Eliciting User-defined Mid-air Hand Gestures For Hybrid Meeting Platform Control: Insights And Design Implications

Elizabete Munzlinger[1,2], Fabricio Batista Narcizo[1,2], **Renata Briet**[3], Mario Tadashi Shimanuki[3], Ted Vucurevich[2], and Dan Witzner Hansen[1]

[1] IT University of Copenhagen [2] GN Advanced Science [3] Federal Institute of São Paulo

Agenda

1 - Problem Contextualization
2 - Related Work
3 - Study Methodology
4 - Results
5 - Conclusions

The Hybrid Meeting **Context**

1 - Problem Contextualization

Effects from **Limitations** of Existing **Interaction** Methods

Advantages of Mid-air **Hand Gesture** Recognition

GPT-4o Edited Image

Related Work

- **Home devices** [45] BUT Two-handed gestures (pointing device+command)
- **Smart cameras** (robotic) [33] BUT Pointing gestures
- **Similar tasks** (volume control) [3] BUT Micro-gestures for small devices
- **UCP context** [21] BUT Emoji mapping

Shared Gestures Between Domains = Low Agreement Rates (AR)

- **Volume control** [44]: Swipe Up/Down movements, and pointing Up/Down
- **Increase and decrease** (a control) [16]: Move hand Up/Down (low *AR*)
- Turn On/Off (device) [16]: Snap fingers, rub hands (low *AR*)
- Etc.

Investigation of **Gestures** for **Control** and **Interaction** in Hybrid Meetings

User-Defined Method

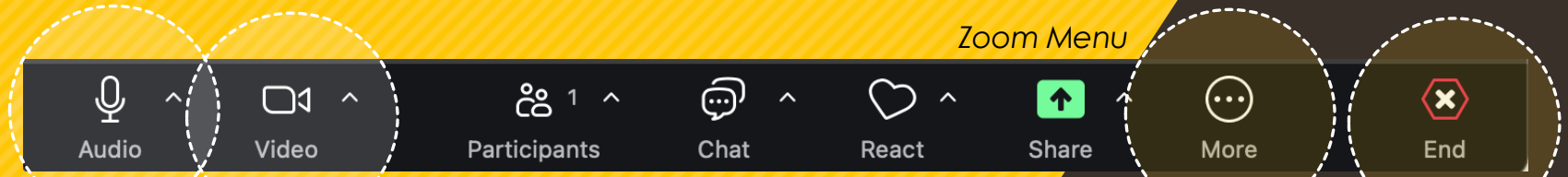
8 frequently used control **commands** for unified communication platforms (UCP)

The Selected Referents

Referents:

R_1 Increase Volume
R_2 Decrease Volume
R_3 Mute Microphone
R_4 Unmute Microphone
R_5 Turn off Camera
R_6 Turn on Camera
R_7 Ask for a Question
R_8 End the Call

Zoom Menu



Microsoft Teams Menu

Scenario-Driven Approach

The Study Setup, Procedures and Instructions

- Pre- and post-elicitation **questionnaires**
- **Think-Aloud** method
- Participants are seated in a **simulated** hybrid meeting room
- Guided through a realistic scenario using **Storytelling**
- Encouraging gestures aligned with the **social and functional norms** of formal meetings
- Limited to **a single** proposal per referent
- **Free sequences** of gestures per command

Diverse, contextually relevant Brazilian users

Participants Recruitment and Demographics

- **103** participants (volunteers)
- **Diverse backgrounds** (TI–36, Admin.–12, Eng.–10, Commerce–10, Education–9, Business–9, and Arts–5, Env. Science-5, Finance-2, Government-1, N/A-3)
- **57** males, **45** females, **1** non-binary, 18-55 years old
- **99%** reported **prior experience** using **video conferencing tools**
- **62%** with **gesture-based interaction** devices (Kinect, VR)

ELICITCAM Dataset

The Gesture Corpus Creation

Each **proposal** consists of one or more hand gestures, performed with one or two hands in static poses or dynamic movements, accommodated in a **gesture unit** (rest-to-rest position)

- **103** recorded **videos** with a resolution of 1920 × 1080 px
- Recordings **segmented** into eight parts and **labelled** (ELAN tool)
- Gesture taxonomy annotation
- **824** gesture **proposals**
- **133 unique** gestures

The Taxonomy Classification

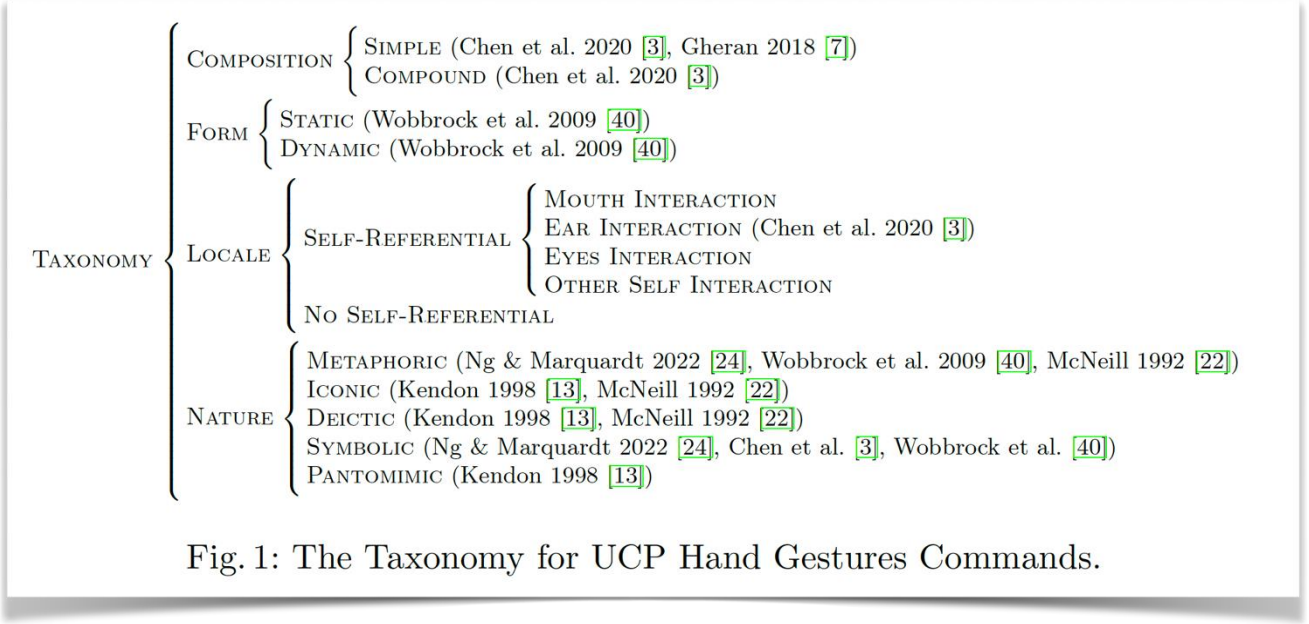


Fig. 1: The Taxonomy for UCP Hand Gestures Commands.

Human-based classification

- Name, Description, Composition, Form

AI-based classification (OpenAI GPT-4o, (2) Google Gemini 2.0 Flash, and (3) Meta Llama 3.3)

- Locale and Nature

Data Analysis and Results

Results of **Taxonomy** Classification

Human-Based and **AI-Based** Approaches

Agreement Between Models:

Inter-rater reliability metric Fleiss' Kappa agreement score:

- *Nature κ = 0.7428*
- *Locale κ = 0.8812*

Pairwise Models Alignment:

Fleiss' Kappa for pairwise comparisons (GPT-4o, Gemini 2.0 Flash, and Llama 3.3):

- *Nature: GPT and Gemini more consistent/aligned κ = 0.7929*
- *Locale: GPT and Gemini κ =0.9167*

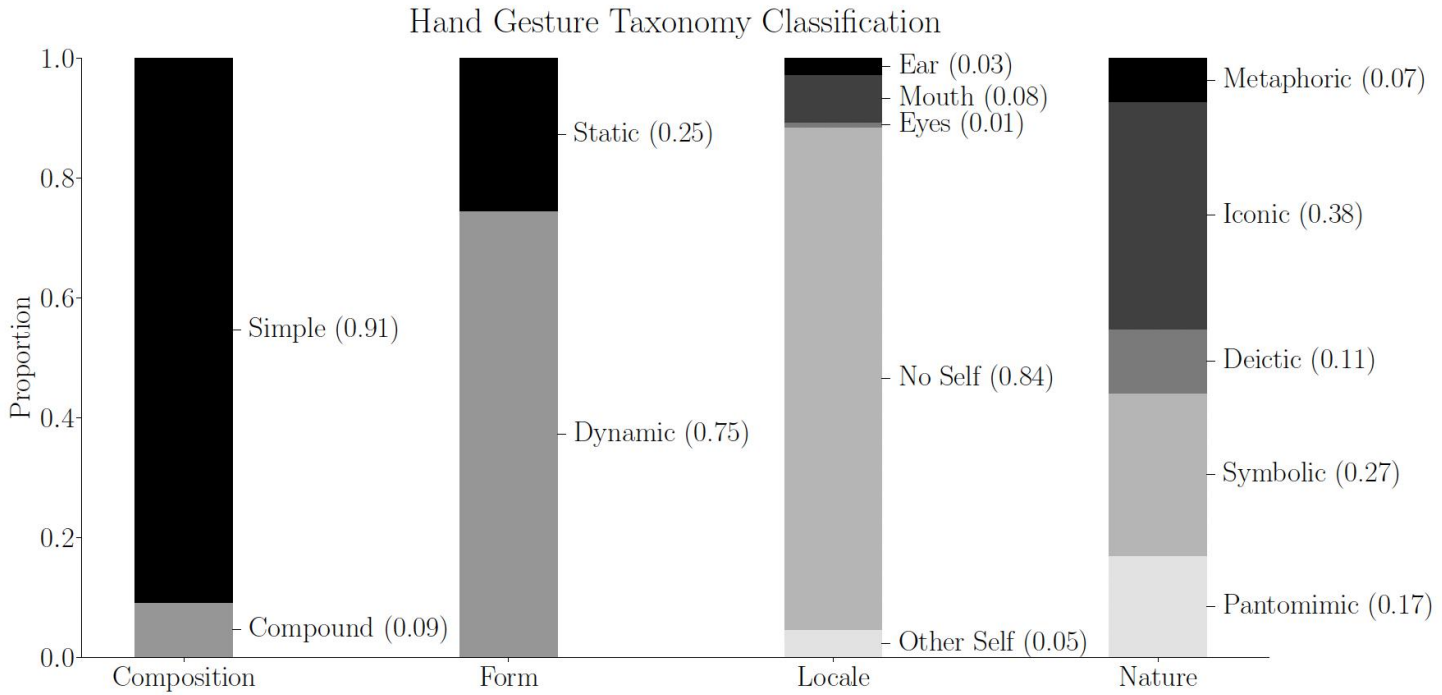
Selection of Taxonomy Classes Assigned by Models

Aggregating classifications from LLMs, selecting the most frequently assigned label.

- *Disagreement applied an expert adjudication approach, where a human rater manually*

Results of **Taxonomy** Classification

Human-Based and **AI-Based** Approaches



Agreement on **Gesture Proposals**

Agreement Rate Metric (*AR*(r)) [Vatavu and Wobbrock]

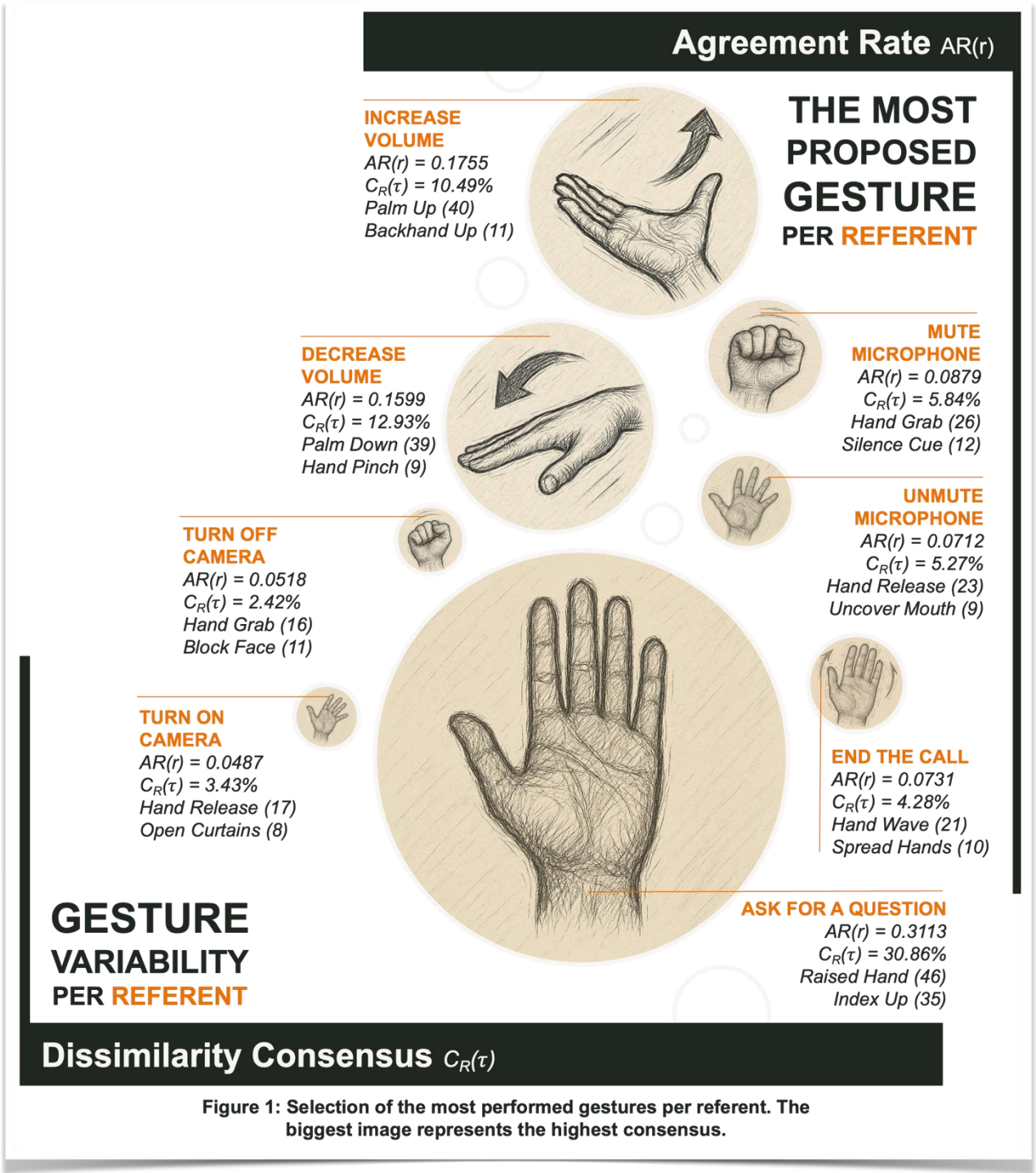


Figure 1: Selection of the most performed gestures per referent. The biggest image represents the highest consensus.

Agreement on **Gesture Proposals**

Findings

- *824 user-defined gesture proposals*
- *133 distinct gestures*
- *Number of proposals per referent ranging from 20 to 41*
- *58 distinct gestures met a threshold of at least two occurrences*
- *90.8% of all proposals consisted of a single gesture*
- *8.3% included two gestures in sequence, and 0.9% comprised three or more*
- *74.39% of proposed gestures are dynamic (movements), while approximately 25.49% are static (poses)*

Dichotomous **Gestures**

Findings

For Volume Controls

- *R1 Increase Volume*
- *R2 Decrease Volume*

Participants proposed upward and downward hand palm movements

- *45.63% of participants used opposite gestures (random)*

For Microphone Controls

- *R3 Mute Microphone*
- *R4 Unmute Microphone*

Participants proposed closing and opening hand movements

- *30.10% of participants used opposite gestures*

Dichotomous **Gestures**

Findings

For Camera Controls

- *R5 Turn off Camera*
- *R6 Turn on Camera*

Participants proposed dispersed gesture choices

Lowest consensus scores among all referents

- *20.39% of participants used opposite gestures*
- *39 distinct gesture proposals for R5*
- *41 distinct gesture proposals for R6*

Effect of Prior Gesture Experience on Agreement Rates

Findings

Possible bias?

*64 participants declared **prior** hand gesture **experience***

*60 proposed gestures that included **common manipulative** VR or Kinect-style actions such as **Swipe**, **Push**, **Point**, or **Grab***

Agreement rates:

HG group and No-HG group:

- *interpretation summarizes a relative consensus between the two groups*
- *prior HG experience does not substantially affect collective consensus*

Effect of Prior Gesture Experience on Agreement Rates

Findings

Referent	Most Frequent	Second Most Frequent
R_1 Increase Volume	Palm Up (16)	Hand Pinch (4)
R_2 Decrease Volume	Palm Down (12)	Hand Pinch (5)
R_3 Mute Microphone	Hand Grab (12)	Silence Cue (4)
R_4 Unmute Microphone	Hand Release (13)	Point Camera (4)
R_5 Turn Off Camera	Block Face (5)	Hand Swipe / Raised Hand (3)
R_6 Turn On Camera	Hand Release (5)	Open Curtains (5)
R_7 Ask for a Question	Index Up (16)	Raised Hand (12)
R_8 End Call	Hand Wave (6)	Hand Grab / Spread Hands (4)

Most/second most proposed gestures among (No-HG)

Referent	Most Frequent	Second Most Frequent
R_1 Increase Volume	Palm Up (24)	Backhand Up (9)
R_2 Decrease Volume	Palm Down (27)	Hand Pinch / Pinch Hold (4)
R_3 Mute Microphone	Hand Grab (14)	Silence Cue (8)
R_4 Unmute Microphone	Hand Release (10)	Uncover Mouth (6)
R_5 Turn Off Camera	Hand Grab (14)	Block Face / Raised Hand (6)
R_6 Turn On Camera	Hand Release (12)	Raised Hand / Hand Wave (4)
R_7 Ask for a Question	Raised Hand (34)	Index Up (19)
R_8 End Call	Hand Wave (15)	Hand Grab / Spread Hands (6)

Most/second most proposed gestures among (HG)

Variations in Gesture Execution

Linear Discriminant Analysis (LDA) to reduce the dimensionality of the gesture description embeddings from 1,024 dimensions to two.

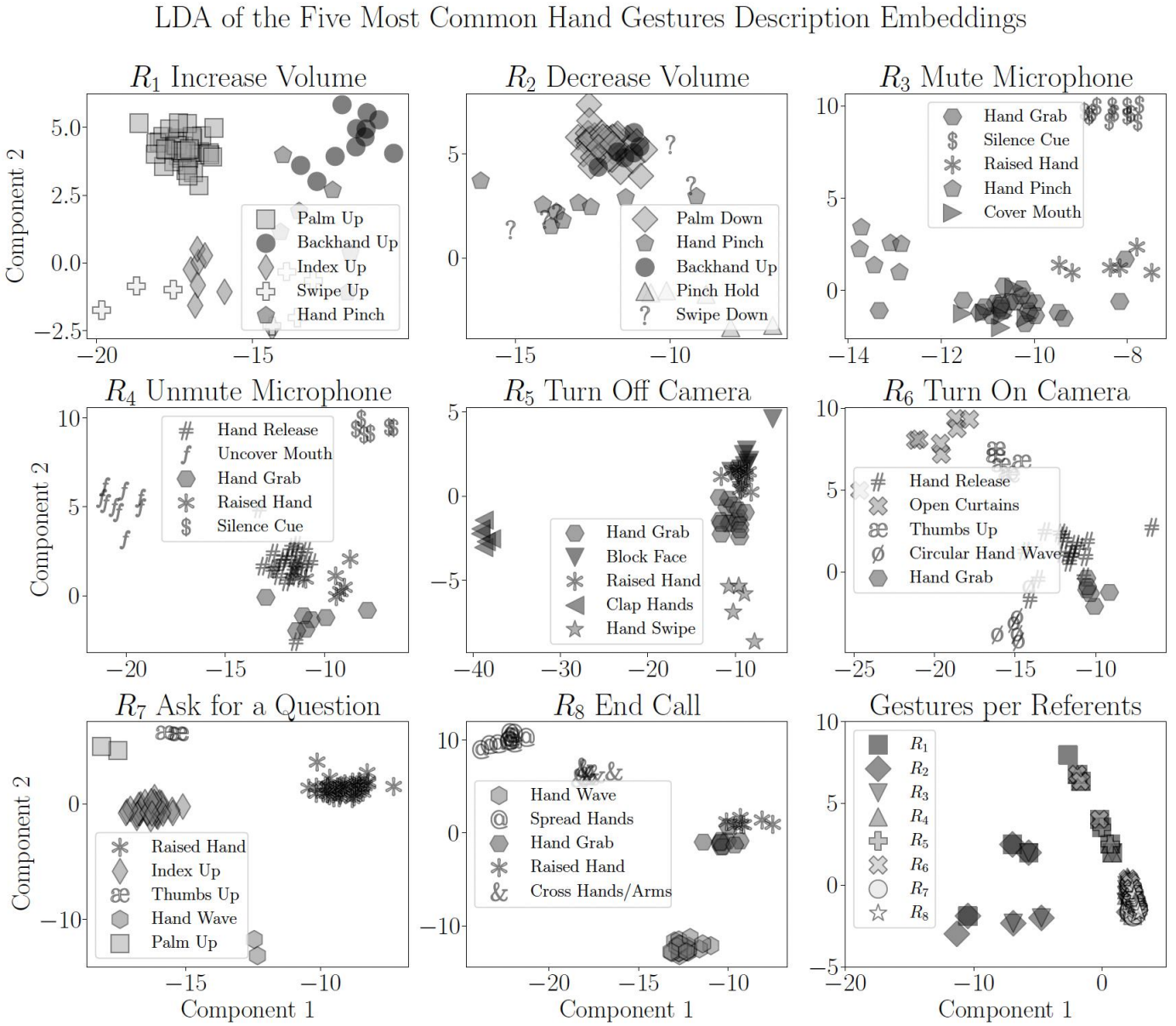


Fig. 5: Clustering of Similar Gestures per Referents

- *Many **unique proposals per referent** (20 to 41 from 103) and generally **low agreement**, except for a single referent that exhibited high agreement (R7).*
- *The remaining seven referents (R1-R6, R8) showed significant variability in gesture proposals.*
- ***Low agreement** rates do not necessarily indicate a lack of usability or intuitiveness but **reflect the expansive nature of free-form gesture creation***

Conclusions

- *Context can influence on gesture proposals*
- *Social Acceptability and Professionalism Constraints*
- *Cognitive Load and Adaptation of Mental Models*
- *Prior Experience and Legacy Bias*

- ***Implications for Gesture Set Design**: Our findings **highlight** the **critical role** of contextual factors in shaping gesture vocabularies for hybrid meeting control.*

Conclusions

Eliciting User-defined Mid-air Hand Gestures For Hybrid Meeting Platform Control: Insights And Design Implications

Elizabete Munzlinger[1,2], Fabricio Batista Narcizo[1,2], **Renata Briet**[3], Mario Tadashi Shimanuki[3], Ted Vucurevich[2], and Dan Witzner Hansen[1]

1 IT UNIVERSITY OF COPENHAGEN

2 GN Advanced Science

Innovation Fund Denmark

3 INSTITUTO FEDERAL São Paulo Câmpus Caraguatatuba

PIBIFESP Programa Institucional de Bolsas de Iniciação Científica e Tecnológica

Thank you Q&A