



## VIII Encontro de Iniciação Científica e Tecnológica

VIII EnICT

ISSN: 2526-6772

IFSP – Câmpus Araraquara

19 e 20 de outubro de 2023



### EYETALK: UMA FERRAMENTA PARA COMUNICAÇÃO ASSISTIDA COM PROCESSAMENTO DE LINGUAGEM NATURAL EM NUVEM.

ALAN GABRIEL PEREIRA DE CARVALHO<sup>1</sup>, LUIZ HENRIQUE NUNES<sup>2</sup>

<sup>1</sup>Discente do Curso Superior em Análise e Desenvolvimento de Sistemas, IFSP Câmpus Araraquara, [carvalho.alan@aluno.ifsp.edu.br](mailto:carvalho.alan@aluno.ifsp.edu.br).

<sup>2</sup>Docente no IFSP Câmpus Araraquara, [henriquenunes@ifsp.edu.br](mailto:henriquenunes@ifsp.edu.br).

**Área de conhecimento** (Tabela CNPq): Sistemas de Informação – 1.03.03.04-9

**RESUMO:** Estatísticas obtidas por meio da Pesquisa Nacional por Amostra de Domicílios, realizada no terceiro trimestre do ano de 2022, pelo IBGE, mostram que aproximadamente 18,6 milhões de indivíduos com 2 anos ou mais vivem com algum tipo de deficiência, o que corresponde a cerca de 8,9% da população. Este grupo abrange uma ampla gama de deficiências, abrangendo tanto as físicas quanto as intelectuais. O objetivo deste projeto é apresentar uma proposta de alternativa comunicacional para pessoas que possuem deficiência motora severa e ausência de fala, buscando utilizar tecnologias como processamento de imagem, reconhecimento e síntese de fala, bem como inteligência artificial.

**PALAVRAS-CHAVE:** chat gpt; comunicação alternativa; opencv; tecnologia assistiva; visão computacional.

### INTRODUÇÃO

Segundo informações do Ministério da Saúde, no Brasil, a cada mil crianças nascidas vivas, duas enfrentam algum grau de paralisia cerebral. Dentro desse grupo, 70% experimentam dificuldades na comunicação (TABITH, 1989), que podem incluir problemas na fala, na escrita e em outras formas de expressão. Estima-se ainda que, em países desenvolvidos, há ocorrência de 30 a 40 mil novos casos por ano, mas em países subdesenvolvidos a ocorrência pode ser ainda maior.

Diante dos desafios enfrentados por essa parcela da população, este projeto tem como objetivo facilitar a realização de tarefas para pessoas com deficiência motora severa que apresentam ausência total na fala. Ele busca promover maior independência em sua interação com o ambiente físico e socialmente, oferecendo um software que utiliza tecnologia de Visão Computacional e Inteligência Artificial. Isso possibilita a aplicação de uma Tecnologia Assistiva, criando uma alternativa viável para superar as barreiras de comunicação impostas pela condição.

Para tanto, o presente trabalho se vale das abordagens anteriormente aplicadas com objetivo em conjunção e compatibilidade com os mesmos, mas buscando otimização e melhorias em usabilidade. A título de exemplo, trabalhos que utilizam a detecção de piscadas de modo a captar interação do usuário, como Kanegae e Nunes (2020) e o subsequente trabalho de Martins e Nunes (2021), apresentam alto custo de tempo ao dar respostas, o que dificulta sua utilização, ainda que o processamento esteja otimizado por ser de processamento local e não online, como a última abordagem de Carvalho e Nunes (2022).

## FUNDAMENTAÇÃO TEÓRICA

De acordo com o Relatório Mundial Sobre a Deficiência, mais de um bilhão de pessoas, o que corresponde a cerca de 15% da população global (com base em estimativas de 2010), enfrentam algum tipo de deficiência. Essa cifra supera consideravelmente a estimativa anterior da Organização Mundial da Saúde (OMS) de cerca de 10% em 1970. Outros dados, como os provenientes da World Health Survey, sugerem que aproximadamente 785 milhões de pessoas com 15 anos ou mais vivem com alguma forma de deficiência, enquanto a Global Burden of Disease estima esse número em torno de 975 milhões de indivíduos. Só no Brasil, em 2022, 18,6 milhões de pessoas apresentavam algum tipo de deficiência.

Em vista disso, surgiram tecnologias assistivas que visam aprimorar a qualidade de vida das pessoas com deficiência e ampliar sua capacidade de interagir com o ambiente por meio da utilização de computadores, sensores e atuadores, conforme destacado por Nunes, Machado e Moraes em seu trabalho de 2014. Essas tecnologias têm como objetivo prevenir, compensar, aliviar ou neutralizar deficiências, incapacidades ou desvantagens, contribuindo significativamente para aumentar a autonomia e a qualidade de vida dos indivíduos, conforme definido pela Norma Internacional ISO 9999 para tecnologias assistivas.

Com o intuito de atender uma demanda específica desse grupo social, como a de pessoas com paralisia cerebral que possuam entraves comunicacionais, diversos trabalhos se utilizam de novas áreas tecnológicas, assim como o presente trabalho utiliza-se de Visão Computacional, Síntese e Reconhecimento da Fala aliado às recentes propostas de Inteligência Artificial, a citar o ChatGPT, que realiza o Processamento de Linguagem Natural necessário para identificação de perguntas e geração de respostas.

Para ajudar na condução de uma cadeira de rodas motorizada por pessoas com dificuldades nas suas habilidades motoras, visuais e cognitivas, Santana (2020), utilizou-se de tecnologias de Visão Computacional para que fosse realizado um processo que estimasse o ponto do olhar ou o movimento dos olhos em relação à cabeça. Assim, esse sistema de rastreamento de olhar (Gaze Tracking) mede a posição e o movimento das pupilas para que correspondam a direções em que a cadeira de rodas deveria ser guiada, como direita, esquerda e frente.

Tendo como proposta também ser uma Tecnologia Assistiva que utiliza tecnologias de Visão Computacional, mas com o intuito de oferecer uma alternativa para comunicação, o presente trabalho pretende atender pessoas sem fala, escrita funcional ou em defasagem entre sua necessidade comunicativa e sua habilidade em falar e/ou escrever.

Para essa finalidade, utiliza-se uma sequência de trabalhos com atualizações em relação ao antecessor, sendo a base de todos a ferramenta COSEM (KANEGAE; NUNES, 2020), com Visão Computacional no reconhecimento facial para detecção de piscadas, para que, posteriormente, uma sequência de piscadas fossem transformadas em ações para a interface de um sistema de seleção.

Retrabalhado por Carvalho e Nunes (2022), a abordagem, utilizando-se do sistema de identificação de piscadas, foi repensada para que seu processamento fosse realizado localmente, sem dependência da conexão com redes de internet e com maior estabilidade na execução, em contraposição ao trabalho de Martins e Nunes (2021), que também utilizou-se da ferramenta COSEM (KANEGAE; NUNES, 2020), mas com processamento online por intermédio de navegadores e tecnologias para sistemas web.

## METODOLOGIA

Foi realizada a análise de uma sequência de trabalhos relacionados por atualizações em relação aos antecessores. Em 2020, foi construído um protótipo de uma tecnologia assistiva intitulada COSEM (COmmunication System based on Eyelid Movement), a qual é responsável por prover um sistema de comunicação baseado no movimento das pálpebras (KANEGAE; NUNES, 2020), base para os demais trabalhos. Em 2021, houve atualizações neste trabalho já que foi desenvolvido um módulo responsável por transformar os comandos emitidos pelas piscadas em sons, possibilitando a integração com assistentes virtuais como a Alexa e ambientes inteligentes (MARTINS; NUNES, 2021). Por fim, em 2022 ocorreu a adaptação da interface gráfica do projeto para ambientes desktops, com o intuito de possibilitar sua utilização sem a dependência da Internet, mas sem integração com assistentes virtuais (CARVALHO; NUNES, 2022).

Entretanto, após a realização dos testes com a nova interface, notou-se que são efetuadas diversas piscadas para o usuário chegar a resposta desejada em uma determinada categoria, o que torna muito custoso no quesito tempo de seleção. Neste sentido, a fim de otimizar a utilização da ferramenta COSEM, o objetivo deste trabalho é desenvolver um módulo capaz de reconhecer perguntas feitas a um usuário por meio de inteligência artificial e fazer um novo módulo de interação com a interface utilizando ainda movimento dos olhos, mas em vez de somente piscadas, também movimento das pupilas, como esquerda e direita.

Também foram definidos novos papéis de usuários para o funcionamento do módulo. Agora, ele atende a dois tipos de perfis: usuário interlocutor e usuário principal.

O usuário interlocutor é o usuário que, de maneira oral, realiza perguntas e questionamentos de modo a ser respondido com alternativas seguindo a escala de Likert. Com sua ação de perguntar, inserindo a entrada necessária, inicia todo o processamento do módulo.

O usuário principal, por sua vez, é a pessoa com deficiência motora, portanto com dificuldade em utilizar sistemas interativos de forma convencional, e que também possui entraves na comunicação falada. Ele deve reagir a pergunta realizada pelo interlocutor utilizando o movimento das pupilas de modo a selecionar as respostas de sua preferência.

Inicialmente, o processo começa com a coleta da pergunta do interlocutor, como ilustrado na Figura 1. Essa etapa utiliza o reconhecimento de fala para transcrever a pergunta em forma de texto. Uma vez obtida a versão textual da pergunta, é chegada a etapa 2, onde o COSEM encaminha a pergunta ao modelo de geração de texto por meio do processamento de linguagem natural.

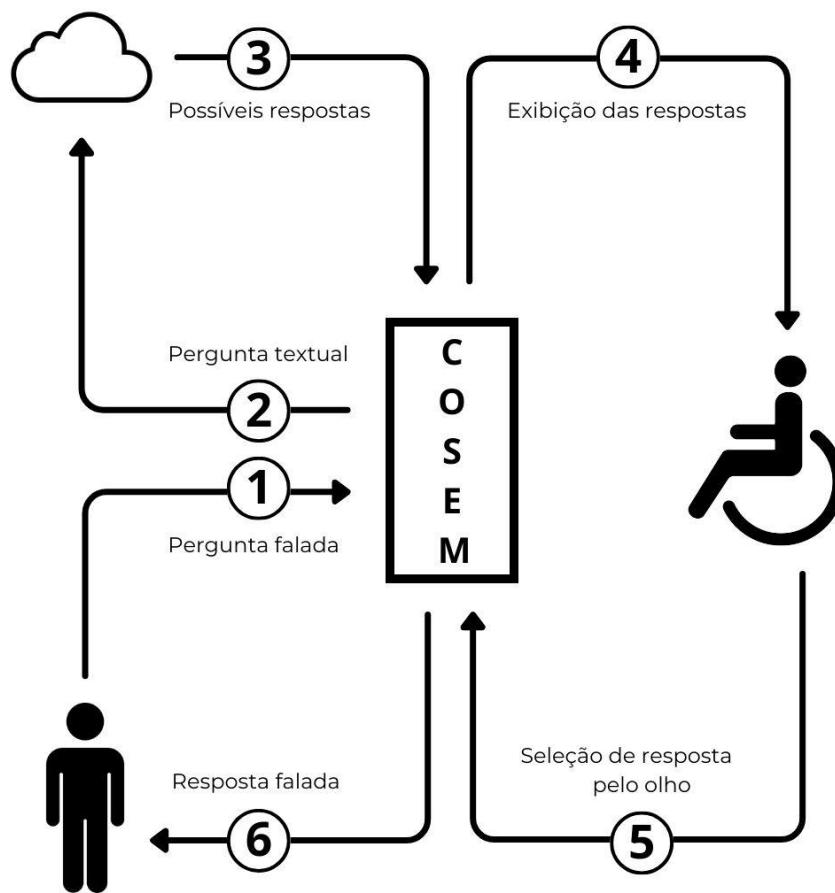
Nesta fase, o processamento de linguagem natural é aplicado para encontrar respostas que sejam compatíveis e coerentes com a pergunta oral. Além disso, a escala de Likert é utilizada para facilitar a geração de respostas e a seleção da resposta mais adequada pelo usuário principal.

Em seguida, na etapa 3, o modelo de chat retorna as possíveis respostas. As respostas são apresentadas uma por uma como opções selecionáveis, acessíveis por meio de um carrossel na interface gráfica, sendo parte da exibição de respostas ao usuário, conforme demonstrado na etapa 4.

A etapa 5 introduz uma nova forma de interação, permitindo que usuários com deficiência motora utilizem o movimento de pupilas (Gaze Tracking) para selecionar algumas das respostas disponíveis. O sistema de detecção de movimento das pupilas identifica as direções como esquerda ou direita e assim, amplia a acessibilidade do sistema permitindo uma forma de interação alternativa aos usuários com deficiência.

Por fim, na etapa 6, o módulo emite a resposta escolhida utilizando a síntese de fala, permitindo que o interlocutor a ouça. Este processo pode ser repetido ciclicamente, conforme necessário.

FIGURA 1. Processo realizado pelo novo módulo COSEM.



Fonte: Própria do autor.

Para o bom funcionamento do novo módulo, seria necessário renovação do sistema de interação por meio do movimento dos olhos. Assim sendo, foi realizada uma pesquisa exploratória sobre diferentes algoritmos e métodos de rastreamento ocular (eye tracking). A pesquisa foi usada para definir o novo módulo que deveria substituir o sistema de detecção de contração de pálpebras.

## RESULTADOS E DISCUSSÃO

Até o presente momento, como resultado deste projeto, pode-se destacar o desenvolvimento de um protótipo de baixa fidelidade do sistema. O protótipo representa a interface do módulo COSEM, subdividido em três áreas.

Na área 1, há a subdivisão onde o usuário vê a captura de seu vídeo em tempo real, o que permite a análise para possíveis ajustes a fim de melhorar a eficiência da detecção do movimento de pupilas. Já a área 2 é a área onde o usuário interlocutor visualiza o resultado do reconhecimento da sua fala e transcrição. A área 3, por sua vez, é responsável por retornar as possíveis respostas que a API do ChatGPT retornará para a pergunta da área 2, posicionando-as em uma sequência para que sejam acessadas por meio do movimento das pupilas para esquerda ou direita. Protótipo exemplificado na Figura 2, abaixo.

FIGURA 2. Protótipo de baixa fidelidade.



Fonte: Própria do autor.

## CONCLUSÕES

Este projeto apresentou um protótipo de baixa fidelidade com a proposta de possibilitar a Comunicação Aumentativa e Alternativa por meio do desenvolvimento de um novo software que permita a pessoas com dificuldade de fala devido a deficiência motora severa uma alternativa para interação com o ambiente físico de forma autônoma por meio do movimento de suas pupilas como forma de Tecnologia Assistiva.

Os próximos passos deste trabalho serão compostos pelo desenvolvimento do protótipo funcional e sua avaliação, para, como forma de trabalho futuro, utilizar Remote Procedure Call para fazer com que o módulo seja na forma distribuída servidor/cliente, e assim, faça todo processamento ser de responsabilidade da nuvem e não mais local.

## REFERÊNCIAS

Brasil, Ministério da Saúde. **Diretrizes de Atenção à Pessoa com Paralisia Cerebral**. Brasília, DF: Ministério da Saúde, 2013. 08 p.

KANEGAE, Yuri; NUNES, Luiz Henrique. **Proposta de aplicação para auxiliar a comunicação por meio do movimento de pálpebras**. [s. l.], 2020. Disponível em: <https://arq.ifsp.edu.br/eventos/index.php/enict/5EnICT/paper/view/498>. Acesso em: 07 set. 2023.

MARTINS, Renan; NUNES, Luiz Henrique. **Desenvolvimento de um software para auxiliar a comunicação por meio do movimento das pálpebras**. [s. l.], 2021. Disponível em: <https://arq.ifsp.edu.br/eventos/enict/6EnICT/paper/view/555/326>. Acesso em: 07 set. 2023.

NUNES, F. L. S.; MACHADO, L. S.; NIRAES, R. M. **Evolution of Virtual and Augmented Reality in Health: A Reflection from 15 Years of SVR**. IEEE, Piata Salvador, Brazil, p. 220-228, Maio 2014. ISSN 978-1-4799-4261-9

SANTANA, E. M. **Condução de cadeira de rodas usando rastreamento ocular por imagem**. Disponível em: <https://repositorio-dev.ufu.br/handle/123456789/30257>. Acesso em: 07 de set. 2023.

Santos RA, Da-Silva VR, dos-Santos JP, Siqueira AN. **Perfil epidemiológico e assistência à saúde de crianças e adolescentes com paralisia cerebral em um município do ES.** Resid Pediatr. 2019;9(3):252-260 DOI: 10.25060/residpediatr-2019.v9n3-10. Acesso em: 07 set. 2023.

TABITH, A. **Foniatria: disfonias, fissuras labiopalatais, paralisia cerebral.** Cortez Editora, São Paulo, caderno 5, p. 51-117, 1989.

**Brasil tem 18,6 milhões de pessoas com deficiência, indica pesquisa divulgada pelo IBGE e MDHC.** Disponível em:

<https://www.gov.br/mdh/pt-br/assuntos/noticias/2023/julho/brasil-tem-18-6-milhoes-de-pessoas-com-deficiencia-indica-pesquisa-divulgada-pelo-ibge-e-mdhc>.