Original Paper

Automatic Classification of Screen Gaze and Dialogue in Doctor-Patient-Computer Interactions: Computational Ethnography Algorithm Development and Validation

Samar Helou[1], PhD; Victoria Abou-Khalil[2], PhD; Riccardo Iacobucci[3], PhD; Elie El Helou[4], MD, MSc; Ken Kiyono[5], PhD

[1]Global Center for Medical Engineering and Informatics, Osaka University, Osaka, Japan

[2]Academic Center for Computing and Media Studies, Kyoto University, Kyoto, Japan

[3]Department of Urban Management, Graduate School of Engineering, Kyoto University, Kyoto, Japan

[4]Faculty of Medicine, Saint Joseph University, Beirut, Lebanon

[5]Graduate School of Engineering Science, Osaka University, Osaka, Japan

Corresponding Author:
Samar Helou, PhD
Global Center for Medical Engineering and Informatics
Osaka University
Osaka Prefecture, Suita, Yamadaoka 2-2
Osaka, 565-0871
Japan
Phone: 81 8056856848
Email: helou.samar@gmail.com

Abstract

Background: The study of doctor-patient-computer interactions is a key research area for examining doctor-patient relationships; however, studying these interactions is costly and obtrusive as researchers usually set up complex mechanisms or intrude on consultations to collect, then manually analyze the data.

Objective: We aimed to facilitate human-computer and human-human interaction research in clinics by providing a computational ethnography tool: an unobtrusive automatic classifier of screen gaze and dialogue combinations in doctor-patient-computer interactions.

Methods: The classifier's input is video taken by doctors using their computers' internal camera and microphone. By estimating the key points of the doctor's face and the presence of voice activity, we estimate the type of interaction that is taking place. The classification output of each video segment is 1 of 4 interaction classes: (1) screen gaze and dialogue, wherein the doctor is gazing at the computer screen while conversing with the patient; (2) dialogue, wherein the doctor is gazing away from the computer screen while conversing with the patient; (3) screen gaze, wherein the doctor is gazing at the computer screen without conversing with the patient; and (4) other, wherein no screen gaze or dialogue are detected. We evaluated the classifier using 30 minutes of video provided by 5 doctors simulating consultations in their clinics both in semi- and fully inclusive layouts.

Results: The classifier achieved an overall accuracy of 0.83, a performance similar to that of a human coder. Similar to the human coder, the classifier was more accurate in fully inclusive layouts than in semi-inclusive layouts.

Conclusions: The proposed classifier can be used by researchers, care providers, designers, medical educators, and others who are interested in exploring and answering questions related to screen gaze and dialogue in doctor-patient-computer interactions.

(J Med Internet Res 2021;23(5):e25218) doi: 10.2196/25218

KEYWORDS

computational ethnography; patient-physician communication; doctor-patient-computer interaction; electronic medical records; pose estimation; gaze; voice activity; dialogue; clinic layout

Introduction

Background

Doctor-patient communication is a combination of verbal and nonverbal expressions and can affect patient satisfaction, adherence, disclosure, and outcomes [1-8]. Health communication researchers have examined various aspects of clinician-patient verbal interactions, such as the content of the clinician's speech and their voice tone [5,9] and the intent that an utterance has in communication [10]. Various nonverbal aspects have also been examined, such as facial expressions, eye contact, body posture, fluency [5,11], and the physical distance between clinicians and patients [12]. With the widespread adoption of electronic medical record systems, computers have become an integral part of clinics. As a result, the traditional 2-way doctor-patient relationship has been replaced by a triadic relationship among doctor, patient, and computer [13]. The use of electronic medical record systems during consultations has been shown to affect doctor-patient verbal [14] and nonverbal [2] communication, and consequently, doctor-patient relationships both positively and negatively [15,16]. Accordingly, the study of doctor-patient-computer interactions has become a key research area for examining doctor-patient relationships [17].

Doctor-Patient-Computer Interactions

Multiple studies, mainly in primary care settings, noted that doctor-patient communication is affected [18-25] and even shaped [26] by the use of computers during clinical encounters. The use of computers was shown to modify or amplify doctors' verbal and nonverbal behaviors [16,21,27-29] that are essential to avoid communication failures and to have effective doctor-patient communication [20]. Examples of negative verbal and nonverbal behaviors that could be amplified by the use of a computer include lack of eye contact, deficient active listening, avoidance, and interruption [30-32].

In addition to studying the effect of computer use on doctor-patient interactions, multiple studies [25,33] examined factors that affect the way these computers are used. Pearce et al [34] described the overarching styles and behaviors of doctors, patients, and computers by studying the orientation of the general practitioners' and patients' bodies as well as their conversations. Chan et al [35] found that doctors spent 50% less time using computers in examinations with psychological components than in examinations with no psychological components. Lanier et al [36] found that consultation content, physicians' gender and level of experience, and whether the consultation was new or a follow-up were modestly related to the way physicians used the computer in primary care settings.

Computational Ethnography Inside Clinics

Researchers studying doctor-patient-computer interactions need to identify which interactions are taking place during the consultations. To do so, researchers have used qualitative methods such as taking notes during live observations [31,37], conducting interviews [37,38], administering questionnaires [39], and sending unannounced standardized patients to collect information [40] and quantitative methods such as videotaping consultations and manually coding the videos [36,41,42] or setting up complex mechanisms for automatic data collection and analysis inside the clinics [43]. Methods that include direct observations are likely to generate more accurate data than clinician or patient reports; however, direct observations are costly in terms of time and human resources, may be obtrusive in a clinical environment, and may cause the participants to knowingly or unknowingly alter their behavior (because of the presence of an observer) [44]. Moreover, they present privacy and ethical concerns for patients and doctors such as concerns about data security and anonymization; changes to the research question that make it different from the one described in initial consent forms, and researchers' inability to take into account all nonpublic information or situations that will be accessed [45].

Given recent technological advancements, computational ethnography has been proposed as an alternative method for studying doctor-patient-computer interaction in depth. Computational ethnography was defined as a new family of methods for conducting human-computer interaction studies in health care settings by using "automated and less obtrusive (or unobtrusive) means for collecting in situ data reflective of real end users' actual, unaltered behaviors using a software system or a device in real-world settings [46]."

Recently, a number of tools that automate the measurement and analysis of specific behaviors in clinical settings were proposed and evaluated: Hart et al. [47] proposed and validated an automated video analysis tool to measure the synchrony and dominance in doctor-patient interactions by analyzing the cross-correlation of the kinetic energy and the frequency spectrum of their motion [47]. Gutstein et al reported developing a system that automatically learns the physician's gaze using their hand positioning [48] or body positioning and optical flow [49]. Weibel et al [43] introduced a solution that enables the capture of multimodal activity in clinical settings [43] to support computational ethnography studies in clinics. Their solution combined computer logging functionality, body motion tracking, audio detection, and eye tracking. By synchronizing data from these sensors, Weibel et al [43] were able to detect the person talking, whether the doctor is looking at the screen, the amount of gesturing, the cognitive load, information searching behavior, workflow interruptions, and the amount of computer activity; however, their solution had some limitations. First, the accuracy of the automatic classification was not reported. Second, they noted that the use of Kinect presents some limitations such as the need to set up the machine, Kinect's inability to reidentify a body once it re-enters the scene, and the occasional transfer of skeletal tracking from human to nonhuman objects. Third, to detect the person who was talking, a Dev-Audio Microcone [50] was used. This means that such a tool may not fit the needs of people looking for a cheap and portable solution with a known robustness level. In this case, recent advancements in pose and voice activity detection algorithms could address some of these limitations. For video consultations, Faucett et al [51] created ReflectLive, a tool that provides real-time feedback to clinicians about speaking contributions, interruptions, eye gaze, and face position. ReflectLive [51] uses an open-source library for audio analysis and a commercial Javascript-based computer-vision

face-tracking software for visual analysis. The real-time feedback provided by ReflectLive was evaluated in terms of its usefulness to the clinicians, but the feedback's accuracy was not reported.

Currently, there are few truly robust and unobtrusive computational ethnography tools for clinical settings, as most tools require researchers to add external artifacts into the clinical environment. Moreover, to our knowledge, none of the existing tools is freely available to the public. To enable human-computer and human-human interaction studies in clinical settings, there is a need for publicly available, robust, unobtrusive, and automated tools for detecting and classifying doctor-patient-computer interactions.

Aims

We aimed to provide a public, robust, unobtrusive, privacy-ensuring, and automated tool for detecting and classifying screen gaze and dialogue in doctor-patient-computer interactions. We chose to focus on screen gaze and dialogue due to recent advancements in machine learning that render the automatic and accurate estimation of pose and voice activity possible.

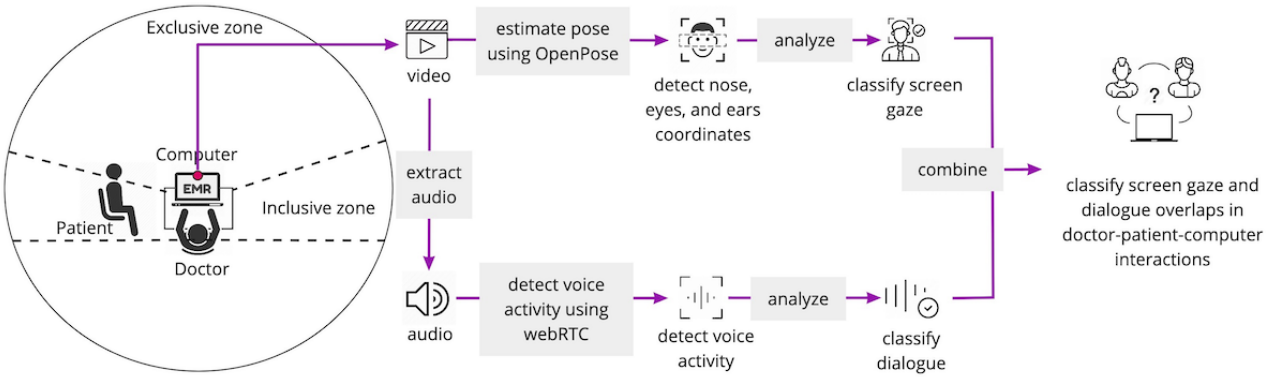
Methods

Overview

The purpose of the classifier (Figure 1) was to detect the following interactions: (1) screen gaze and dialogue: doctor gazing at the computer screen while having a conversation with the patient; (2) dialogue: doctor conversing with the patient while looking away from the computer screen, or (3) screen gaze: doctor gazing at the computer screen without conversing with the patient. Any other type of interaction in which the doctor and the patient were not having a conversation and the doctor is not gazing at their computer screen were considered out of scope.

The code of the proposed classifier is publicly available [52] and can be used by researchers, care providers, designers, medical educators, and others who are interested in exploring and answering questions related to screen gaze and dialogue combinations in doctor-patient-computer interactions.

Figure 1. Overview of the classification process. EMR: electronic medical record.



Screen Gaze Classifier

The purpose of the screen gaze classifier was to detect when the doctor's gaze was aimed at the computer screen. The input of the classifier was the video captured by the doctor's computer camera and the output was a binary classification: *no screen gaze* or *screen gaze*.

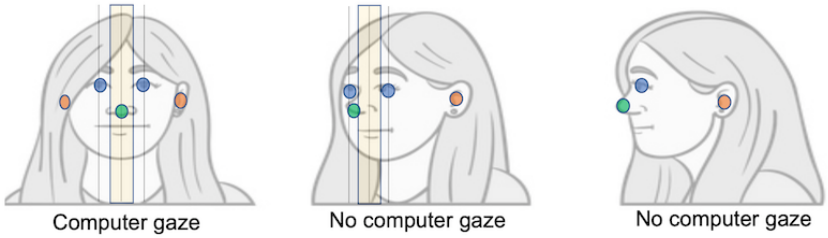
We used the pose estimation library OpenPose [53] as a tool to detect the coordinates of key points of the doctor's face. OpenPose is an open-source library that allows real-time multiperson key point detection for body, face, hands, and feet. We extracted the coordinates of the doctor's eyes (x_{LeftEye}, y_{LeftEye}), (x_{RightEye}, y_{RightEye}), ears (x_{LeftEar}, y_{LeftEar}), (x_{RightEar}, y_{RightEar}), and nose (x_{Nose}, y_{Nose}), and using the coordinates, we assumed that the doctor's gaze was targeting the computer screen if (1) the location of both the doctor's ears could be estimated, and (2) the doctor's nose was centered between the eyes. For the second condition, we allowed a tolerance equal to half the distance between the 2 eyes. We assessed these criteria (Figure 2) using the following equations for each frame in the video:

$$(x_{\text{LeftEar}}, y_{\text{LeftEar}}) \neq null\ AND\ (x_{\text{RightEar}}, y_{\text{RightEar}}) \neq null$$

$$x_{\text{LeftEye}} + \frac{x_{\text{RightEye}} - x_{\text{LeftEye}}}{4} < x_{\text{Nose}} < x_{\text{RightEye}} - \frac{x_{\text{RightEye}} - x_{\text{LeftEye}}}{4}$$

Then we assigned to each 0.5-second interval of video the most frequent classification of its corresponding frames. This results in a binary classification (no screen gaze, screen gaze) for each 0.5 seconds of video.

Figure 2. Classifying the doctor's computer screen gaze using face key point estimation.



Dialogue Classifier

The purpose of the dialogue classifier was to detect when the doctor and patient were engaging in conversation. The input of the classifier was the audio captured by the doctor's computer's microphone, and the output was a binary classification of the doctor-patient conversation: *no dialogue* or *dialogue*.

We used a library based on the webRTC voice activity detection engine (an open source project maintained by the Google WebRTC team [54]). The voice activity detection library allows the detection of voice activity in an audio file by processing audio segments and estimating the probability in each segment.

We set the length of each audio segment to 5 milliseconds, which we found to offer the best results through trial and error. We set the voice activity detection to its highest aggressiveness mode in order to increase the probability of filtering out nonspeech. We assigned to each 0.5-second interval of audio the most frequent classification of its corresponding segments. This results in a binary classification (no dialogue, dialogue) for each 0.5 seconds of audio.

Classifier of Screen Gaze and Dialogue Combinations

By combining the results of the screen gaze classifier and the dialogue classifier described above, we classify doctor-patient-computer interactions into 4 different classes (Table 1). The Screen Gaze and Dialogue (SG+D) class defines interactions wherein the doctor is gazing at the computer screen while conversing with the patient. The Dialogue (D) class defines interactions wherein the doctor is looking away from the computer screen and conversing with the patient, and the Screen Gaze (SG) class defines interactions where the doctor is gazing at the computer screen and not conversing with the patient. An interaction wherein the doctor is neither looking at the computer screen nor conversing with the patient is classified as Other. For each 0.5 seconds of video, the interactions classifier assigns 1 of the 4 classes.

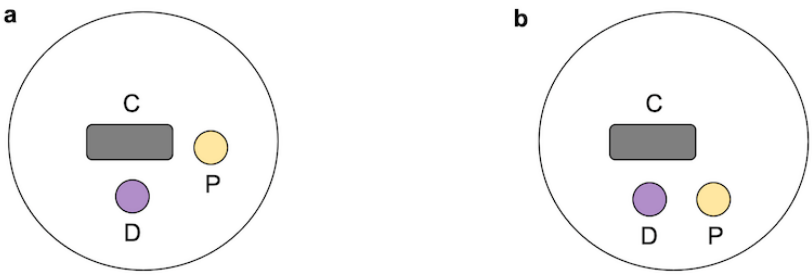
Table 1. Four classes of doctor-patient-computer interactions.

Components		Class	
Screen gaze	Dialogue	Doctor-Patient-Computer interaction	Label
Screen gaze	Dialogue	Screen Gaze + Dialogue	SG+D
No screen gaze	Dialogue	Dialogue	D
Screen gaze	No dialogue	Screen Gaze	SG
No screen gaze	No dialogue	Other	Other

Evaluation of the Classifier

We considered 2 clinical layouts in our evaluation: a semi-inclusive layout, where the patient is seated next to the computer desk, and a fully inclusive layout, where the patient is seated next to the doctor and facing the computer desk (Figure 3). The data that we used to evaluate our classifier consisted of 10 videos provided by 5 physicians. Each physician provided 2 videos—1 video simulating a consultation in a fully inclusive layout and 1 video simulating a consultation in a semi-inclusive layout. Each video was approximately 3 minutes long.

Figure 3. (a) Semi-inclusive and (b) fully inclusive layouts were considered in the evaluation. C: computer; D: doctor; P: patient.



For ground truth data, each video was initially annotated by a human coder. The coder assigned 1 of the 4 interaction classes to each 0.5 seconds of video. The coder reviewed the videos several times and refined the initial annotations until they were satisfied.

To evaluate the classifier, we compared the classifier's performance to that of a different human coder. This second coder was allowed to go through the video only once. This was to simulate a real-world scenario of video coding assigned to an external coder. The performance of the classifier and that of the second human coder were assessed in relation to the ground truth data (generated by the first coder).

The overall performance reflects the performance over the 10 videos including 5 videos in a semi-inclusive layout and 5 videos in a fully inclusive layout. The performances were assessed using an overall accuracy measure in addition to measures of precision, recall, F1 scores for each class. For each class, the support number (ie, the number of its occurrences in the ground truth data set) is reported. Weighted scores of precision, recall, and F1 scores, where the weight of a class is proportional to its support, were also measured. Difference in performance between the classifier and the human coder were assessed using 2-tailed independent *t* tests with *P* values <.05 considered statistically significant. We first report the overall performance, which reflects the performance over the 10 videos. Then, we separately report the performances over the 5 videos in a semi-inclusive layout and the 5 videos in a fully inclusive layout.

Results

Overall Performance

Table 2 shows the overall performances of the classifier and the human coder. The classifier showed a slightly lower overall accuracy than the coder (classifier: 0.83; human coder: 0.85); however, there was no significant difference between the accuracy of the classifier and that of the human coder (t_{18}=0.6, *P*=.55).

The F1 scores of both the classifier and the coder were better when classifying SG+D (classifier: 0.81; human coder: 0.81) and D (classifier: 0.89; human coder: 0.90) than that when classifying SG (classifier: 0.63; human coder: 0.55) and Other (classifier: 0.35; human coder: 0.36) interactions. Since the D class and the SG+D class were the most frequent interactions (D: 2415/3921, 62%; SG+D: 1189/3921, 30%), the overall accuracies mainly reflect performances for these 2 classes.

Confusion matrices for overall performance (Figure 4) show that the classifier and the coder had similar patterns. Both mistook SG+D for D and vice versa, SG for SG+D, and Other for D interactions. The main difference between the classifier and the coder is that the classifier tended to mistake D for Other interactions, whereas the coder tended not to.

Table 2. Overall performance of the classifier.

Classes	Classifier				Human coder				Support, n
	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	
All	—[a]	—	—	0.83	—	—	—	0.85	—
SG+D[b]	0.79	0.82	0.81	—	0.78	0.84	0.81	—	1189
D[c]	0.92	0.86	0.89	—	0.91	0.90	0.90	—	2425
SG[d]	0.64	0.63	0.63	—	0.71	0.45	0.55	—	228
Other	0.24	0.67	0.35	—	0.35	0.38	0.36	—	79
Weighted score	0.85	0.83	0.84	—	0.85	0.85	0.84	—	—

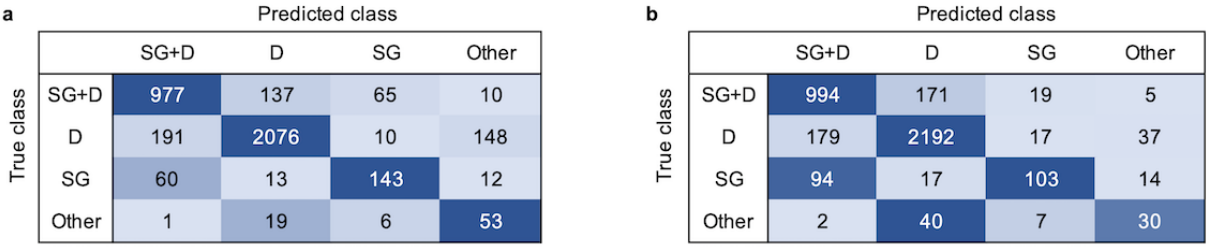
[a]Not calculated or not applicable.

[b]SG+D: Screen Gaze and Dialogue.

[c]D: Dialogue.

[d]SG: Screen Gaze.

Figure 4. (a) Classifier and (b) human coder confusion matrices for overall performance. D: Dialogue; SG: Screen Gaze; SG+D: Screen Gaze + Dialogue.



Performance in a Semi-inclusive Layout

Table 3 shows the performances of the classifier and human coder for a semi-inclusive layout. The classifier had a slightly lower accuracy than the coder (classifier: 0.80; human coder: 0.83); however, there was no significant difference between the accuracy of the classifier and that of the human coder in the semi-inclusive layout (t_8=1.04, P=0.32).

Both the classifier and the coder performed well when classifying D (classifier: F1 score 0.86; human coder: F1 score 0.88) and SG+D (classifier: 0.79; human coder: 0.82) interactions. The classifier had a slightly better F1 score than the coder when detecting SG (classifier: 0.47; human coder: 0.45), but both the classifier and the coder had low F1 scores when classifying Other interactions (classifier: 0.24; human coder: 0.21). The D and the SG+D classes had the most support (D: 1157/1958, 59%; SG+D: 702/1958, 36%) thus the overall accuracies mainly reflect performances for these 2 classes.

Confusion matrices of the classifier and the human coder (Figure 5) for a semi-inclusive layout show somewhat similar patterns for the classifier and the coder. Both mostly mistook D for SG+D and vice versa. The classifier tended to mainly mistake SG for SG+D, whereas the coder mistook SG for Other interactions as well. Finally, the coder exceedingly mistook Other for D.

Table 3. Performance of the classifier in a semi-inclusive layout.

Classes	Classifier				Human coder				Support, n
	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	
All	—[a]	—	—	0.80	—	—	—	0.83	—
SG+D[b]	0.76	0.83	0.79	—	0.81	0.82	0.82	—	702
D[c]	0.92	0.80	0.86	—	0.88	0.88	0.88	—	1157
SG[d]	0.40	0.57	0.47	—	0.61	0.36	0.45	—	69
Other	0.16	0.50	0.24	—	0.17	0.30	0.21	—	30
Weighted score	0.83	0.80	0.81	—	0.84	0.83	0.83	—	—

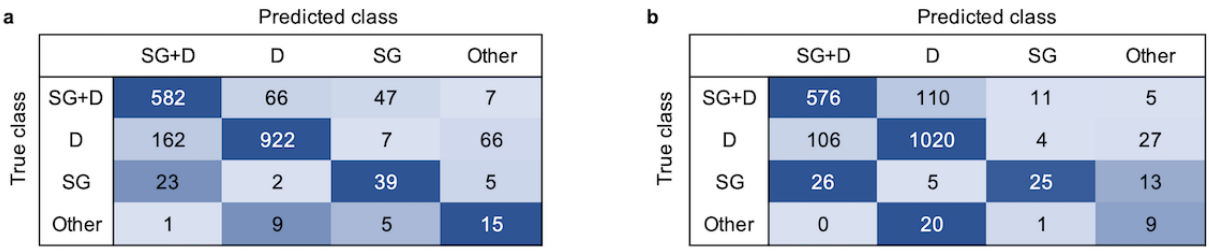
[a]Not calculated or not applicable.

[b]SG+D: Screen Gaze and Dialogue.

[c]D: Dialogue.

[d]SG: Screen Gaze.

Figure 5. (a) Classifier and (b) human coder confusion matrices for semi-inclusive layout. D: Dialogue; SG: Screen Gaze; SG+D: Screen Gaze + Dialogue.



Performance in a Fully Inclusive Layout

Table 4 shows the performances of the classifier and the human coder in a fully inclusive layout. The classifier and the coder showed similar accuracy (both equal to 0.86), and there was no significant difference between the accuracy of the classifier and that of the human coder for a fully inclusive layout (t_8=0.43, P=0.67).

The classifier and the coder had good F1 scores when classifying D (classifier: 0.92; human coder: 0.93) and SG+D (classifier: 0.83; human coder: 0.80). The classifier performed better than the coder for the SG class (classifier: 0.73; human coder: 0.59), but worse for Other interactions (classifier: 0.42; human coder: 0.52). The D and SG+D classes had the most support(D: 1268/1963, 65%; SG+D: 487/1963, 25%), thus the overall accuracy mainly reflects the performance for these 2 classes.

Confusion matrices of the classifier and the human coder for a fully inclusive layout (Figure 6) show similar patterns for the classifier and the coder. Both mistook SG+D for D, SG for SG+D, and Other interactions for D; however, the classifier tended to mostly mistake D for Other interactions, whereas the coder mostly mistook D for SG+D.

Table 4. Performance of the classifier in a fully inclusive layout.

Classes	Classifier				Human coder				Support
	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	
All	—[a]	—	—	0.86	—	—	—	0.86	—
SG+D[b]	0.86	0.81	0.83	—	0.75	0.86	0.80	—	487
D[c]	0.93	0.91	0.92	—	0.93	0.92	0.93	—	1268
SG[d]	0.83	0.65	0.73	—	0.74	0.49	0.59	—	159
Other	0.29	0.78	0.42	—	0.66	0.43	0.52	—	49
Weighted score	0.88	0.86	0.87	—	0.86	0.86	0.86	—	—

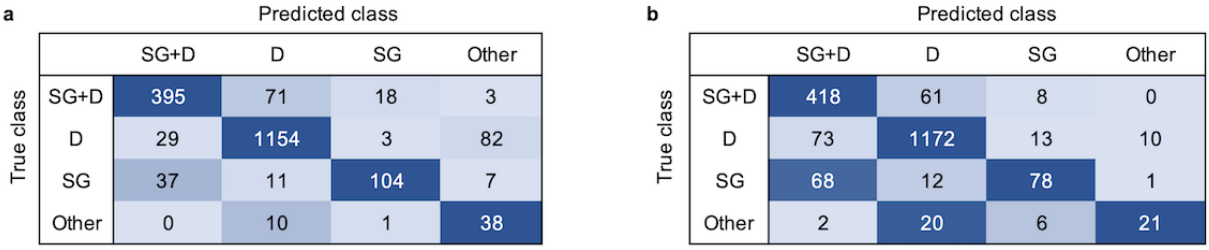
[a]Not calculated or not applicable.

[b]SG+D: Screen Gaze and Dialogue.

[c]D: Dialogue.

[d]SG: Screen Gaze.

Figure 6. (a) Classifier and (b) human coder confusion matrices for fully inclusive layout. D: Dialogue; SG: Screen Gaze; SG+D: Screen Gaze + Dialogue.



Transitions Between Doctor-Patient-Computer Interactions

We found that many errors in the classifications (for both the human coder and our classifier) were due to slight time inconsistencies during transitions. To confirm this hypothesis, we conducted an analysis of the transitions between the interactions. Table 5 reports the frequency of each transition in the ground truth data and shows that most transitions happen from D to SG+D and vice versa. We define a transition timing error as a temporal shift of 0.5 to 1 seconds between the ground truth and the classification. We only included transitions that were preceded and followed by a continuous type of interaction for at least 1.5 seconds. Table 6 reports the absolute and relative number of errors that can be attributed to early or late coding of transitions. We think these errors can be overlooked for any practical purpose.

Table 5. Transitions in ground truth data.

Transition	Semi-inclusive, n	Fully inclusive, n
From SG+D[a]		
...to D[b]	103	65
...to SG[c]	10	16
...to Other	1	0
From D		
...to SG+D	104	67
...to SG	7	15
...to Other	10	16
From SG		
...to SG+D	8	13
...to D	10	17
...to Other	0	2
From Other		
...to SG+D	1	1
...to D	10	14
...to SG	2	4

[a]SG+D: Screen Gaze and Dialogue.

[b]D: Dialogue.

[c]SG: Screen Gaze.

Table 6. Transition-related errors.

Layout	Classifier		Human coder	
	Total errors, n	Transition errors, n (%)	Total errors, n	Transition errors, n (%)
Semi-inclusive	400	65 (16.2)	329	73 (22.2)
Fully inclusive	272	45 (16.5)	274	61 (22.3)

Discussion

Principal Results

We developed an unobtrusive, inexpensive, and automatic classifier of screen gaze and dialogue combinations in doctor-patient-computer interactions. The classifier was evaluated in 2 clinical layouts, semi-inclusive and fully inclusive, and had a performance similar to that of a human coder with an overall accuracy of 0.83. The proposed classifier is unobtrusive since it does not require additional setup in the clinic and only requires that doctors record video using their computer's internal microphone and camera. The proposed classifier is an inexpensive solution since it is built using open-source tools and takes advantage of the internal camera and microphone built into most available computing devices. Finally, the video can be locally processed, thus reducing the risks of handling private and sensitive data off the clinic's premises, and ensuring that no collateral data are collected and used for purposes other than those initially consented to by the participants.

Both the classifier and the coder had better accuracies in a fully inclusive layout (both equal to 0.86) than in the semi-inclusive layout (classifier: 0.80; human coder: 0.83). The difference in performance can be attributed to the different postures that the doctor maintains when interacting in the 2 clinic layouts. In the fully inclusive layout, the doctor has to rotate their head a full 90 degrees away from the screen in order to gaze at the patient, whereas in semi-inclusive scenarios, the head rotation angle is smaller; therefore, it is easier to make distinctions between the interactions in a fully inclusive layout.

Both the classifier and the coder confused SG+D interactions and D interactions. Some instances occurred when classifying near transitions between interactions. Indeed, our analysis showed that 16.4% of the classifier's errors (110/672) and 22.2% of the human coder's errors (134/603) were early or delayed markings of transitions, which can be overlooked in practical use-cases. Unlike the human coder, the classifier tended to mistake D for Other interactions (ie, the doctor is neither looking at the screen nor conversing with the patient). This may be attributed to the fact that human coders overlook small moments of silence and regard them as response offsets that are due to

turn-taking in the conversation [55] or lapses that are expected in multiactivity settings [56,57], whereas our classifier classifies them as an absence of dialogue. Here, the classifier presents an advantage since it easily detects moments of silence that are usually overlooked by human coders. This kind of information may be useful to detect conversational dimensions such as hesitant speech [42]. However, for our purposes, this leads the classifier to overestimate the lack of verbal interaction. To counteract this, further rules are needed to identify which moments of silence are part of a conversation and which are not. These rules need to take into consideration the language and the content of the conversation [55,58], other activities that individuals are engaging in while conversing [56,57], and accompanying nonverbal behavior such as nodding and gaze [55].

As a collateral result, our experiment also confirmed the findings of previous work that highlighted the effect of the clinic layout on doctor-patient-computer interactions [59,60]. Fully inclusive videos contained more D (64% versus 59%) and SG interactions (8% versus 3.5%) and fewer SG+D interactions (25% versus 36%) than those in semi-inclusive videos. In a fully inclusive layout, the doctor has to choose whether to face the computer or their patient, whereas a semi-inclusive layout allows the doctors to maintain a conversation with their patients while looking at the computer screen.

Use Cases

Functionality

Because health communication researchers study various complex behaviors, the coding process is far more complex than a binary classification of screen gaze and dialogue over time. However, the proposed classifier could contribute to reducing the cost of future studies since gaze and dialogue are part of the behaviors that health communication researchers and medical informaticians are often interested in quantifying to examine the relationship between computer use, clinic and tool design, and physician-patient interactions [3,5,19,28,61-65]. Moreover, by detecting these behaviors over short intervals of time, more complex behaviors and patterns could be inferred as shown in our results on transitions between interactions. Therefore, the classifier can be directly used to monitor screen gaze and dialogue or extended and combined with other tools or processes to examine complex interactions in clinical settings.

For Researchers

The proposed tool can be used to study the effects of screen gaze and dialogue combinations in doctor-patient-computer interactions on quality of care and health outcomes. Currently, conducting this type of study would require (1) recording a video of the consultation, (2) transferring the video outside of the clinic, (3) manual coding of screen gaze and presence of dialogue in the video, and (4) assessing the specific outcome. This process can be facilitated by our classifier, which eliminates the need for the second and third steps.

The classifier can also be used in clinics to examine the effect of external factors on screen gaze and dialogue combinations in patient-doctor-computer interactions such as sociodemographic characteristics, clinic layout, and modifications of electronic medical record system's design. Non–self-reported large-scale studies of this kind would be nearly impossible to conduct using current methods and tools.

In addition, the gaze classifier can be used in conjunction with commonly used interaction analysis coding systems, such as the Roter interaction analysis system [42], that do not systematically account for nonverbal behaviors [66]. This would provide useful data for studies examining provider-patient communication in the presence of a computer.

For Practicing Physicians

Since the processing of the video can happen in real time, tools that allow physicians to *reflect in action* [67] can be created. The classifier can be used to create tools that allow physicians to conduct autoethnographies and reflect on their interactions with the patient, or on the role of technology in their care practice [68]. This would allow them to adapt their behavior and level of attention based on feedback. Similar concepts were proposed by Liu et al [69] and Faucett et al [51], who described tools that provide feedback to clinicians about their verbal and nonverbal communication behaviors during online consultations. Their studies highlighted the usefulness of summative [70] and real-time self-reflection tools and the need to design real-time feedback in a way that minimizes intrusiveness and ensure that it does not create extra distractions [51].

For Medical Educators and Students

Medical educators and students can use the classifier to teach and learn the best practices of doctor-patient interactions. The tool could be expanded to detect different interaction categories (eg, listening/ignoring; confronting/avoiding [71,72]) and used during practical learning sessions to provide students with formative feedback [69].

Limitations and Future Work

The first limitation of the classifier is its applicability to certain clinic layouts. Our evaluation explored 2 clinical scenarios: semi-inclusive and fully inclusive. We did not explore exclusive scenarios, even though these scenarios might be encountered in real-world clinical settings. With the proposed classifier, detecting the doctor's computer gaze would not be possible in a fully exclusive scenario since the classifier considers the doctor's head turn to estimate her gaze. We consider this limitation acceptable since inclusive scenarios are already commonplace [25], and we expect an increase in their prevalence to support technology-mediated information-sharing between clinicians and patients. Indeed, previous studies [31,73,74] have shown that clinicians use their computer screens as tools to share information with their patients; therefore, the doctor may turn the screen, along with the camera, toward the patient. If this interaction happens during an ongoing conversation, our tool may classify it as a dialogue between doctor and patient, but the fact that this doctor-patient interaction is mediated by the computer would not be highlighted. Further improvements are needed to detect scenarios where both clinician and patient are interacting with the computer at the same time.

Our work also assumes that clinicians use their computers during consultations, which is not always the case, especially in secondary care settings. Moreover, the classifier is built and evaluated around the premise that the only people in the clinic are the doctor and the patient and that the only screen is the doctor's computer screen. However, it is possible that patients are accompanied by their family members, friends, or partners [75] and that multiple health care staff are involved in the care of 1 patient and present during the consultation. In addition, extra screens might be installed inside clinics to engage the patient in their care and offer them an easy and clear view of their data. These screens may affect the doctor's behavior in various ways; for example, they might use this screen as an explanation support tool while they converse with the patient or even as their main computer screen. Therefore, another limitation of this work is its nonapplicability in scenarios that include patient screens and stakeholders other than the doctor and patient.

Furthermore, the classifier does not allow us to identify the speaker's identity or the content of the doctor-patient dialogue. Therefore, the classifier does not currently support conversation analysis. To identify the speaker's identity, we would have to perform accurate speaker diarization, a hard goal to achieve especially using a single channel for audio recording and without prior training. Advancements in speaker diarization techniques may render this task feasible in the near future [76]. To identify the content of the dialogue, automatic speech recognition solutions can be used. Though automatic speech recognition solutions have become more robust in the last decade, the performance of automatic speech recognition engines remains limited when applied to conversational clinical speech [77]. Future work could explore the feasibility of automatic conversation analysis in doctor-patient-computer interactions through the application of novel speaker diarization and automatic speech recognition tools.

In other respects, although the direction of the head could be considered a proxy for the direction of attention, the head only communicates short-term attention [78,79]. Pearce et al classified physicians as unipolar, those who maintain the lower pole of their body facing the computer, or bipolar, those who repeatedly alternate the orientation of their lower pole between the computer and the patient [34]. Unipolar physicians experience situations where their body segments are not aligned, also referred to as body torque. Body torque communicates an instability of attention where the most strongly projected resolution involves the upper body getting realigned with the lower body. This means that the orientation of the torso communicates longer-term attention than head orientation, and the orientation of the legs communicates longer-term attention than torso orientation and head orientation [79]. Therefore, to examine the attention of a physician in a clinical scenario, we also need to examine the orientation of their torso and lower body. Future work can use the same pose estimation approach to monitor the direction of the clinician’s torso. This is possible because the shoulders and the torso of the clinician are usually visible to their computer’s camera; however, monitoring the lower part of the body would require setting up extra cameras in the clinic.

Finally, our classifier is model-driven as it derives its decisions from the explicit rules that we set. To be able to classify interactions that do not fit strictly into our specified rules, the classifier has to be driven by data or perhaps be a system that combines model-driven and data-driven logic. Our future work will include collecting and annotating more videos of consultations in order to create a data-driven classifier.

Conclusions

To facilitate human-computer and human-human interaction studies in clinical settings, we presented a computational ethnography tool—an automatic unobtrusive classifier of gaze and dialogue combinations in doctor-patient-computer interactions. The classifier only requires that the doctor record video using their computer’s internal camera and microphone. Our evaluation showed that the classifier's performance was similar to that of a human coder when classifying 3 combinations of screen gaze and dialogue in doctor-patient-computer interactions.

Acknowledgments

This research was supported by Japan Society for the Promotion of Science Kakenhi (grant number JP20K20244).We thank the 5 physicians who provided us with video data. We also thank our reviewers for their valuable comments and suggestions, which helped us improve this work.

Conflicts of Interest

None declared.

References

1. Hall JA. Affective and nonverbal aspects of the medical visit. In: The Medical Interview, Frontiers of Primary Care. New York, NY: Springer; 1995:495-503.
2. Robinson JD. Nonverbal communication and physician-patient interaction: review and new directions. In: Manusov V, Patterson ML, editors. The SAGE Handbook Of Nonverbal Communication. Thousand Oaks, CA: SAGE Publications, Inc; 2006:437-460.
3. Roter D, Frankel R, Hall J, Sluyter D. The expression of emotion through nonverbal behavior in medical visits. mechanisms and outcomes. J Gen Intern Med 2006 Jan;21 Suppl 1(1):S28-S34 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00306.x] [Medline: 16405706]

4. Beck RS, Daughtridge R, Sloane PD. Physician-patient communication in the primary care office: a systematic review. J Am Board Fam Pract 2002;15(1):25-38 [FREE Full text] [Medline: 11841136]
5. Griffith CH, Wilson JF, Langer S, Haist SA. House staff nonverbal communication skills and standardized patient satisfaction. J Gen Intern Med 2003 Mar;18(3):170-174 [FREE Full text] [doi: 10.1046/j.1525-1497.2003.10506.x] [Medline: 12648247]
6. Schoenthaler A, Chaplin WF, Allegrante JP, Fernandez S, Diaz-Gloster M, Tobin JN, et al. Provider communication effects medication adherence in hypertensive African Americans. Patient Educ Couns 2009 May;75(2):185-191 [FREE Full text] [doi: 10.1016/j.pec.2008.09.018] [Medline: 19013740]
7. Matusitz J, Spear J. Effective doctor-patient communication: an updated examination. Soc Work Public Health 2014;29(3):252-266. [doi: 10.1080/19371918.2013.776416] [Medline: 24802220]
8. Kelley J, Kraft-Todd G, Schapira L, Kossowsky J, Riess H. The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2014;9(4):e94207 [FREE Full text] [doi: 10.1371/journal.pone.0094207] [Medline: 24718585]
9. Hall JA, Roter DL, Rand CS. Communication of affect between patient and physician. J Health Soc Behav 1981 Mar;22(1):18-30. [Medline: 7240703]
10. Meeuwesen L, Schaap C, van der Staak C. Verbal analysis of doctor-patient communication. Soc Sci Med 1991;32(10):1143-1150. [doi: 10.1016/0277-9536(91)90091-p] [Medline: 2068597]
11. Duggan P, Parrott L. Physicians' nonverbal rapport building and patients' talk about the subjective component of illness. Human Comm Res 2001 Apr;27(2):299-311. [doi: 10.1111/j.1468-2958.2001.tb00783.x]
12. Aruguete M, Roberts C. Participants' ratings of male physicians who vary in race and communication style. In: Psychological Reports. Los Angeles, CA: SAGE Publications; Dec 2002:793-806.
13. Scott D, Purves IN. Triadic relationship between doctor, computer and patient. Interact Comput 1996 Dec;8(4):347-363. [doi: 10.1016/s0953-5438(97)83778-2]
14. Vogel D, Meyer M, Harendza S. Verbal and non-verbal communication skills including empathy during history taking of undergraduate medical students. BMC Med Educ 2018 Jul 03;18(1):157 [FREE Full text] [doi: 10.1186/s12909-018-1260-9] [Medline: 29970069]
15. Alkureishi M, Lee W, Lyons M, Press V, Imam S, Nkansah-Amankra A, et al. Impact of electronic medical record use on the patient-doctor relationship and communication: a systematic review. J Gen Intern Med 2016 May;31(5):548-560 [FREE Full text] [doi: 10.1007/s11606-015-3582-1] [Medline: 26786877]
16. Frankel R, Altschuler A, George S, Kinsman J, Jimison H, Robertson NR, et al. Effects of exam-room computing on clinician-patient communication: a longitudinal qualitative study. J Gen Intern Med 2005 Aug;20(8):677-682 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0163.x] [Medline: 16050873]
17. Chen Y, Cheng K, Tang C, Siek KA, Bardram JE. Is my doctor listening to me? impact of health IT systems on patient-provider interaction. 2013 Presented at: CHI'13 Extended Abstracts on Human Factors in Computing Systems; April 27-May 2; Paris, France p. 2419-2426. [doi: 10.1145/2468356.2468791]
18. Als AB. The desk-top computer as a magic box: patterns of behaviour connected with the desk-top computer; GPs' and patients' perceptions. Fam Pract 1997 Feb 01;14(1):17-23. [doi: 10.1093/fampra/14.1.17] [Medline: 9061339]
19. McGrath J, Arar N, Pugh J. The influence of electronic medical record usage on nonverbal communication in the medical interview. Health Informatics J 2007 Jun;13(2):105-118 [FREE Full text] [doi: 10.1177/1460458207076466] [Medline: 17510223]
20. Hashim MJ. Patient-centered communication: basic skills. Am Fam Physician 2017 Jan 01;95(1):29-34 [FREE Full text] [Medline: 28075109]
21. Silverman J, Kinnersley P. Doctors' non-verbal behaviour in consultations: look at the patient before you look at the computer. Br J Gen Pract 2010 Feb 01;60(571):76-78. [doi: 10.3399/bjgp10x482293]
22. Pearce C, Arnold M, Phillips C, Trumble S, Dwan K. The patient and the computer in the primary care consultation. J Am Med Inform Assoc 2011 Mar 01;18(2):138-142 [FREE Full text] [doi: 10.1136/jamia.2010.006486] [Medline: 21262923]
23. Pearce C, Trumble S. Computers can't listen--algorithmic logic meets patient centredness. Aust Fam Physician 2006 Jun;35(6):439-442 [FREE Full text] [Medline: 16751862]
24. Noordman J, Verhaak P, van Beljouw I, van Dulmen S. Consulting room computers and their effect on general practitioner-patient communication. Fam Pract 2010 Dec 26;27(6):644-651. [doi: 10.1093/fampra/cmq058] [Medline: 20660530]
25. Crampton N, Reis S, Shachak A. Computers in the clinical encounter: a scoping review and thematic analysis. J Am Med Inform Assoc 2016 May;23(3):654-665 [FREE Full text] [doi: 10.1093/jamia/ocv178] [Medline: 26769911]
26. Pearce C, Trumble S, Arnold M, Dwan K, Phillips C. Computers in the new consultation: within the first minute. Fam Pract 2008 Jun;25(3):202-208. [doi: 10.1093/fampra/cmn018] [Medline: 18504254]
27. Booth N, Robinson P, Kohannejad J. Identification of high-quality consultation practice in primary care: the effects of computer use on doctor-patient rapport. Inform Prim Care 2004 May 01;12(2):75-83 [FREE Full text] [doi: 10.14236/jhi.v12i2.111] [Medline: 15319059]
28. Asan O, D Smith P, Montague E. More screen time, less face time - implications for EHR design. J Eval Clin Pract 2014 Dec 19;20(6):896-901 [FREE Full text] [doi: 10.1111/jep.12182] [Medline: 24835678]

29. Asan O, Young HN, Chewning B, Montague E. How physician electronic health record screen sharing affects patient and doctor non-verbal communication in primary care. Patient Educ Couns 2015 Mar;98(3):310-316 [FREE Full text] [doi: 10.1016/j.pec.2014.11.024] [Medline: 25534022]
30. Kee J, Khoo H, Lim I, Koh M. Communication skills in patient-doctor interactions: learning from patient complaints. Health Prof Educ 2018 Jun;4(2):97-106. [doi: 10.1016/j.hpe.2017.03.006]
31. Helou S, Abou-Khalil V, Yamamoto G, Kondoh E, Tamura H, Hiragi S, et al. Understanding the situated roles of electronic medical record systems to enable redesign: mixed methods study. JMIR Hum Factors 2019 Jul 09;6(3):e13812 [FREE Full text] [doi: 10.2196/13812] [Medline: 31290398]
32. Ha J, Longnecker N. Doctor-patient communication: a review. Ochsner J 2010;10(1):38-43 [FREE Full text] [Medline: 21603354]
33. Ventres W, Kooienga S, Vuckovic N, Marlin R, Nygren P, Stewart V. Physicians, patients, and the electronic health record: an ethnographic analysis. Ann Fam Med 2006 Mar 01;4(2):124-131 [FREE Full text] [doi: 10.1370/afm.425] [Medline: 16569715]
34. Pearce C, Dwan K, Arnold M, Phillips C, Trumble S. Doctor, patient and computer--a framework for the new consultation. Int J Med Inform 2009 Jan;78(1):32-38. [doi: 10.1016/j.ijmedinf.2008.07.002] [Medline: 18752989]
35. Chan W, Stevenson M, McGlade K. Do general practitioners change how they use the computer during consultations with a significant psychological component? Int J Med Inform 2008 Aug;77(8):534-538. [doi: 10.1016/j.ijmedinf.2007.10.005] [Medline: 18036885]
36. Lanier C, Cerutti B, Dominicé Dao M, Hudelson P, Junod Perron N. What factors influence the use of electronic health records during the first 10 minutes of the clinical encounter? Int J Gen Med 2018 Oct;11:393-398. [doi: 10.2147/ijgm.s178672]
37. Park S, Chen Y, Raj S. Beyond health literacy: supporting patient-provider communication during an emergency visit. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: ACM Conference on Computer Supported Cooperative Work and Social Computing; February 25- March 1; Portland, Oregon p. 2179-2192. [doi: 10.1145/2998181.2998357]
38. Antoun J, Hamadeh G, Romani M. Effect of computer use on physician-patient communication using interviews: a patient perspective. Int J Med Inform 2019 May;125:91-95. [doi: 10.1016/j.ijmedinf.2019.03.005] [Medline: 30914186]
39. Sustersic M, Gauchet A, Kernou A, Gibert C, Foote A, Vermorel C, et al. A scale assessing doctor-patient communication in a context of acute conditions based on a systematic review. PLoS One 2018 Feb 21;13(2):e0192306 [FREE Full text] [doi: 10.1371/journal.pone.0192306] [Medline: 29466407]
40. Zabar S, Ark T, Gillespie C, Hsieh A, Kalet A, Kachur E, et al. Can unannounced standardized patients assess professionalism and communication skills in the emergency department? Acad Emerg Med 2009 Sep;16(9):915-918 [FREE Full text] [doi: 10.1111/j.1553-2712.2009.00510.x] [Medline: 19673703]
41. Street R, Liu L, Farber N, Chen Y, Calvitti A, Weibel N, et al. Keystrokes, mouse clicks, and gazing at the computer: how physician interaction with the EHR affects patient participation. J Gen Intern Med 2018 Apr;33(4):423-428 [FREE Full text] [doi: 10.1007/s11606-017-4228-2] [Medline: 29188544]
42. Roter D, Larson S. The Roter interaction analysis system (RIAS): utility and flexibility for analysis of medical interactions. Patient Educ Couns 2002 Apr;46(4):243-251. [doi: 10.1016/s0738-3991(02)00012-5] [Medline: 11932123]
43. Weibel N, Rick S, Emmenegger C, Ashfaq S, Calvitti A, Agha Z. LAB-IN-A-BOX: semi-automatic tracking of activity in the medical office. Pers Ubiquit Comput 2014 Sep 28;19(2):317-334. [doi: 10.1007/s00779-014-0821-0]
44. Goodwin MA, Stange KC, Zyzanski SJ, Crabtree BF, Borawski EA, Flocke SA. The Hawthorne effect in direct observation research with physicians and patients. J Eval Clin Pract 2017 Dec 28;23(6):1322-1328 [FREE Full text] [doi: 10.1111/jep.12781] [Medline: 28752911]
45. Zahle J. Privacy, informed consent, and participant observation. Perspect Sci (Neth) 2017 Aug;25(4):465-487. [doi: 10.1162/posc_a_00250]
46. Zheng K, Hanauer D, Weibel N, Agha Z. Computational ethnography: automated and unobtrusive means for collecting data in situ for human-computer interaction evaluation studies. In: Cognitive Informatics for Biomedicine. Cham, Switzerland: Springer; 2015:111-140.
47. Hart Y, Czerniak E, Karnieli-Miller O, Mayo AE, Ziv A, Biegon A, et al. Automated video analysis of nonverbal communication in a medical setting. Front Psychol 2016 Aug 23;7:1130 [FREE Full text] [doi: 10.3389/fpsyg.2016.01130] [Medline: 27602002]
48. Gutstein D, Montague E, Furst J, Raicu D. Hand-eye coordination: automating the annotation of physician-patient interactions. 2019 Presented at: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE); October 28-30; Athens, Greece p. 657-662. [doi: 10.1109/BIBE.2019.00123]
49. Gutstein D, Montague E, Furst J, Raicu D. Optical flow, positioning, and eye coordination: automating the annotation of physician-patient interactions. 2019 Presented at: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE); October 28-30; Athens, Greece p. 943-947. [doi: 10.1109/bibm47256.2019.8983239]
50. Dev-Audio: innovative microphones for groups. Biamp Systems. URL: http://www.dev-audio.com/ [accessed 2021-04-28]
51. Faucett HA, Lee ML, Carter S. I should listen more: real-time sensing and feedback of non-verbal communication in video telehealth. Proc ACM Hum Comput Interact 2017 Dec 06;1(CSCW):1-19. [doi: 10.1145/3134679]

52. Comp-Ethno/gaze-dialogue-doc-pat-comp. GitHub. URL: https://git.io/JL8KJ [accessed 2021-04-28]
53. Cao Z, Hidalgo G, Simon T, Wei S, Sheikh Y. OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans Pattern Anal Mach Intell 2021 Jan 1;43(1):172-186. [doi: 10.1109/tpami.2019.2929257]
54. WebRTC real-time communication for the web. Google Developers. URL: http://www.webrtc.org/ [accessed 2021-04-28]
55. Stivers T, Enfield NJ, Brown P, Englert C, Hayashi M, Heinemann T, et al. Universals and cultural variation in turn-taking in conversation. Proc Natl Acad Sci U S A 2009 Jun 30;106(26):10587-10592 [FREE Full text] [doi: 10.1073/pnas.0903616106] [Medline: 19553212]
56. Hoey E. Lapses: How people arrive at, and deal with, discontinuities in talk. Res Lang Soc Interact 2015 Nov 18;48(4):430-453. [doi: 10.1080/08351813.2015.1090116]
57. Keisanen T, Rauniomaa M, Haddington P. Suspending action: from simultaneous to consecutive ordering of multiple courses of action. In: Multiactivity in Social Interaction: Beyond Multitasking. Amsterdam: John Benjamins Publishing Company; 2014:109-134.
58. Kendrick K, Torreira F. The timing and construction of preference: a quantitative study. Discourse Process 2015 Mar 03;52(4):255-289. [doi: 10.1080/0163853x.2014.955997]
59. Pearce C, Walker H, O'Shea C. A visual study of computers on doctors' desks. Inform Prim Care 2008;16(2):111-117 [FREE Full text] [doi: 10.14236/jhi.v16i2.682] [Medline: 18713527]
60. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. J Gen Intern Med 2009 Mar;24(3):341-348 [FREE Full text] [doi: 10.1007/s11606-008-0892-6] [Medline: 19130148]
61. Choudhury A, Crotty B, Asan O. Comparing the impact of double and single screen electronic health records on doctor-patient nonverbal communication. IISE Trans Occup Ergon Hum Factors 2020 Apr 01;8(1):42-49. [doi: 10.1080/24725838.2020.1742251]
62. Montague E, Xu J, Chen P, Asan O, Barrett B, Chewning B. Modeling eye gaze patterns in clinician-patient interaction with lag sequential analysis. Hum Factors 2011 Oct;53(5):502-516 [FREE Full text] [doi: 10.1177/0018720811405986] [Medline: 22046723]
63. Montague E, Asan O. Physician interactions with electronic health records in primary care. Health Syst (Basingstoke) 2012 Dec 01;1(2):96-103 [FREE Full text] [doi: 10.1057/hs.2012.11] [Medline: 24009982]
64. Asan O, Chiou E, Montague E. Quantitative ethnographic study of physician workflow and interactions with electronic health record systems. Int J Ind Ergon 2015 Sep 01;49:124-130 [FREE Full text] [doi: 10.1016/j.ergon.2014.04.004] [Medline: 26279597]
65. Wetterneck TB, Lapin JA, Krueger DJ, Holman GT, Beasley JW, Karsh B. Development of a primary care physician task list to evaluate clinic visit workflow. BMJ Qual Saf 2012 Jan 06;21(1):47-53 [FREE Full text] [doi: 10.1136/bmjqs-2011-000067] [Medline: 21896667]
66. Miller E, Nelson EL. Modifying the Roter Interaction Analysis System to study provider-patient communication in telemedicine: promises, pitfalls, insights, and recommendations. Telemed J E Health 2005 Feb;11(1):44-55. [doi: 10.1089/tmj.2005.11.44] [Medline: 15785220]
67. Schon D. The Reflective Practitioner: How Professionals Think In Action. New York: Basic books; 1984.
68. Sengers P, Boehner K, David S, Kaye J. Reflective design. 2005 Presented at: Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility; August 20-25; Aarhus, Denmark p. 49-58. [doi: 10.1145/1094562.1094569]
69. Liu C, Scott K, Lim R, Taylor S, Calvo R. EQClinic: a platform for learning communication skills in clinical consultations. Med Educ Online 2016;21:31801 [FREE Full text] [doi: 10.3402/meo.v21.31801] [Medline: 27476537]
70. Liu C, Lim RL, McCabe KL, Taylor S, Calvo RA. A web-based telehealth training platform incorporating automated nonverbal behavior feedback for teaching communication skills to medical students: a randomized crossover study. J Med Internet Res 2016 Dec 12;18(9):e246. [doi: 10.2196/jmir.6299] [Medline: 27619564]
71. Kagan N, Schauble P, Resnikoff A, Danish SJ, Krathwohl DR. Interpersonal process recall. J Nerv Ment Dis 1969 Apr;148(4):365-374. [doi: 10.1097/00005053-196904000-00004] [Medline: 5768914]
72. Werner A, Schneider J. Teaching medical students interactional skills. a research-based course in the doctor-patient relationship. N Engl J Med 1974 May 30;290(22):1232-1237. [doi: 10.1056/NEJM197405302902206] [Medline: 4825853]
73. Asan O, Montague E. Technology-mediated information sharing between patients and clinicians in primary care encounters. Behav Inf Technol 2014 Apr 14;33(3):259-270 [FREE Full text] [doi: 10.1080/0144929X.2013.780636] [Medline: 26451062]
74. Chen Y, Ngo V, Harrison S, Duong V. Unpacking exam-room computing: negotiating computer-use in patient-physician interactions. 2011 Presented at: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 7-12; Vancouver BC Canada p. 3343-3352. [doi: 10.1145/1978942.1979438]
75. Helou S, Abou-Khalil V, Yamamoto G, Kondoh E, Tamura H, Hiragi S, et al. Understanding the EMR-related experiences of pregnant Japanese women to redesign antenatal care EMR systems. Informatics 2019 Apr 04;6(2):15. [doi: 10.3390/informatics6020015]

76. Kanda N, Horiguchi S, Fujita Y, Xue Y, Nagamatsu K, Watanabe S. Simultaneous speech recognition and speaker diarization for monaural dialogue recordings with target-speaker acoustic models. 2019 Presented at: IEEE Automatic Speech Recognition and Understanding Workshop; December 14-18; Singapore p. 31-38. [doi: 10.1109/asru46091.2019.9004009]
77. Kodish-Wachs J, Agassi E, Kenny IIIP, Overhage JM. A systematic comparison of contemporary automatic speech recognition engines for conversational clinical speech. AMIA Annu Symp Proc 2018;2018:683-689 [FREE Full text] [Medline: 30815110]
78. Robinson JD. Getting down to business: talk, gaze, and body orientation during openings of doctor-patient consultations. Human Comm Res 1998 Sep;25(1):97-123. [doi: 10.1111/j.1468-2958.1998.tb00438.x]
79. Schegloff E. Body torque. Social Research 1998;65(3):535-596.

Abbreviations

D: Dialogue
SG: Screen Gaze
SG+D: Screen Gaze and Dialogue

Edited by G Eysenbach; submitted 22.10.20; peer-reviewed by L Seuren, E Bellei; comments to author 13.11.20; revised version received 07.01.21; accepted 14.04.21; published 10.05.21

Please cite as:
Helou S, Abou-Khalil V, Iacobucci R, El Helou E, Kiyono K
Automatic Classification of Screen Gaze and Dialogue in Doctor-Patient-Computer Interactions: Computational Ethnography Algorithm Development and Validation
J Med Internet Res 2021;23(5):e25218
URL: https://www.jmir.org/2021/5/e25218
doi: 10.2196/25218
PMID: