Wirtschaftsinformatik 2023 Proceedings Wirtschaftsinformatik

10-9-2023

A State of the Art Overview on Biosignal-based User-Adaptive Video Conferencing Systems

Julia Seitz
Karlsruhe Institute of Technology, Germany, julia.seitz@kit.edu

Chiara Krisam
Karlsruhe Institute of Technology, Germany, julia.seitz@kit.edu

Ivo Benke
Karlsruhe Institute of Technology, Germany, julia.seitz@kit.edu

Recommended Citation

Seitz, Julia; Krisam, Chiara; and Benke, Ivo, "A State of the Art Overview on Biosignal-based User-Adaptive Video Conferencing Systems" (2023). *Wirtschaftsinformatik 2023 Proceedings*. 27.
https://aisel.aisnet.org/wi2023/27

A State of the Art Overview on Biosignal-based User-Adaptive Video Conferencing Systems

Research Paper

Julia Seitz, Chiara Krisam, and Ivo Benke

Karlsruhe Institute of Technology, Institute of Information Systems and Marketing, Karlsruhe, Germany
{julia.seitz,ivo.benke}@kit.edu

Abstract. Video conferencing systems are widely used in times of distributed teams since they support flexible work arrangements. However, they have negative impacts on users, such as lacking eye gaze or zoom fatigue. Adaptive interventions in video conferences based on user behavior provide interesting solutions to overcome these challenges, for example, by alerting users when looking tired. Specifically, biosignals measured by sensors like microphones or eye-trackers are a promising basis for adaptive interventions. To provide an overview of current biosignal-based user-adaptive video conferencing systems, we conducted a systematic literature review and identified 24 publications. We summarize existing knowledge in a morphological box and outline further research directions. Thereby, a focus on biooptical signals is visible. Current adaptations target audience feedback, expression understanding and eye gaze mostly by image and representation modifications. In future, we recommend including further biosignals and addressing more diverse problems by investigating adaptation capabilities of further software elements.

Keywords: Video Conferencing System, Adaptation, Systematic literature review

1 Introduction

In recent years, we see an increase in the use of video conferencing systems (VCS) when humans want to collaborate and interact across physical distances. For example, the use of VCS for corporate communication arose from approximately 50% in 2018, to 72% in 2022 (Bitkom 2022) and managers predict to more frequently use VCS beyond the pandemic (Skillscouter 2022). However, there exist multiple challenges with VCS for productivity and well-being, such as lacking eye contact, lacking joint attention, reduced audience feedback, or fatigue due to prolonged, frequent meetings (Fauville et al. 2021, Kelly 2020, Karl et al. 2021). Despite several decades of research on video-mediated communication, to date, these problems are not yet solved in practice and research. Research has carved out a potential reason: the fact that individual VCS users differ from each other by individual characteristics, needs, and behavior (Döring et al. 2022, Valkenburg & Peter 2013). Moreover, the situations in which such problems arise are dynamic and not static. For example, they depend not only on the user's characteristics but also on the context and timing in the meeting. Therefore, it is challenging to design a VCS that fits all users and provides holistic support in all situations.

However, we increasingly see suggestions from research and practice, that the interface of the system should be able to react to problematic situations in video conferences and change the presentation of participants (Kaptelinin 2023, Pexip 2023, Seitz 2022, Abramova et al. 2021). One example might be to display personal status information such as the emotion or the stress level of VCS participants. As another example, one may show the presenter how attentive his audience is to allow the presenter to include breaks meaningfully or change the way of presentation when the audience is tired or bored (Sun et al. 2019, Kaptelinin 2023). Or, the other way around, the audience could be informed which part of a slide the presenter is currently gazing at to follow the presentation more closely (Langner et al. 2022). Besides, information about individual users may be useful to document fatiguing effects after a long day of meetings, and issue break recommendations or modify software elements that may increase fatigue, such as the self-view (Epstein 2020, Seitz et al. 2022).

One approach to designing VCS is by making them user-adaptive and biosignal-based. Adaptive systems are systems that are able to change their behavior to meet the needs of users (Feigh et al. 2012). Thereby, this system class is comparable to established systems in information systems that are adaptable to the user, such as websites making use of adaptations based on behavioral data or user group characteristics (Reinecke & Bernstein 2013, White & Morris 2007).

User-adaptive systems require the continuous collection of information about the users. One key input source therefore are biosignals, especially in NeuroIS (Riedl & Léger 2016, Hettinger et al. 2003). Biosignals are defined by Schultz et al. (2013) as autonomous signals produced by living organisms and can be measured energetically using sensors. In VCS, we believe that biosignals may provide meaningful information, as the main input and output channels of the meetings are audio and video signals, which provide a broad range of verbal and non-verbal biosignals. In contrast to more traditional information systems, behavioral log and click data is less frequent available. Therefore, to collect biosignals, various types of sensors can be employed. This may cover existing cameras and microphones used in video conferences or established sensors such as eye-tracking or electroencephalography devices. Moreover, new wearable devices, may be beneficial to collect biosignals such as the heart rate when users are working in a mobile setup (Knierim et al. 2022, Zheng et al. 2014). Despite their enormous potential, however, to the best of our knowledge, there is no structured overview of biosignal-based user-adaptive VCS. We thus aim to answer the following question:

What is the state of the art of biosignal-based user-adaptive video conferencing systems (VCS) and what are future research directions?

To answer this research question, we conducted a structured literature review and analyzed 24 publications. We contribute with a conceptual framework for user-adaptive VCS based on existing definitions, we review and classify existing research, and we propose future research directions.

2 Conceptual Foundations

We aim to identify research that uses biosignals to design user-adaptive systems, particularly VCS. This interplay between human and technology is reflected in physiological computing. Subsequently, we describe the core concepts biosignals and user adaptation.

2.1 Biosignals

According to Schultz et al. (2013), a biosignal is an autonomous signal produced by an organism and measurable through sensors. Biosignals can be distinguished according to their physical properties in different categories: bioacoustic, biochemical, bioelectric, biomagnetic, biokinematic, biooptical, and biothermal (Telaar et al. 2014, Schultz et al. 2013). Sensors for the measurement of these signals differ between these types and also among the particular human signals. Depending on the sensor used for measurement, signals can belong to different categories. Most important in video conferencing are biooptical and bioacoustic signals. Biooptical signals are captured by a camera such as the movement of the eyes measured via eye-tracking or facial expressions extracted from video feeds. According to Schultz et al. (2013), biokinematic signals include signals such as blood pressure (BP) or pulse rate (PR). Also motion of the extremities or facial expressions (if measured via kinematic sensors) can be listed in this category (Telaar et al. 2014, Schultz et al. 2013). Bioacoustic signals cover signals that emanate from the voice. Furthermore, electrical signals are measured by detecting the potential difference between two electrodes, including signals from the peripheral nervous system such as electrodermal activity (EDA) or electrocardiograms (ECG), or from the central nervous system such as brain activity from electroencephalography (EEG).

Based on these signals, the biocybernetic loop describes the process which is underlying physiological computing and NeuroIS (Fairclough 2017, Pope et al. 1995). The biocybernetic loop is an iterative feedback loop that follows three steps: collection, analysis, and translation (Fairclough 2011). During the first step “collection” the signals are accumulated. Depending on the type of signals different measurement methods and sensors, like ECG or eye-tracking, can be applied. In the step “analysis” the signals are processed and artifacts can be identified. Fairclough (2017) mentions for instance the mapping of signals to emotional states. These artifacts can be utilized in the technology in the third step “translation”. One example is the use of artifacts as commands in a brain-computer interface (BCI) to control the technology (Zander & Kothe 2011). Since the model represents a loop, following this step, a new iteration can be initiated.

2.2 Adaptive Systems

The term adaptation was shaped primarily by the field of biology and describes the adjustment of an organism to react to changes in its environment. However, the terminology “adaptive system” is also present in the Information Systems (IS) and Human-Computer Interaction (HCI) community. Feigh et al. define adaptive systems as “the technological component of joint human machine systems that can change their behavior to meet the changing needs of their users, often without explicit instructions from their users” (Feigh et al. 2012, p. 1009). The definition implies that the user can partially contribute to the adaptation. This differs from other definitions of adaptive systems, which describe adaptive systems more narrowly (see e.g., Benyon & Murray (1993)). On top, the paper includes a taxonomy for adaptations, which states that even small adaptations such as the highlighting of information already represents an adaptation.

Considering our use case of VCS, it is interesting to already look at such small adaptations as they might be helpful for users to receive direct feedback on their biosignals. For example, it might be helpful for users if the system recognizes when they are not attentive and provides feedback on the inattentiveness. In this case, the system itself would not have to change, but only support in alerting user behavior. Besides, users could change something in the system themselves after the system made a suggestion. For example, a recommendation to show or hide specific speakers could be helpful to highlight the person who is currently speaking. In this case, there is no direct adaptation by the system itself, but by the user.

In order to cover as many use cases as possible, we will use a very broad definition of adaptive systems in this paper. Based on the biocybernetic loop, we therefore define different types of **adaptation mechanisms** (see Figure 1). Adaptation mechanism is used as an umbrella term for all actions along the biocybernetic loop where an adaption of the interface or functionality of the system takes place based on the biosignals of the users. In the case of VCS, it is important to mention that the adaptation mechanisms for the user are not necessarily based on the user's own biosignals. Rather, the signals of the interlocutor or a group of interlocutors are used for the modification of the system. **Feedback** is the simplest form of an adaptation mechanism. Here, biosignals are collected, processed, and reflected back to the user in a form that the user can understand. During the **analysis**, the biosignals are subjected to an intermediate step. As in the biocybernetic loop, human states may be identified and reported to the user. During **adaptation**, in the ongoing also called deep adaptation, a modification of the system is automatically carried out based on the biosignals.

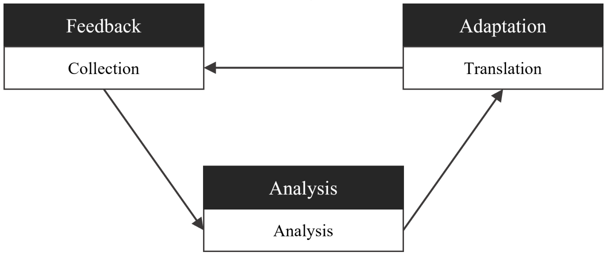

Figure 1. Adaptation Mechanisms (black) based on Biocybernetic Loop (white)

3 Research Method

3.1 Systematic Literature Review

To answer the research question we adopted the approach for systematic literature reviews by Webster & Watson (2002). The identification of the relevant literature involved three steps: First, a search string was utilized to identify literature in relevant databases. In the next step, the literature was screened using previously defined criteria, first on the basis of title and abstract and then on the basis of the full text. In the third step, the remaining literature was subjected to a forward and backward search. We performed the initial search on the three databases IEEE Explore, ACM Digital Library and AIS e-Library which are well established in the HCI, IS and computer science communities. We included only literature in English language, published in a peer-reviewed outlet after the year 2000 in our search.

The search string for the keyword search is based on the three components biosignals, VCS and adaptation mechanisms. Through testing different variations of search strings, several insights were gained on the keywords needed for the string. We noticed that some publications did not mention the term "biosignal" itself, although they were relevant for us. To avoid missing these publications, common biosignals were included in the search string. We also wanted to include systems that provide feedback, either directly or after analyzing the signals. Therefore, in addition, keywords describing feedback systems were included in the search string. Words were iteratively tested and added to the search string based on the results of the search. The final search string, using wildcards and boolean operators, identified 367 publications (in December 2022):

(Bio OR Physio* OR psychophysio* OR behavior* OR heart OR gaze OR eye OR skin OR galvanic OR brain OR electro* OR EEG OR EDA OR ECG OR GSR OR speech OR voice OR facial) AND (adapt* OR feedback OR support* OR aware OR sensit* OR personali* OR coach OR intervention) AND (“video-mediated” OR “video mediated” OR webconferenc* OR videoconferenc* OR teleconferenc* OR “video conferenc*” OR “tele conferenc*” OR “web conference*”)*

Most publications originated from IEEE Explore (281). The search on ACM returned 78 publications, while the search on AISeL returned only eight publications. Afterward, we filtered the retrieved literature based on the following exclusion criteria (C):

C1 The article studies systems that work with biosignals as input.
C2 The article studies video conferencing systems (VCS).
C3 The article studies systems that give feedback on or adapt based on biosignals.

We applied the exclusion criteria to title and abstract of the publication set and after applying C1, 51 publications remained. After applying C2, 37, and after C3, 32 publications remained. Using the full texts as a basis, another 13 publications were sorted out. Next, we performed a forward and a backward search. This identified another five relevant papers, so that the final literature selection consisted of 24 publications.

3.2 Conceptual Framework of Analysis

To analyze the identified studies, we derived a conceptual framework of analysis based on the key concepts of this review. The classes used for the concept matrix correspond to the three concepts that were considered in this review: biosignals (1), VCS (2), and adaptation mechanisms (3). Based on this separation, subclasses were identified top-down from literature and bottom-up using inductive reasoning based on the identified literature (Nickerson et al. 2013). We used an iterative process involving the first and second author and derived subclasses and categories for each subclass. In a first step, the second author screened all publications and coded them in detail with regard to biosignals, VCS, and adaptation mechanism named in the publication. Afterward, we aimed to identify the most distinguishable elements in our publication set by discussing the differences of the papers and drawing on subclasses used in comparable publications (e.g., Loewe & Nadj (2020)). Subsequently, the second author mapped all publications to the aggregated categories based on the codes and the first and second author again verified the correct assignment of the publication to each category.

For **biosignals**, the categories defined by Schultz et al. (2013) were considered as subclasses. In our paper, we only report the signal categories that have been collected at least once. For the class of biosignals, we followed Schultz et al. (2013) and used the categories biokinematic, biooptical, bioelectric, biothermal, and bioacoustic signals. For the class of **video conferencing systems (VCS)**, we observed the use of avatars (2.1) replacing the original video of the meeting partner. Similarly, we categorized whether systems used a three-dimensional (3D) or VR setting (2.2) instead of a two-dimensional (2D) image. For **adaptation mechanisms**, we differentiate the subclasses adaptation mechanism type, the direction, the target state, the representation, and the use of AI. For the subclass adaptation mechanism type (3.1) we identified the categories feedback, analysis, and deep adaptation. Deep adaptation refers to the adaptation in the biocybernetic loop, *deep adaptation* in our framework. The second subclass describes the direction of the adaptation, either whether the adaptation mechanism is applied for both parties (bidirectional) or one party (one directional) (3.2). Third, we distinguish between the target states (3.3) of the adaptation mechanism as an equivalent of the goal the adaptation aims to achieve. The following topics were identified: eye-contact and attention support, emotion and expression understanding, cognitive user state understanding, support for meeting performance, feedback on the audience and technical support for better immersion. Closely related to the target state and the aforementioned type of adaptation, the representation of the adaptation mechanism was examined (3.4). Finally, the intention of the use of artificial intelligence (AI) was classified (3.5). Following the types of adaptation mechanisms in 3.1, we distinguish the use of AI to gather the signals, to perform the analysis, or as actual deep adaptation technique of the representation.

4 Results

In the following, we first provide descriptive results on the publications and afterward describe the conceptual analysis. Therefore, we use the morphological box derived from our conceptual framework depicted in Figure 2 and name exemplary publications for each category. A detailed categorization of all publications is visible here: https://osf.io/m72ny/?view_only=5dcb02caa9914c37887a32426b868acf.

4.1 Descriptive Results

Most articles (79%) were published in the last ten years (since 2012), with most papers published in 2018 and 2022, with respectively four and three publications. 18 articles were published in conference proceedings (75%) and six in journals (25%). All but one paper focus on the artifact development, e.g., the development of an user-adaptive VCS. Furthermore, 22 papers describe the evaluation. In 10 of the 22 papers, a proof of concept test was performed. This involved analyzing existing data to test the feasibility of the prototype. The remaining 13 papers involved a user study in a laboratory. In six cases, 10-19 participants were involved, in five cases there were 20-40 participants and in two cases the study involved more than 100 participants. A between-subject design was used in five cases, and a within-subjects design in the other eight cases. The data collection was carried out by observing the participants in seven cases, and by self-reported measures in 12 cases. A combination of both methods was used in six cases. In all cases, the participants were actual users of the (prospective) systems.

4.2 Conceptual Results

Biosignals Except for two publications (Baig & GholamHosseini 2013, Tan et al. 2014), biooptical signals were involved in all publications (92%). Thereby, most frequently, facial expressions were analyzed (58%, e.g., Carter et al. (2016), Li et al. (2015)). Here, the camera image was mostly analyzed with the help of an AI (e.g., Menychtas et al. (2019), Murali et al. (2021)). In just under half of the studies, eye gaze was measured (42%, e.g., Carter et al. (2016), Otsuki et al. (2018)). In most cases, eye movements were recorded with a standard camera and analyzed later (7 publications, e.g., Pai et al. (2021), Schwartz et al. (2020)). Special eye-tracking devices were employed in three cases (Otsuki et al. 2018, Rae et al. 2011, Roberts et al. 2009). Head gestures were input signal in seven cases (29%, e.g., Murali et al. (2021)). As biokinematic signals, blood pressure and pulse rate were measured (8%, Baig & GholamHosseini (2013), Tan et al. (2014)). Bioelectric signals were recorded in two cases (8%). Thereby, EDA data (Tan et al. 2014) helped meeting participants assess the condition of their communication partner or an ECG (Baig & GholamHosseini 2013) was administered to track the patient's condition. Bioacoustic signals were measured in four cases (17%, e.g., Carter et al. (2016)). Similar to facial expressions, the voice was recorded with standard microphones, processed, and then analyzed, sometimes with the help of an AI (Rojas et al. 2022, Taguchi et al. 2018). In almost half of the studies, only a single kind of signal was recorded (46%, e.g., Li et al. (2015), Takemae et al. (2005)). Almost as often, two different kinds of signals were in use (33%, e.g., Murali et al. (2021), Song et al. (2018)). Among these, prominent combinations involved two different biooptical signals or voice input in addition to a biooptical signal (Rojas et al. 2022, Samrose 2018, Taguchi et al. 2018). Six publications involved three or more kinds of signals (e.g., Carter et al. (2016), Rae et al. (2011)).

Video Conferencing System **Video Format:** There were nine publications (38%) where the image of the person was not displayed directly (e.g., Carter et al. (2016), Nakazato et al. (2014)). Instead, a form of **avatar** (2.1) was displayed that was adapted to the person and their facial expression. Three cases can be distinguished. In the first case, the video of the person was taken and their facial expression was *mapped* to a previously defined avatar (Oh et al. 2016). In the second case, the actual image of the person was *abstracted*. A filter was applied to the image so that the video of the person would appear in a cartoon character style (Carter et al. 2016). In the third case, the face of a person is *reproduced* and placed on an avatar in a 3D environment (e.g., Rae et al. (2011), Roberts et al. (2009)). Often, the facial expression had to be reproduced because the subjects were wearing VR glasses (Schwartz et al. 2020, Song et al. 2018). Besides, in one system instead of facial expressions, an abstract representation of a human indicating its physiological measurements, such as pulsing a heart, was displayed next to the video of a person (Tan et al. 2014).

Dimension: In five publications (21%), the environment for the VCS was 3D (2.2, e.g., Rae et al. (2011), Roberts et al. (2009)). All other publications used 2D software.

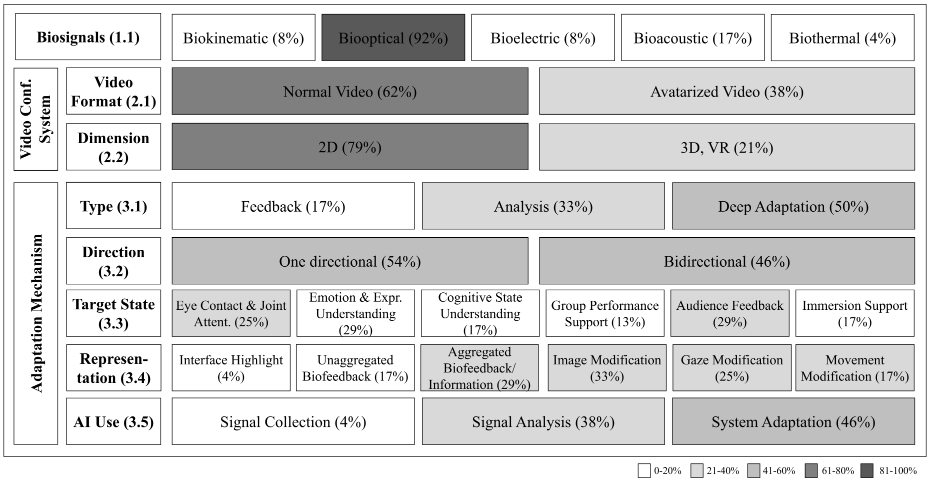

Figure 2. Conceptual Framework (Percentage: share of publications addressing the category)

Adaptation Mechanisms Type: Four systems only gave direct feedback on the biosignals (17%, Otsuki et al. (2018)). Thereby the signals were first summarized in a way that they could be understood by the users. Eight systems analyzed the biosignals in more detail before presenting the results to the users (34%, e.g., Li et al. (2015), Menychtas et al. (2019)). In most cases, facial expressions were used to determine the state of the conversation participants, however, with varying degrees of representation as visible below. A deep adaptation took place in 12 publications (50%, e.g., Carter et al. (2016), Oh et al. (2016)). Here, most systems performed modifications of image, gaze or movement as visible below. In half of the systems, the adaptation took place implicitly (e.g., Nakazato et al. (2014)). This means that the other meeting participants are unaware that an adaptation took place. For example, it is not apparent whether a person's smile was enhanced (Carter et al. 2016) or whether the eye gaze was adapted (Yang & Zhang 2004). Thereby, these implicit adaptations are predominantly present throughout the whole meeting and remain constant modifications of a specific feature.

Direction: Adaptation for every meeting participant took place in several systems bidirectionally (3.2, 46%, e.g., Murali et al. (2021), Oh et al. (2016)). Systems with one directional adaptations were used in online classes (Sartika et al. 2022, Sun et al. 2019) and in therapeutic settings to ease reading facial expressions or the identification of a patient's cognitive or emotional state (Carter et al. 2016, Baig & GholamHosseini 2013). In these publications, a clear distinction between the group that benefits from the adaptation and the group whose signals are used for the adaptation is visible. No systems were investigated that displayed the information only to the user whose signals were collected (e.g., notification of inattentiveness).

Target: A prominent adaptation target (3.3) of the publications is to positively influence the attention of the participants (25%, e.g., Schwartz et al. (2020), Takemae et al. (2004)). Here, systems in which eye contact or eye gaze support play a role are especially common (e.g., Otsuki et al. (2018), Roberts et al. (2009)). Further, publications that focus on the emotional state of the participants and aim to ease emotion and expression understanding are popular (29%, e.g., Rojas et al. (2022), Tan et al. (2014)) and nearly always use facial expressions. Another prominent target was encountering the problem of lacking audience feedback, for example by providing feedback on the presence of the audience (29%, e.g., Murali et al. (2021)). Besides, several publications aim to increase the understanding of the participant's cognitive states, especially the states of other participants (17%, e.g., Pai et al. (2021)). Besides, group performance, such as creativity, is supported by image modifications as well (13%, e.g., Li et al. (2015)). Some authors set the target as the mere feedback of the analysis of the audience (29%, e.g., Murali et al. (2021), Pai et al. (2021)). Finally, there are publications that aim to provide adaptations that support immersion (17%, e.g., Rae et al. (2011), Song et al. (2018)). Nearly all of these studies want to transfer the meeting into a 3D environment.

Representation: Closely related to feedback as the adaptation type, two times, established physiological signals such as PR were displayed in form of unaggregated feedback (17%, Baig & GholamHosseini (2013), Tan et al. (2014)). Additionally, in this category, two applications included signals related to the eyes and displayed the blink rate (Pai et al. 2021) or the eye movement (Otsuki et al. 2018). Moving beyond unaggregated feedback towards analysis, the representation of the user states varied. At the lowest level, in one system, only the interface element was highlighted according to the prevailing emotion (4%, Rojas et al. (2022)). As aggregated biofeedback information (29%), percentages indicated which emotion the person was currently feeling (Menychtas et al. 2019) or systems employed in a classroom setting showed the aggregated states of the students (e.g., percentage of students who are in flow, Sartika et al. (2022), Sun et al. (2019)). Moving beyond analysis towards adaptation, three prominent forms of modifications can be identified. First, systems modified the image or depiction of the person. In one system, the system adaptively changed which person to spotlight, i.e., whose video was displayed prominently by analyzing facial expressions (Murali et al. 2021). Besides, in the remaining cases, the image of the person's facial expression was adapted (e.g., Carter et al. (2016), Oh et al. (2016)). Besides, six systems adapted the eye gaze of the participants, mostly to create eye contact between the participants (e.g., Rae et al. (2011), Roberts et al. (2009)). The movements of the participants, including the head orientation, were adapted in four publications (e.g., (Song et al. 2018, Tewele et al. 2015)) The source data indicates that some systems modified more than one subject. Thus, five systems adapted two elements (e.g., Schwartz et al. (2020), Zhang et al. (2022)), and in one case three elements were undergoing adaptation (Song et al. 2018).

Artificial intelligence: AI (3.5) was applied in almost all systems (20 publications, 83%). One publication specifically describes that a CNN was used to measure the signal (Pai et al. 2021). In this case, the blinking of the eyes is detected. It should be noted that it cannot be ruled out that AI was also used during this step in other studies without explicitly describing it. In nine cases, AI was used to analyze the state of participants, for example, to determine their emotional status based on facial expressions or to measure attention in an online classroom (38%, e.g., Li et al. (2015), Menychtas et al. (2019)). In 11 publications, AI was used to perform an adaptation of the system (46%, e.g., Taguchi et al. (2018)). This included, for example, adjusting eye gaze (Yang & Zhang 2004) to establish eye contact or increasing the smiles of participants (Carter et al. 2016).

5 Discussion

In this work, we provide a systematic literature review including 24 publications on biosignal-based user-adaptive VCS. We categorized our findings in a conceptual framework derived in a top-down and bottom-up approach.

We believe that there are already some opportunities to incorporate biosignals into VCS in practice. The concept developed can provide a basis for practitioners to see how adaptive their systems already are and what existing researched approaches could be implemented to address known issues such as generating eye contact. For inexperienced companies, it can be especially interesting to implement less invasive methods, such as displaying feedback based on devices that are already available. More experienced companies can use the framework to develop new VCS features targeting the future research directions. Building on this practical experience, research can gain field-based insights into how users actually perceive those implementations. In the following, we discuss our review's findings by identifying future research directions, shown in Table 1.

Integrate biosignals beyond biooptical signals: One striking aspect in current research is the primary focus on biooptical signals. Primarily, facial expressions and eye movements were used as input signals. This is most likely because video is recorded anyway and users do not require additional hardware to measure these signals. Incorporating eye gaze can also be easily implemented in the future with novel webcam eye tracking, for example to create artificial eye contact. Therefore, the analysis of video is a practical basis for adaptation. Related to this, we identify future research subdirections. First, we encourage scholars to investigate the use of bioacoustic signals as they are again collected in any way. Second, we recommend the use of wearable sensors which ease the measurement of physiological data provided as bioelectrical, biothermal or biokinematic signals. Building on findings from NeuroIS and passive brain-computer interfaces (BCI), such signals can provide information that video and audio signals are not able to convey as easily, such as detailed information about user's affective or cognitive states (Riedl & Léger 2016). On top, as data privacy is an important issue in signal collection and processing, making use of signals that do not analyze voice data or image data may be of special interest, especially in meetings where sensitive information is shared (i.e., therapy, confidential business meetings). Third, we hint to using these signals as an additional input signal and thereby overcome the problem of many-to-many relationships between signal patterns and user states (Brouwer et al. 2015).

Table 1. Current and Future Research Directions

Area	Research Focus (current)	Research Direction (future)
Biosignal	Biooptical, Unimodal	Bioelectrical, Multimodal
Adaptation Target	Eye gaze, Emotions, Audience feedback	Complex user states, Further problems of video meetings
Adaptation Representation	One interface element (Image of participant)	Multiple interface elements, newly developed features
Evaluation Method	Small scale user studies, proof of concept	Large scale, longitudinal, field studies

Address more complex target states and problems beyond eye gaze, expression understanding, and audience feedback: The target of adaptation in most cases was to influence the attention of participants by supporting eye contact, ease emotion and expression understanding and provide audience feedback. All of these are frequently mentioned problems in research (Karl et al. 2021). However, the identified problems in VCS are not limited to this and the landscape of cognitive and affective user states that arise within a meeting are manifold and not only limited to attention and emotions per se. For example, boredom and the aforementioned fatigue are two common but not yet specifically observed phenomena (Döring et al. 2022, Peper et al. 2021). Also, turn-taking is a common problem in video meetings (Seuren et al. 2021). Thus, we encourage researchers to focus on these more complex user states and engage in identifying if existing applications can already impact these user states as well or whether new systems need to be created for them.

Explore adaptation capabilities of 2D and 3D VCS beyond avatars and image modification and introduce advanced adaptation representations: Of particular interest was the number of modified elements and the representation of the adaptation. Overall, the identified types of adaptation mechanisms were always very specific to single elements. Even though half of the systems belong to the deepest level of adaptation mechanisms, "deep adaptation", the adaptation is represented only by modifying mostly only one aspect of the video element. For example, in 2D, multiple publications augmented video feeds with avatars as a first step towards 3D environments, such as Mozilla Hubs. Gaze and image modifications were thereby performed by using biosignals to mitigate lacking eye contact or enrich facial expressions. This is expected since the video element is a very prominent element of the VCS. Also, using an avatar instead of one's own picture can bring advantages, such as increased privacy (Higgins & McDonnell 2021). In addition, mostly a bidirectional adaptation was used. Thus, the change was visible to the other meeting participants, and in some cases also to the user itself. One directional adaptations were again only visible to the other party. We argue that this results in various underexplored research areas. First, continuing to enrich video elements and avatars with the combined use of biosignals and AI to open up the possibility of reproducing facial expressions or eye gaze for avatars. This can increase the appeal of these environments and should be investigated further. Second, even though sometimes extra elements are added to the standard video as the display of the other participant's user state, the majority of the VCS described do not adapt other VCS features. Thus, future research should also investigate the adaptation capabilities of currently underexplored features (e.g., chat, shared spaces). Third, one directional adaptations only visible for the user whose input is processed are not investigated in our literature sample. As motivated in the introduction, we see potential to support users in overcoming displeasing user states (e.g., nervousness, fatigue) arising in video meetings by adapting the software based on such states, as done in other adaptive systems (Loewe & Nadj 2020).

Conduct large scale studies, field studies, and comparison studies: In terms of evaluation, we identify a strong focus on laboratory studies showing that even small changes in the picture can have a positive influence. As Fairclough (2009) points out, the real-world conditions do differ to laboratory settings, due to uncontrolled extraneous variables. Since scholars aim to introduce such adaptive systems in real-world settings, we argue that more extensive studies including field studies or longitudinal studies should be carried out in future to get real-world user feedback and achieve external validity (Mitchell & Jolley 2013). Besides, studies focus on adapting a single element and exploring its effect isolatedly so far. It would therefore be interesting to see whether a combination of several adaptations could bring advantages (e.g., increased user experience). In addition, inspired by the idea of meta-studies, it is interesting to compare the proposed techniques to identify the most promising one (Gurevitch et al. 2018).

Despite conducting the review rigorously, our results come with limitations. We acknowledge the risk of missing relevant publications due to our search strategy. We selected ACM and AIS as the major databases in IS and added IEEE as a more technical source. We propose including further databases, like Scopus or PubMed, in future to obtain more comprehensive results. Furthermore, the conceptualization is founded in literature but also includes categories derived in a bottom-up approach. We note that especially these subclasses are subject to bias from our review set and due to only involving two authors. We aimed to address these limitations by employing an iterative, structured process and transparent reporting.

6 Conclusion

Our article provides a state-of-the-art overview of VCS with feedback and adaptation via biosignals structured along a conceptual framework based on the biocybernetic loop (Fairclough 2011) and identified 24 publications in a systematic search. We found that most systems are limited to the adaptation of a single element. In the vast majority of cases, this involved the modification of the user's representation (gaze, image, motion). We were also able to show that in most systems the analysis of the video elements like facial expressions and eye gaze was used as input. Besides, we observed that the adaptation target mainly covered the topics eye gaze, audience feedback, and expression understanding. Furthermore, our work provides a basis for further research by outlining four research directions.We encourage researchers to develop new systems that address unsolved problems in VCS or work with biosignals beyond biooptical signals.

Acknowledgements: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK2739/1 – Project Nr. 447089431 – Research Training Group: KD²School – Designing Adaptive Systems for Economic Decisions.

References

Abramova, O., Gladkaya, M. & Krasnova, H. (2021), An unusual encounter with oneself: exploring the impact of self-view on online meeting outcomes, *in* '42nd International Conference on Information Systems', pp. 1–17.

Baig, M. M. & GholamHosseini, H. (2013), A remote monitoring system with early diagnosis of hypertension and hypotension, *in* '2013 IEEE Point-of-Care Healthcare Technologies (PHT)', IEEE, pp. 34–37.

Benyon, D. & Murray, D. (1993), 'Adaptive systems: from intelligent tutoring to autonomous agents', *Knowledge-Based Systems* **6**(4), 197–219.

Bitkom (2022), 'Wie häufig kommen die folgenden Kanäle für die interne und externe Kommunikation zum Einsatz?', *Statista* . Accessed: 2023-03-04.

Brouwer, A. M., Zander, T. O., van Erp, J. B., Korteling, J. E. & Bronkhorst, A. W. (2015), 'Using neurophysiological signals that reflect cognitive or affective state: Six recommendations to avoid common pitfalls', *Frontiers in Neuroscience* **9**(APR).

Carter, E. J., Hyde, J., Williams, D. L. & Hodgins, J. K. (2016), Investigating the influence of avatar facial characteristics on the social behaviors of children with autism, *in* 'Proceedings of the 2016 CHI conference on human factors in computing systems', pp. 140–151.

Döring, N., Moor, K. D., Fiedler, M., Schoenenberg, K. & Raake, A. (2022), 'Videoconference Fatigue: A Conceptual Analysis', *International Journal of Environmental Research and Public Health* **19**(4), 2061 – 2081.

Epstein, H.-A. B. (2020), 'Virtual meeting fatigue', *Journal of Hospital Librarianship* **20**(4), 356–360.

Fairclough, S. H. (2009), 'Fundamentals of physiological computing', *Interacting with Computers* **21**(1-2), 133–145.

Fairclough, S. H. (2011), 'Physiological computing: interfacing with the human nervous system', *Sensing Emotions: The impact of context on experience measurements* pp. 1–20.

Fairclough, S. H. (2017), Physiological computing and intelligent adaptation, *in* 'Emotions and affect in human factors and human-computer interaction', Elsevier, pp. 539–556.

Fauville, G., Luo, M., Muller Queiroz, A. C., Bailenson, J. N. & Hancock, J. (2021), 'Nonverbal Mechanisms Predict Zoom Fatigue and Explain Why Women Experience Higher Levels than Men', *SSRN Electronic Journal* .

Feigh, K. M., Dorneich, M. C. & Hayes, C. C. (2012), 'Toward a characterization of adaptive systems: A framework for researchers and system designers', *Human factors* **54**(6), 1008–1024.

Gurevitch, J., Koricheva, J., Nakagawa, S. & Stewart, G. (2018), 'Meta-analysis and the science of research synthesis', *Nature 2018 555:7695* **555**(7695), 175–182.

Hettinger, L. J., Branco, P., Encarnacao, L. M. & Bonato, P. (2003), 'Neuroadaptive Technologies: Applying Neuroergonomics to the Design of Advanced Interfaces', *Theoretical issues in ergonomics sciences* **4**(1-2), 220–237.

Higgins, D. & McDonnell, R. (2021), A preliminary investigation of avatar use in video-conferencing, *in* '2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)', IEEE, pp. 540–541.

Kaptelinin, V. (2023), 'Interaction Design Solutions for Online Meetings', *Interactions* **30**(2), 42–46.

Karl, K. A., Peluchette, J. V. & Aghakhani, N. (2021), 'Virtual Work Meetings During the COVID-19 Pandemic: The Good, Bad, and Ugly:', *Small Group Research* **53**(53 (3)), 343–365.

Kelly, H. (2020), 'The most maddening part about working from home: Video conferences', *The Washington Post* . Accessed: 2023-03-04.

Knierim, M. T., Schemmer, M. & Bauer, N. (2022), 'A simplified design of a ceegrid ear-electrode adapter for the openbci biosensing platform', *HardwareX* **12**, e00357.

Langner, M., Toreini, P. & Maedche, A. (2022), EyeMeet: A Joint Attention Support System for Remote Meetings, *in* 'CHI Conference on Human Factors in Computing Systems Extended Abstracts', ACM, New York, NY, USA, pp. 1–7.

Li, R., Curhan, J. & Hoque, M. E. (2015), Predicting video-conferencing conversation outcomes based on modeling facial expression synchronization, *in* '2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)', Vol. 1, IEEE, pp. 1–6.

Loewe, N. & Nadj, M. (2020), Physio-adaptive systems-a state-of-the-art review and future research directions., *in* 'ECIS'.

Menychtas, A., Galliakis, M., Tsanakas, P. & Maglogiannis, I. (2019), Real-time integration of emotion analysis into homecare platforms, *in* '2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)', IEEE, pp. 3468–3471.

Mitchell, M. & Jolley, J. (2013), *Research design explained*, 8 edn, Wadsworth Publishing, Belmont, California.

Murali, P., Hernandez, J., McDuff, D., Rowan, K., Suh, J. & Czerwinski, M. (2021), Affectivespotlight: Facilitating the communication of affective responses from audience members during online presentations, *in* 'Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems', pp. 1–13.

Nakazato, N., Yoshida, S., Sakurai, S., Narumi, T., Tanikawa, T. & Hirose, M. (2014), Smart face: enhancing creativity during video conferences using real-time facial deformation, *in* 'Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing', pp. 75–83.

Nickerson, R. C., Varshney, U. & Muntermann, J. (2013), 'A method for taxonomy development and its application in information systems', *European Journal of Information Systems* **22**(3), 336–359.

Oh, S. Y., Bailenson, J., Krämer, N. & Li, B. (2016), 'Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments', *PloS one* **11**(9), e0161794.

Otsuki, M., Maruyama, K., Kuzuoka, H. & Suzuki, Y. (2018), Effects of enhanced gaze presentation on gaze leading in remote collaborative physical tasks, *in* 'Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems', pp. 1–11.

Pai, R., Dubey, A. & Mangaonkar, N. (2021), Real time eye monitoring system using cnn for drowsiness and attentiveness system, *in* '2021 Asian Conference on Innovation in Technology (ASIANCON)', IEEE, pp. 1–4.

Peper, E., Wilson, V., Martin, M., Rosegard, E. & Harvey, R. (2021), 'Avoid Zoom Fatigue, Be Present and Learn', *NeuroRegulation* **8**(1), 47–56.

Pexip (2023), 'PEXIP'.
URL: *https://www.pexip.com/features/adaptive-composition*

Pope, A. T., Bogart, E. H. & Bartolome, D. S. (1995), 'Biocybernetic system evaluates indices of operator engagement in automated task', *Biological psychology* **40**(1-2), 187–195.

Rae, J. P., Steptoe, W. & Roberts, D. J. (2011), Some implications of eye gaze behavior and perception for the design of immersive telecommunication systems, *in* '2011 IEEE/ACM 15th International Symposium on Distributed Simulation and Real Time Applications', IEEE, pp. 108–114.

Reinecke, K. & Bernstein, A. (2013), 'Knowing What a User Likes: A Design Science Approach to Interfaces that Automatically Adapt to Culture', *MIS Quarterly* **37**(2), 427–453.

Riedl, R. & Léger, P.-M. (2016), 'Fundamentals of neurois', *Studies in neuroscience, psychology and behavioral economics* p. 127.

Roberts, D., Wolff, R., Rae, J., Steed, A., Aspin, R., McIntyre, M., Pena, A., Oyekoya, O. & Steptoe, W. (2009), Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together, *in* '2009 IEEE Virtual Reality Conference', IEEE, pp. 135–142.

Rojas, C., Zuccarelli, E., Chin, A., Patekar, G., Esquivel, D. & Maes, P. (2022), Towards enhancing empathy through emotion augmented remote communication, *in* 'CHI Conference on Human Factors in Computing Systems Extended Abstracts', pp. 1–9.

Samrose, S. (2018), Automated collaboration coach for video-conferencing based group discussions, *in* 'Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers', pp. 510–515.

Sartika, N., Sukmana, Y., Eliviani, R. & Awaludin, R. (2022), Smart online class monitoring service design, *in* '2022 8th International Conference on Wireless and Telematics (ICWT)', IEEE, pp. 1–7.

Schultz, T., Amma, C., Heger, D., Putze, F. & Wand, M. (2013), 'Human-Machine Interfaces based on Biosignals', *Automatisierungstechnik* **61**(11), 760–769.

Schwartz, G., Wei, S.-E., Wang, T.-L., Lombardi, S., Simon, T., Saragih, J. & Sheikh, Y. (2020), 'The eyes have it: An integrated eye and face model for photorealistic facial animation', *ACM Transactions on Graphics (TOG)* **39**(4), 91–1.

Seitz, J. (2022), Designing user-adaptive video meeting systems, *in* 'Proceedings of 20th European Conference on Computer-Supported Cooperative Work', European Society for Socially Embedded Technologies (EUSSET).

Seitz, J., Benke, I. & Madche, A. (2022), 'Fatigued by yourself? towards understanding the impact of self-view designs in virtual meeting software', *SIGHCI 2022, HCI/MIS Workshop 2022 - The 21st Annual Pre-ICIS Workshop on HCI Research in MIS, Copenhagen, Denmark* .

Seuren, L. M., Wherton, J., Greenhalgh, T. & Shaw, S. E. (2021), 'Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction', *Journal of Pragmatics* **172**, 63–78.

Skillscouter (2022), ‘Video Conferencing Statistics’.
URL: *https://skillscouter.com/video-conferencing-statistics*

Song, G., Cai, J., Cham, T.-J., Zheng, J., Zhang, J. & Fuchs, H. (2018), Real-time 3d face-eye performance capture of a person wearing vr headset, *in* ‘Proceedings of the 26th ACM international conference on Multimedia’, pp. 923–931.

Sun, W., Li, Y., Tian, F., Fan, X. & Wang, H. (2019), ‘How presenters perceive and react to audience flow prediction in-situ: An explorative study of live online lectures’, *Proceedings of the ACM on Human-Computer Interaction* **3**(CSCW), 1–19.

Taguchi, W., Nihei, F., Takase, Y., Nakano, Y. I., Fukasawa, S. & Akatsu, H. (2018), Effects of face and voice deformation on participant emotion in video-mediated communication, *in* ‘Proceedings of the 20th International Conference on Multimodal Interaction: Adjunct’, pp. 1–5.

Takemae, Y., Otsuka, K. & Mukawa, N. (2004), An analysis of speakers’ gaze behavior for automatic addressee identification in multiparty conversation and its application to video editing, *in* ‘RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)’, IEEE, pp. 581–586.

Takemae, Y., Otsuka, K. & Yamato, J. (2005), Automatic video editing system using stereo-based head tracking for multiparty conversation, *in* ‘CHI’05 extended abstracts on Human factors in computing systems’, pp. 1817–1820.

Tan, C. S. S., Luyten, K., Van Den Bergh, J., Schöning, J. & Coninx, K. (2014), ‘The role of physiological cues during remote collaboration’, *Presence: Teleoperators and Virtual Environments* **23**(1), 90–107.

Telaar, D., Wand, M., Gehrig, D., Putze, F., Amma, C., Heger, D., Vu, N. T., Erhardt, M., Schlippe, T., Janke, M. et al. (2014), Biokit—real-time decoder for biosignal processing, *in* ‘Fifteenth Annual Conference of the International Speech Communication Association’.

Tewele, M. K., Yousefi, S. & Milrad, M. (2015), Supporting video conference communication using a vision-based human facial synthesis approach, *in* ‘2015 SAI Intelligent Systems Conference (IntelliSys)’, IEEE, pp. 807–812.

Valkenburg, P. M. & Peter, J. (2013), ‘The Differential Susceptibility to Media Effects Model’, *Journal of Communication* **63**(2), 221–243.

Webster, J. & Watson, R. T. (2002), ‘Analyzing the past to prepare for the future: Writing a literature review’, *MIS quarterly* pp. xiii–xxiii.

White, R. W. & Morris, D. (2007), Investigating the querying and browsing behavior of advanced search engine users, *in* ‘Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval’, SIGIR ’07, Association for Computing Machinery, New York, NY, USA, p. 255–262.

Yang, R. & Zhang, Z. (2004), ‘Eye gaze correction with stereovision for video-teleconferencing’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(7), 956–960.

Zander, T. O. & Kothe, C. (2011), ‘Towards passive brain-computer interfaces: Applying brain-computer interface technology to human-machine systems in general’, *Journal of Neural Engineering* **8**(2).

Zhang, Y., Yang, J., Liu, Z., Wang, R., Chen, G., Tong, X. & Guo, B. (2022), ‘Virtualcube: An immersive 3d video communication system’, *IEEE Transactions on Visualization and Computer Graphics* **28**(5), 2146–2156.

Zheng, Y.-L., Ding, X.-R., Poon, C. C. Y., Lo, B. P. L., Zhang, H., Zhou, X.-L., Yang, G.-Z., Zhao, N. & Zhang, Y.-T. (2014), ‘Unobtrusive Sensing and Wearable Devices for Health Informatics’, *IEEE Transactions on Biomedical Engineering* **61**(5), 1538–1554.