

“I Should Listen More”: Real-time Sensing and Feedback of Non-Verbal Communication in Video Telehealth

HEATHER A. FAUCETT, Department of Informatics, University of California, Irvine
MATTHEW L. LEE, FX Palo Alto Laboratory, Inc.
SCOTT CARTER, FX Palo Alto Laboratory, Inc.

Video telehealth is growing to allow more clinicians to see patients from afar. As a result, clinicians, typically trained for in-person visits, must learn to communicate both health information and non-verbal affective signals to patients through a digital medium. We introduce a system called ReflectLive that senses and provides real-time feedback about non-verbal communication behaviors to clinicians so they can improve their communication behaviors. A user evaluation with 10 clinicians showed that the real-time feedback helped clinicians maintain better eye contact with patients and was not overly distracting. Clinicians reported being more aware of their non-verbal communication behaviors and reacted positively to summaries of their conversational metrics, motivating them to want to improve. Using ReflectLive as a probe, we also discuss the benefits and concerns around automatically quantifying the “soft skills” and complexities of clinician-patient communication, the controllability of behaviors, and the design considerations for how to present real-time and summative feedback to clinicians.

CCS Concepts: • **Human-centered computing** → **Collaborative and Social Computing** → **Collaborative and social computing theory, concepts and paradigms** → **Computer supported cooperative work**

Keywords: video conference; real-time feedback; patient-clinician communication; nonverbal communication

ACM Reference format:

Heather A. Faucett, Matthew L. Lee, and Scott Carter. 2017. I Should Listen More: Real-time Sensing and Feedback of Non-Verbal Communication in Video Telehealth. *PACM on Human-Computer Interaction*, 1, CSCW, Article 44 (November 2017), 19 pages. DOI: 10.1145/3134679

1 INTRODUCTION

Video-mediated communication enables people to collaborate remotely. A growing domain for video meetings is video telehealth, in which clinicians conduct visits with patients remotely [67]. When clinicians communicate with patients, they must not only exchange information but also convey a sense of sympathy, sensitivity, and attentiveness. Clinicians must often juggle the competing task demands of extracting information from patients’ subjective accounts and documenting the information in medical records while conveying a sense of warmth and care to establish trust and rapport with the patient. Most affective

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

2573-0142/2017/11 – Article 44 \$15.00

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM. <https://doi.org/10.1145/3134679>

communication is conveyed through non-verbal communication such as eye contact, vocal tone, and body posture [35] [64]. Patients look for non-verbal cues from clinicians to corroborate verbal information and to discern the clinician's real feelings about them, particularly given the patient's relative lack of power and control [26] [27].

However, video-mediated communication often is less effective than in-person communication because it is challenging to convey and perceive the non-verbal behaviors that are essential for controlling the floor and managing turns [37] [60]. Moreover, non-verbal behaviors that may be acceptable in in-person business meetings such as looking away at notes may be perceived as being rude or inattentive in a video meeting [12] [50]. Indeed, patients already feel disengaged when clinicians frequently look at medical records instead of at them during in-person visits [20] [58].

Prior work shows that in video visits, clinicians tend to speak more, being more dominant in the conversation while using fewer patient-centric behaviors such as empathy and praise utterances [46], which can lead to poorer patient satisfaction and incomplete information gathering. Further, few clinicians are trained to communicate over a video visit, and many are not always aware of how they present themselves to patients over video. Thus, there is an opportunity to provide feedback about clinicians' communication behaviors on video to help clinicians be more aware of non-verbal behaviors like looking away from the patient or interrupting, while also encouraging more patient-centric, rapport-building communication behaviors.

In this paper, we describe the design and evaluation of ReflectLive, a system that senses and provides real-time feedback about clinicians' communication behaviors during video consultations with patients. In user testing with 10 clinicians, we show that real-time sensing and feedback has the potential to train clinicians to maintain better eye contact with patients and be more aware of their non-verbal behaviors. Our contributions include 1) the design and implementation of ReflectLive, a system that senses and provides real-time and summative feedback in web-based video meetings, 2) the results of a user evaluation with 10 clinicians demonstrating how real-time feedback supported greater eye contact and how summative feedback enabled clinicians to learn new insights into their own communication behaviors, and 3) a discussion of how clinicians' reactions to this approach ranged from enthusiastic to cautious, including how we used ReflectLive as a probe to reveal the limitations of quantifying the complexities of clinical communication with a set of simple metrics, how controllable different behaviors are using real-time feedback, and its use in skills training and practice.

2 BACKGROUND

2.1 Non-Verbal Cues in Patient-Clinician Communication

Non-verbal communication can build rapport, convey unspoken emotions, and reinforce verbal communication to improve collaboration [60]. Patient-clinician communication research has a rich body of observational studies that link non-verbal behaviors with patient/clinical outcomes. Interpersonal behavior can be mapped onto two dimensions: affiliation (e.g., eye contact, facial expressiveness, backchannel) and control (e.g., interruptions, speaking more, vocal or visual dominance) [39]. Patient outcomes such as patient satisfaction, self-disclosure, trust, compliance, and even a reduced likelihood of malpractice claims are more associated with affiliative behaviors than controlling behaviors [3] [44] [51]. Clinicians tend to speak more than patients [55], but increased listening has been found to correlate with greater patient satisfaction [33]. Affiliative eye contact with the patient is associated with higher patient satisfaction [51] and recognizing psychological distress in patients [8]. Indeed, the American Telehealth Association recommends eye gaze be maintained as much as possible [66] which may be the behavior patients notice most [49].

Conducting patient visits over video telehealth can differ from in-person visits. Clinicians in telehealth visits tend to be even more dominant [2] [62], exhibit fewer patient-centered behaviors such as empathy utterances [46], and give a magnified sense of urgency [53]. These affective cues are particularly critical for establishing a "therapeutic alliance" between clinician and patient [30], in tele-mental health, a growing practice of remote psychiatry and therapy sessions to treat the 1 in every 25 adults in the U.S. who have suffered from a serious mental illness [22].

Interpersonal communication training is incorporated into some selective medical training programs. Training includes simulated visits with standardized patient actors that are recorded and manually rated by educators [25] using coding systems that analyze both verbal and non-verbal communication such as turn-taking, interruptions, and empathic expressions [59]. Using an automated approach for quantifying communication behaviors can potentially reduce the high cost of training. Further, prior research suggests the need for specific training for telehealth to train the clinician to stay within the video frame [52].

Based on the prior work that shows how affiliative behaviors are related to positive patient outcomes and the need for more automatic self-feedback and training opportunities for clinicians to communicate better using video telehealth, we focus on sensing and providing feedback for speaking contribution, interruptions, eye contact, and face position.

2.2 Sensing and Feedback Non-Verbal Behavior

Our work builds on previous examples in the CSCW and HCI fields of quantifying and visualizing communication behaviors both in co-located and remote video meeting settings. Numerous examples demonstrate how feedback helps moderate the speaking contributions of over- and under-contributors. Work by Dimicco *et al.* [23] is an early example of measuring and visualizing speaking contribution as bar graphs projected in real-time onto a shared display in a co-located meeting. Similar effects were found with the Conversation Clock [9], a tabletop display that showed both the current and history of speaking contributions over the course of the meeting. Terken *et al.* [65] uses a multimodal approach that measures the eye gaze of individuals using microphones and IR-equipped headbands and visualizes each user's speaking contribution and amount of eye gaze received when the user or others are speaking; however, they did not find any changes in eye gaze behaviors as they found with speaking contribution. Adachi *et al.* [1] used a tablet's mic and camera to track speaking contribution and eye gaze of speakers around a table. The system displayed a score that rewarded under-contributors for speaking more and over-contributors for speaking less. Balaam *et al.* [11] found that a Wizard-of-Oz ambient display that provided feedback helped support greater interactional synchrony in small groups. Wearable examples include SayWAT [4] which provides feedback on a speaker's pitch and volume on a head-mounted display and Logue [21] that uses a Microsoft Kinect and head mounted display to provide feedback on body language and speech rate to people making presentations. Our system similarly uses a multimodal approach that analyzes the audio for speaking behaviors and the video for eye gaze and facial behaviors but instead is designed for use by remote participants.

Prior work in feedback of communication behaviors in remote video meetings show less pronounced effects. The Meeting Mediator [40] is a mobile system that uses wearable sociometric badges to sense each user's voice and display real-time feedback on a mobile phone that highlights speaking balance and interactivity. The system was found to increase interactivity among speakers and make more dominant users less dominant in distributed settings. Byun *et al.* [14] take a similar approach to ours and augment traditional video calls with audio and visual analysis. Their system aggregates lower level signals like speaking rate and interruptions to infer and visualize higher level honest signals [57] such as Activity, Consistency, Influence and Mimicry. In contrast, our system is designed for clinicians and avoids aggregating low-level behaviors into complex, less actionable constructs. Instead, our system directly displays behaviors such as speaking contribution, interruptions, eye gaze, and face positioning so clinicians can use their best judgment to self-regulate these modifiable behaviors. From a technical perspective, our system—ReflectLive—also differs by using a peer-to-peer WebRTC architecture that requires only what is needed for a video call, a device with a webcam and a browser, and is able to process all data streams locally on each individual's client instead of on a server.

2.3 Communication Feedback Systems for Clinicians

Prior work addressing the specific information needs of clinicians for video consultations point to the conceptual value of real-time and summative feedback. The EQClinic platform [46] is a video training tool for medical students that can provide summative reports about speaking contribution, volume, and pitch as well as

facial expressions, head positioning/nodding, and hand-over-face. In a study with medical students, reviewing summaries of non-verbal communication behaviors collected by EQClinic improved student's interview skills. The ReflectLive system uses a similar approach for training but provides real-time, actionable feedback in addition to summative reports so clinicians can adjust their behavior during each video visit. Our work draws particular inspiration from Hartzler *et al.* [31] who conducted an evaluation of ambient, non-verbal communication feedback using a Wizard-of-Oz approach that mapped of non-verbal behaviors to two relational signals important for patient satisfaction: affiliation and control. They found that easily interpretable and actionable feedback needs to be updated in real-time and communicated using appropriate size, color, and visual metaphor. Unlike Hartzler *et al.*'s investigation, our system is fully-functional system that allows us to investigate how clinicians and patients to communicating remotely. We also investigate how clinicians feel about integrating this feedback system into training programs and their regular practice, discussing both benefits and potential barriers. In the next sections, we describe the design and implementation of ReflectLive, how we evaluated the effects of real-time and summative feedback on clinicians' communication behaviors during video telehealth visits, and what benefits, concerns, and opportunities clinicians expressed with using ReflectLive for training and in regular practice.

3 System Design

We designed ReflectLive, a system that senses, analyzes, and provides feedback on four non-verbal behaviors (speaking contribution, eye gaze, conversational overlaps, and centeredness on screen) in real-time. These behaviors have been previously demonstrated in other works to have an impact on video-based conversations [7] [33] [51]. ReflectLive's real-time sensing and feedback capabilities are built on top a WebRTC-based peer-to-peer video conferencing platform similar to [6], which is easy to set up and requires no additional software/hardware beyond an HTML5-capable browser on a device with a webcam and microphone. All audio and video processing is done locally in the client's browser, which minimizes privacy and security risks. ReflectLive analyzes the audio and video streams from the local client and the audio stream from the remote peer to sense non-verbal metrics for Speaking Contribution, Interruptions, Eye Gaze, and Face Position in real-time and visualizes them on a dashboard panel visible during the video meeting (Fig. 1).

3.1 Real-time Audio Analysis

Speaking contributions and interruptions are detected by analyzing of the audio stream in real time. Our software stack uses Hark.js [70], an open-source Javascript module that emits events when a user starts and stops speaking.

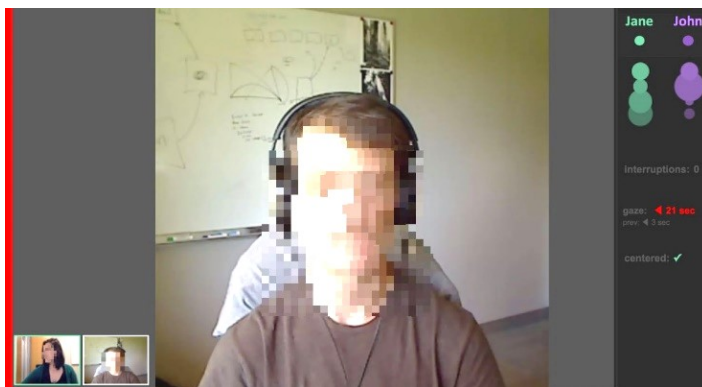


Fig. 1. The ReflectLive video meeting system, with the visualization dashboard on the right showing real-time metrics about non-verbal behaviors. Jane (in the thumbnail) is looking to the left. A red bar flashes on the left of her window as she looks to the side. Faces in figure are blurred for privacy.

3.1.1 Speaking Contribution. To provide live, real-time second-by-second feedback about speaking contribution, the system progressively accumulates the user’s speaking time when a start talking event is received. Speaking contributions by each speaker are visualized as circles whose diameters are proportional to the amount of time a user has spoken. When a user first enters the video chat, they will see two small circles with their name and the name of the other user above (Fig. 2a). As a user speaks, their circle will begin to grow (Fig. 2b). Drawing inspiration from Hartzler *et al.* [31], the visualization also incorporates historical information. Rather than using these growing circles to represent the whole of the conversation, each circle represents the last 60 seconds in a sliding window. Every 60 seconds, a snapshot of the current circles is taken and pushed below the growing and shrinking circles (Fig. 2c). The top circles will continue to grow as a user talks, and shrink when they are silent. The history displays up to the last 5 minutes of conversation. As the conversation moves beyond 5 minutes, the bottom-most circle (representing 5 minutes in the past) will drop off, being replaced with the one above it (Fig. 2d).

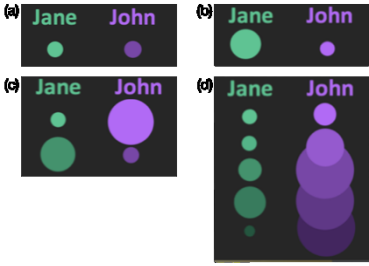


Fig. 2. (a) Speech circles at beginning of meeting, (b) Jane starts speaking more than John. (c) 1 minute into the conversation a snapshot is taken. Jane was speaking more in the last minute but currently John is speaking more. (d) 5 minutes of conversation.

3.1.2 Interruptions. Interruptions are also determined by the audio events. While true conversational interruptions are based on far more than simple overlaps [54], to keep the detection and feedback straightforward, we define interruptions to be conversational overlaps in speech between users. The clinician is considered to have interrupted the patient when the clinician begins to speak while the patient is speaking. The length of the interruption is calculated as the duration of simultaneous speech until either one of the users stops talking. To filter out backchannel utterances such as “mhmmm”, “yeah”, and “I see” [54] (which are not intended as a request for a speaking turn), we ignore overlaps that are less than 1.5 seconds.

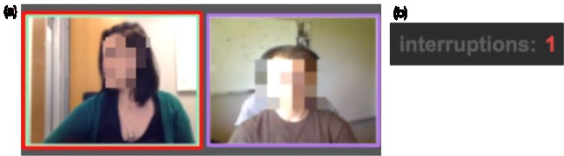


Fig. 3. Jane has interrupted John. Her thumbnail highlights red (a) and her interruption count increases (b). Faces in figure are blurred for privacy.

To visualize interruptions, a simple counter in the dashboard and a thumbnail highlight were implemented (Fig. 3). Only interruptions caused by the clinician, meaning those times when the clinician began talking before the patient had stopped, are displayed in the clinician’s visualization. During an interruption, the interrupter’s thumbnail temporarily highlights red for the duration of the interruption.

3.2 Real-time Visual Analysis

ReflectLive analyzes the client’s video stream to sense eye gaze and body positioning. We use Visage|SDK™ [65], a commercial Javascript-based computer vision face tracking software to detect where a user is looking and where their face is positioned on the screen.

3.2.1 Eye Gaze. Note-taking during a medical consultation is one of the most common reasons to look away from the patient, and in the case of video visits, either to the left or right (towards a secondary display). The ReflectLive system provides real-time feedback when the clinician is looking left or right of the display with the webcam.



Fig. 4. (a) Jane has been looking to the left for 21 seconds, (b) Jane is looking at the center of the screen. The last time she looked away it was to the left for 3 seconds.

Using the Visage|SDK™ gaze tracking module, we first determined the thresholds for left, right, and center gaze. These thresholds were based on the edges of a 27” monitor (typical upper limit) and a user sitting approximately 30” from the webcam mounted above and centered on the monitor [68]. A user was determined to be looking away if their eye gaze averted (with or without orienting the head) past the edges of the monitor to the left or right. If a user’s gaze stays within the screen edges, or looking above the monitor, the system considers this centered eye gaze.

The eye gaze indicator on the dashboard uses icons (directional triangles) to represent whether the user’s gaze is centered, right or left. Once a user looks away from the center, the icon will change to a triangle pointing in the direction of their gaze and a timer will start counting up the number of seconds of sustained non-centered gaze (Fig. 4a). At the same time, an animated red bar is drawn on the edge of the screen in the direction the user is looking (Fig. 1). The animated/flashing bar leverages the sensitivity of the user’s peripheral vision to motion to make this indicator noticeable, even though the user is looking away from the main screen. To reduce noise and provide less intrusive feedback, eye gaze indicators would only be shown after the user has been looking away for more than 2 seconds. Once the gaze is centered on the screen, the icon changes to the centered icon (Fig. 4b) and the previous gaze direction and time looking away are shown below.

3.2.2 Face Position. Similar to gaze tracking, face position is determined by the Visage|SDK™ face tracking module. A user was determined to be un-centered if their face moved to the rightmost or leftmost 15% of the screen width (Fig. 5b). If a user’s face was determined to be within the middle 70% of the screen’s width (Fig. 5a), the user was considered centered. To visualize this, a simple centered/not-centered indicator was shown in the dashboard.

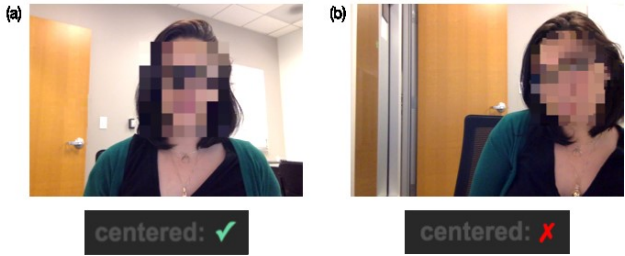


Fig. 5. (a) The user’s face is centered on screen, (b) The user’s face is not centered on screen.

3.3 Summary Result Page

To complement the real-time feedback during the meeting, we designed a summary report web page to be reflected on by the user after the meeting. The report summarizes the metrics accumulated during the live meeting.

3.3.1 Speaking Contribution Summary. Similar to the real-time visualization, the summary page shows the speaking circles as they were displayed during the conversation (Fig. 6). However, instead of a 5-minute history, it shows up to 15 minutes (the maximum meeting length in our study). The top circle represents the last minute of conversation, with each circle beneath representing 1 minute further in the past. Next to this visualization is a pie chart giving the cumulative percentages each user spoke.

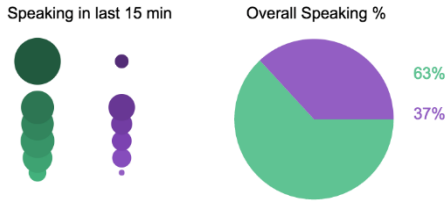


Fig. 6. Speaking contribution summary example. Here the clinician is green and the patient is purple

3.3.2 Interruptions Summary. The summary page shows a total count of interruption by the clinician and a timeline summary (Fig. 7) showing when each interruption occurred relative to the start and end of the meeting.

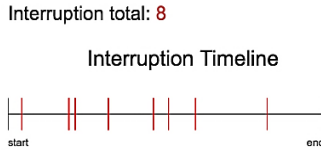


Fig. 7. Interruption summary example. Here the clinician has interrupted throughout the entire conversation.

3.3.3 Eye Gaze Summary. For eye gaze, the summary shows which direction the user looked away towards most often, the average time they looked away for, and a heat map (Fig. 8) showing the percentage of time they spent looking in each direction.

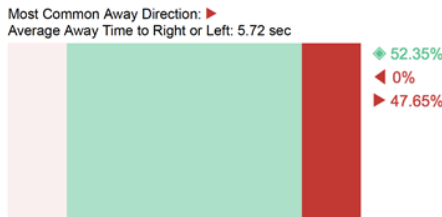


Fig. 8. Eye gaze summary that shows the clinician has spent a significant amount of time looking sideways to the right.

3.3.4 Face Position Summary. Finally, the summary shows a pie chart showing the total time a user's face was positioned centered or un-centered (Fig. 9).

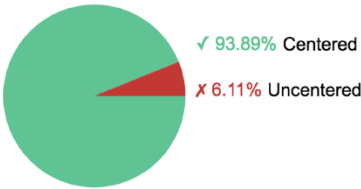


Fig. 9. Face position on screen summary example. Here the clinician’s face was centered on the screen for most of the visit

4 EVALUATION

We conducted an evaluation to investigate the feasibility of using ReflectLive for video consultations and what impact, if any, the real-time and summative feedback has on communication behaviors. We recruited clinicians (experts who have used video telehealth in their practice or novices interested in starting to use it) through a combination of convenience and snowball sampling [17]. We recruited a total of 10 clinicians (9 MDs, 1 psychologist Ph.D.) (Table 1), all well-trained in their respective specialties. We use the term “novice” to refer specifically to their experience with the telehealth medium for communication. Participants were volunteers and did not receive compensation.

Table 1. Participants (practicing clinicians) in the evaluation.

Participant	Telehealth Experience	Medical Specialty
P1 (M)	novice	clinical psychology
P2 (M)	expert	pulmonology
P3 (F)	expert	psychiatry
P4 (M)	novice	internal medicine
P5 (M)	novice	cardiology
P6 (M)	novice	urology
P7 (M)	expert	internal medicine
P8 (M)	expert	psychiatry
P9 (M)	expert	psychiatry
P10 (F)	expert	psychiatry

Clinicians either visited our lab or participated remotely from their home/office (a benefit of ReflectLive being a fully-functional system), with sessions lasting 60-90 minutes. Researchers followed the same procedure for both local and remote participants, using the phone initially to give instructions to remote participants and video call for the interview at the end of the study.

Clinicians were told they were going to test two versions of our video call system, by performing two medical consultations with a standardized patient actor [10] who was trained to exhibit specific symptoms and connected over video from her home computer. Clinicians were given the patient’s name, age, and chief complaint. The clinician’s goal was to conduct a diagnostic interview and make a treatment plan for the patient

in under 15 minutes. Clinicians were also asked to take notes in their usual manner, either by typing on a computer on a secondary screen or hand-writing notes.

In the first video visit (without visualization), clinicians used a version of the prototype with the visualization dashboard hidden, similar to a standard video meeting interface like Skype. Face and speech data were still processed and collected in the background, unbeknownst to the clinician.

Immediately after this first video visit, we introduced participants to the metrics and real-time visualizations on the ReflectLive dashboard and how to read them. We verified their understanding using teach-back. Participants were allowed to play with a demo and test how the eye gaze and face position feedback worked.

After they were satisfied with the demo, they entered the second video visit and performed a medical interview with the visualization dashboard visible, showing them live feedback about their speaking contribution, interruptions, eye gaze, and face position. The same patient actor portrayed a different patient with a slightly different set of symptoms, accent and clothes/hairstyle to give the clinician the feeling of a different patient. In both visits, the symptoms were a combination of psychological and physical but with different backstories, PTSD + Irritable Bowel Syndrome in the first visit and PTSD + Migraines in the second visit. The professionally trained patient actor was instructed to act in a similar manner in both cases, and in a way consistent with how she would perform a typical patient-actor scenario. To keep the actor consistent in her performance, the order of characters portrayed was the same for all participants. Furthermore, the visit without feedback always followed the visit with real-time feedback to avoid priming clinicians with knowledge of their own conversational metrics. We observed/listened to each visit, either in the lab or over the phone as the clinician conducted their video consultation. Visits were recorded for later analysis. Immediately following this second video visit, researchers conducted semi-structured interviews about the clinician's experience with the live real-time feedback. Finally, we showed the clinician the summary results web page for the second visit and conducted semi-structured interviews about their reactions to the summary metrics.

Using an iterative, inductive approach, emergent phenomena were identified, named, and categorized following techniques similar to those employed in grounded theory [18]. We used an inductive method of analysis to understand how participants perceived their behaviors in each visit, their reaction to the quantified results, and their feelings and interpretations of this kind of feedback. Using open coding, we met both directly after each interview and followed up weekly to discuss recurring themes in the data. As dominant codes emerged, they were incorporated into the categories presented in this paper. Additionally, through directed coding we identified the dimensions and degrees of variation around these behaviors and experiences. We used affinity diagramming and axial coding to understand the relationship between, across, and within these codes.

4 RESULTS

The results of the user evaluation address the questions of how real-time and summary feedback about communication behaviors affects 1) clinicians' communication behaviors during a video consultation, and 2) clinicians' self-awareness of their communication behaviors after reviewing summary data. Additionally, we report on how we used ReflectLive as a probe to reveal the difficulties in quantifying the complexities in clinical communication. We first report how participants felt about the feasibility of the design and how it might be used.

4.1 Feasibility and Usability of Real-Time Feedback

All participants initially expected the real-time feedback to be distracting when they first saw the demo, but after using the system with real-time feedback, they all reported that the visualizations were not distracting, with the exception of P1, who reported that the blinking gaze bar was distracting because it was inaccurate and triggered so frequently. However, the patient actor reported P1 did not make eye contact throughout the meeting and reexamination of the video from the meeting by the researchers showed gaze warning was correctly calibrated and performed accurately. P1 further reported, however, that the rest of the visualizations were not distracting, but only after he got more familiar with it, saying *“My initial reaction with it was a little off-putting, but I got more comfortable as I got to play with it.”* P8 echoed this sentiment, explain that while he

did not find the feedback itself distracting, once he became aware of his actions from the feedback it became “...difficult to do two things at once... focusing on the patient but at the same time focusing on, you know, ‘am I communicating correctly, giving the right body language, am I talking too much, am I interrupting?’” However, he continued to say that he felt that this was something he would get used to over time and felt paying attention to these metrics was important. Other clinicians felt the feedback was immediately helpful, “*Surprisingly they were not distracting, I thought they were helpful. (P2)*” P3 said it was not distracting because “*the only distracting thing was when I looked away*” and valued that it was a salient reminder to re-orient her eye gaze towards the patient, even to the point of looking directly at the camera rather than at the screen. P5 replied with a simple “No.” when we asked him whether the visualizations got in the way of having a meaningful conversation. P4 mentioned, “*It didn’t interfere with my work or my ability to ask questions or deliver information. It was interesting but not distracting.*”

All participants reported that they paid most attention to the speaking contribution circles and gaze feedback. All participants reported that they understood the circle visualization and could interpret it in real-time, at a glance. All participants also said they paid the least attention to the interruption counter and face position indicator. The participants explained that the speaking contribution and gaze indicators were larger, and as P3 put it, “*more in your face*” than the interruption counter and centeredness feedback. P3, P4, P5, P6, P7, and P8 all suggested making these visualizations larger and more noticeable. P9 compared the dashboard to an existing feedback UI element, the thumbnail showing his own video feed. He expressed how seeing his own thumbnail is distracting but alerts in the dashboard when his gaze or face position becomes uncentered are timely, actionable, and appropriately noticeable.

4.2 Effect on Conversational Behaviors

Overall, each video visit lasted an average of 9 minutes and ended with the clinician suggesting a referral or a follow-up, which is typical for initial consultations. In the following sections, we report how real-time sensing and feedback of speaking contribution, eye gaze, interruptions, and face position affected clinicians’ communication behaviors by examining the differences between the visits with and without feedback. Only the percentage of centered eye gaze was statistically significantly different between visits, so we focus on the qualitative analysis of how real-time feedback had on clinicians’ communication behaviors.

4.2.1 Speaking Contribution Behavior. All clinicians spoke more (a higher percentage) than the patient in both visits, which is consistent with studies of in-person visits that show clinicians generally speak more than patients [55]. Comparing across visits with and without feedback (Table 2a), 7 out of 10 clinicians decreased their speaking percentage with feedback to be more balanced (less dominant) with the patient.

Clinicians who were able to speak less (and thus were more balanced) with feedback said they used the strategy of listening more to the patient, using the real-time feedback to know when they should listen more actively. Clinicians who increased their speaking percentage with feedback reported that they had difficulty with getting the patient to speak more, and adopted the strategy of asking more and longer questions which increased their speaking contribution. For example, P3 perceived that the patient actor in the second visit appeared to give more terse responses, which required more follow-up question-asking to uncover the important information. To build rapport and elicit more disclosure from the patient, P5 tried to “*have a conversation*” by being more casual and talkative himself.

4.2.2 Eye Gaze Behavior. All participants, except P6, increased (or maintained, in the case of those already near the 100% ceiling) the amount of time their gaze was centered (Table 2b). Indeed, a sign test using the differences within subjects across visits as values shows that the visit with feedback had more centered eye gaze than the visit without feedback ($p=0.035$). The visual salience of the red flashing bar (as mentioned in section 4.1) was likely one important factor for why clinicians were able to notice and use the real-time feedback to improve their eye gaze.

Table 2. Measured behaviors in the visits without and with real-time feedback (using ReflectLive) and the difference.

	(a) Speaking Contribution			(b) Eye Gaze Centered			(c) Interruption Count		
	No Feedback	With Feedback	Diff	No Feedback	With Feedback	Diff	No Feedback	With Feedback	Diff
P1	66%	60%	-6%	30.4%	60%	+29.6%	8	1	-7
P2	62%	66%	+4%	99.4%	100%	+0.6%	1	0	-1
P3	59%	65%	+5%	95.6%	99.3%	+3.7%	7	8	+1
P4	71%	69%	-2%	98.7%	98.9%	+0.2%	3	2	-1
P5	62%	71%	+9%	90.4%	94.9%	+4.5%	0	0	0
P6	64%	54%	-10%	80.0%	52.4%	-27.6%	1	3	+2
P7	63%	61%	-2%	100%	100%	0%	8	0	-8
P8	68%	62%	-7%	99.4%	100%	+0.6%	3	4	+1
P9	69%	64%	-5%	100%	100%	0%	2	1	-1
P10	63%	56%	-8%	58.2%	70.4%	+12.2%	5	0	-5

Only P6, a telehealth novice, reported that he felt gaze was unimportant for him, and decided to focus more on improving the speaking contributions by listening and writing more. As a result, P6 spent more time looking to the right at his notes in ReflectLive meeting, saying “*I need to write what I need to write,*” indicating that “*there was nothing he could do about [sideward gaze at his notes].*” The other clinicians, who had room to improve and were interested in improving (P1, P3, P5, P10), used the real-time feedback to remind them when their gaze was not directed at the patient.

4.2.3 Interruption Behavior. Comparing between visits with and without feedback, 6 out of 9 participants (excluding P5 who had zero interruptions) decreased the number of interruptions (Table 2c). 3 participants managed to use the live feedback to reduce their interruptions quite dramatically, with P1 going from 8 to 1, P7 going from 8 to 0, and P10 from 5 to 0. These clinicians explained that they tried to wait longer after the patient’s turn to begin speaking, to avoid potential overlaps. We did not observe any instances of similarly dramatic increases in interruptions. P6 tried to pay more attention to the interruption feedback saying, “*The effort was with the first two points, giving her more time to talk and then the interruptions.*” However, he actually interrupted 2 more times than in the visit without feedback, indicating controlling interruptions may be difficult because it depends on both speakers.

4.2.4 Face Position Behavior. Participants had their faces centered and large enough on the screen mostly throughout both visits, with only P1, P5, and P7 not being centered 100% of the time. P1 was face-centered 98% in both visits. P5 was 94% centered in the visit without feedback and 100% centered in the visit with feedback. P5 noticeably sat more upright and centered his face at the beginning of the visit with feedback and maintained that posture. All participants reported that they did not pay much attention to the live face position indicator, if at all, as it was too small, and felt they could maintain their face position after being reminded about it before a meeting.

4.3 Reviewing Summary of Conversational Metrics

Following the two video visits, clinicians were shown a summary report (Fig. 6, 7, 8, 9) that illustrated the conversational metrics captured during the second visit. In contrast to the live, real-time feedback, the summary report shows the cumulative values of the metrics for the entire second visit. All participants never previously had their clinical communication behaviors quantified with such precision. Next, we report on their reactions to these quantified behaviors and any actions or intentions triggered.

4.3.1 Reflections on Speaking Contribution Summary. Prior to seeing the summary report, most clinicians reported that they thought they had spoken less than the patient. Upon seeing their results, clinicians were surprised at the amount they spoke and reported that this is a behavior they would like to change in the future. P1, upon seeing the speaking contribution summary immediately exclaimed: “*Wow, I talk too much!*” P9 also was surprised how much he spoke and said he would in future try asking more open-ended questions.

All participants reported that they found the snowman summary (Fig. 6) easy to understand, with several physically pointing to circles in different parts of the conversation to illustrate where different topics were discussed. Participants speculated as to why they may have spoken more than they expected, offering many potential explanations for their speaking contribution and how it might influence their ability to change their behaviors. For example, P3, P5, P8, and P9 hypothesized that some of this imbalance may be due to the patient actor’s reticence, and speculated as to how much control they really have in managing their proportion of speaking time. P7 noted that “*Patients are all over the map: some you need to draw out answers, and other go on and talk about their grandkids.*” While all participants agreed that the clinician should, in an ideal situation, listen more than they talk, after reflecting on their data, several clinicians arrived at a realization that their proportional contribution is not completely in their control, and wanted more data points over time to reinforce whether their dominance in speaking amount was a serious problem.

Clinicians also pointed out predictable patterns in the speaking contributions summary. For example, P10 pointed out how the data showed her speaking quite a bit at the end of the visit and explained that “*most of the time, I want to hear from the patient—ideally more patient in the beginning, but toward the end, the doctor would talk a little more.*” P4, P8, and P9 also identified this pattern where clinicians speak more at the end to summarize the plan to justify their behavior. The clinician’s progression in a video visit echoes the six phases identified by Byrne and Long [13] during in-person visits, including establishing the relationship, discovering the patient’s reasons, and terminating the visit. Each phase may require a different speaking contribution goal for the clinician. This suggests that, rather than reducing speaking contribution to a single percentage, the system should show patterns of contribution over the duration of the visit and across multiple visits. Indeed, the different expectations of speaking contribution highlights a design opportunity to make ReflectLive more context-sensitive to support the valid and desired behaviors at different phases of the clinical visit, for example, showing a higher “cost” of interrupting the patient early in the visit or allowing freedom for clinicians to speak more at the end to summarize the care plan.

4.3.2 Reflections on Interruptions Summary. While most participants reported not noticing when interruptions occurred in real time, several found the interruption timeline provided in the summary report allowed them to reflect on *how*, *why*, and *when* interruptions may have occurred in their conversations. Some behaviors participants identified were expected—several participants speculated that backchanneling might trigger an interruption cue (even though we filter overlaps <1.5 seconds), as well as some common conversational quirks, such as “tripping over each other” when unsure if the other person is finished or about to begin speaking. These conversational overlaps are not unique to the digital environment.

Several participants, however, did identify potential causes of their conversational overlaps that are distinct to online communications. In particular, multiple participants (P3, P7, P8) identified internet connection and speed as a factor in conversational overlaps, consistent with general video meetings [38]. P3 explained that having an indicator of when these interruptions occurred gave her the ability to, in a way, “*see what the patient was hearing.*” She elaborated that in an in-person conversation, you know that the person was able to hear you, and you know that they were able to hear you immediately. However, in a digital environment, you cannot always tell if you were heard or if there was a delay. P8 also identified connection issues as a potential source of overlaps, noting that he had an excellent internet connection during the study, which he felt allowed for less of a chance of accidental overlap. Further, by reflecting on these identified causes through examination of the interruption timeline, these participants identified how they might change this behavior in the future by slowing their speaking rate and reactions.

4.3.3 Reflections on Eye Gaze Summary. Participants overall found the heat-map style summary of gaze throughout the conversation to be easy to understand and used the results to reflect on their behaviors and why they may have happened. Most participants reported being surprised at their results, thinking that they had looked away from the patient for less time than they had. However, most came to terms with the results,

reporting that they must have looked at their notes more often than they had realized. The behavioral results in the previous section show that most were able to improve on their eye gaze with the help of the feedback, but many expressed a limit on how much eye contact they could maintain, given the need to read and document notes during the visit. Some experts (P3, P7, P9) suggested that they should mitigate the perception of inattentiveness by telling the patient that they will be looking to the side to take notes.

P3, P7, P8, and P10 reflected on their behaviors and what might be seen as “looking away” or poor eye contact in an online meeting, though they had different perspectives on how these behaviors are interpreted by the patient. Expert P3, for example, reported that she intentionally looks at the camera, rather than at the image of the patient on screen to maintain the illusion of eye contact. The patient actor commented on this case, reporting how well she felt the clinician had maintained eye contact throughout the conversation, a response consistent with direct gaze awareness [19] [24]. Direct eye contact has been shown to improve recall of information over video conferencing, an important outcome for patients who need to remember the clinician’s instructions [28]. P3 reported that she felt gaze was extremely important, being the most useful of the metrics. She further explained that in an online setting, looking away at all can be perceived more negatively by the patient than when in person, because in a face-to-face meeting, the patient can see what you are looking at. P7 echoed this sentiment, saying *“This medium does not compensate for any shortcomings. It enhances shortcomings.”* P8, however, felt that the video medium might, in fact, allow the clinician to appear as though they are still looking at the patient, even when they are looking at their notes, saying *“there is something special about the video visit medium in that it allows the clinician to look like he/she is actually looking at the patient, compared with in the office when the doc has to look at a computer screen.”* However, he later continued, admitting that *“Eye contact towards the camera is not natural in a video chat, so having this feedback is important.”* To make it easier for clinicians to maintain critical eye contact with the patient without having to look directly at the camera, techniques such as [29] [44] can be applied in the clinical setting to computationally adjust the video image of the clinician’s face/eyes to appear to look directly at the patient.

4.3.4 Reaction to Quantifying Communication Behaviors. Finally, one major theme that arose from reflections on the summary report was the variance in reactions to the act of quantifying communication behaviors and how they may be interpreted by others. These reactions ranged from interested and enthusiastic to skeptical and distrusting. How one reacted to these metrics depended largely on the clinician’s experience with telehealth and whether their role included managing other clinicians.

For example, P6 (a telehealth novice) reacted strongly to the very idea of these behaviors being quantified, questioning their validity and importance. He expressed concern about how these data might be used against him and his colleagues by insurance companies and administrators in unfairly mischaracterizing his effectiveness as a physician. P6 explained, *“The patient-doctor interaction in a closed room is the last sanctuary where a doctor can really, for better or worse, be unmonitored.”* Even though he felt these metrics would improve the art of interviewing, P5 also expressed concern about possible misinterpretation of his communication metrics. As a cardiologist, P5 said he has a focused and task-oriented interview style, which he judges to be appropriate and effective, but may not always measure up according to this limited set of conversational metrics.

However, those who had more experience with telehealth or who acted as administrators (that is, managed or oversaw other clinicians who use video telehealth) did not find the feedback invasive but instead expressed how excited they were about these metrics for either self-improvement or developing skills of clinicians in their organization. P7, a telehealth expert himself and a manager a panel of physicians who regularly see patients using video telehealth said,

“A tool like this would be invaluable. For [my physicians] to see that. To know what they are doing. To be able to get that report at the end of the clinic and how did a physician do in comparison to their past and also compared to everyone in my practice...This is great. Not only from a training purpose, but a CQI purpose...I would run it with every clinic. Were you having a good or bad day? Build a historical average. Stratify my physicians - good / poor / average, and redirect anyone below a level to retraining.”

When we raised the possibility of pushback from individual clinicians not wanting to be monitored, P7 responded, “It’s healthcare, it’s the definition of big brother...Everything I am doing is monitored...I don’t feel it’s intrusive in at all. It depends on how [I] present it. This is a tool to help you [the clinician] communicate in the best way possible and the patient is seeing you in the best light as possible. It’s not about looking over your shoulder. You should want to use it.” P3, P8, P9 all echoed their belief that the individual clinicians they manage would not likely have problems with receiving feedback because they all want to be the best doctor possible. From their managerial perspective, they felt confident that they would not place too much importance in any one or set of metrics and would consider these as part of a more holistic evaluation.

Additionally, all participants noted instances in which the measures do not fully capture the complexities of a good clinical consultation. For example, P3 commented that in her field of psychology, there are times when you may need to interrupt in order to break a patient’s negative self-talk and redirect them. In the case of speaking contribution, as discussed above, several participants recognized patterns in their own behaviors that do not fit the expectation of balanced contribution between clinician and patient, for example when giving the patient detailed information at the end of a meeting. Thus, clinicians were quick point that even though the behaviors measured by ReflectLive are important for good communication, they do not represent all the cases of good patient-clinician communication.

All experts and administrators reported being enthusiastic and excited about future applications, often expressing that they would like to use this design in their own practice. Novices and non-administrators, however, tended to be more skeptical in its application in their own practice, but felt it would be very useful in medical student or resident training where trainees’ motivation for self-improvement is high and the stigma of performing poorly on these measures might be lower. P2 said *“It’s a really nice feedback...we train our students to not talk so much and listen!”* Likewise, expert P3 also said she wanted to use this to teach her staff. Novices P4, P5, P6 agreed that this would be a good training tool, with P5 saying that this system would be good for *“...med students interviewing, it’s a big part of medical teaching. Diagnosis is only as good as the info you extract from the patient...This is definitely going to improve the art of interviewing. [a] very central part of patient care.”* P1 mentioned that the automatically quantified data would reduce the need and cost of experts to observe and rate the communication performance of trainees. Thus, the quantified feedback from ReflectLive was considered useful for augmenting conventional communication training programs guided by experts as well as enable individual clinicians to engage in self-training over time.

5 LIMITATIONS

Our study included only 10 participants, though still within 1 standard deviation of the mean for studies that combine in-person and remote participants [15]. However, all were busy practicing medical professionals which gives our results a level of external validity. The setup also included a realistic remote scenario that connected clinicians in their office to the patient (actor) in her home over video. One limitation of our study procedure is that participants always used the system without feedback first, followed with real-time feedback. The particular symptoms portrayed by the patient actor also followed the same order, which helped the actor provide a consistent performance. It is possible the improvements in eye gaze can be attributed to being “warmed up” after the first meeting; however, all clinicians were actively practicing and have experience with speaking with patients remotely either over video or phone. Having the visit without feedback first allowed us to capture the clinicians’ default behaviors without worrying about carryover effects from any intentions to improve formed by participants if they had started with the visit with feedback. Finally, our study focused on only four metrics that were selected as being both important for building rapport and were feasible to sense and act upon. We realize that these metrics do not fully cover the complexities of clinician-patient communication, and our future work will continue to investigate other cues important for feedback.

6 DISCUSSION

Our results show that the real-time feedback provided by ReflectLive was easily interpreted and not distracting to clinicians. The results provide some evidence that real-time feedback helped clinicians maintain eye gaze

better. Furthermore, a majority of clinicians in our study were able to use the real-time feedback to be more aware of how much they spoke and interrupted the patient. Further, reviewing summary data about their conversational metrics after the visit helped individuals gain insight into their own behaviors, explain when behaviors are judged as good or bad, and form intentions to improve in future video visits with patients.

While prior work has identified the value of shared feedback of nonverbal behaviors in group dynamics [4] [23] [31] [43] [65], our results demonstrate the feasibility of using a system that automatically senses and provides real-time and summative feedback useful for improving clinician-patient communication. We next discuss some of the factors that influenced the participant’s receptiveness to quantifying their clinical soft skills and adopting a tool like ReflectLive in clinical practice.

6.1 Controllability of Communication Behaviors

Our results suggest that some communication behaviors are more controllable than others. Participants were able to more actively control their own eye gaze than their speaking contribution and interruptions. Dyadic behaviors (speaking contribution and interruptions) depend on the second speaker’s behavior and thus is more difficult to change by one party alone. In fact, in many cases, participants reported they needed to speak *more* to elicit answers from the patient, causing a higher speaking contribution, the opposite of what they had intended. For interruptions, some clinicians also pointed to external factors such as limited audio/video quality caused them to accidentally interrupt the patient. Not only did the quantitative behavioral data show that clinicians were less successful at moderating their speaking contribution and interruptions than eye gaze, but their explanations reinforced the relatively lower control over these dyadic behaviors and their desire to improve over time rather than instantly. Thus, when designing feedback, immediate feedback can be useful for behaviors self-dependent behaviors like eye gaze that are more easily changed, and progress in more difficult metrics including dyadic behaviors like speaking contribution are better reflected in data across multiple visits.

6.2 Designing Real-Time Feedback

To support live communication, designers must balance the need for informing the user and minimizing intrusiveness. The feedback visualizations in ReflectLive erred on the side of minimizing intrusiveness and were designed to be glanceable and even ignorable. Some participants suggested that some of these glanceable metrics could be turned into more noticeable customized alerts that do not rely on the user interpreting the data in the dashboard. For example, an alert can be set for data events of interest such as “speaking contribution is over 80% in the first 2 minutes” or “interruption count is increasing and patient turns are short” which provide actionable notifications for clinicians to improve their communication. Furthermore, the system can automatically learn each clinician’s baseline communication behaviors over time and provide personalized alerts when they deviate from that baseline.

Real-time feedback is valuable when clinicians want to self-regulate a behavior that is hard to self-monitor. However, clinicians, in their evidence-based thinking, often only realize they need to change a behavior after seeing their behaviors quantified in post-hoc summative feedback. Thus, system designers should consider showing summative feedback first to make the clinician aware of potential behaviors to change and then providing the specific real-time feedback during video visits to help them self-regulate and change that behavior. Events in summary visualizations can be linked directly back to recorded audio or video of the conversation to allow the clinician to review their own behaviors.

Designers can also employ more subtle “nudges” [63] to automatically induce non-verbal communication behavior change over a video meeting platform. For example, some clinicians suggested placing the dashboard towards the top edge of the screen close to the webcam, which would allow them to more easily glance at the dashboard without averting their eye gaze from the patient. Another clinician suggested that the system artificially narrow the edges of his own thumbnail so that the clinician would automatically centered his face in the video feed. Manipulating the speaker’s own voice feedback volume can automatically help clinicians not speak too loudly, a common habit for telehealth novices.

6.3 Expertise and Experience in Video Telehealth

The participants who found both the real-time and summary data most useful were those who had more experience in performing video telehealth visits, whom we label as “experts” (Table 1). Compared with telehealth novices, experts could more clearly express behavioral goals in communication and attributed greater value to meeting those goals to ensure the patient felt listened to. With these goals already internalized, experts found the real-time feedback and summary feedback most useful for improving their own behaviors. Experts also were noticeably more facile and at ease with speaking with patients over the video medium, indicated by affiliative language and automatically adopting an overtly sympathetic tone. Experts likely felt less cognitively burdened (than novices) and had the cognitive bandwidth to monitor the real-time feedback more closely and adjust their own non-verbal communication behaviors on the fly during the visit.

In contrast, novices were noticeably more focused on gathering clinical data from the patient rather than on controlling their non-verbal behaviors, likely because they were less familiar with the video medium. This observation is consistent with well-studied phenomena in group collaboration [17] [42] that attending to both self-feedback and the task can be difficult. When juggling the competing attentional demands of information gathering, the new video medium, and the feedback about their own self-behaviors, novices were less able to adjust their own behaviors dynamically even if they felt more aware of them. If they are unable to change their behaviors to improve their metrics, the focus on the self can be aversive, leading to reducing the dissonance through de-prioritizing rapport building in the clinical visit. Moreover, telehealth novices in our study (except P1) also were from specialties (e.g., cardiology, urology) that tend to have more task-oriented rather than rapport-building styles found in other specialties (e.g., psychiatry) [36]. As a result, novices valued the non-verbal communication skills less, did not have pre-defined goals for communication metrics, and did not have as much versatility to break their habits of looking away to take notes and interrupting the patient. This finding parallels results by He *et al.* that show how native speakers (experts) could use feedback of their non-verbal behaviors to engage in impression management better than non-native speakers [32]. Thus, critical to the success of using feedback (real-time or summative) includes having defined goals, feeling these goals are important (for good patient care), and having the skills to improve upon these metrics.

6.4 Need for Evidence-Based Norms

The metrics we investigated in our evaluation (speaking contribution, interruptions, eye gaze, and face position) were selected based on literature reviews linking these behaviors to patient satisfaction [7] [31] [41] [51]. Even though clinicians agreed that these metrics were important, they were not always convinced that incrementally improving in these metrics would lead to commensurate improvements in patient outcome, or may, even worse, be at odds with gathering relevant medical data from the patient. Many years of studies of patient-clinician communication have indeed shown that its structure, sequence, and roles are complex [34], so a limited set of metrics cannot comprehensively represent all good communication patterns. Nonetheless, clinicians are in need for guidelines or norms for non-verbal communication behaviors over video that are known to have a meaningful impact on patient satisfaction and outcomes [6]. Telehealth experts are already individually developing their own subjective best practices. Thus, we see an opportunity to codify evidence-based norms using a data-driven approach with ReflectLive’s technique that can automatically sense, extract, and aggregate data about non-verbal behaviors from clinical encounters and link that data with clinical and patient outcomes. Our future work aims to deploy ReflectLive in clinical settings with actual patients at scale to automatically quantify the impact of adjusting non-verbal behaviors in real-time on clinical and patient outcomes. Moreover, we aim to integrate our automated real-time analysis of non-verbal behavior in medical training of students and residents to assess how training can overcome the barriers posed by video technology.

7 CONCLUSION

Non-verbal behaviors are essential to effective clinician-patient communication and can be particularly challenging when using video telehealth--a medium still new to many clinicians. We introduced ReflectLive, a

system that senses and visualizes feedback about clinician’s non-verbal behaviors in real-time such as speaking contribution, interruptions, eye gaze, and face position during video consultations. Through a user evaluation with 10 clinicians, we found evidence that real-time feedback helps doctors self-regulate and maintain eye gaze with their patients while giving clinicians a greater sense of awareness of positive and negative behaviors in video consultations without being distracting. Using ReflectLive as a probe, we also identified that certain non-verbal behaviors that depend on the other speaker(s) like speaking contribution and interruptions are less easily controlled than behaviors than self-directed behaviors like eye gaze. Furthermore, clinical communication is complex, and even though metrics like balanced speaking contribution and centered eye gaze are viewed by clinicians as good behaviors to measure and improve, a limited set of metrics will not fully represent what is considered a good clinical consultation. We showed that ReflectLive has the potential to automatically quantify, analyze, and provide real-time feedback about non-verbal communication behaviors useful for training clinicians. With this tool, clinicians are empowered to be more aware of their own behaviors and also to regulate their behaviors both in skills training and in their regular practice.

REFERENCES

- [1] Hiroyuki Adachi, Seiko Myojin, and Nobutaka Shimada. 2015. ScoringTalk: a tablet system scoring and visualizing conversation for balancing of participation. In *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications*, 9.
- [2] Zia Agha, Debra L. Roter, and Ralph M. Schapira. 2009. An Evaluation of Patient-Physician Communication Style During Telemedicine Consultations. *Journal of Medical Internet Research* 11, 3: e36. <https://doi.org/10.2196/jmir.1193>
- [3] Mara S. Aruguete and Carlos A. Roberts. 2000. Gender, Affiliation, and Control in Physician–Patient Encounters. *Sex Roles* 42, 1–2: 107–118. <https://doi.org/10.1023/A:1007036412480>
- [4] Madeline Balaam, Geraldine Fitzpatrick, Judith Good, and Eric Harris. 2011. Enhancing Interactional Synchrony with an Ambient Display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’11)*, 867–876. <https://doi.org/10.1145/1978942.1979070>
- [5] Rashid L. Bashshur. 1997. Critical issues in telemedicine. *Telemedicine Journal* 3, 2: 113–126.
- [6] Rashid L. Bashshur, Jay H. Sanders, and Gary W. Shannon. 1997. *Telemedicine: Theory and practice*. Charles C. Thomas Springfield, IL.
- [7] Rainer S. Beck, Rebecca Daughtridge, and Philip D. Sloane. 2002. Physician-patient communication in the primary care office: a systematic review. *The Journal of the American Board of Family Practice* 15, 1: 25–38.
- [8] Jozien M. Bensing, Jan J. Kerssens, and Marja van der Pasch. 1995. Patient-directed gaze as a tool for discovering and handling psychosocial problems in general practice. *Journal of nonverbal behavior* 19, 4: 223–242.
- [9] Tony Bergstrom and Karrie Karahalios. 2007. Conversation Clock: Visualizing audio patterns in co-located groups. In *Proceedings of the 40th Annual Hawaii International Conference on Social Systems (HICSS 2007)*, 78–78.
- [10] J. Beullens, J. J. Rethans, Jo Goedhuys, and Frank Buntinx. 1997. The use of standardized patients in research in general practice. *Family Practice* 14, 1: 58–62.
- [11] LouAnne E. Boyd, Alejandro Rangel, Helen Tomimbang, Andrea Conejo-Toledo, Kanika Patel, Monica Tentori, and Gillian R. Hayes. 2016. SayWAT: Augmenting Face-to-Face Conversations for Adults with Autism. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI ’16)*, 4872–4883. <https://doi.org/10.1145/2858036.2858215>
- [12] Jed R. Brubaker, Gina Venolia, and John C. Tang. 2012. Focusing on shared experiences: moving beyond the camera in video communication. In *Proceedings of the Designing Interactive Systems Conference*, 96–105.
- [13] Patrick S. Byrne and B. E. Long. 1976. Doctors talking to patients. A study of the verbal behaviour of general practitioners consulting in their surgeries. Exeter [England]: Royal College of General Practitioners, 1984, c1976.
- [14] Byungki Byun, Anurag Awasthi, Philip A. Chou, Ashish Kapoor, Bongshin Lee, and Mary Czerwinski. 2011. Honest signals in video conferencing. In *Proceedings of 2011 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.
- [15] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI ’16)*, 981–992. <https://doi.org/10.1145/2858036.2858498>
- [16] Scott Carter, Laurent Denoue, and Matthew Cooper. 2015. Searching and Browsing Live, Web-based Meetings. In *Proceedings of the 23rd ACM International Conference on Multimedia (MM ’15)*, 791–792. <https://doi.org/10.1145/2733373.2807964>
- [17] Charles S. Carver and Michael F. Scheier. 2012. *Attention and self-regulation: A control-theory approach to human behavior*. Springer Science & Business Media.
- [18] Kathryn Charmaz. 2006. *Constructing Grounded Theory: A Practical Guide to Qualitative Analysis*. Sage Publications Ltd.
- [19] Milton Chen. 2002. Leveraging the Asymmetric Sensitivity of Eye Contact for Videoconference. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’02)*, 49–56. <https://doi.org/10.1145/503376.503386>
- [20] Yunan Chen, Victor Ngo, Sidney Harrison, and Victoria Duong. 2011. Unpacking exam-room computing: negotiating computer-use in patient-physician interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3343–3352.
- [21] Ionut Damian, Chiew Seng (Sean) Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. 2015. Augmenting Social Interactions: Realtime Behavioural Feedback Using Social Signal Processing Techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI ’15)*, 565–574. <https://doi.org/10.1145/2702123.2702314>
- [22] Epstein Becker Green. 2016. 50-State Survey of Telemental/Telebehavioral Health.
- [23] Joan Morris DiMicco, Katherine J. Hollenbach, Anna Pandolfo, and Walter Bender. 2007. The impact of increased awareness while face-to-face. *Human–Computer Interaction* 22, 1–2: 47–96.
- [24] Paul Dourish, Annette Adler, Victoria Bellotti, and Austin Henderson. 1996. Your place or mine? Learning from long-term use of audio-video communication. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (CSCW ’96)*, 1: 33–62.
- [25] F. Daniel Duffy, Geoffrey H. Gordon, Gerald Whelan, Kathy Cole-Kelly, Richard Frankel, and others. 2004. Assessing competence in

- communication and interpersonal skills: the Kalamazoo II report. *Academic Medicine* 79, 6: 495–507.
- [26] Susan T. Fiske. 1993. Controlling other people: The impact of power on stereotyping. *American Psychologist* 48, 6: 621–628. <https://doi.org/10.1037/0003-066X.48.6.621>
 - [27] Howard S. Friedman. 1979. Nonverbal communication between patients and medical practitioners. *Journal of Social Issues* 35, 1: 82–99.
 - [28] Chris Fullwood and Gwyneth Doherty-Sneddon. 2006. Effect of gazing at the camera during a video link on recall. *Applied Ergonomics* 37, 2: 167–175. <https://doi.org/10.1016/j.apergo.2005.05.003>
 - [29] Dominik Giger, Jean-Charles Bazin, Claudia Kuster, Tiberiu Popa, and Markus Gross. 2014. Gaze correction with a single webcam. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, 1–6.
 - [30] Dehra Glueck, K. Myers, and C. L. Turvey. 2013. Establishing therapeutic rapport in telemental health. *Telemental health: Clinical, technical and administrative foundations for evidence-based practice*: 29–46.
 - [31] Andrea L. Hartzler, Rupa A. Patel, Mary Czerwinski, Wanda Pratt, Asta Roseway, N. Chandrasekaran, and Anthony Back. 2014. Real-time feedback on nonverbal clinical communication. *Methods of Information in Medicine* 53, 5: 389–405. <https://doi.org/10.3414/ME13-02-0033>
 - [32] Helen Ai He, Naomi Yamashita, Ari Hautasaari, Xun Cao, and Elaine M. Huang. 2017. Why Did They Do That?: Exploring Attribution Mismatches Between Native and Non-Native Speakers Using Videoconferencing. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*, 297–309. <https://doi.org/10.1145/2998181.2998205>
 - [33] Stephen G. Henry, Andrea Fuhrer-Forbis, Mary A. M. Rogers, and Susan Eggly. 2012. Association between nonverbal communication during clinical interactions and outcomes: A systematic review and meta-analysis. *Patient Education and Counseling* 86, 3: 297–315. <https://doi.org/10.1016/j.pec.2011.07.006>
 - [34] John Heritage and Douglas W. Maynard. 2006. *Communication in medical care: Interaction between primary care physicians and patients*. Cambridge University Press.
 - [35] Clara E. Hill and James Gormally. 1977. Effects of reflection, restatement, probe, and nonverbal behaviors on client affect. *Journal of Counseling Psychology* 24, 2: 92.
 - [36] Mohammadreza Hojat, Joseph S. Gonnella, Thomas J. Nasca, Salvatore Mangione, Michael Vergare, and Michael Magee. 2002. Physician empathy: definition, components, measurement, and relationship to gender and specialty. *American Journal of Psychiatry* 159, 9: 1563–1569.
 - [37] Ellen A. Isaacs and John C. Tang. 1994. What video can and cannot do for collaboration: a case study. *Multimedia Systems* 2, 2: 63–73.
 - [38] Matthew Jackson, Anne H. Anderson, Rachel McEwan, and Jim Mullin. 2000. Impact of Video Frame Rate on Communicative Behaviour in Two and Four Party Groups. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00)*, 11–20. <https://doi.org/10.1145/358916.358945>
 - [39] Donald J. Kiesler and Stephen M. Auerbach. 2003. Integrating measurement of control and affiliation in studies of physician-patient interaction: the interpersonal circumplex. *Social Science & Medicine* (1982) 57, 9: 1707–1722.
 - [40] Taemie Kim, Agnes Chang, Lindsey Holland, and Alex (Sandy) Pentland. 2008. Meeting Mediator: Enhancing Group Collaboration with Sociometric Feedback. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*, 3183–3188. <https://doi.org/10.1145/1358628.1358828>
 - [41] Ann King and Ruth B. Hoppe. 2013. “Best practice” for patient-centered communication: a narrative review. *Journal of Graduate Medical Education* 5, 3: 385–393.
 - [42] Stuart T. Klapp, Martin D. Hill, John G. Tyler, Zeke E. Martin, Richard J. Jagacinski, and Mari Riess Jones. 1985. On marching to two different drummers: perceptual aspects of the difficulties. *Journal of Experimental Psychology: Human Perception and Performance* 11, 6: 814.
 - [43] Olga Kulyk, Jimmy Wang, and Jacques Terken. 2005. Real-time feedback on nonverbal behaviour to enhance social dynamics in small group meetings. In *International Workshop on Machine Learning for Multimodal Interaction*, 150–161.
 - [44] Claudia Kuster, Tiberiu Popa, Jean-Charles Bazin, Craig Gotsman, and Markus Gross. 2012. Gaze correction for home video conferencing. *ACM Transactions on Graphics (TOG)* 31, 6: 174.
 - [45] Wendy Levinson, Debra L. Roter, John P. Mullooly, Valerie T. Dull, and Richard M. Frankel. 1997. Physician-patient communication. The relationship with malpractice claims among primary care physicians and surgeons. *Journal of the American Medical Association* 277, 7: 553–559.
 - [46] Chunfeng Liu, Renee L. Lim, Kathryn L. McCabe, Silas Taylor, and Rafael A. Calvo. 2016. A Web-Based Telehealth Training Platform Incorporating Automated Nonverbal Behavior Feedback for Teaching Communication Skills to Medical Students: A Randomized Crossover Study. *Journal of Medical Internet Research* 18, 9.
 - [47] Xiao Liu, Yoshie Sawada, Takako Takizawa, Hiroko Sato, Mahito Sato, Hironosuke Sakamoto, Toshihiro Utsugi, Kunio Sato, Hiroyuki Sumino, Shinichi Okamura, and others. 2007. Doctor-Patient Communication: A Comparison between Telemedicine Consultation and Face-to-Face Consultation. *Internal Medicine* 46, 5: 227–232.
 - [48] Marcial Losada, Pedro Sanchez, and Elizabeth E. Noble. 1990. Collaborative Technology and Group Process Feedback: Their Impact on Interactive Sequences in Meetings. In *Proceedings of the 1990 ACM Conference on Computer-supported Cooperative Work (CSCW '90)*, 53–64. <https://doi.org/10.1145/99332.99341>
 - [49] Ludmila Marciniowicz, Jerzy Konstantynowicz, and Cezary Godlewski. 2010. Patients’ perceptions of GP non-verbal communication: a qualitative study. *British Journal General Practice* 60, 571: 83–87.
 - [50] Jennifer Marlow, Eveline van Everdingen, and Daniel Avrahami. 2016. Taking Notes or Playing Games?: Understanding Multitasking in Video Communication. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*, 1726–1737. <https://doi.org/10.1145/2818048.2819975>
 - [51] Marianne Schmid Mast. 2007. On the importance of nonverbal communication in the physician–patient interaction. *Patient Education and Counseling* 67, 3: 315–318.
 - [52] Paul McLaren, Chris J. Ball, A. B. Summerfield, M. J. P. Watson, and M. Lipsedge. 1995. An evaluation of the use of interactive television in an acute psychiatric service. *Journal of Telemedicine and Telecare* 1, 2: 79–85.
 - [53] Edward Alan Miller. 2003. The technical and interpersonal aspects of telemedicine: effects on doctor-patient communication. *Journal of Telemedicine and Telecare* 9, 1: 1–7. <https://doi.org/10.1258/135763303321159611>
 - [54] Martin Montgomery. 2013. *An introduction to language and society*. Routledge.
 - [55] L. M. Ong, J. C. de Haes, A. M. Hoos, and F. B. Lammes. 1995. Doctor-patient communication: a review of the literature. *Social Science & Medicine* (1982) 40, 7: 903.
 - [56] Michael Quinn Patton. 1990. *Qualitative evaluation and research methods*. SAGE Publications, Inc.

- [57] Alex Pentland and Tracy Heibeck. 2010. *Honest signals: how they shape our world*. MIT press.
- [58] Debra L. Roter, Richard M. Frankel, Judith A. Hall, and David Sluyter. 2006. The expression of emotion through nonverbal behavior in medical visits. *Journal of General Internal Medicine* 21, S1: S28–S34.
- [59] Debra Roter and Susan Larson. 2002. The Roter interaction analysis system (RIAS): utility and flexibility for analysis of medical interactions. *Patient Education and Counseling* 46, 4: 243–251.
- [60] Allison Sauppé and Bilge Mutlu. 2014. How Social Cues Shape Task Coordination and Communication. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*, 97–108. <https://doi.org/10.1145/2531602.2531610>
- [61] Abigail J. Sellen. 1992. Speech patterns in video-mediated conversations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 49–59.
- [62] Richard L. Street Jr, E. Jay Wheeler, and William T. McCaughan. 2000. Specialist–primary care provider–patient communication in telemedical consultations. *Telemedicine Journal* 6, 1: 45–54.
- [63] Richard Thaler and Cass Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.
- [64] Donald T. Tepper and Richard F. Haase. 1978. Verbal and nonverbal communication of facilitative conditions. *Journal of Counseling Psychology* 25, 1: 35.
- [65] Jacques Terken and Janienke Sturm. 2010. Multimodal support for social dynamics in co-located meetings. *Personal and Ubiquitous Computing* 14, 8: 703–714.
- [66] American Telehealth Association. 2009. *Evidence-Based Practice for Telemental Health*. <http://dev.americantelemed.org/docs/default-source/standards/evidence-based-practice-for-telemental-health.pdf>
- [67] American Telehealth Association. 2017. *Executive Leadership Survey*. <http://thesource.americantelemed.org/resources/telemedicine-executive-leadership-survey>
- [68] OSHA. https://www.osha.gov/SLTC/etools/computerworkstations/components_monitors.html
- [69] VisageSDK. <http://visagetechnologies.com/products-and-services/visagesdk/>
- [70] Hark.js. <https://github.com/otalk/hark/>

Received June 2017; revised July 2017; accepted November 2017