhuma
Número de quadros/segundo	30
Tipo da imagem	truecolor (RGB)
Resolução vertical (altura)	480 pixels
Resolução horizontal (largura)	680 pixels

4.3 Resultados do módulo PDI

A (i) extração e classificação dos quadros dos vídeos gerados resultaram na base de imagens ‘Alfabeto’, cujas características técnicas são listadas na Tabela 2. Além das 28 classes correspondentes ao *grid* 4 x 7 da interface gráfica, 2 classes extras também foram criadas: ‘piscado’ e ‘sacádicos longos’ – esta última a partir de imagens de desvios longos, como quebras de linha. A Figura 4 ilustra amostras da base gerada.

Tabela 2 Características técnicas da base de imagens “Alfabeto”.

PARÂMETROS	VALORES
Formato de imagem	.bmp
Compressão de imagem	nenhuma
Tipo da imagem	truecolor (RGB)
Resolução vertical (altura)	480 pixels
Resolução horizontal (largura)	680 pixels
Número de classes	30
Número total de imagens (aprox.)	5.040
Número médio de imagens por classe	168



Figura 4. Amostras das imagens originais de diferentes classes da base ‘Alfabeto’.

Após as fases de (ii-iv), as imagens finais obtidas apresentaram-se binarizadas (i.e. *pixel* 1 para branco, *pixel* 0 para preto) e com resolução espacial de 13 x 9 pixels. Amostras finais de diferentes classes são ilustradas na Figura 5.



Figura 5. Amostras das imagens finais de diferentes classes do módulo PDI.

A última fase (v) resultou na conversão das imagens finais em uma base binária ‘alfabeto.arff’, para treinamento, validação e teste da RNA do módulo IA no WEKA®. Cada *pixel* da imagem representa um atributo de instância da base binária. A última base do módulo PDI é descrita na Tabela 3.

Tabela 3. Características técnicas da base binária ‘alfabeto.arff’

PARÂMETROS	VALORES
Nome da base (<i>relation</i>)	alfabeto
Formato da base	.arff
Número de atributos	117 (13 x 9) pixels
Número de classes	29 (4x7 + piscado)
Número de instâncias	300 imagens
Ordenação das instâncias	Não (aleatória)

4.4 Resultados do módulo IA

As entradas (imagens) e saídas desejadas (classes) foram apresentadas à rede durante o treinamento, afim de ajustar os pesos e encontrar uma relação entre os pares entrada-saída, para generalização de novos dados.

A seleção da arquitetura e parâmetros da rede MLP empregada foi baseada nos resultados obtidos *ad hoc* das fases de treinamento, validação e teste das MLPs estudadas. A melhor configuração de parâme-

tros obtida para a RNA, utilizando a base 'alfabeto.arff', é listada na Tabela 4.

Tabela 4. Melhor configuração de parâmetros obtidos para a MLP.

PARÂMETROS	VALORES
Dados da base para:	
• treinamento	230
• validação	40
• teste	30
Neurônios da camada de entrada	117 (pixels)
Neurônios da camada de saída	29 (classes)
Neurônios da camada escondida	29 (prop. à saída)
Taxa de aprendizado	0,3
<i>Momentum</i>	0,2
Épocas de treinamento	60 (máximo)
<i>Threshold</i> de validação	10
Repetições dos experimentos	30 (vezes)

A Tabela 5 ilustra os resultados obtidos a partir de 30 simulações para essa rede.

Tabela 5. Resultados da melhor rede MLP configurada.

	Média	Desvio padrão	Mediana
Número de acertos	29,13	1,06	29,00
Porcentagem de acerto	96,13	3,53	96,66
Número de erros	1,17	1,06	0,00
Porcentagem de erro	3,86	3,53	0,00
Erro Absoluto Médio (MAE)	0,02	0,00	0,02
Raíz do Erro Quadrático Médio (RMSE)	0,06	0,00	0,06
Erro Absoluto Relativo (RAE)	30,28	2,97	30,09
Raíz do Erro Quadrático Relativo (RRSE)	36,35	4,80	36,12
Duração do treinamento (s)	6,46	3,71	4,73

5 Conclusões

Este trabalho apresentou e validou com sucesso uma prova de conceito, propondo a implementação de uma Interface Humano-Máquina inteligente para comunicação escrita de pacientes com Síndrome *Locked-in*.

HCIs baseadas no rastreamento dos movimentos oculares podem ser uma alternativa de comunicação para muitos dos casos de LIS clássica. Os requisitos especiais desses pacientes tanto podem facilitar o projeto de tais sistemas, como tornar seu desenvolvimento uma tarefa peculiar.

Em especial, técnicas de Visão Computacional e Inteligência Artificial podem garantir HCIs dedicadas e robustas, até mesmo a um custo mais baixo. A combinação dessas técnicas não necessariamente precisa incluir algoritmos novos, em estado da arte, para garantir resultados satisfatórios.

As imagens finais obtidas apresentaram padrões representativos em baixa resolução espacial, dispensando a utilização de imagens de alta resolução ou coordenadas tridimensionais no globo ocular.

A utilização de Redes Neurais Artificiais teve grande importância no projeto, em particular, no reconhecimento dos diferentes padrões obtidos, possi-

bilizando adaptabilidade e flexibilidade na interação humano-máquina. As taxas de acertos das redes simuladas apresentaram resultados muito bons, próximos de 100% nos casos de teste final. E a curta duração dessas simulações implica na possibilidade de utilização, em tempo real, de tais redes ou até mesmo de re-treinamento se necessário.

Como trabalhos futuros, propõe-se (i) a adição de um módulo de calibração ao sistema; (ii) a configuração automática dos parâmetros dos módulos PDI e IA; (iii) a análise de precisão do sistema, a fim de incluir mais classes no *grid*; (iv) a execução de testes reais; e (v) a implementação de um produto final em linguagem de programação compilada, a fim de otimizar a integração e o custo computacional do sistema.

Agradecimentos

Os autores agradecem ao professor Eduardo Fontana e ao doutorando Tiago Vieira, do Grupo de Fotônica da UFPE, pela colaboração neste trabalho, possibilitando as gravações de dados realizadas através do dispositivo desenvolvido pelos mesmos [10].

Referências Bibliográficas

- [1] A. Sutcliffe, *Human-computer interface design*, Springer-Verlag, 1989.
- [2] E. Smith and M. Delargy, *Clinical Review: Locked-in Syndrome*, British Medical Journal - BMJ Publishing Group Ltd., 2005.
- [3] All and WHO, *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association, 1994.
- [4] E.R. Kandel, J. Schwartz, and T.M. Jessell, *Principles of Neural Science*, McGraw-Hill Medical, 2000.
- [5] J.D. Bauby, *O escafandro e a Borboleta*, São Paulo, Martins Fontes, 1997.
- [6] M.J. Hogan, J.A. Alvarado, and J.E. Weddell, *Histology of the human eye: an atlas and textbook*, WB Saunders Company, 1971.
- [7] R. Wilson, *Eye Movements and Visual Attention - The MIT Encyclopedia of the Cognitive Sciences (MITECS)*, The MIT Press, 2001.
- [8] J. Gomes and L. Velho, *Computação Gráfica: Imagem*, Rio de Janeiro, RJ: Instituto de Matemática Pura e Aplicada-IMPA, 2002.
- [9] S. Haykin, *Neural Networks: a comprehensive foundation*, Prentice Hall, 2008.
- [10] T.F. Vieira and E. Fontana, "Dispositivo de rastreamento de movimentos oculares baseado em webcam e iluminação com led infravermelho," *XXI Congresso Brasileiro de Engenharia Biomédica*, Bahia, Salvador: 2008, pp. 669-672.
- [11] Metrovision, "Visionboard," 1997, pp. 2-3. *Medical and Biological Engineering and Computing*, vol. 35, 416, 1997.