

Received December 27, 2021, accepted January 8, 2022, date of publication January 14, 2022, date of current version January 27, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3143814

# Subjective Influence of Camera-Gaze Angular Offset

JAN HOLUB<sup>1</sup>, (Member, IEEE), SCOTT ISABELLE<sup>2</sup>, (Member, IEEE),

ADÉLA KRYLOVÁ<sup>1</sup>, AND HAKOB AVETISYAN<sup>1</sup>, (Member, IEEE)

<sup>1</sup>Faculty of Electrical Engineering, Czech Technical University in Prague, 16627 Prague, Czech Republic

<sup>2</sup>Amazon, Alexa Voice Services, Santa Clara, CA 95054, USA

Corresponding author: Jan Holub (holubjan@fel.cvut.cz)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Faculty Dean's Advisory Committee, and performed in line with the Declaration of Helsinki.

**ABSTRACT** The study aimed to evaluate the difficulty in maintaining eye contact during a teleconference for different camera-gaze angular offsets. Videoconferencing systems may compromise the eye contact between participants due to imperfect angular alignment between the center of the screen and the camera. During a teleconference, difficulty maintaining eye contact may be perceived as uncomfortable or unsatisfying by participants in the call. An experiment deploying 33 test subjects was performed and evaluated. Groups of three test subjects performed a videoconference call. The effect of camera-gaze angular offset was quantified by assessing Effectiveness and Engagement ratings. It was shown that even 10 degrees of misalignment cause a statistically significant drop in Effectiveness for business-type teleconference calls. Camera-gaze angular offset is an important parameter to be considered during videoconferencing equipment design or setup. Higher misalignments may cause significant drops in perceived call quality (Quality of Experience). Our research results can be practical in conferencing equipment design and setup, modern telecommunication equipment design, workplace ergonomics, or distance learning equipment setup.

**INDEX TERMS** Product design, interface evaluation, usability/acceptance measurement, distance learning, communication micro design features.

## I. INTRODUCTION

Eye contact is an essential part of human-to-human communication that carries a significant amount of non-verbal communication [1]. New technologies allow people to interact with each other in any situation using audio or video communications. The recent lockdowns due to the COVID-19 virus spread forced many to work and study from home and communicate online using various services such as Zoom, Skype, Google Meet, and Microsoft Teams. However, many professionals use videoconferencing routinely, even during their regular work, including military and public safety commanders and employees. Achieving the highest possible communication quality and experience there is essential, and any problem potentially can cause fatal consequences.

Videoconferencing systems may compromise eye contact between participants due to imperfect angular alignment

The associate editor coordinating the review of this manuscript and approving it for publication was Liang-Bi Chen.

between the center of the screen and the camera. Difficulty in maintaining mutual eye contact (or inability to maintain mutual eye contact at all) during a teleconference may be perceived as uncomfortable by users of the teleconferencing system (participants in the call) [2]. Various software and hardware solutions are implemented to compensate for the gaze-angular offset [3]. To assess the human perception of this phenomenon, user QoE (Quality of Experience) tests may be used.

## II. STATE OF THE ART

Psychologists have studied the significance of person-to-person eye contact since the 1960s. [2] found that the average angular value perceived by two persons as a breach of eye contact is 2.8°. Such angle corresponds approximately to 10 cm at the distance of 2m between the persons. A similar experiment affirming this result was performed by [4]. In this experiment, subjects' heads were fixed by headrests, and the distance between the looker and the observer was 122 cm.

Both of these experiments also indicated that the acuity of eye contact perception is as accurate as the [5] visual acuity. This finding was reaffirmed by further investigations, e.g., by [6].

Several systematic inaccuracies in human gaze perception were identified [2]. Furthermore, [4] found that the observer may underestimate the rotation angle if the looker's head is turned away from the sightline. Another interesting point is that eye contact is often overestimated for longer distances [7], [8].

The influence of eye contact in videoconference technology was first studied in Bell Labs [9], where the threshold for loss of eye contact was identified as  $4.5^\circ$  for horizontal misalignment and  $5.5^\circ$  for vertical misalignment. However, it is unclear whether the difference between the angular thresholds measured during the face-to-face experiment and the videoconference was caused by the videoconferencing technology and its technical quality (resolution, frame rate, etc.) or by some difference in the test methodology. The experiment described in [9] and other consequent experiments have attempted to resolve whether the camera should be positioned below or above the screen [10]. However, no clear answer was identified.

The difference between the accuracy of direct eye identification in a face-to-face meeting and a videoconference meeting was investigated in [11], and only a slight difference was found. Also, [12] suggests that the physical presence or proximity of the other person is not necessary for typical psychophysiological responses to eye contact like autonomic arousal and facial reactions. [13] has identified that the sensitivity to eye contact is asymmetric, being less sensitive when people look below our eyes than when they look to the left, right, or above our eyes. He also concluded that the critical angle at which the eye contact is lost depends on the observer's expectations and viewing conditions but is identified between  $5^\circ$  and  $10^\circ$  in most experiments.

More general non-verbal features and their influence on teleconferencing were analyzed by [14]. A significant difference between head-only framing and upper-body framing and between head-only framing and face-to-face has been found. However, only an insignificant difference was observed between upper-body framing and face-to-face communication. Similarly, [15] concluded that the placement of video cameras influences the final Effectiveness of videoconferencing, suggesting that the camera should be set up in such a way that not only the face but also the upper body is included in the picture. Attention was also drawn to the importance of so-called "gaze awareness" during a teleconference.

The gaze-angular offset effect can be compensated using various software- and hardware-based methods. [3] suggest a software-based technique of eye contact correction by redirecting gaze from an arbitrary angle to the center, alternating the video using a deep convolutional neural network. Another method [16] suggests changing the conversation members' gaze with 3D modeled eyeballs. A hardware-related method using dual (stereo) cameras has been proposed by [17]. [18] suggest a solution by using depth sensors such as Kinect.

A newer DIY (Do it yourself) technique allows eye-to-eye video communication using a semitransparent acrylic mirror [19].

### III. MOTIVATION

Research analysis has shown that gaze awareness plays during teleconferences an essential role. However, it is not clear how this phenomenon influences the overall QoE, as perceived by a user of teleconferencing equipment during a conference call. Therefore, an experiment was set up to investigate the influence of angular camera horizontal misplacement on the overall impression, expressed by two measured parameters: Call Effectiveness and Call Engagement. The Call Engagement parameter is adopted from [20]. Based on [9], [10], and [13], it was decided to investigate the influence of horizontal camera misalignment, keeping the vertical misalignment constant. Assuming the symmetry in the perceptual impact of camera horizontal misalignment as found in [13], we decided to move the camera between the center and right positions only to decrease the number of tests needed.

### IV. METHODOLOGY

#### A. TEST SETUP

The test was designed as follows: three separate rooms, each with identical tablet PCs with an external webcam and headphones, were used. For each conversation, the position of the webcam was changed synchronously in each of the test rooms. Three camera positions were tested - corresponding to horizontally measured  $0^\circ$  (middle),  $5.7^\circ$  (slightly right), and  $11.3^\circ$  (right). Vertically, the cameras were always placed 5 cm above the upper screen border. Figure 1 illustrates the different camera positions. The camera positions were proposed based on [13] results.

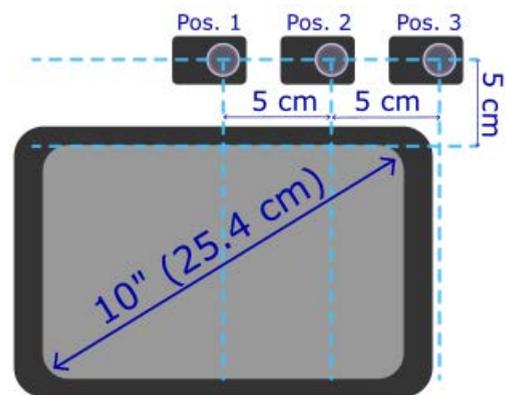
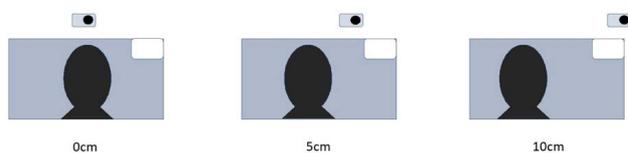


FIGURE 1. Tablet PC and camera positions.

Each subject's camera position was tested three times, so there were nine conversations among all three subjects. The cameras' positions were changing randomly, so it was harder for subjects to get used to the camera positions. The camera positions were changed only during the breaks between the conversations, and these changes were

synchronous in all three rooms, so each conversation was made with a fixed camera position, identical for all three participants.

Prior to the testing, it was determined by a short expert test that the “speaker’s view” mode is a more suitable layout for the test than the default grid layout, where all the participants are seen. Speaker’s view is the setting where only the currently speaking participant is seen; the videoconferencing tool switches the speakers automatically. The advantage of this layout is that the subjects can concentrate on a single person, and they are not disturbed by other participants who are not currently speaking. Figure 2 depicts the view of the subjects for each of the three various positions of the camera. It can be seen that with increasing camera misplacements (identical on all three computers), the speaker’s position on the display is moved moderately to the left while the camera itself is moved to the right. This makes concurrent watching of both the speaker and the camera more difficult for higher displacement.



**FIGURE 2.** Speaker’s view for different camera positions 0cm, 5cm, and 10 cm (head silhouettes are in scale with the display size, viewing camera angle, and viewing distance). The white object in the upper right display corner is the self-preview cover.

It is natural for the subjects to correct their position (or the camera position) if they see how they appear on the partner’s screen. Even this potential own position adjustment does not affect the mutual eye contact (as it does not change the angular misalignment between the camera and their displayed conversational partner eye position) it was decided to hide their own preview on the screen by placing a mechanical cover on the relevant area of the screen to avoid subjects being distracted from the conversation by observing their own preview. The subjects were not explicitly instructed to watch the speaker on the screen or to look into the camera; they were just instructed to maintain the position of their head to keep the distance from the screen within the required limits (see Test Environment chapter for details). In this way, they could not decrease the angular displacement of the camera, so with higher misplacement angle, they were less able to watch the speaker and the camera at the same time, so mutual eye contact was more difficult to maintain (if they watched the speaker, the speaker saw they are looking aside, and if they tried to mimic the eye contact by looking into the camera, they were not able to see the speaker with his/her mimics, emotions, lip movements, etc.)

To simulate natural conversations, a set of 9 scenarios was prepared. These scenarios were based on real-life situations, such as planning a birthday party, arranging a business meeting, making an appointment, etc. In each conversation, each

tested subject played a different role to avoid monotony or to avoid repeating a scenario and to stimulate the interest of the participants. Each participant played each role once to reduce the risk of a spoiled evaluation due to lack of understanding or personal discomfort caused by the theme of the scenario (playing the role of an opponent, acting like a boss, expressing a minority opinion, etc.).

The order of the scenarios was randomized. In the entire experiment, three different randomizations were used to minimize possible inter-scenario dependencies affecting the voting.

## B. TEST ENVIRONMENT

Tests were performed in the acoustically treated critical watching/listening rooms that fully conformed with the requirements of [21] ITU-T Rec. P.800 “Methods for subjective determination of transmission quality.” The light and acoustic conditions followed [22] ITU-T Rec. P.910 “Subjective video quality assessment methods for multimedia applications” and [23] ITU-T Rec. P.920 “Interactive test methods for audiovisual communications,” as appropriate.

The listening equipment used in the tests consisted of closed headphones with a closely spaced boom microphone with a directional pickup microphone (Plantronics Blackwire 5000 series).

The video device was a tablet PC with a 10.1” screen and Windows OS. The brightness of the screen was set up following [23] and was checked before each session, as required there. By expert pre-testing deploying five expert testers running multiple conversational test scenarios and following recommendations in [22] and [23], the viewing distance was defined as 18, with a maintained accuracy of  $\pm 2''$ . This distance approximately equals two diagonals of the tablet screen, or 4H (H = screen height). This value is suitably within the range of 1 to 8 H as recommended by [22] for viewing distance. The distance was maintained by marks indicating the subject’s head position and checked by the experiment supervisor during the entire session (a session denotes a set of 9 conversations among three subjects). The headsets were calibrated before and verified after the experiment.

It was impossible to use the built-in webcams due to their fixed position. Therefore, additional external webcams (Ausdom AW635 with Full HD resolution and viewing angle 60°) were installed for the tests.

A Skype videoconferencing environment was chosen for the testing, as it provides a reasonably stable connection with various anti-distortion algorithms [24].

Printed questionnaires were used for collecting the votes to avoid subjects’ confusion by using another electronic device for vote collection.

## C. TEST SUBJECTS

[25] shows that a sufficient number of participants for tests employing native subjects is usually considered to be 24 – 32, which is also stated in [21]. This number of participants usually proves or disproves the hypothesis with sufficient

accuracy, of course, only after assuring enough test samples are available for each tested condition. We decided to employ the number of test subjects exceeding the upper limit of this interval to achieve for each of the (three) positions of the cameras enough votes (the exact numbers are given in the chapter Test Flow). It was preferred that the subjects do not know any of the other participants within the session group, as it could affect the results. People of different ages, education, and telecommunication experience, were recruited. Each panel was assigned at least one male and at least one female participant randomly so that the M: F gender ratio was maintained between 1:2 and 2:1 (our of total 11 groups of subjects, ten groups were 2F+1M and one group was 2M+1F). The conversations were held in the Czech language, and only Czech native speakers with no visual or hearing, or speaking impairments were accepted as participants.

Before the test, each subject's normal visual acuity was tested using the Snellen test [5], and her/his normal color vision was tested using the Ishihara test [26]. The subjects were also asked to self-report about their hearing ability and their visual ability. In paper form, questions were also asked about the subject's age, gender, and teleconference experience.

Each test session started with a joint instruction session, in which the subjects had an opportunity to interact face-to-face to get acquainted with each other and set their expectations about natural response times and conversational behaviors. This research complied with the tenets of the Declaration of Helsinki. Informed consent was obtained from each participant.

#### D. TEST FLOW

Test subjects were familiarized with the test scenarios and procedures but not with the aim of the test. Disclosing the test items would have influenced subjects to focus more on the camera settings than on the overall teleconference experience.

During each teleconference session, an experiment supervisor was present in each test room. Her/his task was to supervise the test, i.e., to check the subject's head position and the camera positioning, to check that the subject filled in the questionnaire, and to be ready to help if any unexpected problems arose. The supervisors did not disturb the subject or intervene in any way in the test, except for checking the subject's head position during the conversations and changing the camera position between the conversations. They were ready to pause the experiment in case of any subject moving his/her head out of the desired viewing distance range. In the end, it was not necessary (subjects were aware that this parameter was crucial, and the supervisors were focused during the conversations on this test aspect).

After each round of the test, the subjects were asked to answer two questions on the printed questionnaire. The first question asked about the conversation Effectiveness, and the second question asked about the subject's Engagement. The participants answered each question by voting on the usual

**TABLE 1. Rating scales for the effectiveness criteria.**

Opinion Score	How would you rate the video call effectiveness?
5	Excellent / Very effective
4	Good / Effective
3	Fair / Somehow effective
2	Poor / Almost ineffective
1	Bad / Ineffective at all

**TABLE 2. Rating scales for the engagement criteria.**

Opinion Score	How would you rate the video call engagement?
5	Excellent / Very engaging
4	Good / E engaging
3	Fair / Somehow engaging
2	Poor / Almost not engaging
1	Bad / Not engaging at all

opinion scale adopted from [21]. Rating scales of the tests are shown in Table 1 and Table 2. The subjects could also comment in detail on each teleconference call in a separate comment column.

Each conversation lasted approximately three minutes and was terminated by the subjects when they agreed on the solution to the scenario. In total, 33 subjects participated in the tests, and 11 test sessions were run. The average age of the subjects was 33.3 years, with a standard deviation of 12.83 years. There were 12 male subjects and 21 female subjects. Nineteen of the subjects classified themselves as frequent users of teleconference tools (e.g., Skype, Google Meets, WhatsApp, Zoom, Telegram). Ten subjects classified themselves as rare users, and the rest stated that they did not use video calls at all. The impact of experience with teleconferencing systems is examined in the section Result Analysis.

#### V. RESULTS

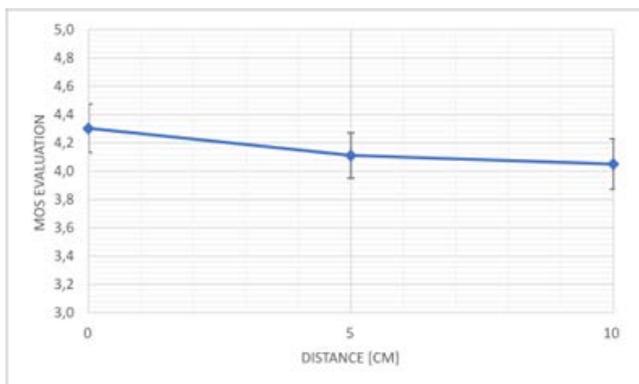
The subjects' answers expressed in an opinion scale of 1-5 were averaged per condition (means per each viewing angle) to obtain the Mean Opinion Score (MOS). The averaging for each condition was done across all three participants, three scenarios, and across all 11 groups of subjects; it means, each of the three examined conditions (for each of the two examined parameters) has been represented by an average of 99 votes. As expected, with the increasing distance of the camera from the center of the field of view, the results show decreasing user ratings. Figures 3 to 6 show the dependence of the user's evaluation of call effectiveness and call Engagement on the camera distance from the horizontal center of the field of view. In Tables 3 and 4, the STD column indicates the standard deviation of the arithmetical mean and the CI95 column shows the 95% confidence interval for the obtained results. Also median of each condition is reported.

**TABLE 3.** Average user evaluation and median of effectiveness, according to the distance of the camera from the horizontal center of the field of view.

Horizontal misplacement	MOS	STD	CI95	median
0cm	4.303	±0.870	±0.171	EXCELLENT
5cm	4.111	±0.811	±0.160	GOOD
10cm	4.051	±0.910	±0.179	GOOD



**FIGURE 3.** Stacked bars visualisation of the user’s evaluation of call Effectiveness versus camera horizontal distance from the center of the field of view.



**FIGURE 4.** The dependence of the user’s evaluation of call Effectiveness on the camera horizontal distance from the center of the field of view.

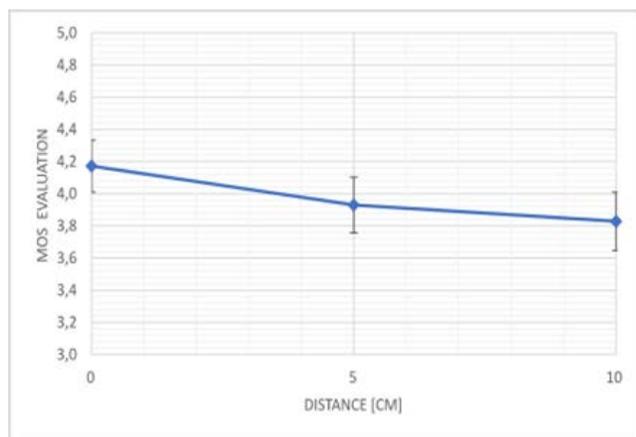
**VI. RESULTS ANALYSIS**

**A. FRIEDMANN TEST**

A non-parametric Friedmann Test [27] was carried out to compare the sums of scores from each test subject for positions 0cm, 5cm and 10cm. There was found to be a significant difference between the positions for the Effectiveness parameter,  $\chi^2(3) = 7.091, p < 0.029$ . Though similar trends of lower assessments for higher camera displacements are seen in the results of Engagement, too, no significant differences have been found in assessments of Engagement parameter for different camera positions  $\chi^2(3) = 2.561, p < 0.278$ .



**FIGURE 5.** Stacked bars visualisation of the user’s evaluation of call Engagement versus camera horizontal distance from the center of the field of view.



**FIGURE 6.** The dependence of the user’s evaluation of call Engagement on the camera horizontal distance from the center of the field of view.

**TABLE 4.** Average user evaluation and median of engagement, according to the distance of the camera from the horizontal center of the field of view.

Horizontal misplacement	MOS	STD	CI95	median
0cm	4.172	±0.821	±0.162	GOOD
5cm	3.929	±0.879	±0.173	GOOD
10cm	3.828	±0.920	±0.181	GOOD

**B. WILCOXON SIGNED-RANKS TEST**

Wilcoxon signed-ranks test identifies individual differences between judgements of two camera positions and has been applied to Effectiveness results to identify which positions are statistically different [28]. Based on the Wilcoxon signed-ranks test of the Effectiveness results, a statistically significant difference between the middle camera position (0cm) and right camera position (10cm) was found ( $p = 0.029$ ) and middle camera position (0cm) and slightly right (5cm) camera position ( $p = 0.037$ ). In Table 5, Wilcoxon signed-ranks test results for all pairs of tested camera positions are reported. The difference between groups is

**TABLE 5.** Wilcoxon signed-ranks test  $T/T_{crit}$  for Effectiveness for all tested position pairs.

	0 cm	5 cm	10 cm
0 cm	-	55.5/58*	81.5/89*
5 cm		-	104/73
10 cm			-

**TABLE 6.** Benjamini-Hochberg correction procedure for all tested position pairs, effectiveness test parameter,  $Q = 0.057$ .

	$(i/m)Q$	$p$	Significant?
0cm, 10cm	0.019	0.029298	YES
0cm, 5cm	0.038	0.037028	YES
5cm, 10cm	0.057	0.301085	NO

**TABLE 7.** Subjective assessment of Effectiveness, based on the subject's experience with the teleconferencing.

	MOS	STD	CI95
Not at all	<b>4.11</b>	±0.222	±0.126
Rare	<b>3.88</b>	±0.553	±0.198
Frequent	<b>4.32</b>	±0.542	±0.141

**TABLE 8.** Subjective assessment of Engagement, based on the subject's experience with the teleconferencing.

	MOS	STD	CI95
Not at all	<b>3.75</b>	±0.516	±0.292
Rare	<b>3.82</b>	±0.628	±0.225
Frequent	<b>4.11</b>	±0.634	±0.165

considered significant for  $T < T_{crit}$ . The significant values are marked by an \* (asterisk).

To control the false discovery rate (Type 1 Error), the Benjamini-Hochberg correction [29] has been applied as the traditionally used conservative Bonferonni's correction led to inconclusive results. The Benjamini-Hochberg correction proved the statistical significance of the comparisons (0cm, 10cm) and (0cm, 5cm) at the Q level 0.057, thus also at both commonly used acceptable false discovery rates 0.15 or even 0.25 [30], see Table 6.

**C. OTHER FACTORS**

Each subject's interpretation of the MOS scale differs; some subjects voted more critically, whereas other subjects voted very leniently. The achieved results suggest the rate of leniency increases with the subject's teleconference experience though these are not statistically significant. Tables 7 and 8 depict the difference in voting based on the subject's experience with the teleconference platform for the Effectiveness and Engagement criteria. The difference could be caused by the more experienced subjects adapting better to a low level of eye contact. Novices may be less able to

deal with the differences between face-to-face interaction and interaction via the teleconference platform.

**VII. CONCLUSION**

An experiment identifying subjects' sensitivity to camera angular misplacement was designed, performed, and evaluated. The tested parameters were Effectiveness and Engagement. While Engagement showed only insignificant sensitivity to the tested angular misalignment, it was shown that even 11.3° of horizontal misalignment causes a drop in Effectiveness of approximately 0.25 MOS for business-type teleconference calls. Eventhough the results are not totally conclusive, there is a clear trend that should be confirmed by further studies. The identified difference in the 5-grade MOS scale is fully comparable with differences between speech coder candidates for the NATO coder selection procedure, which resulted in MELPe adoption as STANAG 4591 in 2001, or performance differences between 3G and LTE speech coders. The identified dependencies between Effectiveness and angular horizontal misalignment give design rules for multimedia hardware designers or integrators and enable existing or future HW solutions optimization (office or home office communication equipment placement, cockpit designs deploying embedded cameras, etc.).

A higher degree of horizontal misalignment may have a larger impact - quantification is needed by future research. Other limitations of the presented experiment are the limited number of camera positions, as well as asymmetric (center and right only) misalignment positions. Another interesting aspect could be vertical misalignment influence, as our experiment investigated horizontal misalignment only. To study the impact of misalignment on Effectiveness may also require larger misalignment values.

During the experiment, the self-view has been disabled. Therefore, the results may differ in teleconferencing environments which provide this feature. It has to be also noted that the tested parameters - Effectiveness and Engagement - are not necessarily related to transmission quality only; for example, one can have an excellent transmission quality but little eye contact and thus a probable lower engagement.

The type of task may have an impact: Personal usage with a possibly higher degree of emotional connection may be impacted differently (e.g., reduction in eye contact may be perceived as a more personal disengagement).

The impact of misalignment may be higher for newer or less-experienced users of video conferencing. Devices intended for professional use may require a different design than devices intended for family or more casual/intermittent use or targeted to users with less experience with this technology.

**ACKNOWLEDGMENT**

The authors would like to thank Mesaqin.com Ltd. subjective test laboratory for providing the test premises, the test equipment, and support for the test procedure (installation, verification, and calibration of equipment).

## REFERENCES

- [1] M. Argyle and J. Dean, "Eye contact, distance, and affiliation," *Sociometry*, vol. 28, pp. 289–304, May 1965.
- [2] J. J. Gibson and A. D. Pick, "Perception of another person's looking behavior," *Amer. J. Psychol.*, vol. 76, no. 3, p. 386, Sep. 1963, doi: [10.2307/1419779](https://doi.org/10.2307/1419779).
- [3] L. F. Isikdogan, T. Gerasimov, and G. Michael, "Eye contact correction using deep neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 3307–3315, doi: [10.1109/WACV45572.2020.9093554](https://doi.org/10.1109/WACV45572.2020.9093554).
- [4] M. G. Cline, "The perception of where a person is looking," *Amer. J. Psychol.*, vol. 80, no. 1, pp. 41–50, 1967, doi: [10.2307/1420539](https://doi.org/10.2307/1420539).
- [5] H. Snellen, *Probuchstaben zur Bestimmung der Sehschärfe*. Rendsburg, Germany: H. Peters, 1873. [Online]. Available: <https://books.google.cz/books?id=zzJAAAIAAJ>
- [6] J. Jaspars, "Het observeren van ogencontact," *Nederlands Tijdschrift Voor Psychol.*, vol. 28, pp. 67–81, Oct. 1973.
- [7] D. J. Knight, D. Langmeyer, and D. C. Lundgren, "Eye-contact, distance, and affiliation: The role of observer bias," *Sociometry*, vol. 36, no. 3, p. 390, 1973, doi: [10.2307/2786340](https://doi.org/10.2307/2786340).
- [8] G. M. Stephenson and D. R. Rutter, "Eye-contact, distance and affiliation: A re-evaluation," *Brit. J. Psychol.*, vol. 61, no. 3, pp. 385–393, Aug. 1970, doi: [10.1111/j.2044-8295.1970.tb01257.x](https://doi.org/10.1111/j.2044-8295.1970.tb01257.x).
- [9] R. Stokes, "Human factors and appearance design considerations of the mod II PICTUREPHONE station set," *IEEE Trans. Commun.*, vol. COM-17, no. 2, pp. 318–323, Apr. 1969, doi: [10.1109/TCOM.1969.1090060](https://doi.org/10.1109/TCOM.1969.1090060).
- [10] B. Stapley, *Visual Enhancement of Telephone Conversations*. London, U.K.: Imperial College, Univ. London, 1972.
- [11] S. M. Anstis, J. W. Mayhew, and T. Morley, "The perception of where a face or television 'Portrait' is looking," *Amer. J. Psychol.*, vol. 82, no. 4, p. 474, Dec. 1969, doi: [10.2307/1420441](https://doi.org/10.2307/1420441).
- [12] J. O. Hietanen, M. J. Peltola, and J. K. Hietanen, "Psychophysiological responses to eye contact in a live interaction and in video call," *Psychophysiology*, vol. 57, no. 6, 2020, Art. no. e13587, doi: [10.1111/psyp.13587](https://doi.org/10.1111/psyp.13587).
- [13] M. Chen, "Leveraging the asymmetric sensitivity of eye contact for videoconference," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. Changing World, Changing Ourselves*, 2002, pp. 49–56, doi: [10.1145/503376.503386](https://doi.org/10.1145/503376.503386).
- [14] D. T. Nguyen and J. Canny, "More than face-to-face: Empathy effects of video framing," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2009, pp. 423–432, doi: [10.1145/1518701.1518770](https://doi.org/10.1145/1518701.1518770).
- [15] P. Slovak, *Effect of Videoconferencing Environments on Perception of Communication Cyberpsychology*. Accessed: Jul. 30, 2020. [Online]. Available: <https://scholarlychology.eu/article/view/4205/3246>
- [16] E. Wood, T. Baltrušaitis, L. P. Morency, and P. Robinson, "Bulling: GazeDirector: Fully articulated eye gaze redirection in video," *Comput. Graph. Forum*, vol. 37, no. 2, pp. 217–225, 2018, doi: [10.1111/cgf.13355](https://doi.org/10.1111/cgf.13355).
- [17] A. Criminisi, J. Shotton, A. Blake, and P. H. Torr, "Gaze manipulation for one-to-one teleconferencing," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 1, 2003, pp. 191–198, doi: [10.1109/iccv.2003.1238340](https://doi.org/10.1109/iccv.2003.1238340).
- [18] C. Kuster, T. Popa, J. C. Bazin, C. Gotsman, and M. Gross, "Gaze correction for home video conferencing," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–6, 2012, doi: [10.1145/2366145.2366193](https://doi.org/10.1145/2366145.2366193).
- [19] D. Schneider, "The zoom box: Look people in the eye while videoconferencing—[Hands on]," *IEEE Spectr.*, vol. 57, no. 8, pp. 14–16, Aug. 2020, doi: [10.1109/MSPEC.2020.9150545](https://doi.org/10.1109/MSPEC.2020.9150545).
- [20] B. Herrmann and I. S. Johnsrude, "A model of listening engagement (MoLE)," *Hearing Res.*, vol. 397, Nov. 2020, Art. no. 108016, doi: [10.1016/j.heares.2020.108016](https://doi.org/10.1016/j.heares.2020.108016).
- [21] *Methods for Subjective Determination of Transmission Quality*, document ITU-T Rec 800, ITU-T Recommendation. 1996.
- [22] *Subjective Video Quality Assessment Methods for Multimedia Applications Subjective Video Quality Assessment Methods for Multimedia Applications*, document ITU-T Rec. P.910, 1996.
- [23] *Interactive Test Methods for Audiovisual Communications*. document ITU-T Rec. P.920, ITU-T Recommendation. 2000.
- [24] X. Zhang, Y. Xu, H. Hu, Y. Liu, Z. Guo, and Y. Wang, "Modeling and analysis of Skype video calls: Rate control and video quality," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1446–1457, Oct. 2013, doi: [10.1109/TMM.2013.2247988](https://doi.org/10.1109/TMM.2013.2247988).
- [25] M. H. Pinson, L. Janowski, R. Pepion, Q. Huynh-Thu, C. Schmidmer, P. Corriveau, A. Younkin, P. Le Callet, M. Barkowsky, and W. Ingram, "The influence of subjects and environment on audiovisual subjective tests: An international study," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 640–651, Oct. 2012, doi: [10.1109/JSTSP.2012.2215306](https://doi.org/10.1109/JSTSP.2012.2215306).
- [26] J. H. Clark, "The ishikara test for color blindness," *Amer. J. Physiol. Opt.*, vol. 5, pp. 269–276, Feb. 1924.
- [27] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Statist. Assoc.*, vol. 32, no. 200, pp. 675–701, Dec. 1937, doi: [10.1080/01621459.1937.10503522](https://doi.org/10.1080/01621459.1937.10503522).
- [28] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945, doi: [10.2307/3001968](https://doi.org/10.2307/3001968).
- [29] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc., B*, vol. 57, pp. 289–300, Jan. 1995. [Online]. Available: <http://www.jstor.org/stable/2346101>
- [30] J. McDonald, *Handbook of Biological Statistics*. Baltimore, MD, USA: Sparky House, 2014, pp. 254–260.



**JAN HOLUB** (Member, IEEE) received the Ph.D. degree, in 1999. He became a Professor, in 2016. He currently leads the Human Factors Laboratory, Department of Measurement, Faculty of Electrical Engineering, Czech Technical University in Prague. He regularly contributes to 3GPP (ETSI) standardization and authored two international patents and 38 journal articles. His research interests include speech, audio, and multimedia transmission quality, subjective testing deploying parallel task methodologies, and objective user experience estimators.



**SCOTT ISABELLE** (Member, IEEE) received the Ph.D. degree from Boston University. Now, he works as a Principal Solutions Engineer at Amazon Alexa. He contributes to the standardization effort of ITU-T and ETSI, where he currently chairs the Speech and Multimedia Transmission Quality Group (STQ). Before joining the Amazon Alexa Research Team, he worked as a Distinguished Researcher at Motorola and Knowles Intelligent Audio afterward. He has authored more than 20 journal articles and over 40 international recommendations. His research interests include human hearing, psychoacoustics, audio signal processing, and audio and video quality assessment methods.



**ADÉLA KRYLOVÁ** is currently pursuing the degree in computer science. She works as a Laboratory Assistant with the Human Factors Laboratory, Department of Measurement, Faculty of Electrical Engineering, Czech Technical University in Prague. Her primary research interests include new subjective testing methods, test tools development, programming and validation, and advanced statistical data analysis. Her focus on human factors was accelerated, in 2018, when she participated in a simulated mission to mars, being isolated in a spaceship simulator for 101 hours, which was the longest simulated student mars mission in Europe.



**HAKOB AVETISYAN** (Member, IEEE) received the Ph.D. degree from the Faculty of Electrical Engineering, Czech Technical University in Prague, in 2021. He is the author or coauthor of multiple conference contributions and six journal articles in the area of subjective user experience measurements.