Mapping the Landscape of Non-Verbal Communication Tracking in Videoconferencing: A Systematic Literature Review

Carlota Vazquez Gonzalez
Department of Informatics
King's College London, UK
Carlota.vazquez_gonzalez@kcl.ac.uk

Rita Borgo
Department of Informatics
King's College London, UK
Rita.borgo@kcl.ac.uk

Timothy Neate
Department of Informatics
King's College London, UK
Timothy.neate@kcl.ac.uk

Non-verbal cues help people communicate emotion, attention, and intent, but are harder to perceive in videoconferencing because of restricted framing, camera use, and audio-visual limitations. We present a systematic review of 35 papers on systems that track, interpret, or represent non-verbal communication cues in videoconferencing. We map how these systems capture cues, present outputs, and are evaluated. Most rely on camera and microphone inputs, provide real-time feedback, and use bespoke labels or representations. Evaluations commonly use controlled studies and convenience samples, with limited participatory design and little attention to users with disabilities. We derive a minimum reporting structure for cue-to-feedback design and identify priorities for accountable labelling, ecological and participatory evaluation, and accessible, user-controlled systems.

Systematic literature review, non-verbal communication, videoconferencing, dataset

1. INTRODUCTION

Non-verbal communication extends and shapes the meaning of verbal communication through facial expressions, gestures, posture, gaze, and paralinguistic features such as tone, pitch, pauses, interruptions, and speech rate (7; 16; 8; 9). These cues may be instinctual, learned, imitated, or intentional, but their meaning is often implicit, ambiguous, and context-dependent (8). Unlike structured systems such as sign language or symbol sets, which have more predefined meanings (55; 56), non-verbal cues are interpreted through shared understanding, social norms, culture, and situational context (10). This ambiguity can be particularly challenging for some users, including autistic and disabled people, for whom interpreting or producing non-verbal cues may involve additional barriers (59; 5; 6).

Videoconferencing makes these cues harder to perceive and interpret. Even when cameras and microphones are enabled, restricted framing, limited visual fields, and audio delays can constrain the exchange of gaze, gesture, posture, facial expression, and paralinguistic cues (18; 15). Participants also manage their on-screen presence by turning cameras off or reducing expressiveness because of fatigue, privacy, or self-presentation concerns (19; 20; 22; 21). As a result, users may rely more heavily on substitute cues such as voice tone, chat messages, or explicit reactions, which can affect engagement, comprehension, and interaction in professional, educational, and social settings (23; 16; 17). For example, camera-off participation has been associated with multitasking, which can make it harder for speakers to gauge attention and engagement (24; 54).

Recent developments in computer vision, machine learning, and multimodal sensing have enabled systems that track, interpret, or present non-verbal cues in videoconferencing contexts (12; 11; 13; 14). However, existing research remains fragmented across application areas, input modalities, output formats, cue labels, and evaluation methods. This systematic literature review maps 35 studies on non-verbal communication tracking in videoconferencing. It examines how systems capture, label, present, and evaluate non-verbal cues, and identifies gaps in current design and testing practices.

© Vazquez Gonzalez et al.
Published by BCS Learning and Development Ltd.
Proceedings of BCS HCI 2026, UK

This paper makes two contributions. First, it provides a systematic map and taxonomy of 35 papers on non-verbal communication tracking in videoconferencing, showing how systems capture, label, present, and evaluate non-verbal cues. Second, it synthesises implications for future work, particularly around transparent cue labelling, participatory and ecological evaluation, accessibility, and responsible use.

1.1. Research questions

This systematic literature review provides an overview of papers on non-verbal tracking in videoconferencing identified through a bounded search and snowballing process. It is guided by the following research questions:

RQ1: How do videoconferencing systems present tracked non-verbal communication cues to users?

RQ2: What input modalities, recorded features, output modalities, and cue labels are used in these systems?

RQ3: How are these systems evaluated, and which users, settings, and methods are represented?

2. BACKGROUND

Non-verbal cues are used across healthcare, professional, and educational settings to infer emotion, attention, understanding, engagement, and interactional intent (23; 25; 26; 14; 13; 28; 27). In videoconferencing, these cues are mediated through cameras, microphones, screen layouts, and platform features, making their capture and interpretation dependent on both social context and technical design (8; 7; 29). Earlier work on non-verbal cue recognition has often focused on robotics or computer vision more broadly (33; 30), while videoconferencing-specific systems are still emerging.

Existing systems range from simple platform-level reactions to research prototypes. Widely used tools such as Microsoft Teams and Apple FaceTime now use gesture recognition to trigger visual reactions, animations, or effects (35; 34). These features provide lightweight ways to express reactions, but they rely on a small set of predefined gestures and can still be misunderstood when visual effects are interpreted differently by users (36). More advanced systems, such as Meeting Coach, Emodash, and TeamSpiritous, use automated analyses of gestures, emotions, or paralinguistic behaviour to provide feedback during or after video-mediated interaction (14; 13; 28; 27). Many of these systems are multimodal, meaning that they combine inputs such as video, audio, or sensor data rather than relying on a single source (11).

Together, these examples show an emerging design space for non-verbal communication tracking in videoconferencing. However, the field remains fragmented in what systems capture, how they label or present cues, and how they are evaluated. This motivates a systematic literature review that maps current approaches and identifies gaps in design, evaluation, and accessibility.

3. METHODS

We conducted a systematic mapping literature review to classify and synthesise how non-verbal communication tracking has been studied in videoconferencing systems. This approach is appropriate for organising a heterogeneous technical and HCI literature, identifying recurring design and evaluation patterns, and making gaps visible through a transparent search, screening, and coding process. We followed PRISMA reporting guidance where applicable, while focusing on descriptive mapping rather than meta-analysis or assessment of intervention effectiveness (37). The review involved defining the scope, identifying and screening candidate papers, supplementing the dataset through snowballing, and coding the final papers qualitatively and quantitatively.

3.1. Scope

The review focused on systems that track, interpret, or represent implicit non-verbal communication cues in videoconferencing. We excluded sign language because it involves explicit, lexically structured signs and has already been covered by dedicated reviews (1; 2; 3). We recognise, however, that sign language AI research has also raised important concerns around bias and Deaf-led research agendas (61). Our focus was instead on more ambiguous and context-dependent cues, such as gestures, affective signals, gaze, facial expressions, posture, and paralinguistic features.

We also excluded systems centred on co-located, robotic, or VR-only interaction without a videoconferencing component, as these areas have distinct bodies of related work (4; 33). Included papers had to meet three requirements:

- **Non-verbal communication:** the paper described tracking, interpretation, or representation of implicit non-verbal cues, such as gestures, affective signals, gaze, facial expressions, posture, or paralinguistic features.

- **Videoconferencing:** the system was used in a hybrid or online meeting, call, or collaboration context where users interacted remotely through computers.
- **Contribution:** the paper presented the creation, demonstration, or evaluation of an artefact that supported non-verbal communication tracking or representation.

3.2. Dataset

Following PRISMA reporting principles (37), we created a dataset of 35 papers on non-verbal communication tracking in videoconferencing: 28 papers identified through the ACM Digital Library query and seven through snowballing. Figure 1 summarises the identification, screening, eligibility, and snowballing process.

3.2.1. Identification

We first conducted a preliminary search to identify key papers and terminology used in this area. These papers informed a targeted ACM Digital Library query combining non-verbal communication and feedback terms with videoconferencing, remote meeting, online learning, and remote work terms. The query searched titles and keywords, applied no publication date restrictions, and was last run on 7 May 2026. It returned 1,205 candidate papers published between 1992 and 2026. The full query is provided in Appendix A.

3.2.2. Screening

After removing 35 duplicate records, 1,170 papers remained. We screened titles, keywords, and abstracts against the review scope. Exclusion terms included robot, machine learning, algorithms, MOOC, virtual reality, streaming, natural language, video game, neural network, linear regression, speech recognition, LLM, deep learning, model, and YouTube. These terms were developed iteratively and were used as flags for manual review rather than automatic exclusion criteria. Papers were excluded only when the flagged topic was the primary focus rather than a method used to support videoconferencing-based non-verbal communication tracking. All records were manually assessed, and no record was excluded solely because a flagged term appeared.

3.2.3. Eligibility

The remaining papers from the screening were reviewed in full against four eligibility criteria. Papers had to: (EC1) be peer-reviewed research articles or papers; (EC2) explicitly relate to non-verbal communication in online communication systems; (EC3) present, demonstrate, or evaluate an interaction involving non-verbal communication; and (EC4) focus on implicit or non-universally defined non-verbal cues rather than structured sign language. Papers outside this scope were excluded from the dataset. The first author assessed all full texts. A reduced set of candidate papers was then independently reviewed by all three authors, with disagreements resolved through discussion and consensus. Twenty-eight papers from the ACM query met all scope and eligibility criteria.

3.2.4. Snowballing

To supplement the database search, we conducted backwards and forward snowballing using Google Scholar. We reviewed references and citations from the initial dataset and foundational papers, applying the same scoping and eligibility criteria. The first author identified candidate papers, and all three authors confirmed their inclusion. Seven papers not returned by the ACM query were added, producing a final dataset of 35 papers.

3.2.5. Use of AI tools

AI-assisted tools were used to support writing, editing, and checking descriptive statistical summaries of the systematic coding. The author retained responsibility for all intellectual work: inclusion and exclusion decisions, coding decisions, statistical outputs, interpretations, and the final manuscript were reviewed, verified, and approved by the author.

3.3. Analysis

We developed a coding system to address the research questions and map patterns across the dataset. Three researchers independently coded the papers. Disagreements were discussed and resolved through consensus before producing the final coded dataset.

The final codebook, shown in Appendix C, contains 15 categories with 2 to 11 sub-codes. Four categories were adapted from Mack et al. (38), focusing on user study methods and target communities. The contribution category was adapted from Wobbrock and Kientz (39). The remaining categories were developed iteratively from the research questions and covered design context, inputs, outputs, cue labels, interaction features, and evaluation practices.

Most categories allowed multiple codes per paper, so some percentages exceed 100%. To assess coding consistency, we calculated inter-rater reliability using Cohen's Kappa (40) and interpreted agreement using the Landis and Koch benchmark scale (43). The final codebook and agreement values are reported in Appendix C.

3.3.1. Qualitative and quantitative analysis

We qualitatively coded all 35 papers and calculated descriptive frequencies for each code.

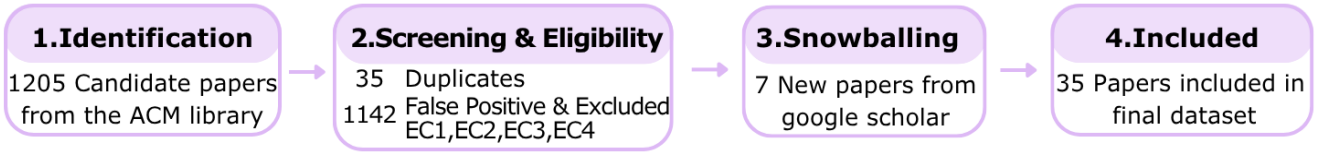

Figure 1: *Flowchart summarising the process used to create the dataset.*

4. RESULTS

We report the results across three areas: design, system configuration, and evaluation. The codes were not mutually exclusive, so percentages indicate the proportion of papers assigned each code rather than exclusive categories. Table 1 in Appendix B summarises the main frequencies.

4.1. Design

4.1.1. Presentation

Presentation formats varied across the reviewed systems. Most systems used visual elements to communicate non-verbal feedback, including graphs (54%), text (49%), and images (43%), while 63% used bespoke or less common formats. Feedback was usually given in real-time (83%) rather than retrospective (29%). Most systems also combined multiple cues or modalities (83%), reflecting the multimodal nature of non-verbal communication. For example, Kumar et al. (48) introduced "gaze bubbles" to visualise shared gaze attention during online meetings. MeetCues provides another example of combined real-time and retrospective feedback, using likes, clarify clicks, comments, emoji-cloud visualisations, timelines, and post-meeting audio snippets to support awareness and reflection during online meetings (31).

4.1.2. Context

The reviewed systems were most often designed for industry meetings (37%) or education (26%), while 40% addressed broader or multi-context settings, including family communication (50), general communication (58), and call centres (57). Target demographics were frequently unspecified (40%), followed by students (31%) and workers (31%). People with disabilities were rarely identified as a target group (3%).

Collaboration designs were mainly one-to-one (43%) or many-to-many (31%). In this review, one-to-one refers to dyadic video calls, one-to-many to settings such as a presenter or educator addressing an audience, many-to-one to multiple participants providing cues to one focal person or system, and many-to-many to group interactions where cues from several participants are tracked or represented. These categories show that systems were usually designed either for focused dyadic interaction or broader group awareness.

4.2. System

4.2.1. Inputs

Camera input dominated the reviewed systems (80%), followed by microphone input (69%). This reflects the emphasis on visually observable cues and paralinguistic cues such as tone, pitch, pauses, or speech rhythm. External physical sensors were less common (20%). The most frequently recorded features were context-specific or bespoke features (49%), followed by facial expressions (46%), gestures (34%), and eyes (31%). For example, Ou et al. (62) used a digital pen to capture drawings as expressive gestures over video streams. KAIROS illustrates this broader sensing approach by combining mobile, smartwatch, earable, and audio data to capture body cues during virtual meetings (32).

4.2.2. Outputs

The most common output modalities were video (66%), sound (57%), and image (34%). Cue labels were highly fragmented: 83% of papers used other or context-specific labels, while smaller groups used paralanguage (29%), valence/arousal (14%), or Ekman's six basic emotions (11%). In this paper, Ekman refers to the use of six basic emotion categories rather than facial action coding (41). For example, Watanabe et al. (49) used a gauge interface to visualise self and group engagement in real time during online meetings. Overall, the high use of bespoke labels suggests that systems often adapt their representations to specific contexts rather than relying on standardised non-verbal cue vocabularies.

4.3. Evaluation

Evaluation practices were dominated by controlled experiments (60%) and user studies (60%). These commonly involved predefined tasks, interaction with the system, and post-use feedback on usability or functionality. Only one paper used participatory or co-design methods (3%), involving target users earlier in the design process (50). This suggests that most systems were evaluated after development rather than shaped through iterative participation.

Participant samples were also narrow. Although many systems were designed for workplace, educational, or broader communication settings, students were the most common participant group (54%). People with disabilities were rarely included as participants (3%), aligning with their limited presence as a target demographic. For example, some systems addressed accessibility-related scenarios (47), but disability-focused design and testing remained uncommon across the dataset.

Testing locations mainly reflected either remote use or controlled research settings. Online studies were most common (66%), followed by studies conducted in or near research laboratories (46%). The reviewed papers were primarily empirical and artefact-oriented, with all papers contributing empirical findings and artefacts. This indicates a field focused on building and testing systems, with less emphasis on theory-building, methodological development, or long-term deployment.

5. DISCUSSION

5.1. Design Diversity and Labelling

The reviewed systems show a field that is diverse by necessity rather than simply inconsistent. Non-verbal cue tracking tools are shaped by their setting, users, interaction model, and communication goals, which makes a single universal design framework unlikely to fit all cases (44; 45). However, this does not mean that the field should remain difficult to compare. The value of the taxonomy is therefore not to impose a universal model of non-verbal communication, but to make design choices comparable across otherwise heterogeneous systems. At minimum, future papers should report what cue is captured, what it is assumed to mean, how it is transformed into a label or representation, who receives the output, when feedback is shown, and whether users can refuse, correct, or contest the interpretation.

This is particularly important for cue labels. The reviewed systems often used bespoke labels, spectrum-based representations, or context-specific categories rather than predefined frameworks such as Ekman's six basic emotions or valence-arousal models (41; 42; 46). This flexibility is appropriate because emotion and non-verbal meaning are not only internal states to be detected, but also social and situated phenomena that emerge through interaction (65). Yet this flexibility also creates a responsibility problem. Labels such as engagement, attention, stress, or emotion may be read as objective measurements unless systems make uncertainty and context explicit. Future systems could therefore treat ambiguity as a design condition by showing confidence, allowing multiple interpretations, preserving user control, and avoiding labels that imply unjustified certainty (66). This is especially important in educational and workplace contexts, where inferred labels may influence judgements about participation, performance, or professionalism.

5.2. System Complexity and Modality

The dominance of complex and multimodal systems suggests that researchers recognise non-verbal communication as distributed across face, body, voice, timing, and interactional context. However, adding more inputs does not automatically make a system more meaningful or useful. A camera, microphone, gaze tracker, or sensor may increase the amount of data captured, but the design challenge is deciding which cues should be surfaced, to whom, and for what purpose. In this sense, multimodality should not be treated only as a technical achievement, but as an interaction design problem.

The taxonomy therefore helps shift attention from simply adding more sensing to assessing input-output appropriateness. A system that detects facial expression, speech rhythm, and gaze still has to decide whether these signals become a graph, a warning, a dashboard, an emoji, a score, or no visible output at all. These design choices shape how users understand themselves and others during a call. Future systems should therefore justify why each input modality is needed and report the social role of each output: whether it supports private self-reflection, group awareness, accessibility, moderation, training, or assessment. This would make it easier to compare systems that use different technical methods but serve similar interactional purposes.

5.3. Evaluation and Testing Approaches

The reviewed literature is strongly artefact-oriented, but evaluation often happens after systems have already been designed. This limits what studies can reveal. Post-hoc usability testing can show whether a tool is understandable or acceptable in a short task, but it is less able to reveal whether inferred non-verbal cues are useful, uncomfortable, ignored, contested, or misused in everyday communication. Because non-verbal meaning is situated, evaluations should examine how feedback changes interaction over time, not only whether participants can use a prototype during a controlled session (67).

This suggests three methodological priorities. First, participatory and co-design approaches should be used earlier, particularly when systems target

groups who may be affected by misinterpretation or monitoring (51; 52; 53). Second, studies should move beyond one-off laboratory tasks towards more ecological settings, including real meetings, classrooms, telehealth sessions, or longitudinal deployments. Third, evaluations should examine social consequences as well as usability: who benefits from the feedback, who feels exposed by it, and whether the system changes power dynamics between speakers, students, managers, educators, clinicians, or participants. These questions would help move the field from prototype validation to evaluating the interactional consequences of non-verbal feedback.

5.4. Inclusivity and Accessibility

Accessibility should not be treated as a separate application area for non-verbal cue tracking, but as a condition for whether these systems are valid and fair. The issue is not only that the reviewed work rarely targeted or included people with disabilities, but also that many input and output choices assume that communication is visible, audible, stable, and interpretable through standard cues (63; 64). Those assumptions may not hold for Deaf or hard-of-hearing users, blind or low-vision users, autistic users, people who stutter, or people with aphasia (47; 52; 59; 60).

Ability-based design offers a useful direction because it asks systems to adapt to users' abilities rather than expecting users to conform to a narrow model of interaction (68). Applied to videoconferencing, this means that future systems should not only track more cues, but also support different ways of producing, receiving, hiding, correcting, and interpreting cues. For example, some users may benefit from additional channels that make tone, turn-taking, or attention more explicit, while others may need stronger privacy controls or the ability to disable inference altogether. Designing with diverse access needs from the start would therefore improve accessibility while also exposing broader assumptions about visibility, attention, expressiveness, and social judgement in video calls.

Overall, this systematic review shows that non-verbal cue tracking in videoconferencing is being actively developed on a technical level; it remains conceptually fragmented. The field now needs to move in three directions: from cue detection to accountable interpretation, from isolated prototypes to situated evaluation, and from assumed norms of communication to accessible and user-controlled design. The taxonomy contributes to this shift by making the design space visible and by identifying the reporting dimensions future work should make explicit.

6. LIMITATIONS

This systematic review has three main limitations. First, the dataset was primarily built from the ACM Digital Library and supplemented through Google Scholar snowballing. This produced a focused HCI-oriented dataset, but searches in IEEE Xplore, Scopus, Web of Science, Springer, or other databases may identify additional relevant papers. The review should therefore be read as a bounded map of this research area rather than an exhaustive account of all related work.

Second, the search terms and exclusion filters may have shaped the dataset. Terminology varies across videoconferencing, CSCW, affective computing, accessibility, education, and computer vision. As a result, relevant studies using terms such as video call, online call, body cues, social signals, or meeting analytics may not have been captured. Exclusion terms were used as flags for manual review rather than automatic removal, but the search strategy may still have influenced which papers were included.

Third, the coding framework involved interpretive judgement. We mitigated this through multiple coders, inter-rater reliability calculation, and consensus discussion, but other researchers may classify some systems differently, particularly for categories such as cue labels, collaboration design, or bespoke output formats. The taxonomy should therefore be understood as a transparent analytical map rather than a fixed classification of the field.

7. CONCLUSION

This systematic review maps 35 papers on non-verbal communication tracking in videoconferencing. The reviewed work shows a technically active but conceptually fragmented field: systems often combine multiple input modalities and provide real-time feedback, but they use diverse labels, bespoke representations, controlled evaluations, and limited participatory design. Future work would benefit from clearer reporting of what is captured and inferred, more ecological and participatory evaluation, and stronger attention to accessibility and user control. Moving beyond cue detection towards accountable, context-sensitive support will be essential if these systems are to meaningfully improve mediated interaction.

REFERENCES

[1] Madhiarasan, M. & Roy, P. A Comprehensive Review of Sign Language Recognition: Different Types, Modalities, and Datasets. (2022), https://arxiv.org/abs/2204.03328

[2] Renjith, S. & Manazhy, R. Sign language: a systematic review on classification and recognition. *Multimedia Tools And Applications*. **83**, 77077-77127 (2024,2)

[3] Bhatia, P. & Wadhawan, A. Sign Language Recognition Systems: A Decade Systematic Literature Review. *Archives Of Computational Methods In Engineering*. pp. 785-813 (2019,1)

[4] McColl, D., Hong, A., Hatakeyama, N., Nejat, G. & Benhabib, B. A Survey of Autonomous Human Affect Detection Methods for Social Robots Engaged in Natural HRI. *Journal Of Intelligent & Robotic Systems*. **82**, 101-133 (2015,8)

[5] Akter, T., Cha, Y., Figueira, I., Branham, S. & Piper, A. "If I'm supposed to be the facilitator, I should be the host": Understanding the Accessibility of Videoconferencing for Blind and Low Vision Meeting Facilitators. (Association for Computing Machinery,2023), https://doi.org/10.1145/3597638.3608420

[6] Wong, A., Bryck, R., Baker, R., Hutt, S. & Mills, C. Using a Webcam Based Eye-tracker to Understand Students' Thought Patterns and Reading Behaviors in Neurodivergent Classrooms. (Association for Computing Machinery,2023), https://doi.org/10.1145/3576050.3576115

[7] Shafique, Z., Wang, H. & Tian, Y. Nonverbal Communication Cue Recognition: A Pathway to More Accessible Communication. *2023 IEEE/CVF Conference On Computer Vision And Pattern Recognition Workshops (CVPRW)*. pp. 5666-5674 (2023)

[8] Knapp, M. & Hall, J. Nonverbal communication in human interaction. (Thomson Wadsworth,2006)

[9] Gordon, R. & Druckman, D. Nonverbal Behaviour as Communication: Approaches, Issues, and Research. *The Handbook Of Communication Skills*. pp. 76-130 (2018)

[10] Pang, H., Zhou, X. & Chu, M. Cross-cultural differences in using nonverbal behaviors to identify indirect replies. *Journal Of Nonverbal Behavior*. **48** (2024,2), https://link.springer.com/article/10.1007/s10919-024-00454-z

[11] Zhou, Q., Suraworachet, W. & Cukurova, M. Detecting non-verbal speech and gaze behaviours with multimodal data and computer vision to interpret effective collaborative learning interactions. *Education And Information Technologies*. **29**, 1071-1098 (2023,11)

[12] Singh Tomar, P., Mathur, K. & Suman, U. Unimodal approaches for emotion on: A systematic review. *Cognitive Systems Research*. **77** pp. 94-109 (2023), https://www.sciencedirect.com/science/article/pii/S1389041722000626

[13] Samrose, S., Zhao, R., White, J., Li, V., Nova, L., Lu, Y., Ali, M. & Hoque, M. CoCo: Collaboration Coach for Understanding Team Dynamics during Video Conferencing. *Proceedings Of The ACM On Interactive, Mobile, Wearable And Ubiquitous Technologies*. **1**, 1-24 (2018)

[14] Samrose, S., McDuff, D., Sim, R., Suh, J., Rowan, K., Hernandez, J., Rintel, S., Moynihan, K. & Czerwinski, M. Meetingcoach: An intelligent dashboard for supporting effective & inclusive meetings. *Proceedings Of The 2021 CHI Conference On Human Factors In Computing Systems*. (2021)

[15] Fauville, G., Luo, M., Muller Queiroz, A., Bailenson, J. & Hancock, J. Nonverbal mechanisms predict zoom fatigue and explain why women experience higher levels than men. *Available At SSRN 3820035*. (2021)

[16] Tomprou, M., Kim, Y., Chikersal, P., Woolley, A. & Dabbish, L. Speaking out of turn: How video conferencing reduces vocal synchrony and collective intelligence. *PLOS ONE*. **16** (2021,3)

[17] Belt, E. & Lowenthal, P. Synchronous video-based communication and online learning: an exploration of instructors' perceptions and experiences. *Education And Information Technologies*. **28** (2022,10)

[18] Bailenson, J. Nonverbal overload: A theoretical argument for the causes of zoom fatigue. *Technology, Mind, And Behavior*. **2** (2021,2)

[19] Fauville, G., Queiroz, A., Luo, M., Hancock, J. & Bailenson, J. Impression formation from video conference screenshots: The role of gaze, camera distance, and angle.. *Technology, Mind, And Behavior*. **3** (2022,1), https://assets.pubpub.org/bmfkdqen/51641509251997.pdf

[20] Sun, Y., Guo, W. & Wu, X. Camera Shy in Online Synchronous Class: A Qualitative Study in

College Students. *Blended Learning : Lessons Learned And Ways Forward* . pp. 111-123 (2023)

[21] Shockley, K., Gabriel, A., Robertson, D., Rosen, C., Chawla, N., Ganster, M. & Ezerins, M. The fatiguing effects of camera use in virtual meetings: A within-person field experiment.. *Journal Of Applied Psychology*. **106**, 1137-1155 (2021,8), https://psycnet.apa.org/fulltext/2021-77825-003.pdf

[22] Sederevičiūtė-Pačiauskienė, Ž., Valantinaitė, I. & Asakaviciute, V. 'Should I Turn On My Video Camera?' The Students' Perceptions of the Use of Video Cameras in Synchronous Distant Learning. *Electronics*. **11** pp. 813 (2022,3)

[23] Faucett, H., Lee, M. & Carter, S. I Should Listen More: Real-time Sensing and Feedback of Non-Verbal Communication in Video Telehealth. *Proc. ACM Hum.-Comput. Interact.*. **1** (2017,12), https://doi.org/10.1145/3134679

[24] Karl, K., Peluchette, J. & Aghakhani, N. Virtual work meetings during the COVID-19 pandemic: The good, bad, and ugly. *Small Group Research*. **53**, 343-365 (2022,5)

[25] Zhu, Q., Zhang, X., Dai, S., Satake, N. & Wang, H. ZoombaTogether: A Video Conference Add-on for Generating Interactive Visual Feedback for Online Group Exercise Through On-The-Fly Pose Tracking. *Companion Publication Of The 2023 ACM Designing Interactive Systems Conference*. pp. 261-265 (2023), https://doi.org/10.1145/3563703.3596653

[26] Aggarwal, D. Supporting Bodily Communication in Video-based Clinical Consultations. *Proceedings Of The 2016 CHI Conference Extended Abstracts On Human Factors In Computing Systems*. pp. 188-192 (2016), https://doi.org/10.1145/2851581.2859019

[27] Benke, I., Schneider, M., Liu, X. & Maedche, A. Teamspiritous - a retrospective emotional competence development system for video-meetings. *Proceedings Of The ACM On Human-Computer Interaction*. **6**, 1-28 (2022)

[28] Ez-zaouia, M., Tabard, A. & Lavoué, E. Emodash: A Dashboard supporting retrospective awareness of emotions in online learning. *International Journal Of Human-Computer Studies*. **139** pp. 102411 (2020)

[29] Montepare, J. Nonverbal Behavior in the Digital Age: Meanings, Models, and Methods. *Journal Of Nonverbal Behavior*. **38**, 279-281 (2014,5)

[30] Costea, D., Marcu, A., Lazar, C. & Leordeanu, M. Maia: A Real-time Non-Verbal Chat for Human-AI Interaction. *ArXiv.org*. (2024,2), https://arxiv.org/abs/2402.06385

[31] Aseniero, B. A., Constantinides, M., Joglekar, S., Zhou, K. and Quercia, D. 2020. MeetCues: Supporting Online Meetings Experience. In *Proceedings of the 2020 IEEE Visualization Conference (VIS)*. IEEE, 236–240. DOI: https://doi.org/10.1109/VIS47514.2020.00054

[32] Choi, J.-H., Constantinides, M., Joglekar, S. and Quercia, D. 2021. KAIROS: Talking Heads and Moving Bodies for Successful Meetings. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications (HotMobile '21)*. ACM, New York, NY, USA, 30–36. DOI: https://doi.org/10.1145/3446382.3448361

[33] Urakami, J. & Seaborn, K. Nonverbal Cues in Human-robot Interaction: A Communication Studies Perspective. *ACM Transactions On Human-Robot Interaction*. **12** (2022,12)

[34] Support, Apple. Use Reactions, Presenter Overlay and other effects when video conferencing on Mac – Apple Support (UK). *Apple Support*. (2024,12), https://support.apple.com/en-gb/105117

[35] Support, Microsoft. Express yourself in Microsoft Teams meetings with live reactions - Microsoft Support. *Microsoft.com*. (2019), https://support.microsoft.com/en-gb/office/express-yourself-in-microsoft-teams-meetings-with-live-reactions-a8323a40-3d07-4129-934b-305370a36e21?utm

[36] Petridis, S. & Chilton, L. Human Errors in Interpreting Visual Metaphor. *Proceedings Of The 2019 On Creativity And Cognition*. (2019,6)

[37] Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., Mulrow, C., Shamseer, L., Tetzlaff, J., Akl, E., Brennan, S. *The Prisma 2020 Statement: An Updated Guideline For Reporting Systematic Reviews*. (2020)

[38] Mack, K., McDonnell, E., Jain, D., Lu Wang, L., E. Froehlich, J. & Findlater, L. What Do We Mean by "Accessibility Research"? A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. *Proceedings Of The 2021 CHI Conference On Human Factors In Computing Systems*. (2021), https://doi.org/10.1145/3411764.3445412

[39] Wobbrock, J. & Kientz, J. Research contributions in human-computer interaction. *Interactions*. **23**, 38-44 (2016,4), https://doi.org/10.1145/2907069

[40] Fleiss, J. & Cohen, J. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational And Psychological Measurement*. **33**, 613-619 (1973,10)

[41] Ekman, P. 1992. An argument for basic emotions. *Cognition and Emotion*, 6(3–4), 169–200. DOI: https://doi.org/10.1080/02699939208411068

[42] Citron, F., Gray, M., Critchley, H., Weekes, B. & Ferstl, E. Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework. *Neuropsychologia*. **56** pp. 79-89 (2014,4)

[43] Landis, R. & Koch, G. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. **33**, 159-174 (1977,3)

[44] Vazquez Gonzalez, C., Neate, T. & Borgo, R. Trusting Tracking: Perceptions of Non-Verbal Communication Tracking in Videoconferencing. (Association for Computing Machinery,2025), https://doi.org/10.1145/3706598.3714306

[45] Beyan, Ç., Vinciarelli, A. & Bue, A. Co-Located Human-Human Interaction Analysis using Nonverbal Cues: A Survey. *ACM Computing Surveys*. (2023,10)

[46] Hoemann, K. Chapter Two - What's in an emotion word? The multiple roles of labeling in emotional functioning and well-being. *The Intersection Of Language With Emotion, Personality, And Related Factors*. **80** pp. 33-72 (2024), https://www.sciencedirect.com/science/article/pii/S007974212400001X

[47] Oomori, K., Shitara, A., Minagawa, T., Sarcar, S. & Ochiai, Y. A Preliminary Study on Understanding Voice-only Online Meetings Using Emoji-based Captioning for Deaf or Hard of Hearing Users. (Association for Computing Machinery,2020), https://doi.org/10.1145/3373625.3418032

[48] Kumar, C., Saini, B. & Staab, S. Enhancing Online Meeting Experience through Shared Gaze-Attention. *Extended Abstracts Of The CHI Conference On Human Factors In Computing Systems*. (2024), https://doi.org/10.1145/3613905.3651068

[49] Watanabe, K., Dengel, A. & Ishimaru, S. Metacognition-EnGauge: Real-time Augmentation of Self-and-Group Engagement Levels Understanding by Gauge Interface in Online Meetings. (Association for Computing Machinery,2024), https://doi.org/10.1145/3652920.3653054

[50] Hunter, S., Maes, P., Tang, A., Inkpen, K. & Hessey, S. WaaZam! supporting creative play at a distance in customized video environments. (Association for Computing Machinery,2014), https://doi.org/10.1145/2556288.2557382

[51] Treasure-Jones, T. & Joynes, V. Co-design of technology-enhanced learning resources. *The Clinical Teacher*. **15**, 281-286 (2017,11)

[52] Li, J., Wu, S. & Leshed, G. Re-envisioning Remote Meetings: Co-designing Inclusive and Empowering Videoconferencing with People Who Stutter. *Proceedings Of The 2024 ACM Designing Interactive Systems Conference*. pp. 1926-1941 (2024), https://doi.org/10.1145/3643834.3661533

[53] Stefaniak, J., Dousay, T., Asino, T. & Bagdy, L. Promoting Co-Design in Learning Design Research: Integrating Design-Based Research, Decision-Making, and Ethnographic Approaches. *TechTrends*. **69**, 645-655 (2025,4)

[54] H. Cao, C.-J. Lee, S. Iqbal, M. Czerwinski, P. N. Y. Wong, S. Rintel, B. Hecht, J. Teevan, and L. Yang, "Large Scale Analysis of Multitasking Behavior During Remote Meetings," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*, Yokohama, Japan, 2021, Article 448, 13 pages. doi: https://doi.org/10.1145/3411764.3445243.

[55] Grove, N. & Walker, M. The Makaton Vocabulary: Using manual signs and graphic symbols to develop interpersonal communication. *Augmentative And Alternative Communication*. **6**, 15-28 (1990,1), https://www.tandfonline.com/doi/abs/10.1080/07434619012331275284

[56] Carnett, A., Devine, B., Ingvarsson, E. & Esch, B. A Systematic and Quality Review of Augmentative and Alternative Communication Interventions that use Core Vocabulary. *Review Journal Of Autism And Developmental Disorders*. **12** (2023,8)

[57] Wu, J., Hagiya, T., Tang, Y. & Hoashi, K. Effects of objective feedback of facial expression

recognition during video support chat. *Proceedings Of The 16th International Conference On Mobile And Ubiquitous Multimedia*. pp. 293-297 (2017,11)

[58] Chan, A., Quek, F., Yamauchi, T. & Seo, J. Co-Verbal Touch: Enriching Video Telecommunications with Remote Touch Technology. (2021,10)

[59] Neate, T., Kladouchou, V., Wilson, S. & Shams, S. "Just Not Together": The Experience of Videoconferencing for People with Aphasia during the Covid-19 Pandemic. (2022,4)

[60] Nevsky, A., Bircanin, F., Cruice, M., Wilson, S., Simperl, E. & Neate, T. "I Wish You Could Make the Camera Stand Still": Envisioning Media Accessibility Interventions with People with Aphasia. *City Research Online (City University London)*. pp. 1-17 (2024,10)

[61] Desai, A., De Meulder, M., Hochgesang, J., Kocab, A. & Lu, A. Systemic Biases in Sign Language AI Research: A Deaf-Led Call to Reevaluate Research Agendas. *ArXiv (Cornell University)*. (2024,3)

[62] Ou, J., Fussell, S., Chen, X., Setlock, L. & Yang, J. Gestural communication over video stream. *International Conference On Multimodal Interfaces*. (2003,11)

[63] Seymour, W., Zhan, X., Cote, M. & Such, J. Who are CUIs Really For? Representation and Accessibility in the Conversational User Interface Literature. *arXiv preprint arXiv:2306.05228*. (2023), https://arxiv.org/abs/2306.05228

[64] Seaborn, K., Barbareschi, G. & Chandra, S. Not Only WEIRD but "Uncanny"? A Systematic Review of Diversity in Human-Robot Interaction Research. *arXiv preprint arXiv:2304.08750*. (2023), https://arxiv.org/abs/2304.08750

[65] Boehner, K., DePaula, R., Dourish, P. and Sengers, P. 2007. How emotion is made and measured. *International Journal of Human-Computer Studies*, 65(4), 275–291. DOI: https://doi.org/10.1016/j.ijhcs.2006.11.016

[66] Gaver, W. W., Beaver, J. and Benford, S. 2003. Ambiguity as a resource for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 233–240. DOI: https://doi.org/10.1145/642611.642653

[67] Suchman, L. A. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, Cambridge, UK.

[68] Wobbrock, J. O., Kane, S. K., Gajos, K. Z., Harada, S. and Froehlich, J. 2011. Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing*, 3(3), Article 9. DOI: https://doi.org/10.1145/1952383.1952384

Appendices

A. ACM SEARCH QUERY

```
[[[Title: feedback] OR [Title: non-verbal] OR [Title: "emotional
competence"] OR [Title: support*] OR [Title: understand*] OR [Title:
collaboration]] AND [[Title: video conferenc*] OR [Title: "remote
meeting"] OR [Title: video-conferenc*] OR [Title: "video-meeting"] OR [Title:
meeting] OR [Title: "online learning"] OR [Title: "remote work"]]] OR
[[[Keywords: non-verbal] OR [Keywords: feedback] OR [Keywords:
"emotional competence"] OR [Keywords: support*] OR [Keywords:
understand*] OR [Keywords: collaboration]] AND [[Keywords: video
conferenc*] OR [Keywords: "remote meetings"] OR [Keywords:
video-conferenc*] OR [Keywords: "video-meeting"] OR [Keywords: meeting] OR
[Keywords: "online learning"] OR [Keywords: "remote work"]]]
```

B. SUMMARY OF RESULTS

Table 1: *Summary of main coding frequencies across the 35 reviewed papers. Codes were not mutually exclusive. Percentages are rounded to the nearest whole number.*

Area	Category	Most frequent codes
Design	Visual presentation	Other/bespoke formats (63%), graphs (54%), text (49%), images (43%)
	Feedback timing	Real-time (83%), retrospective (29%), other (6%)
	Complexity	Complex combination of cues or modalities (83%), one feature (17%)
Context	Focus	Other/multi-context (40%), industry meetings (37%), education (26%), health (11%)
	Target demographic	No specific group/anyone (40%), students (31%), workers (31%), people with disabilities (3%)
	Collaboration design	One-to-one (43%), many-to-many (31%), one-to-many (20%), many-to-one (6%)
System	Inputs	Camera (80%), microphone (69%), other (23%), external physical sensor (20%)
	Recorded features	Other/context-specific features (49%), facial expression (46%),gestures (34%), eyes (31%)
	Outputs	Video (66%), sound (57%), image (34%), other (34%)
	Cue labels	Other/bespoke labels (83%), paralanguage (29%), valence/arousal (14%), Ekman six basic emotions (11%)
Evaluation	Participants	Students (54%), other (37%), professionals (31%), people with disabilities (3%)
	Methods	Controlled experiments (60%), user studies (60%), surveys (51%), interviews (31%)
	Location and contribution	Online/remote studies (66%), research laboratories (46%), empirical and artefact contributions (100%)

C. CODEBOOK

Category	Codes	Multiple codes	Pairwise Agreement	IRR	Level
Visual Presentation	Graphs; Text; Icon/emoji/emoticons; Images; Other	Yes	0.766	0.488	Moderate
Feedback	Real-time; Retrospective; Other	Yes	0.828	0.636	Substantial
Complexity	Simple; Complex	No	0.759	0.517	Moderate
System Inputs	Camera; Microphone; External physical sensor; Other	Yes	0.836	0.668	Substantial
Recorded Features	Eyes; Gestures; Facial expressions; Other	Yes	0.845	0.661	Substantial
System Outputs	Sound; Video; Image; External actuator; Other	Yes	0.862	0.714	Substantial
Focus	Industry meeting; Education; Entertainment; Health; Other	Yes	0.814	0.453	Moderate
Non-verbal communication Labels	Valance and Arousal; Paralanguage; Ekman 6 emotions; Plutchik; Other	Yes	0.828	0.439	Moderate
Target Demographic	No one in particular/anyone; Students; Workers; People with disabilities; Other	Yes	0.8	0.498	Moderate
Collaboration Design	One-to-one; One-to-many; Many-to-one; Many-to-many; Other	Yes	0.779	0.311	Fair
Participants of study	No user study; Own testing; People with disabilities; Teachers; Professionals; Students; Other	Yes	0.879	0.258	Fair
User study Method	Controlled experiment; Usability testing;User study; Interviews;Case study; Focus groups; Field study; Workshop/design; Other; None	Yes	0.763	0.488	Moderate
Participatory Design	Yes; No	No	0.862	0.277	Fair
User study Location	Online/remote; No user study, Near/at research's lab; Participant's home, residence, or school; Neutral location; Other	Yes	0.924	0.787	Substantial
Type of Contribution	Empirical; Artifact; Methodological; Theoretical; Benchmark/Deataset; Survey; Opinion	Yes	0.854	0.643	Substantial

Figure 2: *The final codebook and calculated pairwise agreement, Cohen's Kappa IRR and Landis & Koch level.*