

Reflecting on the Status Quo of Nonverbal Human-Machine Interaction

Fábio Barros^{1,2}^a, António Teixeira^{1,2,3} and Samuel Silva^{1,2,3}

¹DETI – Department of Electronics, Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal

²IEETA – Institute of Electronics Engineering and Informatics of Aveiro, University of Aveiro, Aveiro, Portugal

³LASI – Intelligent Systems Associate Laboratory, Aveiro, Portugal

Keywords: Nonverbal Communication, Human-Computer Interaction, Multimodal Interaction.

Abstract: Among humans, speech plays a central role in providing a form of communication that is efficient and natural. But communication goes beyond the verbal component harnessing a wide range of nonverbal aspects ranging from how words are said to posture, gestures, and facial movements. These, can complement or reinforce the message increasing the adaptiveness of our communication abilities to different contexts, a desirable characteristic to also have in our interaction with machines. Nevertheless, nonverbal communication cues are still scarcely considered for human-machine interaction not only motivated by the complexity of understanding and tackling them, but also by difficulties in translating them into a broader range of scenarios. In this context, this article examines the current state of research on nonverbal interaction and reflects on the challenges that need to be addressed in a multidisciplinary effort to advance the field.


1 INTRODUCTION

Communication is a very important process among humans, since it enables interaction to exchange ideas and experiences. In this context, speech is our most direct and natural channel of communication, but the communication process is often enriched by a set of nonverbal contributions. Body posture, face expression, voice intonation, appearance, touch, distance and, time are some of nonverbal features that establish multiple channels that humans also use to communicate and interact with others (de Gelder et al., 2015). It is this diversity of choices that makes human-human communication so efficient, natural and adaptive, in a wide variety of contexts.

In a world where interaction between users and machines is ubiquitous and continues to grow, at a fast pace, it is necessary to respond with innovative systems that can embrace the users' needs as well as new attractive methods of interaction that can reach different audiences. In the past few years, speech has been widely considered for interaction with smart environments, boosted by the advances in several supporting technologies, with special attention to conversational assistants, such as Google Assistant, Siri or Alexa (Seaborn et al., 2021).

The adoption of conversational assistants potentially provides addressing the complexity of the environment by transforming many of the interactions with it in a conversation moving away from more tangible or graphical interfaces, for some purposes. This brings users closer to a context that resembles human-human communication creating expectations on how systems will understand them along the way. Additionally, from the interactive system's perspective, every additional information that can be gathered to better establish the user's intentions, disambiguate content, or provide increased levels of adaptiveness is important. In this regard, the consideration of nonverbal communication channels can potentially foster a greater level of naturalness and efficiency and move the interaction further towards an increasingly human-centred perspective (Guzman, 2018).

The research on considering multiple forms of interaction with machines has travelled a long way, but while the literature already addresses, to different extents, some nonverbal cues (e.g., for gestures or facial cues), several challenges still remain concerning not only the evolution of the base technology, but also how to support their integration in novel systems. In this regard, this article looks into the overall research devoted to tackling interaction with a focus on supporting nonverbal channels and reflects on the challenges it needs to address to make nonverbal aspects

^a <https://orcid.org/0000-0001-8392-6922>

a more pervasive feature considered during human-machine interaction. These challenges highlight a range of research opportunities that should mobilize a multidisciplinary community including behavioral sciences, human activity processing and analysis, software engineering, and human-machine interaction. To this end, the remainder of this article is organized as follows: section 2 briefly reviews human-human communication broadly identifying relevant nonverbal channels; section 3 takes on the identified channels and overviews current trends regarding nonverbal cues in interaction identifying major challenges for its wider consideration; section 4 takes on the identified challenges and proposes a set of routes to follow towards advancing the topic; finally, section 5 presents the overall conclusions.

2 THE MULTIMODAL NATURE OF HUMAN COMMUNICATION

Since birth, humans use their physical abilities and behavior, such as sounds, movements, or expressions to communicate with others and exchange ideas and experiences. Communication becomes something natural to perform but it is not a simple process given that humans use a plethora of nonverbal contributions that play a pivotal role in conveying a message, providing redundant or complementary information, and, sometimes, being the message itself. Additionally, nonverbal aspects can support backchannel cues, keeping the communication channel open and providing a confirmation to the speaker that it is being listened (Mueller et al., 2015).

Verbal communication can be defined as a form of interaction through the use of words, or messages in linguistic form (oral, written communication and, also, sign language) while the nonverbal communication is a process of generating meaning using behavior other than words (Chandler and Munday, 2011; Andersen, 2008). Additionally, the literature adopts a systematic way to define the different components according to the communication channel distributing them over a set of different groups of codes: Kinesics, Vocalics, Proxemics and Haptics, Chronemics, and Artifacts (Steinberg, 2007) (figure 1), standing for visual, auditory, contact, time, and place characteristics respectively. These codes are explained, in more detail, in what follows.

Kinesics — Kinesics concerns the interpretation of body motion in communication through facial expressions, gestures, eyes, body posture, i.e. pertaining the behaviors that are related to movement of any part of the body or the body as a whole (Key, 1975).

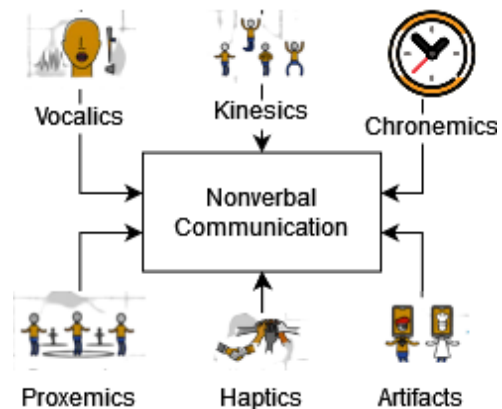


Figure 1: Multiple codes of nonverbal communication.

Vocalics – Vocalics or Paralanguage encompasses nonverbal communication that is related to the vocalized but nonverbal part of a message. This includes several characteristics, such as speaking rate, pitch, volume, tone of voice, and vocal quality (Rankin et al., 2009). These aspects reinforce the meaning of verbal communication, allow us to emphasize particular parts of a message or, in some cases, contradict the verbal message (e.g., by giving an ironic tone to what is said).

Haptics and Proxemics – Haptics refers to touch behaviors that convey meaning during interactions, i.e., how the strength, placement and duration of touch can be used to communicate aspects such as empathy or power. On the other hand, Proxemics refers to the use of space and distance within communication. Proxemics also studies territoriality, or how people take up and defend personal space during an interaction between human and environments (Hans and Hans, 2015).

Chronemics – Chronemics refers to how time affects communication and includes how different time cycles affect our interaction with other human beings (Pirjol and Ionesc, 2019). Naturally, time can influence communication, for instance, deriving from our daily schedules and availability, but what mostly interests us, in this regard, is how time is a component of communication and how it informs how the message is interpreted. Interesting components, in this regard, pertain, e.g., how much time people are willing to listen, the time taken to respond, and the amount of time someone speaks.

Artifacts – Artifacts are related to the environment serving as scenario for the interaction. For example, a comfortable chair can help facilitate interactions between a therapist and a patient or lighting and sound can help set the tone of the communication. While a person can have an active role in setting certain artifacts as context for the communication, it is often

something that does not result from explicit action, as with the other codes. For instance, a darker room may elicit a softer voice tone (Fatt, 1998).

3 NONVERBAL HUMAN-COMPUTER INTERACTION

The interaction between humans and machines is part of our everyday life, creating a demand for more natural and efficient ways of interaction as well as more user friendly systems (Bachmann et al., 2018) for which nonverbal communication can play a key role. The codes described in the previous section provide a reference for what to look for when trying to address nonverbal communication for interactive systems. In what follows we look into the literature to understand the extent to which they are being considered for human-machine interaction and discuss the challenges that are potentially hindering its advances.

3.1 Nonverbal Interaction

The literature gathered in this section does not aim to be a comprehensive account of all works pertaining nonverbal interaction. Considering the nature of the proposed reflection, the goal is to provide recent or otherwise notable references that enable an overall idea about what nonverbal aspects are being covered, to what extent, what are the overall approaches to do it, and how they are being integrated in interactive systems. In this regard, it is important to distinguish two important stages of the research: 1) how the technology supports the core aspects, e.g., detecting gestures; and 2) how it harnesses it to address nonverbal communication. So, while the first is important to understand the overall capabilities of the existing technology, providing an important context, here, we are mostly aiming for greater insight on the latter. Additionally, we are focusing on the machine perspective, i.e., how they might be able to tackle the use of nonverbal channels by humans and not how they might use nonverbal communication to convey a message (e.g., through an avatar).

Speech – Speech has gained prominence given the massive emergence of conversational agents, social robots and chatbots with the rich nature of speech serving various contexts (e.g., smart homes (Seaborn et al., 2021)). The nonverbal aspects of speech have received a strong attention from the community in application scenarios directly or indirectly related to interaction, e.g., speech therapy (Cassano et al., 2022),

speech emotion recognition (Huang et al., 2019), speaker verification/identification, and the detection of language, age, and gender (Atmaja et al., 2022), among others.

Gestures – Interaction with gestures is, perhaps, along with speech, a communication channel that has been strongly explored in human-machine interaction. Although it is valuable to look at how literature approaches gesture recognition, in general, it is important to note that most of the works (e.g., robot control and interaction (Gao et al., 2021), interaction with home devices (Kshirsagar et al., 2020)) are based in pre-established gestures that are assigned a meaning or action. For nonverbal interaction we are mostly interested in gestures that are naturally associated with a meaning, e.g., assigned by culture or social contexts (e.g., making a gesture for "no"), and convey a message during communication. Additionally, systems proposed for sign language, such as for American Sign Language, belong to verbal communication (Islam et al., 2018).

Gaze and Head Posture – Two nonverbal contributions that are considered, regarding kinesics, are head-pose and gaze. These two are closely related and are usually jointly explored for recognizing the focus of attention of the users (something we do, e.g., to identify to whom we are speaking, in a group) and also communicative acts such as interest or attentiveness. These aspects enable natural and "hands-busy" approaches and allow non-invasive interaction (Chen et al., 2019; Brammi et al., 2020). The application areas for this form of nonverbal interaction include the detection of drivers' focus of attention (Naqvi et al., 2018), studying natural interactions in multi-person communication (Müller et al., 2018) and, even, in healthcare scenarios (Luo et al., 2021).

Facial Expressions and Cues – The human face also plays an important role in many aspects of verbal and nonverbal interaction. In terms of interaction, the human face does not only involve the expression of facial emotions, a topic that is increasingly explored, but also human communicative acts such as eye winking or eyebrow raising (Lyons and Bartneck, 2006). Nevertheless, the integration of this nonverbal channel has not been much considered even though its potential advantages are widely recognized (Wimmer et al., 2008). The involuntary nature of micro-expressions, for instance, can inform a context for inferring, e.g., the veracity of what is being said (Oh et al., 2018; Xia et al., 2019). Additionally, there are also works that focus their research on recognition and detection of expressions using the FACS system as rational (Baltrusaitis et al., 2018).

Environment Artifacts – The environment also af-

ffects how our nonverbal communication unfolds while it adapts to the communicative needs, e.g., circumventing communication barriers. For example, if we are in a large or noisy room trying to interact with a conversational agent which is far away from us, we raise our tone of voice or go close to it. In this regard, an understanding of these aspects may help as a context for the communication. Notable examples might be the adaptation to conversation distance (Weisser et al., 2021) or the consideration of the very arrangement of objects in a physical space as context to interaction systems if, for example, an artefact is an interaction partner (Stephanidis et al., 2019).

Multimodality and Nonverbal Cues – Between humans, communication is often multimodal and may consider several nonverbal aspects (e.g., speech, gestures, facial expressions) articulated during the interaction. This same concept has also emerged in interaction with machines through the use of multiple inputs, to provide more robust interaction (Rafael, 2021; Xie et al., 2021). In fact, interaction research has profited from the adoption of multiple modalities of interaction in order to make interaction between humans and systems more efficient and also much more attractive by combining some nonverbal cues, e.g., gaze and gestures (Kim et al., 2019), gestures and speech (Yongda et al., 2018), gaze and facial expressions (Su et al., 2021), focus and gestures (Aftab et al., 2021), and speech, facial cues, and gestures (Strazdas et al., 2022).

3.2 Discussion

By considering the panorama provided by the overviewed literature there are a few aspects that can be identified as challenges for a more pervasive presence of a wide range of nonverbal communication capabilities in human-computer interaction.

Limited Coverage of Nonverbal Channels – The diversity of communication channels, verbal and nonverbal, is what makes human-human so efficient and adaptive, in a wide variety of contexts and the body of work for this topic shows contributions in adopting a few forms of nonverbal communication towards a potentially more natural interaction between humans and machines. However, the literature still explores only a small range of those features, mostly providing works on nonverbal interaction considering gestures or voice cues.

Tightly Coupled Solutions Hindering Reuse – While these works make important contributions to advance the technology and inform further research, they typically target a specific purpose or application. While having a well defined scenario can be important

to focus and evaluate the research, this often yields tightly coupled approaches to the proposed methods – i.e., with interaction developed as part of the core of the applications – meaning that third parties that want to apply them to other scenarios will have to mostly implement them and master all the technologies involved. Overall, this affects how these features can reach a broader set of systems and how cumulative improvement can occur.

Scarce Exploration of Nonverbal Synergies – Additionally, what makes human-to-human communication so efficient, diverse, and even more complex is the way nonverbal aspects are fused with verbal and other nonverbal aspects during the interaction between humans. While the literature shows the adoption of multiple channels in interaction scenarios using voice with gestures or gaze and gestures, for example, this exploration is limited to just a few codes and does not yet include some of the richness that can arise from the redundancy or complementarity between the verbal and nonverbal aspects of communication. These limitations also stem from the fact that their implementation across multiple channels is somewhat complicated in itself and requires knowledge of the implicit and explicit dynamics of these aspects motivating approaches that often address limited tasks for specific contexts.

Scarcity of Datasets – Finally, the creation of sophisticated methods for interaction often relies on the amount and quality of data available, e.g., to train models to recognize particular actions. However, data for nonverbal cues is scarce and most work relies on limited sets of data that are costly to obtain, mostly result from isolated efforts, and are not available to the community. In addition, existing data-sets are often hard to expand due to difficulties in adding data that is acquired in a manner that is coherent with the data-set properties (e.g., environment lighting, point-of-view). Furthermore, defining what data might inform the design and development of these methods is not trivial and often requires complex acquisition protocols entailing, for instance, methods and tasks to elicit nonverbal cues, during communication.

4 EVOLVING NONVERBAL INTERACTION

Interaction with verbal and non-verbal communication signals can foster increased naturalness and efficiency to our interaction with smart ecosystems. However, their potential is still behind some challenges that need to be addressed. In this regard, this section conceptualizes the different aspects that need

to be tackled moving from the challenges previously identified into a roadmap for future research.

4.1 Vision

Nowadays, interactive systems are increasingly multimodal, so the approach to supporting nonverbal communication cues would profit from embracing current evolutions on this matter. Noteworthy, in this scope, is the proposal of standards by the World Wide Web Consortium (W3C) for multimodal interactive architectures (Dahl, 2013), as well as, frameworks and architectures that implement them (e.g., (Almeida et al., 2019)). These architectures approach multimodality considering a decoupled design with each modality existing independently from the applications. Such characteristic should help move nonverbal interaction research into easier reuse and refinement in a variety of new scenarios. By not having to master the technologies and complexities for each of these modalities, a range of new developers may integrate them, in their works, as off-the-shelf modules. This is akin to the concept of generic modalities, i.e., modalities that are part of the framework supporting the multimodal architecture over which applications can be developed. For instance, considering a smart home, nonverbal interaction features would not be integrated in applications, but as part of the home infrastructure dealing with interaction and to which the applications would connect.

All these things considered, figure 2 depicts the overall modules of an interactive system serving as grounds to this vision. Nonverbal interaction modalities connect to an interaction manager – the core of the multimodal architecture (e.g., see (Almeida et al., 2019)) – along with other modalities and applications. While this would already address a part of the challenges identified earlier – regarding coupling and reuse –, it mostly serves as the laying grounds for the routes of action argued in what follows.

4.2 Roadmap

The evolution of nonverbal interaction provides a plethora of research opportunities regarding not only their integration as part of human-machine interaction, but also the evolution of the underlying technologies for processing and analysing a set of multimodal biological data for human behavior. In line with the vision presented above, and in order to advance the technologies and integration of nonverbal interaction in smart ecosystems, several routes need to be taken:

- **Expand the range of nonverbal cues that is considered for interaction** – This entails the de-

velopment and evolution of sensing, processing, and analysis methods to detect nonverbal communicative actions from humans along with their deployment as generic modalities encapsulating the complexity of the methods and enabling the consideration of verbal and nonverbal aspects for interaction, off-the-shelf. For instance, regarding vocalics, the research on the nonverbal aspects of speech is very strong, but its consideration on interactive systems has yet to fully harness it.

- **Advance the exploration of synergies among verbal and nonverbal communication** – Along with exploring more nonverbal communicative cues, exploring how they work together, e.g., with speech, would help unravel their full potential for a more natural and adaptive interaction. The fusion engine in figure 2 alludes to precisely this aspect, with a wide range of opportunities open to explore different levels of fusion (e.g., feature or semantic level) and unimodal/multimodal models to optimize the outcomes based on the available data, at each time. Additionally, aspects such as proxemics might help build a context – hence the reference to it, in the figure – for the interaction, e.g., establishing the system as a focus of the message.
- **Create novel datasets for nonverbal interaction** – There is no denying that nonverbal communication can entail complex (or subtle) action or behavior. In this domain the exploration of the different communication channels would profit on data that can provide, in a first instance, grounds for a basic understanding of the communicative role of different cues, and allow building and testing methods to serve their detection. In this regard, having a systematic approach to the data collection that might allow multi-site acquisition or later expansion (e.g., detailing a protocol and acquisition conditions) would potentially allow larger datasets. The datasets can also boost the contributions by a wider range of researchers.
- **Embrace a stronger drive for multidisciplinary approaches** – The task of building novel datasets and understanding nonverbal communication would strongly profit from a multidisciplinary approach to define what aspects are relevant, how they can be elicited for collection, what sensors are adequate, and what processing and analysis is required for the different communication codes.
- **Place emphasis on scenario-driven research** – A multidisciplinary setting should also contribute to a better understanding of the communication

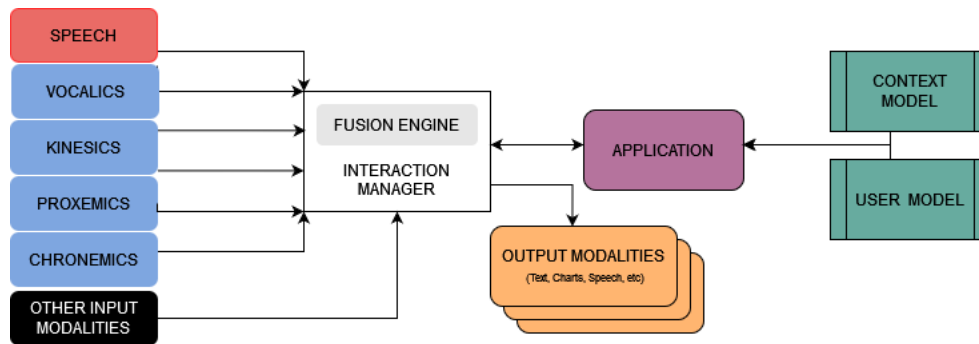


Figure 2: Diagram depicting the main modules of proposed vision for integration of nonverbal communication channels in smart ecosystems.

phenomena in the definition of relevant and meaningful scenarios, providing important clues to the conceptualization of the solutions. And these scenarios help to establish clear research objectives informing the validation of the proposed approaches, potentially allowing the research outcomes to be more focused and measurable. Furthermore, the scenarios establish credible grounds to define when and how to evaluate whether what has been developed is useful, an overarching principle that should not be forgotten.

5 CONCLUSIONS

The role of nonverbal communication cues in human-human communication is indisputable and its consideration for our interaction with machines could foster an increasingly natural, efficient, and adaptive interaction. To this end, this article argues that several challenges need to be tackled and proposes several lines of action that, by also providing grounds for discussion, may help raise awareness for the importance and range of multidisciplinary research opportunities that it entails. In this discussion we aimed for how machines may profit from nonverbal communication cues, but the understanding of how humans do it, to evolve these methods, can also be an important route towards their use, as a communication channel, by machines, e.g., through avatars in virtual environments (Aburumman et al., 2022).

REFERENCES

Aburumman, N., Gillies, M., Ward, J. A., and Hamilton, A. F. d. C. (2022). Nonverbal communication in virtual reality: Nodding as a social signal in virtual interactions. *International Journal of Human-Computer Studies*, 164:102819.

- Aftab, A. R., von der Beeck, M., Rohrhirsch, S., Diotte, B., and Feld, M. (2021). Multimodal fusion using deep learning applied to driver's referencing of outside-vehicle objects. *arXiv preprint arXiv:2107.12167*.
- Almeida, N., Teixeira, A., Silva, S., and Ketsmur, M. (2019). The am4i architecture and framework for multimodal interaction and its application to smart environments. *Sensors*, 19(11):2587.
- Andersen, P. (2008). *Nonverbal Communication: Forms and Functions*. Waveland Press Incorporated.
- Atmaja, B. T., Sasou, A., and Akagi, M. (2022). Survey on bimodal speech emotion recognition from acoustic and linguistic information fusion. *Speech Comm.*
- Bachmann, D., Weichert, F., and Rinkenauer, G. (2018). Review of three-dimensional human-computer interaction with focus on the leap motion controller. *Sensors*, 18(7):2194.
- Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 59–66. IEEE.
- Brammi, J., Rb, D., Tb, P., and Rb, A. (2020). HCI based input device for differently abled. *Intelligent Systems and Computer Technology*, 37:304.
- Cassano, F., Pagano, A., and Piccinno, A. (2022). Supporting speech therapies at (smart) home through voice assistance. In *International Symposium on Ambient Intelligence*, pages 105–113. Springer.
- Chandler, D. and Munday, R. (2011). Verbal communication.
- Chen, W., Cui, X., Zheng, J., Zhang, J., Chen, S., and Yao, Y. (2019). Gaze gestures and their applications in human-computer interaction with a head-mounted display. *arXiv preprint arXiv:1910.07428*.
- Dahl, D. A. (2013). The w3c multimodal architecture and interfaces standard. *Journal on Multimodal User Interfaces*, 7(3):171–182.
- de Gelder, B., De Borst, A., and Watson, R. (2015). The perception of emotion in body expressions. *Wiley Interdisc. Rev.: Cognitive Science*, 6(2):149–158.
- Fatt, J. P. T. (1998). Nonverbal communication and business success. *Management Research News*.

- Gao, Q., Chen, Y., Ju, Z., and Liang, Y. (2021). Dynamic hand gesture recognition based on 3d hand pose estimation for human-robot interaction. *IEEE Sensors*.
- Guzman, A. L. (2018). *Human-Machine Communication*. Peter Lang, Bern, Switzerland.
- Hans, A. and Hans, E. (2015). Kinesics, haptics and proxemics: Aspects of non-verbal communication. *IOSR Journal of Humanities and Social Science (IOSR-JHSS)*, 20(2):47–52.
- Huang, K.-Y., Wu, C.-H., Hong, Q.-B., Su, M.-H., and Chen, Y.-H. (2019). Speech emotion recognition using deep neural network considering verbal and non-verbal speech sounds. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5866–5870. IEEE.
- Islam, M. R., Mitu, U. K., Bhuiyan, R. A., and Shin, J. (2018). Hand gesture feature extraction using deep convolutional neural network for recognizing american sign language. In *4th Int. Conf. on Frontiers of Signal Processing (ICFSP)*, pages 115–119.
- Key, M. R. (1975). Paralanguage and kinesics (nonverbal communication).
- Kim, J.-H., Choi, S.-J., and Jeong, J.-W. (2019). Watch & do: A smart iot interaction system with object detection and gaze estimation. *IEEE Transactions on Consumer Electronics*, 65(2):195–204.
- Kshirsagar, S., Sachdev, S., Singh, N., Tiwari, A., and Sahu, S. (2020). Iot enabled gesture-controlled home automation for disabled and elderly. In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, pages 821–826. IEEE.
- Luo, W., Cao, J., Ishikawa, K., and Ju, D. (2021). A human-computer control system based on intelligent recognition of eye movements and its application in wheelchair driving. *Multimodal Technologies and Interaction*, 5(9):50.
- Lyons, M. J. and Bartneck, C. (2006). HCI and the face. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, pages 1671–1674.
- Mueller, M., Leuschner, D., Briem, L., Schmidt, M., Kilgour, K., Stueker, S., and Waibel, A. (2015). Using neural networks for data-driven backchannel prediction: A survey on input features and training techniques. In *International conference on human-computer interaction*, pages 329–340. Springer.
- Müller, P., Huang, M. X., Zhang, X., and Bulling, A. (2018). Robust eye contact detection in natural multi-person interactions using gaze and speaking behaviour. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, pages 1–10.
- Naqvi, R. A., Arsalan, M., Batchuluun, G., Yoon, H. S., and Park, K. R. (2018). Deep learning-based gaze detection system for automobile drivers using a nir camera sensor. *Sensors*, 18(2):456.
- Oh, Y.-H., See, J., Le Ngo, A. C., Phan, R. C.-W., and Baskaran, V. M. (2018). A survey of automatic facial micro-expression analysis: databases, methods, and challenges. *Frontiers in psychology*, 9:1128.
- Pirjol, F. and Ionesc, D.-E. (2019). Communication, chronemics, silence.
- Rafael, S. (2021). The contribution of early multimodal data fusion for subjectivity in HCI. In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6. IEEE.
- Rankin, K. P., Salazar, A., Gorno-Tempini, M. L., Solberger, M., Wilson, S. M., Pavlic, D., Stanley, C. M., Glenn, S., Weiner, M. W., and Miller, B. L. (2009). Detecting sarcasm from paralinguistic cues: anatomic and cognitive correlates in neurodegenerative disease. *Neuroimage*, 47(4):2005–2015.
- Seaborn, K., Miyake, N. P., Pennefather, P., and Otake-Matsuura, M. (2021). Voice in human-agent interaction: A survey. *ACM Computing Surveys (CSUR)*, 54(4):1–43.
- Steinberg, S. (2007). *An introduction to communication studies*. Juta and Company Ltd.
- Stephanidis, C., Salvendy, G., Antona, M., Chen, J. Y., Dong, J., Duffy, V. G., Fang, X., Fidopiastis, C., Fragomeni, G., Fu, L. P., et al. (2019). Seven HCI grand challenges. *International Journal of Human-Computer Interaction*, 35(14):1229–1269.
- Strazdas, D., Hintz, J., Khalifa, A., Abdelrahman, A. A., Hempel, T., and Al-Hamadi, A. (2022). Robot system assistant (RoSA): Towards intuitive multi-modal and multi-device human-robot interaction. *Sensors*, 22(3):923.
- Su, Z., Zhang, X., Kimura, N., and Rekimoto, J. (2021). Gaze+ lip: Rapid, precise and expressive interactions combining gaze input and silent speech commands for hands-free smart tv control. In *ACM Symposium on Eye Tracking Research and Applications*, pages 1–6.
- Weisser, A., Miles, K., Richardson, M. J., and Buchholz, J. M. (2021). Conversational distance adaptation in noise and its effect on signal-to-noise ratio in realistic listening environments. *The Journal of the Acoustical Society of America*, 149(4):2896–2907.
- Wimmer, M., MacDonald, B. A., Jayamuni, D., and Yadav, A. (2008). Facial expression recognition for human-robot interaction—a prototype. In *International Workshop on Robot Vision*, pages 139–152. Springer.
- Xia, Z., Hong, X., Gao, X., Feng, X., and Zhao, G. (2019). Spatiotemporal recurrent convolutional networks for recognizing spontaneous micro-expressions. *IEEE Transactions on Multimedia*, 22(3):626–640.
- Xie, B., Sidulova, M., and Park, C. H. (2021). Robust multimodal emotion recognition from conversation with transformer-based crossmodality fusion. *Sensors*, 21(14):4913.
- Yongda, D., Fang, L., and Huang, X. (2018). Research on multimodal human-robot interaction based on speech and gesture. *Computers & Electrical Engineering*, 72:443–454.