

## 16º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2025

### RECONHECIMENTO DE GESTOS COMO ALTERNATIVA EFICIENTE PARA EXECUÇÃO DE COMANDOS EM REUNIÕES VIRTUAIS

Renata Rodrigues dos Santos Briet<sup>1</sup>, Elizabete Munzlinger<sup>2</sup>,  
Fabricio B. Narcizo<sup>3,4</sup>, Mario T. Shimanuki<sup>5</sup>

<sup>1</sup> Cursando Tecnologia em Análise e Desenvolvimento de Sistemas, Bolsista PIBIFSP, IFSP, Câmpus Caraguatatuba, [renata.briet@aluno.ifsp.edu](mailto:renata.briet@aluno.ifsp.edu).

<sup>2</sup> Industrial Ph.D. student, Computer Science Department, IT University of Copenhagen, København S, Denmark, [munzlinger@itu.dk](mailto:munzlinger@itu.dk).

<sup>3</sup> AI Research Scientist, GN Advanced Science, GN One, [fbnarcizo@jabra.com](mailto:fbnarcizo@jabra.com).

<sup>4</sup> Part-time Lecturer, Computer Science Department, IT University of Copenhagen, København S, Denmark, [narcizo@itu.dk](mailto:narcizo@itu.dk).

<sup>5</sup> Professor no Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, IFSP, Campus Caraguatatuba, [mario@ifspcaragua.net](mailto:mario@ifspcaragua.net)

Área de conhecimento (Tabela CNPq): 1.03.02.00-1 – Sistemas de Computação.

**RESUMO:** Este estudo tem como objetivo identificar quais são os gestos de mãos mais comuns propostos por participantes para o controle de algumas funções em ambientes de reuniões virtuais e híbridas em plataformas como Zoom Meeting e Google Teams, de modo a contribuir para o aumento da produtividade e da inclusão, e proporcionar uma interação mais natural, humanizada e eficiente aos participantes. A pesquisa, realizada por meio de um estudo de elicitación com 103 participantes de diferentes faixas etárias, os convidou a participar de uma simulação de reunião híbrida, na qual cada um executou gestos de sua preferência para representar os comandos estudados. A análise dos dados possibilitou a elaboração de um vocabulário de gestos, cuja integração às plataformas de comunicação unificada apresenta potencial para otimizar o controle de funções. Os resultados indicam que os participantes apresentam relativo baixo acordo entre os comandos, contudo apresentando tendência de preferência por gestos dinâmicos.

**PALAVRAS-CHAVE:** interação humano-computador; controle por gestos; estudo de elicitación; plataformas de comunicação unificada; vocabulário de gestos.

### GESTURE RECOGNITION AS AN EFFICIENT ALTERNATIVE FOR EXECUTING COMMANDS IN VIRTUAL MEETINGS

**ABSTRACT:** This study aims to identify the most common hand gestures proposed by participants to control some functions in virtual and hybrid meeting environments on platforms such as Zoom Meeting and Google Teams, in order to contribute to increased productivity and inclusion, and provide a more natural, humanized and efficient interaction to participants. The research, carried out through an elicitation study with 103 participants of different age groups, invited them to participate in a hybrid meeting simulation, in which each one performed gestures of their preference to represent the commands studied. Data analysis made it possible to develop a gesture vocabulary, whose integration with unified communication platforms has the potential to optimize function control. The results indicate that the participants have relatively low agreement between the commands, however showing a tendency to prefer dynamic gestures.

**KEYWORDS:** human-computer interaction; gesture control, elicitation study, unified communication platforms; gesture vocabulary.

### INTRODUÇÃO

Plataformas de reuniões virtuais e híbridas se tornaram essenciais para a comunicação a distância. Um grande aumento no uso de plataformas de comunicação unificada, como Google Teams e Zoom Meeting, se dá ainda nos dias de hoje, sendo utilizadas em reuniões corporativas, entrevistas de emprego, aulas remotas, entre outros. Em uma era em que a comunicação e a tecnologia são extremamente importantes, essas plataformas ainda

apresentam limitações quanto ao uso, como a execução de comandos de controle e configuração, como por exemplo o controle de áudio e vídeo exclusivamente por meios de dispositivos convencionais, como teclado e mouse, e, em raros casos, por voz.

Ao integrarmos gestos de mãos como uma forma de execução de comandos dentro dessas plataformas, como aumentar o volume ou desligar a câmera, tornamos a comunicação mais natural ao passo que se evita que o fluxo da atividade seja interrompido e, consequentemente a torna mais eficiente. De acordo com Susan Goldin-Meadow (Goldin-Meadow, 1999), a comunicação gestual é um fenômeno robusto, presente até mesmo em indivíduos cegos de nascença. Ela destaca que, quando utilizados junto à fala, os gestos adquirem uma forma visual e representativa, transmitindo informações que podem complementar ou até substituir a comunicação verbal, funcionando simultaneamente como ferramenta de comunicação para ouvintes e como auxílio cognitivo para quem fala.

De forma complementar, David McNaill (McNaill, 1992) reforça a importância dos gestos como parte integrante da linguagem e do pensamento. Em suas obras, como *Hand and Mind* (1996) e *Gesture and Thought* (1933), o autor explica que gestos e fala formam um único sistema integrado de expressão, no qual cada um carrega informações únicas e interdependentes. Ele propõe o conceito de *growth point*, a unidade cognitiva mínima onde a linguagem verbal e representação gestual se unem para estruturar o pensamento e a comunicação. Sob essa perspectiva, a adoção de comandos gestuais em plataformas de reunião não apenas amplia as formas de interação disponíveis, mas também explora um mecanismo natural do processo comunicativo humano, capaz de tornar a execução de tarefas mais intuitiva e alinhada à forma como as pessoas já processam e transmitem informações no dia a dia.

Com a evolução tecnológica os gestos deixaram de ser apenas elementos de expressão comunicativa e passaram a ser capturados e interpretados digitalmente por sistemas que utilizam reconhecimento de gestos. Essa tecnologia, baseada em visão computacional e aprendizado de máquina, possibilita transformar gestos em interfaces de comando, ampliando as formas de interação além dos dispositivos convencionais (Munzlinger et al., 2025). Quando integrados a outros recursos, como a fala, esses sistemas configuram uma interação multimodal, na qual diferentes canais comunicativos se complementam, tornando a experiência mais natural e menos propensa a ambiguidades.

Este estudo tem como objetivo fornecer uma base consistente para a integração dos gestos às plataformas de reuniões virtuais, por meio de um vocabulário de oito gestos principais no cenário de uma reunião. Esse vocabulário foi desenvolvido com base nos dados coletados em um estudo de elicitação, realizado com participantes de diferentes idades e níveis de experiência com tecnologia. Os participantes responderam a um questionário e participaram de uma simulação de reunião híbrida, na qual foram instituídos a realizar gestos de sua preferência para cada um dos comandos propostos neste trabalho. A partir dos resultados é possível dar início à aplicação da tecnologia, que tem como principal objetivo promover uma comunicação eficaz e humanizada em ambientes virtuais e híbridos, reduzindo barreiras de interação e tornando a experiência mais próxima de um encontro presencial ao aproximar a interação humano-computador de processos comunicativos típicos da vida real, fortalecendo sua aplicação em cenários de reuniões virtuais (Briet et al., 2024).

## MATERIAL E MÉTODOS

Esta pesquisa foi conduzida como um estudo qualitativo, utilizando duas abordagens principais para a coleta de dados: (1) aplicação de um questionário estruturado para identificar percepções e necessidades dos participantes em relação ao uso de gestos como forma de execução de comandos em reuniões, e (2) simulação de reunião híbrida, com foco em oito comandos específicos.

O participante foi inserido no cenário de simulação participando de uma reunião híbrida com colegas do trabalho distante do laptop conectado à reunião, precisando usar gestos com as mãos para executar os seguintes comandos: Aumentar Volume, Diminuir Volume, Mutar Microfone, Desmutar Microfone, Ligar Câmera, Desligar Câmera, Pedir para Falar e Encerrar Chamada.

Durante a simulação, foi aplicada a técnica *think aloud* (pense em voz alta), na qual os participantes foram orientados a explicar o motivo de cada gesto realizado, se o gesto remetia alguma associação ou lembrança, e o que representava para eles. Essa abordagem permitiu coletar informações valiosas sobre quais gestos são mais intuitivos e como os usuários normalmente escolheriam os gestos baseados em experiências anteriores com outras tecnologias.

Os vídeos gravados foram analisados manualmente, quadro a quadro, permitindo uma visualização detalhada dos movimentos executados e a identificação de padrões. Cada gesto foi nomeado conforme o movimento das mãos, por exemplo, *raised hand* foi a denominação escolhida para o movimento de levantar as mãos, como ao indicar que deseja fazer uma pergunta. Os gestos foram classificados em dinâmicos, aqueles que apresentavam movimento, e estáticos, os que permaneciam parados, também chamados de poses. Além disso, todos os gestos receberam uma descrição detalhada que possibilitou a identificação do movimento apenas por meio da leitura.

Para medir a concordância dos gestos usados pelos participantes, foi usada a fórmula de acordo (*agreement*) criada por Jacob O.Wobbrock (Wobbrock, 2022). Essa fórmula mostra o grau de concordância na escolha dos gestos pelos participantes. O nível de concordância para cada comando é classificado em três grupos: alto ( $AR \geq 0,7$ ), médio ( $0,4 \leq AR < 0,7$ ) e baixo ( $AR < 0,4$ ).

## RESULTADOS E DISCUSSÃO

Grande parte dos gestos coletados no estudo qualitativo foram simples e intuitivos. Muitos gestos se repetiam para diferentes participantes, e isso nos mostra que são gestos com grande potencial para serem adotados nas plataformas de reuniões, por seu fácil entendimento e execução.

No gráfico abaixo nota-se que o comando *Pedir para Falar* teve maior uniformidade se comparado aos outros, sendo o gesto de levantar uma das mãos o escolhido pela grande maioria dos participantes. Isso se deu possivelmente por ser um comando frequentemente presente no nosso dia a dia, como ao levantar a mão ao precisar fazer uma pergunta ao professor na escola. Desse modo, o gesto parece ser familiar e intuitivo, com grande potencial de ser usado para execução de comandos em ambientes virtuais. Podemos ver uma grande diferença quando comparado com os comandos *Desligar Câmera* e *Ligar Câmera*, que possuem um baixo nível de concordância, indicando maior diversidade nos gestos coletados.

No total foram 103 vídeos com 8 comandos em cada, totalizando 824 gestos coletados e estudados. No gráfico abaixo vemos que o gesto com maior incidência foi *Levantar a Mão*, majoritariamente presente no comando *Pedir para Falar*. Ele foi sugerido 81 vezes pelos participantes, seguido por *Fechar a Mão* que apareceu 66 vezes. Esses resultados nos mostram escolhas intuitivas e familiares dos usuários, frequentemente associadas a outras tecnologias e experiências anteriores. Ambos são gestos claros e de fácil execução, o que pode ter influenciado a escolha dos participantes.

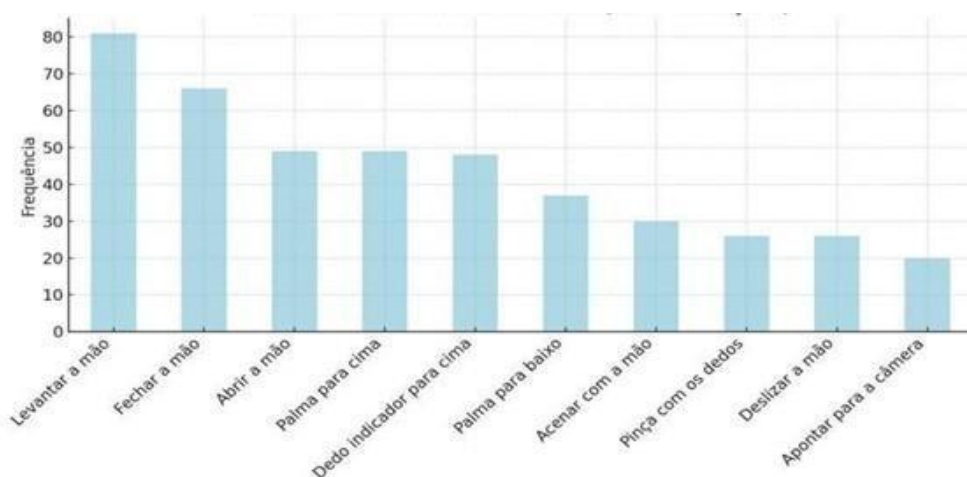


Gráfico 1 – Gestos com maior incidência. Fonte: Elaborado pelos autores.

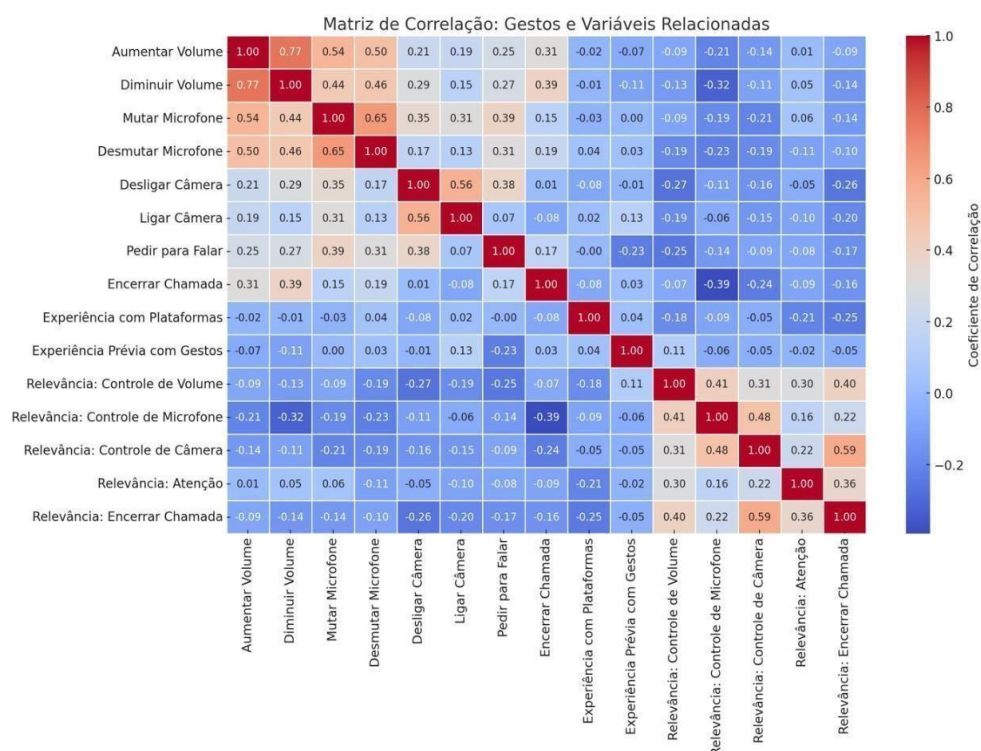


Figura 1 – Matriz de correlação. Fonte: Elaborado pelos autores

Na matriz de correlação acima o azul representa correlações negativas e o vermelho positivas. Os valores podem variar de -1 (correlação negativa forte) e 1 (correlação positiva forte).

Os comandos *Diminuir Volume* e *Aumentar Volume* apresentaram uma correlação positiva forte de 0.77, que pode indicar uma semelhança funcional entre os comandos, por serem opostos complementares e comumente usados no mesmo contexto para regular o volume. Na grande maioria dos casos os participantes tenderam a executar gestos opostos para comandos opostos.

Também foi observada correlação alta entre os comandos *Mutar Microfone*, *Desmutar Microfone*, *Ligar Câmera* e *Desligar Câmera*, fortalecendo a análise anterior, dada a semelhança funcional dos comandos. Já os comandos que representam intenções de uso distintas, como entre *Ligar Câmera* e *Encerrar Chamada*, apresentaram correlação quase nula de -0.08, outro exemplo é entre *Desligar Câmera* e *Encerrar Chamada* com correlação de 0.01, isso indica que a execução de um desses comandos não influencia a execução do outro, indicando que são ações independentes na simulação.

Tabela 1 – Gestos mais usados para cada comando da simulação

| Comando 1:<br>Aumentar Volume                                                              | Comando 2:<br>Diminuir Volume                                                                  | Comando 3: Mutar<br>Microfone                                                      | Comando 4:<br>Desmutar Microfone                                                       |
|--------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
|           |               |   |     |
| Palm Up                                                                                    | Palm Down                                                                                      | Hand Grab                                                                          | Hand Release                                                                           |
| Gesto dinâmico, com a palma de uma das mãos aberta, volta para cima, movendo-se para cima. | Gesto dinâmico, com a palma de uma das mãos aberta, voltada para baixo, movendo-se para baixo. | Gesto dinâmico, a palma de uma das mãos inicia aberta e se fecha para dentro.      | Gesto dinâmico, a palma de uma das mãos se inicia fechada e em seguida abre para fora. |
| Comando 5: Desligar<br>Câmera                                                              | Comando 6: Ligar<br>Câmera                                                                     | Comando 7: Pedir para<br>falar                                                     | Comando 8: Encerrar<br>Chamada                                                         |
|           |               |  |     |
| Hand Grab                                                                                  | Hand Release                                                                                   | Raised Hand                                                                        | Hand Wave                                                                              |
| Gesto dinâmico, a palma de uma das mãos inicia aberta e se fecha para dentro.              | Gesto dinâmico, a palma de uma das mãos se inicia fechada e em seguida abre para fora.         | Gesto estático, uma das mãos levantadas, com a palma aberta voltada para frente.   | Gesto dinâmico, palma de uma das mãos aberta, movendo-se de um lado para o outro.      |

Fonte: Elaborado pelos autores.

Na tabela acima, observam-se os gestos mais utilizados para cada um dos comandos analisados. Nota-se que a maioria envolve movimentos, ou seja, são dinâmicos, como *Palm Up*, *Hand Grab* e *Hand Wave*, evidenciando uma preferência por gestos que implicam ação, possivelmente por serem mais intuitivos. O único comando que apresentou maior frequência de gestos estáticos foi *Pedir para Falar*, o que pode indicar uma tendência a utilizar gestos mais simples e diretos em situações que exigem formalidade ou interação com os demais participantes da reunião.

Observa-se também que o gesto *Hand Grab* foi predominante tanto para o comando de *Mutar Microfone* quanto para *Desligar Câmera*, enquanto *Hand Release* se destacou em *Desmutar Microfone* e *Ligar Câmera*, sugerindo uma associação consistente entre gestos de fechamento das mãos e ações de desativação, e entre gestos de abertura das mãos e ações de ativação, reforçando a lógica intuitiva na escolha dos gestos pelos participantes. Apesar da aplicação da técnica *Think Aloud*, a maioria dos participantes não descreveu seus pensamentos sobre os gestos realizados, apenas executaram e apresentaram os gestos de forma direta, sem verbalizar o raciocínio por trás de suas escolhas.

Os dados coletados neste trabalho permitem melhor compreensão da maneira como os usuários de plataformas de reuniões virtuais, como Zoom Meeting, se adaptam de maneira eficiente ao uso de gestos para executar comandos nesses ambientes, com escolhas de fácil compreensão e execução.

Com base nas informações coletadas no questionário que antecedeu a simulação, foi possível identificar que a grande maioria dos participantes possui experiência prévia com plataformas de comunicação e já fizeram uso de gestos em alguma outra tecnologia, o que possivelmente influenciou suas escolhas. Ainda assim, os gestos analisados foram majoritariamente intuitivos e familiares para cada um dos comandos propostos, se mantendo dentro do objetivo da pesquisa.

Esses achados possibilitam a criação de vocabulários de gestos intuitivos e adaptados ao uso em sistemas multimodais, como as plataformas de reuniões virtuais, tornando a experiência do usuário mais eficiente, inclusiva e acessível.

Para futuras pesquisas, é crucial buscar a robustez da tecnologia, integrando o reconhecimento de fala para confirmar a execução dos comandos gestuais e, assim, minimizar erros de interpretação do sistema.

## CONTRIBUIÇÕES DOS AUTORES

Renata R. S. Briet: Investigação, coleta de dados, análise de dados e escrita do artigo.

Elizabete Munzlinger e Fabricio B. Narcizo: Conceitualização, metodologia, análise dos dados, escrita, revisão e supervisão.

Mario T. Shimanuki: Orientação, revisão na análise dos dados e validação.

## AGRADECIMENTOS

Os autores agradecem ao Instituto Federal de Educação, Ciência e Tecnologia de São (IFSP), por ter celebrado o *Memorandum of Understanding* (MOU) entre IFSP e GN Audio A/S, e ao Programa Institucional de Bolsas de Iniciação Científica (PIBIFSP 2025), pelo apoio e pela concessão da bolsa de Iniciação Científica, que viabilizou a realização deste projeto.

## REFERÊNCIAS

BRIET *et al.* Interação por Gestos de Mãos em Plataformas de Comunicações Unificadas. In: **SICLN –14.**, 2024, Caraguatatuba. Disponível em: < <https://ocs.ifspcaraguatatuba.edu.br/sicln/XIVSICLN/paper/view/515>>. Acesso em: 16 ago. 2025.

MUNZLINGER *et al.* Eliciting User-Defined Mid-Air Hand Gestures for Hybrid Meeting Platform Control: Results, Insights, and Design Implications. In: INTERACT 2025, 20., 2025, Belo Horizonte. Proceedings. **Lecture Notes in Computer Science**. Springer, 2025. Disponível em: < [https://link.springer.com/chapter/10.1007/978-3-032-05005-2\\_22](https://link.springer.com/chapter/10.1007/978-3-032-05005-2_22)>. Acesso em: 11 out. 2025.

MCNEILL, David. **Hand and mind: What gestures reveal about thought**. University of Chicago press, 1992. Disponível em: <<https://doi.org/10.2307/1576015>>. Acesso em: 16 ago. 2025.

GOLDIN-MEADOW, Susan. The role of gesture in communication and thinking. **Trends in cognitive sciences**, v. 3, n. 11, p. 419-429, 1999. Disponível em: <<https://www.cell.com/trends/cognitive-sciences/home>>. Acesso em: 16 ago. 2025.

TER BEKKE, Marlijn; DRIJVERS, Linda; HOLLER, Judith. Co-speech hand gestures are used to predict upcoming meaning. **Psychological Science**, v. 36, n. 4, p. 237-248, 2025. Disponível em: <<https://journals.sagepub.com/doi/10.1177/09567976251331041>>. Acesso em: 16 ago. 2025.

VATAVU, R.D., WOBBEROCK, J.O.: Clarifying Agreement Calculations and Analysis for End-User Elicitation Studies. **ACM Transactions on Computer-Human Interaction** 29(1), 5:1–5:70 (Jan 2022). <Disponível em: <https://dl.acm.org/doi/10.1145/3476101>>. Acesso em: 16 ago. 2025.

YASEEN *et al.* Next-gen dynamic hand gesture recognition: Mediapipe, inception-v3 and lstm-based enhanced deep learning model. **Electronics**, v. 13, n. 16, p. 3233, 2024. Disponível em: <<https://www.mdpi.com/2079-9292/13/16/3233>>. Acesso em: 16 ago. 2025.