

INTERAÇÃO POR GESTOS DE MÃOS EM PLATAFORMAS DE COMUNICAÇÕES UNIFICADAS

RENATA RODRIGUES DOS SANTOS BRIET¹, ELIZABETE NARCIZO²,
FABRICIO B. NARCIZO^{3,4}, MARIO T. SHIMANUKI⁵

¹ Cursando Tecnologia em Análise e Desenvolvimento de Sistemas, Bolsista PIBIFSP, IFSP, Câmpus Caraguatatuba, renata.briet@aluno.ifsp.edu

² Industrial Ph.D. student, Computer Science Department, IT University of Copenhagen, København S, Denmark, munzlinger@itu.dk

³ AI Research Scientist, GN Advanced Science, GN One, fbnarcizo@jabra.com

⁴ Part-time Lecturer, Computer Science Department, IT University of Copenhagen, København S, Denmark, narcizo@itu.dk

⁵ Professor no Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, IFSP, Campus Caraguatatuba, mario@ifspcaragua.net

Área de conhecimento (Tabela CNPq): Sistemas de Computação – 1.03.04.00-2

RESUMO: Os gestos foram uma das primeiras formas de comunicação da humanidade, e embora tenhamos desenvolvido a fala ao longo dos anos e nos comuniquemos predominantemente por meio dela, os gestos não caíram em desuso, pelo contrário, se tornaram um complemento a comunicação verbal, é por meio deles que nos expressamos com maior clareza e até mesmo confiança. Diante disso, e com o aumento no uso das Plataformas de Comunicação Unificadas (PCU), como Google Teams e Zoom Meeting para reuniões virtuais, este estudo propõe o uso de uma tecnologia de reconhecimento de gestos para a execução de comandos específicos dentro das PCU, se baseando em aspectos da cultura e da psicologia sobre como os gestos influenciam na percepção e interação social. Ao integrar essa funcionalidade as PCU, os usuários poderão se beneficiar de uma experiência mais rica, que combina elementos visuais e verbais, essenciais para a colaboração em equipe. Além disso, a integração dessas tecnologias pode auxiliar na inclusão de pessoas com dificuldades verbais e motoras específicas, oferecendo alternativas de interação que tornam o ambiente virtual mais acessível. Os resultados aqui apresentados foram obtidos por meio de um experimento realizado com voluntários de diferentes idades, gêneros e etnias, onde foi possível notar padrões na comunicação gestual e encontrar gestos intuitivos para todos, independentemente da idade e experiência anterior com tecnologias de interação ou videoconferências. Essa inovação representa um passo importante na evolução das interações digitais, transformando a forma como nos conectamos e colaboramos no mundo moderno, permitindo que os usuários se expressem de maneira mais eficiente, natural e intuitiva, utilizando movimentos com as mãos para transmitir emoções, enfatizar pontos e tornar as reuniões mais colaborativas e inclusivas.

PALAVRAS-CHAVE: interação humano-computador, detecção e reconhecimento de gestos.

1 INTRODUÇÃO

Nos últimos anos, o reconhecimento de gestos de mãos tem se mostrado uma tecnologia inovadora, transformando a forma como interagimos com dispositivos eletrônicos, desde videogames, como o Kinect, até aparelhos domésticos inteligentes.

A capacidade de controlar sistemas por meio de gestos naturais proporciona uma experiência mais envolvente e intuitiva e ao implementar essa tecnologia às PCU, é possível que movimentos simples das mãos sejam interpretados como comandos, eliminando a necessidade de interfaces físicas, como mouse e teclado.

A comunicação por meio de gestos visa superar as restrições das interfaces convencionais, oferecendo uma interação mais dinâmica e acessível às reuniões virtuais, permitindo que os participantes se expressem de maneira mais rica, simplificando os comandos que atualmente necessitam de contato físico ou verbal. Com base nisso, este estudo se propõe a analisar como a integração desta tecnologia às Plataformas de Comunicação Unificadas pode torná-las mais eficientes, criando um ambiente dinâmico e inclusivo para os usuários.

Ao permitir que os comandos sejam executados através de gestos, dispensando o uso das interfaces tradicionais, que podem ser desafiadoras para alguns, o ambiente de reunião ganha grande potencial em beneficiar usuários que apresentam determinadas dificuldades motoras ou verbais, como o Transtorno do Espectro Autista, que pode causar desafios ao utilizar o teclado por conta do comprometimento de coordenação motora fina, que se trata da capacidade de realizar movimentos precisos com os músculos menores, como os das mãos (HEIDRICH, 2022). O estudo busca estabelecer um vocabulário comum de gestos, intuitivos e simples, para garantir que as necessidades específicas desses usuários sejam atendidas, tornando a comunicação remota mais inclusiva e refletindo a diversidade das interações humanas.

O estudo desta tecnologia foi enriquecido através de duas entrevistas e um experimento de caráter observacional, realizados com 103 voluntários, para identificar suas preferências e necessidades específicas. Todo o processo, desde a coleta até a análise e uso dos dados seguiu os princípios éticos da pesquisa envolvendo seres humanos, sendo realizada somente com os voluntários que consentiram em participar. Nenhuma das fases do experimento apresentava qualquer risco aos participantes e o anonimato e privacidade de todos que participaram é garantido.

Os resultados aqui apresentados evidenciam as vantagens da integração do reconhecimento de gestos de mãos às PCU e oferece informações que podem colaborar no aprimoramento contínuo da tecnologia, tornando possível que no futuro a comunicação remota, por meio de videoconferências, não supere somente barreiras geográficas, mas se transforme em um ambiente mais inclusivo e eficiente, refletindo as vantagens da comunicação face a face.

2 TEORIA

O reconhecimento de gestos de mãos é uma tecnologia emergente que visa transformar a forma como interagimos com dispositivos. No livro *Handbook of Human-Computer Interaction*, editado por M.G. Helander, T.K. Landauer e P. Prabhu (1997), o capítulo “Gesture-Based Interaction” explora como os gestos são percebidos, modelados, reconhecidos e interpretados por sistemas de computador interativos para aproveitar a expressividade das habilidades comunicativas humanas, permitindo meios de comunicação naturais, intuitivos e fluentes para os usuários. Essa perspectiva teórica

é altamente relevante para esta pesquisa, por sustentar a ideia de interfaces que possibilitam uma comunicação expressiva e eficaz por meio da implementação do reconhecimento de gestos de mãos.

A tecnologia aqui discutida é essencial na melhoria das interações em reuniões virtuais, tornando-as mais intuitivas, eficientes e inclusivas, permitindo que os usuários controlem funções através de gestos naturais, diminuindo a necessidade de interfaces físicas e aumentando a produtividade, permitindo uma experiência de reunião mais fluida e envolvente. Na pesquisa de McNeill (1985), intitulada “Hand and Mind: What Gestures Reveal about Thought”, temos uma base teórica essencial para o estudo. De acordo com McNeill, a ocorrência de gestos ao longo da fala indica que dois tipos de pensamentos, o imagístico e o sintático, estão sendo coordenados, sugerindo que os gestos são uma parte integral do processo de comunicação humana, e não meramente um complemento. Ao aplicar essa perspectiva ao reconhecimento de gestos de mãos, observa-se que essa tecnologia possibilita uma forma de interação mais natural e expressiva, reforçando a ligação entre pensamento e expressão. Assim, essa abordagem não apenas facilita a participação de usuários com certas limitações motoras ou dificuldades de comunicação verbal (por meio da fala), mas também reflete a complexidade e riqueza da comunicação face a face, alinhando-se às observações de McNeill sobre sinergia entre gestos e fala. As tecnologias que permitem o rastreamento e a interpretação de gestos estão em constante evolução, mas enfrentam obstáculos quanto à precisão e confiabilidade. A qualidade das câmeras, condições de iluminação do ambiente e a capacidade de processamento são alguns dos fatores que afetam significativamente essas limitações. De acordo com o artigo “Sistema de Interação por Gestos em Ambientes Inteligentes” (Goulart, 2019), esses fatores podem afetar diretamente a eficácia dos sistemas de reconhecimento de gestos. Além disso, é crucial considerar aspectos humanos, tendo em vista que a comunicação pode ser afetada por questões emocionais, psicológicas e contextuais, como diferenças culturais e barreiras linguísticas, as quais também se aplicam ao uso de gestos em PCU. Portanto, a implementação de instruções claras e de fácil compreensão sobre os gestos a serem utilizados é de extrema importância para que os usuários possam se adaptar rapidamente e utilizar a tecnologia de forma eficiente e eficaz.

3 MATERIAL E MÉTODOS

Para ser possível a elaboração de um vocabulário comum de gestos para interação nas PCU, foi elaborado um “estudo de elicitación”, que consistiu em um experimento observacional e dois questionários. Antes de iniciar a pesquisa, os voluntários receberam uma explicação de como o estudo seria conduzido e foram convidados a assinar um termo de consentimento, garantindo a privacidade das informações coletadas, o estudo seguiu para as próximas fases somente com os participantes que estavam de acordo com os termos, no total foram 103.

O questionário inicial teve como intuito coletar informações referentes aos participantes e a experiência anterior dos mesmos com as Plataformas de Comunicação Unificadas.

Tabela 1 Perguntas do questionário inicial

Numeração	Perguntas
1	Qual o seu nome?
2	Data de nascimento
3	Com qual gênero você se identifica?
4	Em quais áreas você já atuou ou estudou?
5	Qual seu país de origem?
6	Quantos anos você morou no país de origem?
7	Já morou em outro país? Se sim, qual?
8	Qual sua língua nativa?
9	Qual sua segunda língua? (Se tiver)
10	Quão proficiente você é em inglês?
11	Quão proficiente você é em utilizar um computador?
12	Quais plataformas de reuniões virtuais você já utilizou?

Na fase seguinte os voluntários participaram de uma simulação de reunião híbrida, onde eles não alcançavam o computador conectado a chamada e precisavam executar os comandos propostos por meio de gestos com as mãos.

Tabela 2 Comandos executados na simulação

Numeração	Comandos
1	Aumentar volume
2	Diminuir volume
3	Mutar microfone
4	Desmutar microfone
5	Desligar câmera
6	Ligar câmera
7	Pedir para falar
8	Encerrar chamada

Para calcular a concordância de gestos entre os usuários foi aplicada a fórmula de *agreement* proposta por Jacob O. Wobbrock, que avalia o grau de concordância entre participantes para a escolha de gestos em estudos de elicitación. O nível de consenso dos gestos entre os participantes para cada comando, juntamente com a classificação de interpretação, vai de alto *agreement* ($AR \geq 0.7$), médio *agreement* ($0.4 \leq AR < 0.7$) e baixo *agreement* ($AR < 0.4$).

Ao finalizar a simulação os voluntários responderam ao questionário final, que coletou informações relacionadas ao uso anterior das PCU e de tecnologias de reconhecimento de gestos, este questionário também contou com duas perguntas opcionais onde os participantes poderiam informar ideias de comandos e de melhorias para a tecnologia estudada.

Tabela 3 Perguntas do questionário final

Numeração	Perguntas
1	De 1 a 5, como você avaliaria a experiência anterior com o uso das plataformas de reuniões virtuais?
2	Você já utilizou interação por gestos de mãos em alguma tecnologia? Se sim, qual?
3	De 1 a 5, como você avaliaria sua experiência anterior utilizando interação por gestos de mãos?
4	De 1 a 5, quão relevante você considera a adoção de gestos de mãos para regular o volume?
5	De 1 a 5, quão relevante você considera a adoção de gestos de mãos para ativar e desativar o microfone?
6	De 1 a 5, quão relevante você considera a adoção de gestos de mãos para ativar e desativar a câmera?
7	De 1 a 5, quão relevante você considera a adoção de gestos de mãos para realizar interações com outros participantes?
8	De 1 a 5, quão relevante você considera a adoção de gestos de mãos para encerrar a chamada?
9	De 1 a 5, quão inclinado você estaria em executar comandos com gestos de mãos em uma reunião, caso eles estivessem disponíveis?
10	Gostaria de compartilhar ideias de comandos que poderia ser útil a adoção de gestos de mãos para executá-los?
11	Gostaria de compartilhar ideias de melhoria para a tecnologia?

Para garantir a padronização e precisão na coleta dos dados, os voluntários foram posicionados em um local demarcado, sentados de frente para a câmera, que teve a altura regulada com o auxílio de um tripé, de modo que os gestos fossem capturados corretamente, evitando possíveis cortes causados pelo enquadramento. O local de gravação teve um fundo branco, livre de objetos que pudessem atrapalhar o reconhecimento dos gestos executados.

Durante a simulação, foi aplicada a técnica “think aloud”, ou literalmente “pensar em voz alta”, na qual os participantes foram instruídos a verbalizar seus pensamentos enquanto performavam os gestos, explicando o significado e o motivo pelo qual o haviam escolhido. Os métodos de pesquisa de pensamento em voz alta têm uma base teórica sólida e fornecem uma fonte válida de dados sobre o pensamento do participante, que nos auxilia a entender com maior precisão suas ações e escolhas (CHARTERS, 2003).

A análise dos gestos foi realizada manualmente com o auxílio da extração de frames, que permitiu uma visualização mais clara dos movimentos executados, facilitando a identificação de padrões. Foram nomeados com base no movimento das mãos e classificados de duas formas, dinâmicos para os gestos que continham movimento, e estáticos para aqueles que se mantinham estacionários.

Todos os gestos receberam uma descrição detalhada, informando quantas mãos foram usadas na execução do movimento, para qual direção estavam voltadas e se moviam, qual era o movimento realizado, se ele se repetia e todas as outras informações necessárias para que fosse possível visualizar com clareza qual foi o gesto performed através da leitura da descrição. Com isso se torna possível encontrar semelhanças e padrões gestuais, definindo quais seriam mais intuitivos e acessíveis para os usuários.

4 RESULTADOS E DISCUSSÃO

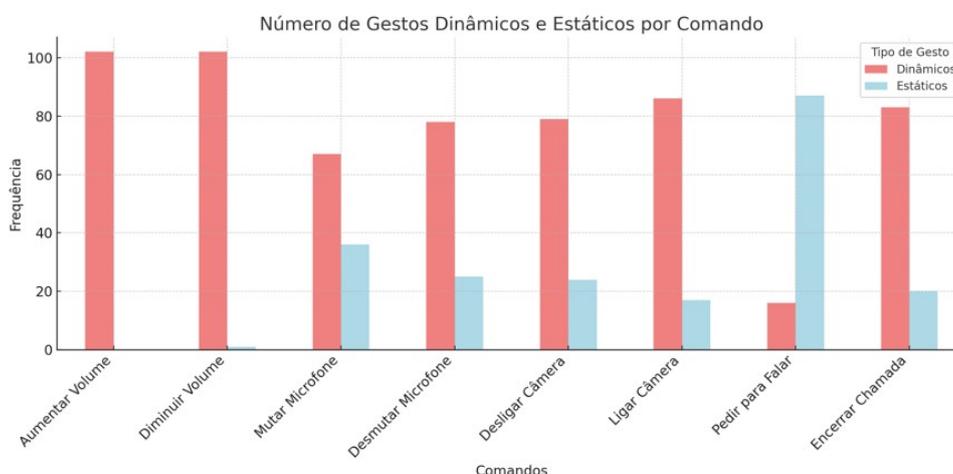
Neste estudo, tanto os dados coletados na simulação de reunião híbrida quanto nos questionários foram utilizados com o intuito de encontrar os gestos que ao serem integrados às plataformas virtuais de reunião, trariam maior conforto e acessibilidade para todos, sendo movimentos intuitivos e simples.

Os gestos registrados foram classificados como dinâmicos (movimento) ou estáticos (posições fixas das mãos), em sua maioria os gestos foram dinâmicos, sendo muito mais frequentes que os estáticos.

Observa-se que os comandos de Aumentar Volume e Diminuir Volume são dominados por gestos dinâmicos, com quase nenhum uso de gestos estáticos. Ligar e Desligar Câmera, assim como Encerrar Chamada mostram uma preferência maior por gestos dinâmicos, mas também apresentam uma quantidade considerável de gestos estáticos. Enquanto Pedir para Falar é o único comando em que os gestos estáticos têm uma representação relativamente maior, destacando a possível necessidade de precisão para este comando.

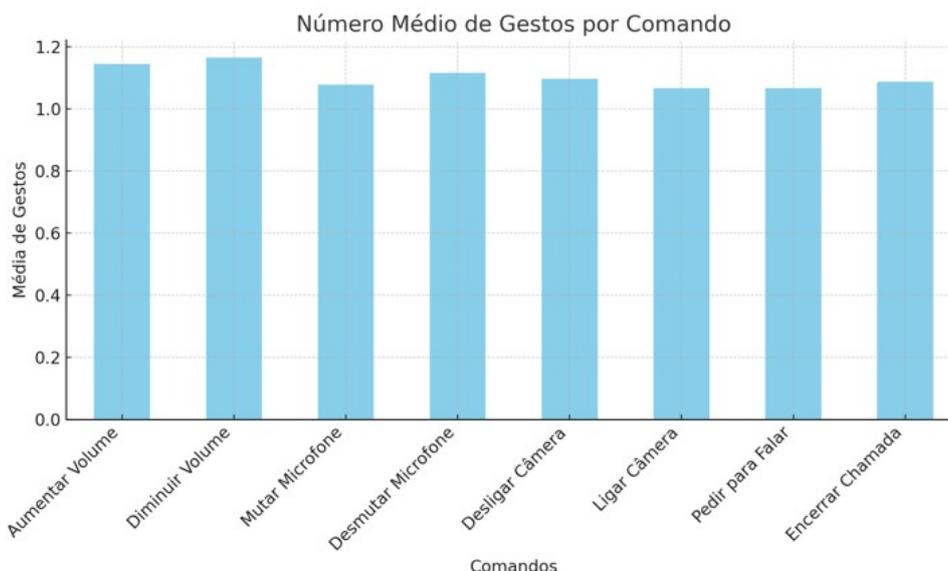
Esse padrão pode indicar que os participantes preferem gestos que envolvam movimento, talvez por acharem mais intuitivo para expressar comandos ou por serem mais adequados ao contexto das interações propostas. Os gestos dinâmicos são amplamente preferidos para comandos relacionados ao controle de áudio e vídeo, enquanto gestos estáticos são usados em contextos que podem demandar maior precisão ou formalidade, como pedir para falar.

Figura 1 – Número de Gestos Dinâmicos e Estáticos por Comando



Foi notado que para cada um dos oito comandos analisados no estudo (como aumentar/diminuir volume, silenciar/desmutar microfone etc.). Os participantes geralmente utilizavam um único gesto para realizar cada comando, com apenas variações mínimas entre cada comando, sugerindo que os gestos escolhidos são simples e diretos, havendo poucos casos de combinações ou sequências de múltiplos gestos para executar uma única ação.

Figura 2 – Número Médio de Gestos por Comandos



Foram raros os participantes que fizeram uso de dois ou mais gestos para um único comando. Para melhor entendimento, nas figuras abaixo é possível notar um exemplo desta combinação de gestos no comando referente a aumentar o volume do volume.

Figura 3a - Gesto inicial aumentar volume

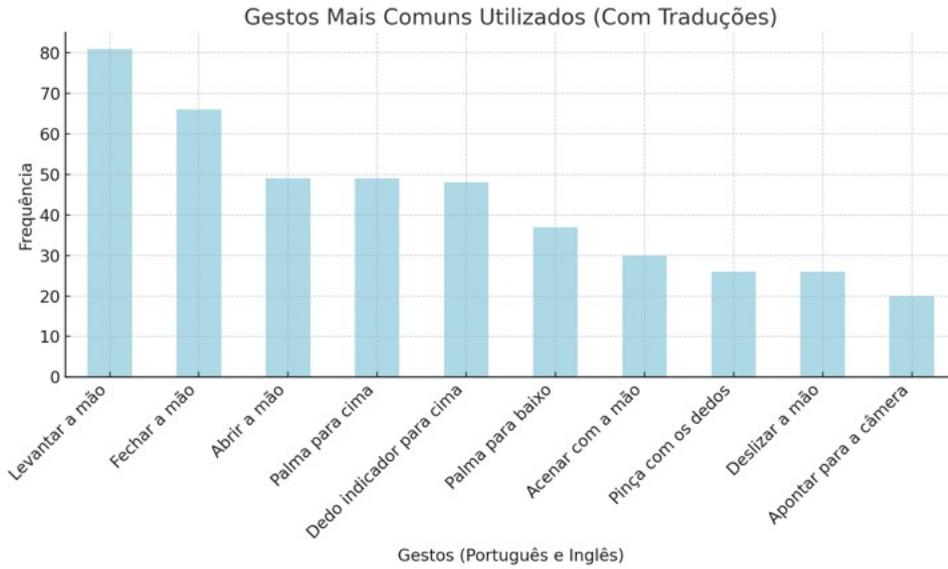


Figura 3b – Gesto final aumentar volume



O gráfico abaixo apresenta quais foram os gestos mais frequentemente utilizados pelos participantes durante o estudo, no total foram 824 gestos. Ele evidencia a predominância de gestos como “levantar a mão”, sendo o mais comum, com 81 ocorrências. O segundo mais realizado foi “fechar a mão”, 66 vezes. “Abrir a mão” e “palma para cima” foram ambos usados 49 vezes.

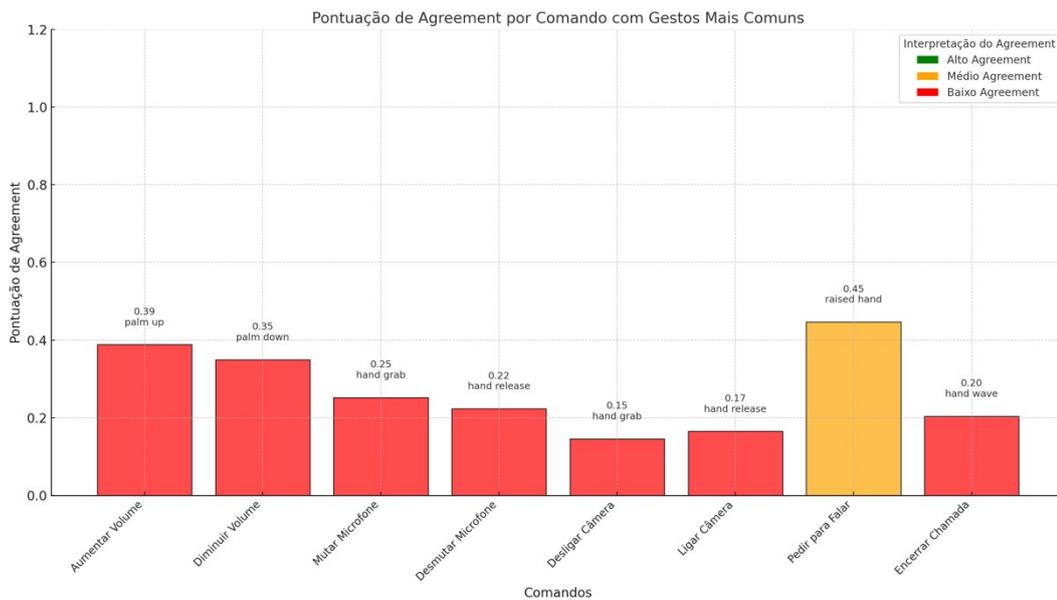
Figura 4 – Gráfico de Gestos mais Comuns Utilizados



Esses gestos refletem escolhas intuitivas que podem ser baseadas na interação natural e no contexto das funções controladas. Gestos como “levantar a mão” e “fechar a mão” são fáceis de executar e visualmente claros, o que pode justificar sua popularidade. Os gestos mais usados são simples e intuitivos, com movimentos que são fáceis de realizar e reconhecer visualmente, reforçando sua popularidade para comandos em plataformas de interação.

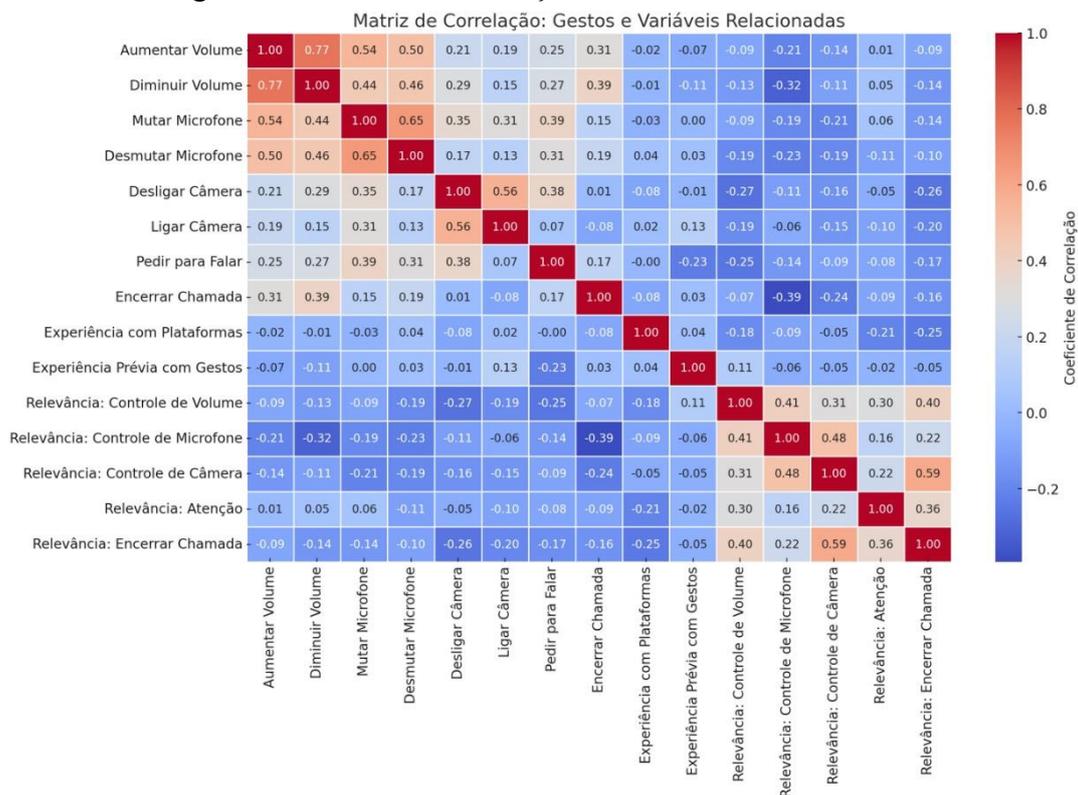
Abaixo temos os valores de *agreement* score para cada comando, junto com os gestos mais comuns utilizados pelos participantes. As informações foram simplificadas para garantir clareza.

Figura 5 – Gráfico de Pontuação de Agreement por Comando com Gestos Mais Comuns



Percebe-se que os comandos como “Aumentar Volume” e “Diminuir Volume” apresentam consenso médio, sugerindo maior uniformidade nas escolhas dos gestos, mas ainda assim baixo. Comandos como “Mutar Microfone”, “Desmutar Microfone” e “Desligar Câmera” possuem um baixo nível de acordo, indicando maior diversidade ou falta de padrão nos gestos escolhidos.

Figura 6 – Matriz de Correlação: Gestos e Variáveis Relacionadas



A imagem acima apresenta a matriz de correlação com linhas e colunas apresentando os comandos (gestos) e variáveis relacionadas. As cores indicam a força e a direção da correlação, sendo tons de azul correlações negativas e tons de vermelho correlações positivas. Os coeficientes são os valores dentro das células e variam entre -1 (correlação negativa forte) e 1 (correlação positiva forte).

A correlação de 0.77 entre os comandos “Diminuir Volume” e “Aumentar Volume” indica uma correlação positiva forte. O valor pode apresentar uma similaridade funcional, pois ambos os comandos são opostos complementares e frequentemente utilizados no mesmo contexto (controle de volume). Participantes que realizam gestos mais variados para um comando provavelmente têm um comportamento similar ao realizar gestos para outro comando. Outra correlação alta que pode ser observada, acontece entre “Mutar Microfone” e “Desmutar Microfone”, “Ligar Câmera” e “Desligar Câmera”, sugerindo a mesma análise, dada a semelhança funcional dos comandos.

Com relação aos questionários, os participantes compartilharam sugestões interessantes sobre o uso de gestos. Algumas ideias destacadas incluem:

1. Compartilhamento de tela: Vários participantes sugeriram gestos específicos para compartilhar a tela, como separar as mãos ou criar formas retangulares.

2. Zoom: Gestos como movimentos de pinça para ampliar ou reduzir a visualização foram mencionados.
3. Reações e interações sociais: As ideias incluem bater palmas para aparecer um ícone na tela ou usar gestos para mandar reações como “curtir” ou “coração”.
4. Controle avançado: Sugestões como silenciar outros participantes ou ajustar brilho e volume separadamente com diferentes mãos.
5. Captura de tela: Diversas variantes de gestos para “tirar print” foram propostas.

Algumas preocupações foram relatadas:

1. Confirmação de ações: Os participantes sugeriram que o sistema peça confirmação antes de executar um comando baseado em gestos.
2. Movimentos espontâneos vs predefinidos: A distinção entre gestos acidentais e intencionais foi levantada como um possível problema.

Essas ideias e preocupações reforçam a necessidade de desenvolver gestos intuitivos, mas com um sistema robusto para evitar erros de interpretação. Além disso, funcionalidades que integrem confirmação ou modos de aprendizagem adaptativos podem ser relevantes.

O perfil demográfico dos participantes destaca que a maioria possui experiência prévia com plataformas de comunicação e gestos, o que pode ter influenciado suas escolhas. Apesar disso, a análise sugere que os gestos escolhidos foram amplamente intuitivos e alinhados às expectativas funcionais dos comandos. Esses achados fornecem uma base sólida para desenvolver vocabulários de gestos mais intuitivos e inclusivos, especialmente em contextos de controle multimodal para plataformas de comunicação.

5 CONSIDERAÇÕES FINAIS

Este estudo de elicitación revelou informações valiosas sobre os padrões de gestos utilizados pelos participantes para comandos específicos em uma plataforma online. Identificamos forte correlação entre gestos complementares, como "Aumentar Volume" e "Diminuir Volume", refletindo a consistência funcional nas escolhas dos participantes.

Entretanto, alguns comandos apresentaram baixo nível de consenso (baixo *agreement score*), como "Mutar Microfone" e "Desligar Câmera", indicando maior diversidade ou falta de padronização nas escolhas dos gestos. Esses resultados sugerem que comandos mais críticos e intuitivos têm maior potencial de uniformidade, enquanto aqueles com menor consenso podem exigir intervenções de design ou treinamento para melhorar a consistência.

A distribuição semântica dos gestos em um espaço 2D também apontou clusters claros para alguns comandos, reforçando a necessidade de explorar padrões culturais e contextuais em estudos futuros.

Em resumo, este projeto contribuiu de forma significativa para o campo da interação humano-computador, destacando o papel central que a intuição e a familiaridade desempenham na criação de interfaces mais inclusivas e acessíveis.

Agradecimento: Os autores desejam agradecer ao IFSP-CAR e ao Programa de Bolsas de Iniciação Científica do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (PIBIFSP 2023) pela concessão da bolsa de Iniciação Científica, o que tornou possível a realização deste projeto.

REFERÊNCIAS

CHARTERS, Elizabeth. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. **Brock Education Journal**, v. 12, n. 2, 2003.

HELANDER, Martin G. (Ed.). **Handbook of human-computer interaction**. Elsevier, 2014.

BARRETT, Paul H. **The Works of Charles Darwin: Vol 23: The Expression of the Emotions in Man and Animals**. Routledge, 2016.

LIMA, Etelvina do Rosário Silva; CRUZ-SANTOS, Anabela. Aquisição dos gestos na comunicação pré-linguística: uma abordagem teórica. **Revista da Sociedade Brasileira de Fonoaudiologia**, v. 17, p. 495-501, 2012.

HEIDRICH, Taís Elena et al. Content validity of an instrument for motor assessment of youth with autism. **Fisioterapia em Movimento**, v. 35, p. e35135, 2022.

OUDAH, Munir; AL-NAJI, Ali; CHAHL, Javaan. Hand gesture recognition based on computer vision: a review of techniques. **journal of Imaging**, v. 6, n. 8, p. 73, 2020.

GOULART, Eduardo Freitas; PEROZZO, Reiner Franchesco. Sistema de Interação por Gestos em Ambientes Inteligentes, 2019.

DO SANTOS, Clebeson Canuto et al. RECONHECIMENTO DE GESTOS DINÂMICOS PARA A INTERAÇÃO HUMANO-ROBÔ. **Cep**, v. 29075, p. 910, 2019.

ANDRIC, Michael; SMALL, Steven L. Gesture's neural language. **Frontiers in psychology**, v. 3, p. 99, 2012.

MCNEILL, David. **Gesture and thought**. University of Chicago press, 2019.

STUDDERT-KENNEDY, Michael. Hand and Mind: What Gestures Reveal About Thought. **Language and Speech**, v. 37, n. 2, p. 203-209, 1994.

WU, Ying Choon; COULSON, Seana. How iconic gestures enhance communication: An ERP study. **Brain and language**, v. 101, n. 3, p. 234-245, 2007.

VATAVU, Radu-Daniel; WOBROCK, Jacob O. Formalizing agreement analysis for elicitation studies: new measures, significance test, and toolkit. In: **Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems**. p. 13251334, 2015.