Improving Medical Students' Awareness of Their Non-Verbal Communication through Automated Non-Verbal Behavior Feedback

Chunfeng Liu[1], Rafael A. Calvo[1]* and Renee Lim[1,2]

[1] School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW, Australia, [2] Sydney Medical School, The University of Sydney, Sydney, NSW, Australia

The non-verbal communication of clinicians has an impact on patients' satisfaction and health outcomes. Yet medical students are not receiving enough training on the appropriate non-verbal behaviors in clinical consultations. Computer vision techniques have been used for detecting different kinds of non-verbal behaviors, and they can be incorporated in educational systems that help medical students to develop communication skills. We describe EQClinic, a system that combines a tele-health platform with automated non-verbal behavior recognition. The system aims to help medical students improve their communication skills through a combination of human and automatically generated feedback. EQClinic provides fully automated calendaring and video conferencing features for doctors or medical students to interview patients. We describe a pilot (18 dyadic interactions) in which standardized patients (SPs) (i.e., someone acting as a real patient) were interviewed by medical students and provided assessments and comments about their performance. After the interview, computer vision and audio processing algorithms were used to recognize students' non-verbal behaviors known to influence the quality of a medical consultation: including turn taking, speaking ratio, sound volume, sound pitch, smiling, frowning, head leaning, head tilting, nodding, shaking, face-touch gestures and overall body movements. The results showed that students' awareness of non-verbal communication was enhanced by the feedback information, which was both provided by the SPs and generated by the machines.

Keywords: non-verbal communication, non-verbal behavior, clinical consultation, medical education, communication skills, non-verbal behavior detection, automated feedback

INTRODUCTION

Over the last 10 years, we have witnessed a dramatic improvement in affective computing (Picard, 2000; Calvo et al., 2015) and behavior recognition techniques (Vinciarelli et al., 2012). These techniques have progressed from the recognition of person-specific posed behavior to the more difficult person-independent recognition of behavior in "the-wild" (Vinciarelli et al., 2009). They are considered robust enough that they are being incorporated into new applications. For example, new learning technologies have been developed that detect a student's emotions and use this to guide the learning experience (Calvo and D'Mello, 2011). They can also be used to support reflection by

generating visualizations and feedback that a learner can interpret and use to improve their skills. This approach has been used, for example, in helping students improve their written communication skills (Calvo, 2015).

Similar techniques could, but have rarely been used to support the development of professional communication skills. The development of such skills is an essential part of professional development in areas such as medicine, since it is known that the quality of the clinical interviews and effective patient–doctor communication can lead to better health outcomes (Stewart, 1995).

Existing research in clinical communication skill teaching mainly focuses on verbal communication (Mast et al., 2008). However, in the past three decades, non-verbal communication, which accounts for approximately 80% of essential communication between individuals (Gorawara-Bhat et al., 2007), has been attracting increasing attention from the medical education community. Specific non-verbal behaviors, including facial expression and body movement, have been proven to be related to patient satisfaction (DiMatteo et al., 1980) and health outcomes (Hartzler et al., 2014). Within medical education, the performance of certain non-verbal behaviors by medical students, such as "maintaining adequate facial expressions, using affirmative gestures and limiting both unpurposive movements and hand gestures, had a significant positive effect on perception of interview quality during (an) Objective Structure Clinical Examination" (Collins et al., 2011). Ishikawa et al. (2006) also mentioned that students who nodded when listening to standardized patients' (SPs) talk and spoke at a similar speed and voice volume to the SPs achieved higher ratings on their performance in an OSCE. However, manually annotating the non-verbal behavior in these studies is a time consuming process and prevents them from being widely incorporated into the teaching curriculum, even though some annotation tools, such as Roter Interaction Analysis System (Roter and Larson, 2002), have been developed to ease the work of annotators.

Non-verbal behavior includes every communicative human act other than speech and it can be generally classified into four categories: (1) kinesics: head and body movements, such as facial expressions and gestures, (2) vocalics: non-linguistic vocal cues, such as volume and sound pitch, (3) haptics: body contact, such as handshakes, (4) proxemics: spatial cues, such as doctor–patient distance and body orientation (Hartzler et al., 2014; Burgoon et al., 2016). At this time, there is no consensus on which category of non-verbal behavior is most important in clinical consultations.

Kinesics behavior includes the movements of the face, head, trunk, hands, and limbs (Hartzler et al., 2014). Face and head movements are among the most salient non-verbal behaviors in patient–doctor consultations. Smiling often relates to rapport (Duggan and Parrott, 2001). Frequent nodding signals affiliation and head shaking indicates authority (Burgoon, et al., 2016). A doctor leaning their body forward toward the patient often expresses willingness of involvement (Martin and Friedman, 2005), and backward leaning can decrease patient satisfaction (Larsen and Smith, 1981). Doctors' hand gestures have been related to "warmth and expressive likability" (Harrigan et al., 1985).

Experienced clinicians use variant volume, sound pitch, speaking rate, and silence time to shape the conversation. Buller and Street, (1992) mentioned that louder speech and lower pitch is associated with the rapport. Silence is another powerful tool for some clinicians. They use different lengths of silence to control the progress of the consultation (Gibbings-Isaac et al., 2012). Normally, increased speaking time is a sign of dominance (Burgoon et al., 2016).

Touch, such as a pat on the patient's shoulder, is a communication gesture used by clinicians to express support and sympathy (Hartzler et al., 2014). In tele-conferencing, physical contact is not possible between clinicians and patients. However, clinicians successfully, albeit unconsciously, use self-touch to build rapport and to indicate that they are paying attention to the conversation (Harrigan et al., 1985).

In general, close conversational distance indicates an intimate consultation (Hartzler et al., 2014). Some studies shown that patients' satisfaction was greater if the doctor lean toward them (Larsen and Smith, 1981) or sit in a closer distance (Weinberger et al., 1981) in clinical consultations. However, in tele-consultations, measuring the distance between the patient and the doctor is difficult. Instead, the distance between the camera and the participant can be represented by the size of face. For example, the distance increases when the face size decreases. It is still an open question if such behaviors contribute to a sense of intimacy in tele-consultations.

With the fast developments of technology, there have been some systems developed to recognize non-verbal behavior using computer algorithms. For example, Pentland and Heibeck (2010) used a digital sensor to identify non-verbal behavior (proximity and vocalics) and classify these behaviors into one of four "honest signals": activity, consistency, influence, and mimicry.

The ROC Speak framework was developed for sensing non-verbal behavior during public speaking and providing feedback for the user (Fung et al., 2015). The framework detects multiples types of non-verbal behavior including smile intensity, body movement, volume, sound pitch, and word prosody. Body gestures were also detected by other systems for helping public speakers. AutoManner used Microsoft Kinect to detect the body gestures of speakers, and automatically generate feedback based on the patterns of their body gestures (Tanveer et al., 2016).

My Automated Conversation coacH (MACH) (Hoque et al., 2013) is a social skills training platform that allows users to communicate with a virtual actor. By automatically sensing the non-verbal behavior of the users, such as facial expressions, head movements, etc., the system can analyze these behavior and provide feedback for the user to reflect on.

However, very few studies have applied these novel technologies into educational applications that for example, improve the communication skills of medical students or other medical professionals. Hartzler et al. (2014) made some efforts to provide medical professionals with real-time non-verbal behavior feedback. They did not implement automatic systems to identify non-verbal behaviors. Instead, they conducted a "Wizard-of-Oz" (Kelley, 1984) study to implicate that real-time, non-verbal behavior feedback facilitated the empathy of doctors.

The way each of the above of non-verbal behavior are used, and how effective they are at improving patients satisfaction and health outcomes, depends on the type of doctor–patient interview. Medical students learn to perform for "Taking a medical history," "Breaking bad news," and "Informed consent" interactions, etc. (Dent and Harden, 2013). The appropriate type and amount of each category of non-verbal behavior is therefore dependent on the scenario and even the personal style of the doctor. This makes it challenging, if not impossible, to provide summative feedback (i.e., assessment) using automated techniques. A better way to use these techniques is in reflective feedback.

We contribute a new method of identifying non-verbal behaviors in clinical consultations using behavior recognition techniques in this paper. Our proposed platform, EQClinic, is an e-learning platform, which allows medical students to have recorded tele-consultations with SPs. EQClinic was designed to (1) automatically identify medical students' non-verbal behaviors; (2) promote students' reflection around the different categories of non-verbal behavior; and (3) improve students' communication skills.

In this paper, we address the following research questions (RQ) through a user study:

RQ1: Is students' communication skills improved through using the EQClinic platform? If so, which learning aspects have been enhanced?

RQ2: Is the EQClinic platform acceptable to medical students?

In the user study, participants (medical students and SPs) were required to complete tele-consultations on EQClinic platform. After the consultation, the system automatically identified students' non-verbal behaviors, such as pitch, facial expressions, and body movements, using audio processing and computer vision techniques. This was then provided as graphical representations for students. In addition, the platform allowed the SP to assess and comment on the performance of the student. The assessment results and comments were provided to the student for reflection.

The paper is organized as follows: The Section "Materials and Methods" describes the main components of EQClinic and the algorithms for detecting non-verbal behaviors. In Section "Results," we evaluate how EQClinic can help medical students, through a pilot user study. The Section "Discussion" describes the results and conclusions, together with the limitations of the EQClinic evaluation (and ways to address them in future studies).

MATERIALS AND METHODS

According to social cognitive theory, students acquire competence through practice and feedback (Mann, 2011); therefore, to develop clinical communication skills, medical students need to practice with real or SPs and receive feedback from patients and tutors. Kolb defines "Learning is the process whereby knowledge is created through the transformation of experience" (Kolb, 1984, p. 38). His experiential learning cycle theory defined learning as an integrated process with a cycle of four stages: concrete experience, reflective observation, abstract conceptualization, and active experimentation (Kolb, 1984). In the concrete experience

stage, students should have a real experience such as attending a workshop. Then students should reflect on this experience by keeping a log in the reflective observation stage. Third, students should be provided some abstract concepts through research and analysis. Lastly, student can use the concepts in new situations and generate some new experiences (McLeod, 2013).

The design of EQClinic is based on Kolb's learning cycle and attempts to encompass the first two stages. In order to allow medical students to achieve concrete experience of having clinical consultations, EQClinic provides students with concrete experiences through opportunities to communicate with SPs in a tele-conference. After the tele-consultations, multiple types of feedback including video recordings, SPs' assessments and comments and automated non-verbal behavior feedback facilitate the reflective process. The goal of the project is to develop a platform where students learn about their non-verbal communication from the reflective feedback, and implement this in their future interviews.

In designing each specific stage, we considered cognitive load theory (Young et al., 2014; Leppink et al., 2015). The platform provides moderate fidelity, low complexity, and minimal instructional support to students to maximize learning at their stage of study. As EQClinic is a student learning platform and we want students to receive multiple types of feedback, the SP's experience of the platform is more complex – with moderate to high complexity, moderate fidelity, and minimal instructional support. While this may decrease the user experience of the SP, we believe it more than proportionately increases the student learning experience.

EQClinic

EQClinic (Liu et al., 2015) is a web application developed by the Positive Computing lab at the University of Sydney Australia in collaboration with medical schools at the University of Sydney and UNSW Australia. **Figure 1** illustrates the four components of EQClinic, which are a personal calendar, a real-time interaction component, a non-verbal behavior detector and a feedback generator, and their interactions. In the following sections, we separately describe these components.

Personal Calendar

The time and human resources needed to organize large numbers of practice consultation sessions is a challenge for many medical schools. EQClinic solves this problem by providing students and SPs with an automated personal calendar system to book consultations. SPs can offer their available time slots on the calendar to allow students to make a booking. EQClinic uses different colors to label the changing status of the appointments, so that students can easily request and check the status of any available time slots. Email and SMS notifications are sent from the system when users request or confirm appointments.

Real-Time Interaction Component

After the appointment has been confirmed, students and SPs can have tele-consultations through the real-time interaction component. This component includes the video conferencing component and online assessments.

FIGURE 1 | System architecture of EQClinic.

FIGURE 2 | SP's video conferencing page.
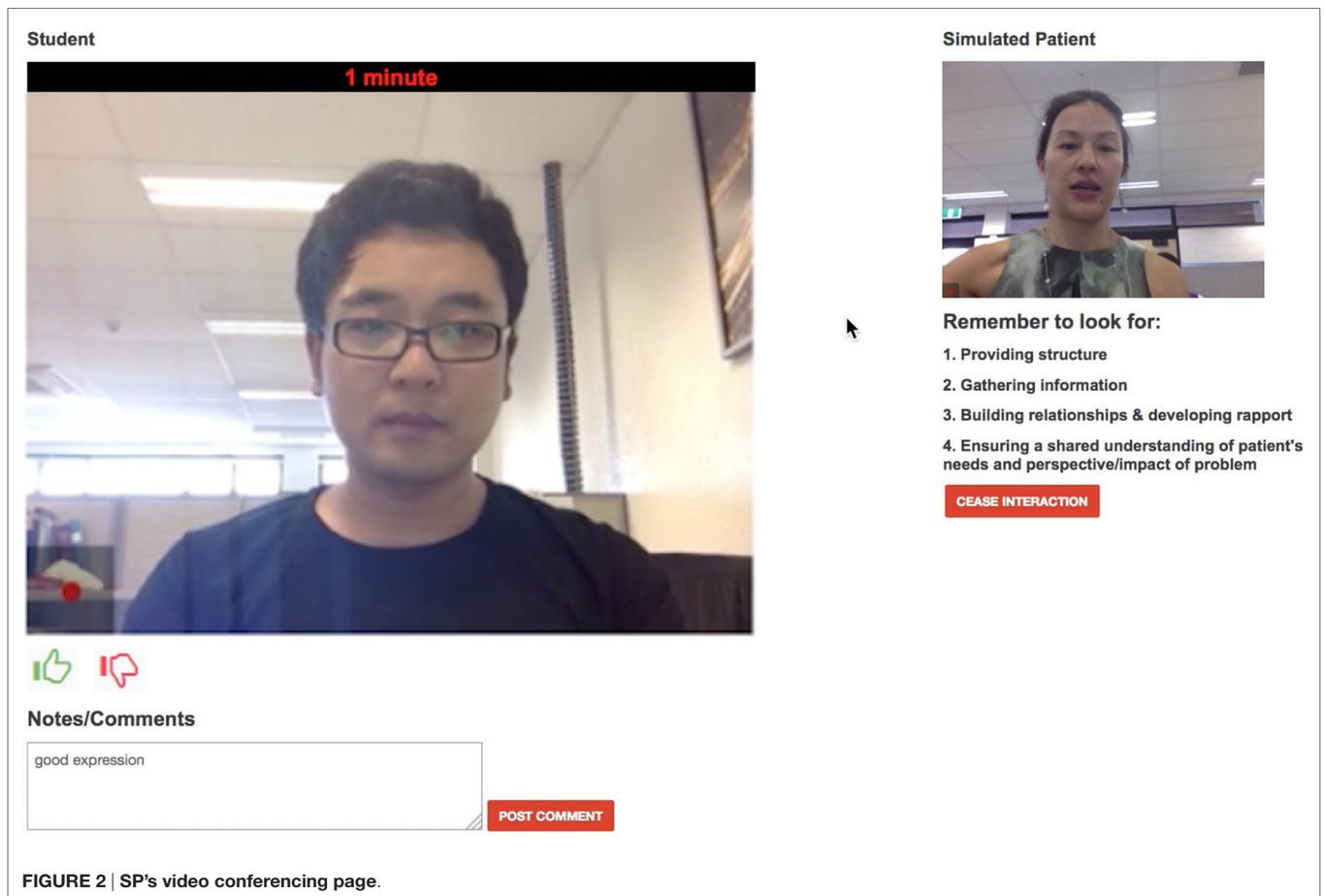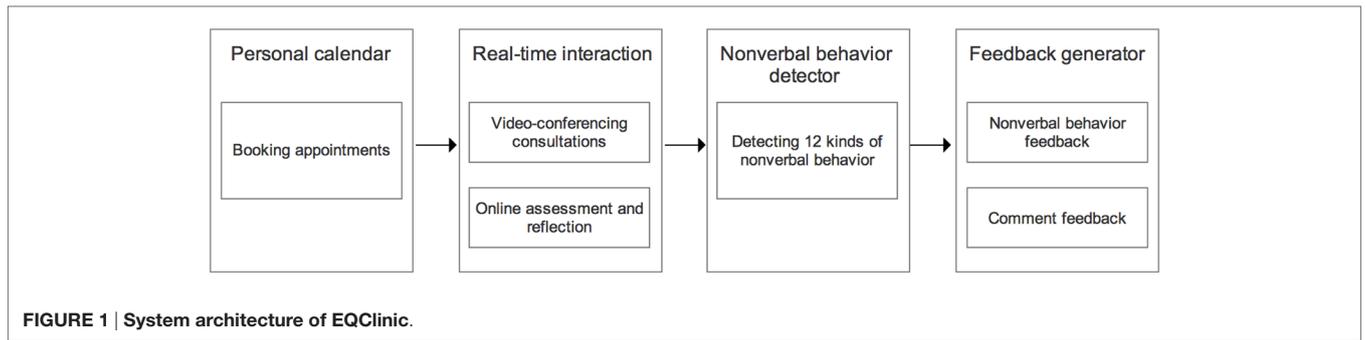
The video conferencing component works on most web browsers of a PC or an Android tablet. Once both participants are connected, the system automatically records the consultation. This component uses OpenTok, a Web Real-Time Communication (WebRTC) service, to provide real-time video communication. In traditional face-to-face clinical communication skills training, SPs provide overall assessments and comments on the student's performance after the consultations. However, using this method, it is difficult to track SPs' thoughts during the consultation. EQClinic attempts to solve this problem by providing SPs with two tools: a thumbs tool, which provides a simple indication of positive (thumbs up) and negative (thumbs down) moments in the consultation, and a comments box (see bottom left hand corner of **Figure 2**). Both forms of feedback are stored with timestamps and can be seen by the students when they review the recording. A timer is also included on the video conferencing page to indicate the duration of the interview. The students' video conferencing page is similar to the SPs' except that the two tools, and the SP video image enlarges to full screen to increase the visual engagement of the interview, and decrease distractions.

In order to facilitate the reflection process for students, we developed the online assessment interactions, where SPs evaluate the performance of the students through an assessment form immediately after the tele-consultation, and the students also assess themselves, and reflect on the interview. When the SP finishes the assessment form, this result is reviewed by the student immediately, and the student reflects on the assessment. The reflection procedure is a compulsory step for students.

Non-Verbal Behavior Detector

When the tele-consultation is completed, EQClinic will automatically analyze the video recordings using audio processing and computer vision techniques. The system mainly detects three categories of students' non-verbal behavior: vocalics, kinesics, and haptics behaviors. Vocalics behavior includes: voice properties (volume and sound pitch) and speech patterns (turn taking and speaking ratio changes); kinesics behavior includes: head movements (nodding, head shaking and head tilting), facial expressions (smiling and frowning), and body movements [body leaning and overall body movements (OBM)]; haptics behavior includes face-touch gestures. Each 15-min video recording takes 35 min to be analyzed on a personal computer with 3.40 GHz CPU and 16 gb RAM. The details of the algorithms that are used in EQClinic for detecting these behaviors will be introduced in Non-Verbal Behavior Detection.

Feedback Generator

The feedback generator is responsible for generating different kinds of feedback for students to reflect on. The feedback information includes: computer-generated non-verbal behavior feedback and comment feedback from the SP.

EQClinic visualizes students' non-verbal behavior using two types of feedback reports: single-feature and combined-feature. The single-feature feedback report illustrates each form of non-verbal behavior separately. **Figure 3** is an example of a single-feature feedback report that describes the speaking ratio of the student. Comments (the "C" labels) from the SP are also shown on the graphs. On clicking a point on the graph, the video moves to that particular timestamp. From this report, students can easily observe the variations of a particular kind of non-verbal behavior during a consultation.

The single-feature feedback report helps students to focus on one aspect of non-verbal behavior, but interactions between different types of non-verbal behavior are also useful. For example, from the single-feature feedback report, it is difficult to determine whether a student often shakes his head while smiling. Thus, EQClinic also provides a combined-feature feedback report that displays multiple kinds of non-verbal behavior on one page (turn-taking and smile intensity in **Figure 4**). Students are able to combine different types of non-verbal behavior according to their needs.

The comment feedback provides students a report that contains all the comments from the SP and tutor. As shown in **Figure 5**, the system displays the comments along the timeline. Students can easily review the video associated with the comments. By clicking on a comment, the video will jump to that particular timestamp.

Non-Verbal Behavior Detection

We selected 12 types of non-verbal behaviors (mentioned in Accuracy) to detect in EQClinic. This decision was based on a combination of knowledge regarding non-verbal behaviors in previous clinical consultation literature and the capability of current technology. In the following sections, we describe the details of the algorithms we adopted for detecting those non-verbal behaviors.

Head Movements

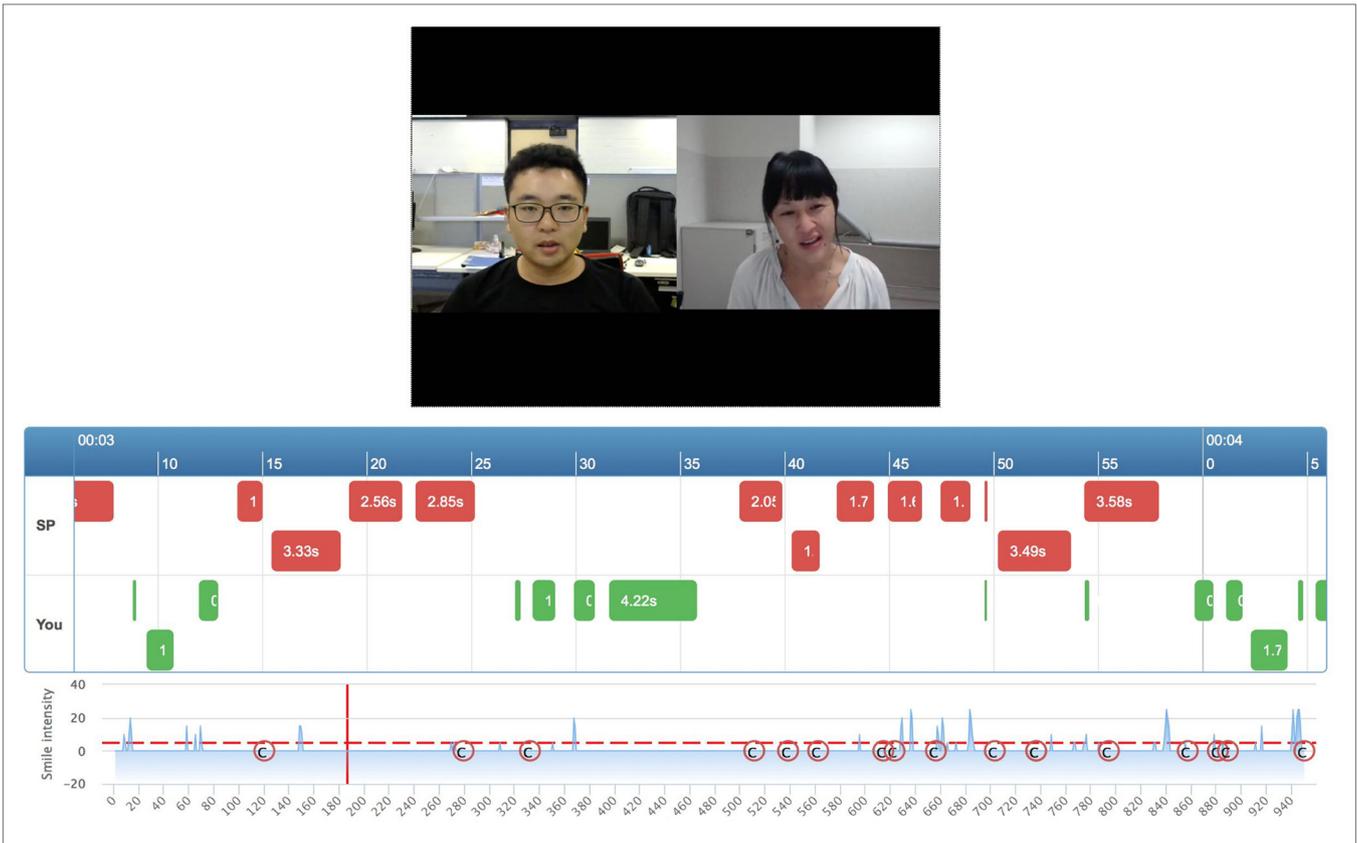EQClinic detects three kinds of head movements: nodding, head shaking, and head tilting.

FIGURE 3 | Single-feature non-verbal behavior feedback.
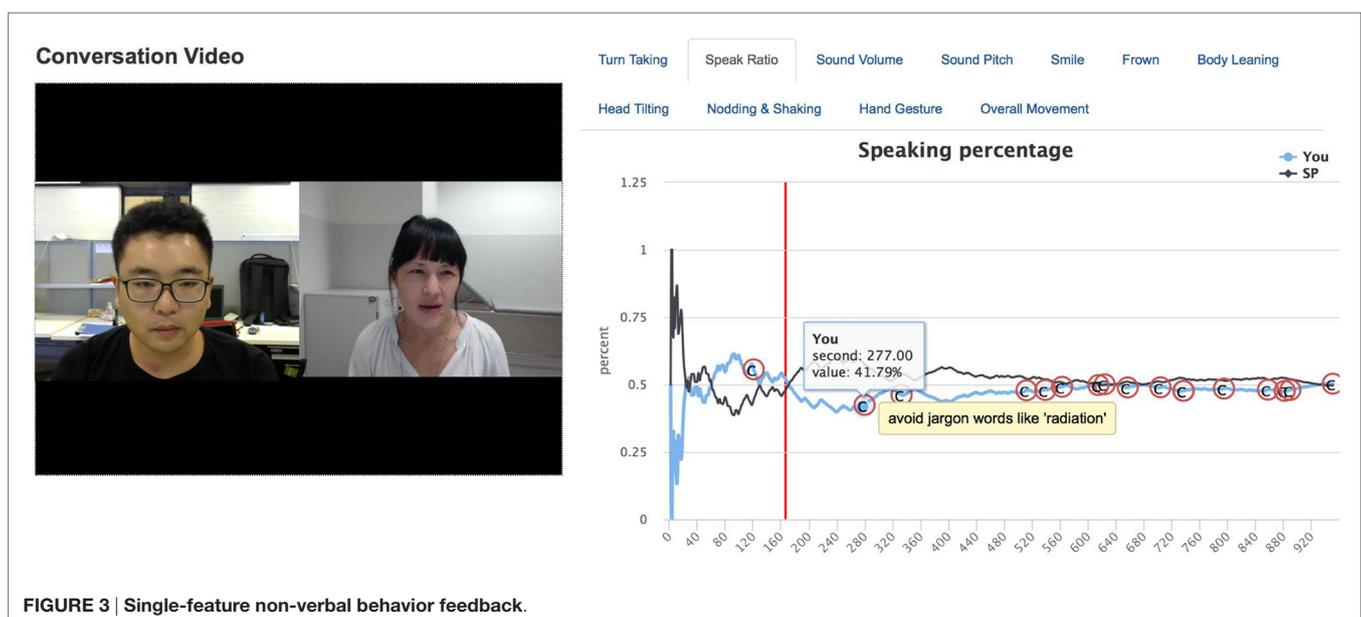
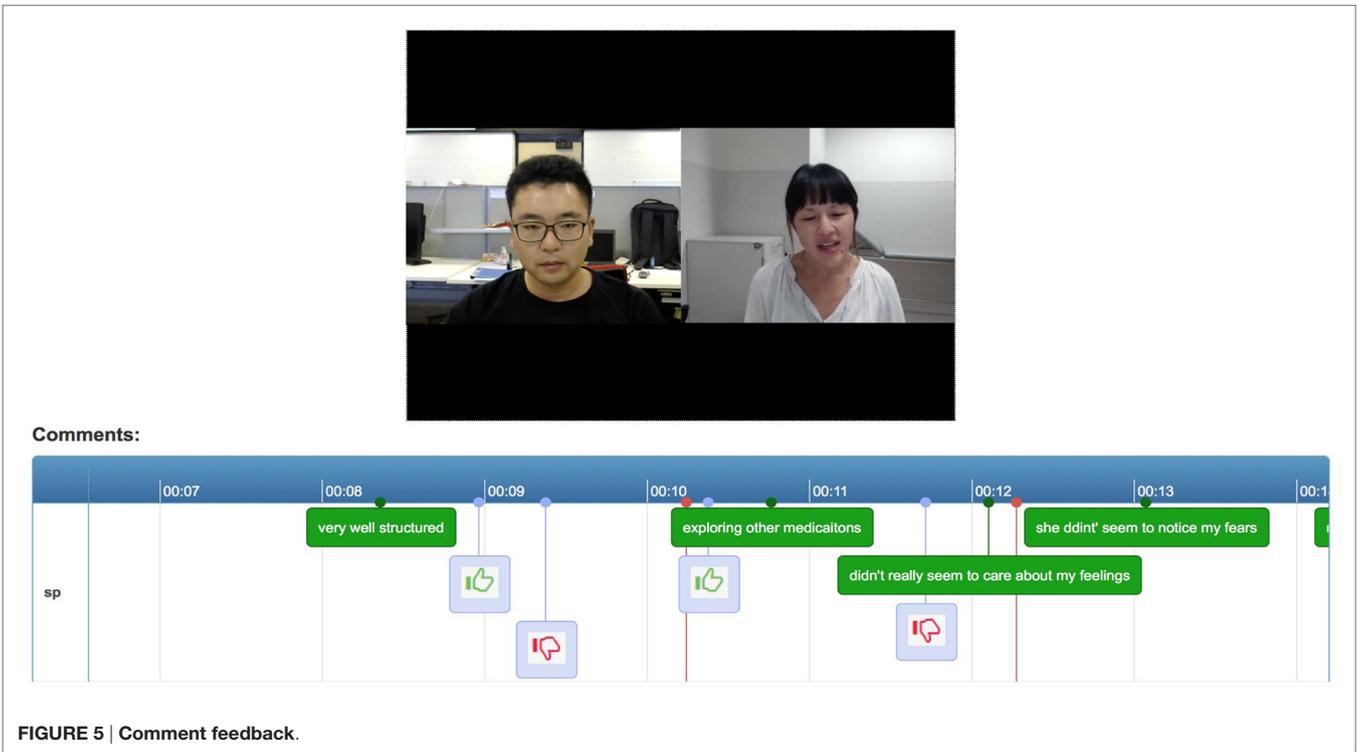**FIGURE 4** | Combined-feature non-verbal behavior feedback.

FIGURE 5 | Comment feedback.

The fundamental procedure of detecting head movements is identifying the landmarks of a face. ASMLibrary (Wei, 2009), which is developed based on the ASM method (Milborrow and Nicolls, 2008) and able to locate 68 facial landmarks, is adopted by EQClinic. In order to evaluate the performance of ASMLibrary, BIOID face dataset (BioID, 2016) is selected to test. BIOID face dataset contains 1521 frontal view facial gray images from 23 people, and the resolution of the image is 384×286 pixels. In addition, the researchers of BIOID dataset manually identified 20 landmarks, including right eye pupil, left eye pupil, etc., in each image. These manual makeups were stored in text files and included in BIOID dataset. We ran ASMLibrary to locate five landmarks of each image in BIOID dataset, and the five landmarks include inner end of left eyebrow, inner end of right eyebrow, inner corner of left eye, and inner corner of right eye and tip of nose. Then, we compared the locations of these landmarks, which are located by ASMLibrary with those are provided by the BIOID dataset.

As a result, ASMLibrary successfully detected 1462 faces from 1521 images with 96.12% detection rate. **Table 1** lists the mean errors on X and Y axes and shows that the performance of ASMLibrary is acceptable at identifying the landmarks on a face.

EQClinic senses nodding and head shaking by tracking the point (mid-point) between the two inner corners of the eyes (Kawato and Ohya, 2000). Using this method, every frame of the video recording can be classified into one of three states: stable state, extreme state, and transient state. Head nodding and head shaking are very similar movements for this algorithm, except that nodding appears in the y_i movement, while head-shaking appears in the x_i movement. Here, x_i represents the value on the X axis in ith frame, and y_i represents the value on the Y axis. The definitions of these three states when detecting head shaking are defined by the following:

A. If $\max(x_{i+n}) - \min(x_{i+n}) \leq 2$ ($n = -2, \ldots, +2$), then the frame is in a stable state.
B. If $x_i = \min(x_{i+n})$ ($n = -2 \ldots, +2$) or $x_i = \max(x_{i+n})$ ($n = -2, \ldots, +2$), then the frame is in an extreme state.
C. Otherwise, the frame is in a transient state.

The core idea of this algorithm is that, if there are more than two extreme states between the current stable frame and the previous stable frame, and all adjacent extreme frames differ by more than 2 pixels in the x coordinate, then the algorithm assumes that a head shaking has occurred. The previous stable frame and current stable frame are the starting point and ending point of the head shaking.

After locating the five important landmarks, it is easy to calculate the tilting angle of the head. EQClinic represents this angle as the angle between the vertical line and the line contains the tip of node point and the mid-point.

Facial Expression

EQClinic detects two kinds of facial expression: smiling and frowning.

Two steps are performed for detecting smiles: training and classifying. In the training step, 100 smile images and 100 non-smile images, which were from the GENKI4K dataset (GENKI-4K, 2016) were selected as the training set. First, the detector extracted the mouth areas of training images using haar-cascade classifier provided by OpenCV (Viola and Jones, 2001), and then regulated the size of each picture as 100×70 pixels. In the classifying step, Principal Components Analysis (PCA) is the core method for classifying images (Smith, 2002). By comparing an image with the training data, the algorithm returns 1 if the image contains a smile expression, otherwise it returns 0. Every frame of the video recording will be classified using this algorithm. CK+ dataset (Lucey et al., 2010) was selected as the testing set to evaluate the accuracy of this smile detection algorithm. Sixty-nine smile images and 585 non-smile images are included in the CK+ dataset. The precision, recall and F1-score of this algorithm were 0.78, 0.96, and 0.86, respectively.

Three steps are performed to identify whether the person is frowning in an image. The first step is locating the landmarks of a face. Then the detector extracts the area between two eyebrows. Lastly, the detector processes the extracted area using SOBEL filter and classifies whether the image contains frowning (Chung et al., 2014). There will be more white pixels in the frowning images after being processed by SOBEL filter. The processed image, which contains more than 1% white pixels is considered as a frown image (Chung et al., 2014). Thirty-eight frown images (in anger category) and 38 non-frown images (in neutral category) were selected from CK+ dataset (Lucey et al., 2010) to evaluate the accuracy of this algorithm. These images came from 38 different people, and each person contributed one frown image and one non-frown image. As a result, the precision, recall and F1-score of this algorithm were 0.79, 0.71, and 0.75, respectively.

Body Movement

EQClinic detects two kinds of body movements: body leaning and OBM.

EQClinic detects body leaning by observing the size of the face. When the size of the face dramatically increases within a period of time, a forward body leaning has happened. Similarly, a backward body leaning happened when the size of the face decreases.

Overall body movement detection does not focus on a particular kind of movement. It represents the overall level of the participant's body movements. The higher the OBM value is, the more dramatically the participant moves. The core idea of OBM detection is calculating the difference in pixels between every adjacent pair of video frames (Fung et al., 2015). As we assume

TABLE 1 | Mean absolute error of locating five landmarks by ASMLibrary (compared with manual makeups on BIOID dataset).

	X error mean	X error SD	Y error mean	Y error SD
Inner end of left eyebrow	2.95	2.46	1.80	1.73
Inner end of right eyebrow	3.63	2.57	1.48	1.72
Inner corner of left eye	1.95	1.56	1.12	1.60
Inner corner of right eye	2.19	1.84	1.20	1.61
Tip of nose	1.76	1.66	3.62	3.79

All numbers are in pixels.

that the background of the images is constant and only the participant moves in the video frames, the difference value is able to represent the overall level of the participant's body movement.

Voice Properties

The volume and sound pitch are the two voice properties detected by EQClinic.

An open source Matlab library is used by EQClinic to detect these properties (Jang, 2016). EQClinic represents the volume in decibels (db). The value is calculated according to the equation (Jang, 2016):

$$volume = 10 * \log_{10} \sum_{i=1}^{n} S_i^2$$

where S_i is the value of i-th sample and n is the frame size.

EQClinic represents the sound pitch in Semitones rather than Herz. According to the equation: semitone = $69 + 12 * \log_2$(frequency/440), we can easily convert the frequency value to semitone. Average Magnitude Difference Function (AMDF) is the method used for sound frequency detection (Ross et al., 1974).

Speech Pattern

EQClinic detects two kinds of speech pattern: turn-taking and speaking ratio.

Turn taking illustrates all the speaking and silence periods of both participants. This section also concludes the longest speaking period and average length of each speaking turn. Speaking ratio describes the cumulative percentage of total time; the student and the SP each spoke within a given time frame.

A key procedure of detecting these two features is End Point Detection (EPD). The aim of EPD is identifying the starting and ending point of an audio segment. A time domain method, which contains a volume threshold and a zero crossing rate is used for EPD (Jang, 2016).

Face-Touching Gesture

EQClinic detects the moment when the student touches his facial area with his own hands.

Even though many algorithms have been developed to recognize hand gestures, no existing algorithm was found for effectively sensing the face-touching gestures. Thus, we propose a new one to do so. The core idea of detecting face-touching gestures is tracking the proportion of skin color in each video frame. Otsu thresholding method (Otsu, 1975) is used in EQClinic to dynamically calculate the Cr thresholds, and any pixel within the image, whose Cr value locates within the ranges of the threshold, is considered as a skin color pixel.

Figure 6 illustrates the main steps of detecting face-touch gestures. We first extract the facial area (the first row of **Figure 6**) and the enlarged facial area (the second row of **Figure 6**), which is four times larger than the original facial area. Then, we convert the original and enlarged facial areas to binary images (third row of **Figure 6**). The white pixels in these binary images represent skin color pixels, and black pixels are non-skin color ones. Lastly, we calculate the difference of white pixel in the enlarged binary image and the original binary image, and divide this difference by the size of enlarged face area to generate the final value of each frame. The algorithm assumes that there is a face-touching gesture if the difference between two adjacent frames less than -3%.

Pilot Study

Participants

Participants were 3 SPs and 18 medical students from year 1 to 4 of an Australian medical school. All participants in this study were volunteers that signed informed consent forms. The study was approved by the University of Sydney Human Ethics Research Committee (protocol 2015/151).

Trainings

At the beginning of this study, three SPs received a 2-h face-to-face training session on EQClinic platform from a researcher of this project. During the training session, the researcher demonstrated the procedures of this study using the main components of the platform: booking appointments, having consultations with students, providing real-time comments, and evaluating student's performance.
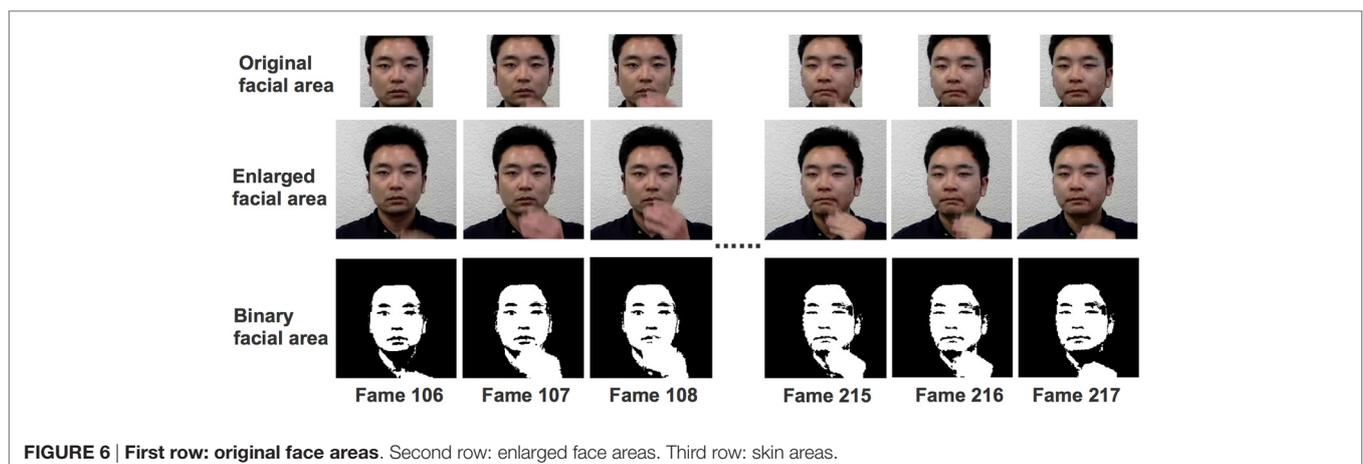
FIGURE 6 | First row: original face areas. Second row: enlarged face areas. Third row: skin areas.

In this study, all sessions focused on history taking, to ensure a structured and consistent interaction. The SPs were given a patient scenario in the training session and were asked to simulate being a patient. The scenario mentioned the main symptoms of the patient included having chest pain on and off, and an unproductive cough for 5 days. As all the SPs were experienced in the task, so no training on performance was provided.

Students were expected to be more proficient with technology; therefore, we did not provide them with face-to-face training. Instead, we provided a training video with an introductory email. In that email, we described the details of the study, confirmed consent for the study, and asked them to watch the training video. We also informed them that once they felt comfortable and confident with the system, they could start requesting consultation time slots on their personal calendar. After their consultation requests had been confirmed, they would have an online conversation with the SP. During the training, all SPs and students were told that each consultation would take approximately 15 min.

Surveys

Five surveys were used in the study:

Pre-Interview Survey

Five questions were included to ascertain a student's existing understanding about communication skills in general, and of non-verbal behaviors in particular. The first three questions were on a seven-point Likert scale, and the last two questions were free-text questions:

1. How confident do you feel now about your communication skills?
2. Do non-verbal behaviors have a significant effect on medical communication?
3. Is verbal content the main factor that affects medical communication?
4. What do you think are the three main non-verbal behaviors that affect medical communication?
5. Are there any specific communication skills you would like to work on?

Post-Interview Non-verbal Behavior Reflection Survey

This survey aimed to help students to reflect on their non-verbal behavior in the consultation. Ten questions were included in this survey, and they all asked how often the students engaged in certain behaviors (smiling, frowning, etc.), such as "Please estimate how often you smiled during the interview." All the questions were on a scale of one to seven. One represented "none," and seven represented "very often."

Student-Patient Observed Communication Assessment Form

A Student-Patient Observed Communication Assessment (SOCA) form, an edited version of the Calgary Cambridge Guide (Simmenroth-Nayda et al., 2014), was used by SPs to assess student performance in clinical consultations. It evaluated students from four aspects: providing structures, gathering information, building rapport, and understanding the patient's needs.

Each aspect was scored on a four-point scale in which one was the lowest and four the highest. Besides the overall score for each aspect, assessors can also select the specific criteria that the student needs to focus on. A free-text comment box was also included in this assessment form. The student was also asked to perform a personal SOCA prior to receiving the assessment SOCA.

Reflection Survey

This aimed to help students to reflect on the consultations. Students completed it after reviewing the SPs' assessments. The free-text questions were:

1. How do you feel the interview went for you at the time?
2. How does this compare with the grade and comments entered by the assessor?
3. How will you continue to develop your communication skills?

Post-Interview Survey

This survey was provided for students after all the extended feedback was provided – including the opportunity to review the video and SP's real time comments, the SOCA form and their reflection answers, and the non-verbal behavior feedback. The aim of this survey was evaluating the usefulness of non-verbal behavior feedback for students and the system acceptability of EQClinic. In addition, this survey contained all the questions of the Pre-interview Survey to help us analyze the influence of EQClinic on students' understanding of communication skills.

Six questions were also included in this survey for evaluating the non-verbal behavior feedback. The first three questions were on a scale of seven. Then there were two multiple-choice questions. The last question was a free-text question:

1. The feedback information was informative.
2. The system feedback information was clear.
3. I will change my communication behavior with patients after seeing the feedback information.
4. Which sections of the feedback information were most helpful for you?
5. Which sections of the feedback information were least helpful for you?
6. Do you have any further comments to help us improve the feedback information?

We selected five questions from the Computer System Usability Questionnaire (Lewis, 1995) for evaluating the system acceptability. These five questions evaluated the system from the aspects of structure, information and user interface of the system. All these questions were on a scale from one to seven.

Procedure

Each student was asked to conduct one consultation with a SP. The SPs and students were allowed to have the consultation anywhere as long as there was: (1) a web browser on a PC (windows or Mac) or an Android tablet with external or build-in camera and microphone; (2) a good Internet connection; (3) good lighting.

We divided each consultation into two parts: interviewing and assessing and each part took around 20 min (**Figure 7**). In the interviewing section, the student filled out the Pre-interview Survey first and then the student and the SP had a 15-min interview through the tele-conference component. In the assessing section, the student and the SP completed the online assessments. After each interview, the SP assessed the performance of the student using the SOCA Form. The student estimated their non-verbal behavior using Post-Interview Non-verbal Behavior Reflection Survey, completed a personal SOCA, reviewed the SP's Assessment Form and reflected on the interview using the Reflection Survey.

As EQClinic took time to analyze the video and identify non-verbal behaviors, the students were asked to return to the system 24 h after the consultation to review different kinds of feedback, which included the video recording, the comments from the SP and the automated non-verbal behavior feedback, and fill out the Post-interview Survey.

RESULTS

In total, 18 sessions were completed by 18 medical students (7 males and 11 females). Thirteen of them finished all the surveys of the study, and five students did not complete the Post-interview Survey. We recorded 314 min of consultations. At the start of each consultation was a short introduction between participants. The average length of the consultations was 16.6 min (SD = 2.1).

Students' Learning

In the Pre- and Post-interview Survey, we asked students the same question "Are there any specific communication skills you would like to work on?" We found 3 of 18 students mentioned that they would like to work on some behaviors which related to non-verbal communication in their Pre-interview Survey. Interestingly, all of these three students only mentioned the vocalics behavior such as speaking ratio and tone of voice. However, in the Post-interview Survey, which was completed after reviewing computer-generated non-verbal behavior feedback, 7 of 13 students mentioned that they would like to work on non-verbal communication. In addition, the seven students not only mentioned the vocalics behavior but also mentioned the kinesics behavior such as facial expressions and body movements to work on. For example, one student mentioned that "to make more facial expressions, mainly to smile a bit more and be more responsive in terms of head

movements and gesturing," and another student mentioned that "being comfortable with silence, controlling my facial expressions more, and controlling nervous laughter."

On average, students agree more on the statement of "Non-verbal behaviors have a significant effect on medical communication" (non-verbal statement) than "Verbal content is the main factor that affects medical communication" (verbal statement) (**Figure 8**). In addition, students agree on the non-verbal statement more in their Post-interview Survey (increased from 5.9/7 to 6.2/7). More specifically, 6 of 13 students provided a higher score on this statement in their Post-interview Survey than the Pre-interview Survey, and 5 of 13 students lowered their score in the Post-interview Survey. In contrast, on average, students were less in agreement on the verbal statement in their Post-interview survey (decreased from 4.1/7 to 3.6/7). More specifically, 5 of 13 students decreased their scores on this statement in their Post-interview survey, and only 2 of 13 increased their scores. These data indicated that, after reviewing the non-verbal behavior feedback, students paid more attention to the importance of non-verbal behavior.

In addition, according to the student feedback for answering the question "What do you think are the three main non-verbal behaviors that affect medical communication?", most students changed their answers in the Post-interview survey. One student changed two of the three non-verbal behavior, and nine students changed one. Only three students' answers remained the same.

The students' feedback in the Reflection Survey indicated that most of them agreed with the SPs' assessment and comments. The consultation and feedback helped them to identify skills that needed improvement. More than half of the students mentioned that they felt rushed during the interview and did not have enough time to ask questions or clarify their questions, and they would have liked to have improved their time management. Some students felt confused about organizing the structure of the consultation. For example, some students mentioned that they were unclear about which questions should be asked and in which order, so they spent too much time on unnecessary topics.

User Experience

Students were positive about the system usability and felt comfortable using it (average = 5.9, SD = 0.9). They reported that the structure (average = 5.4, SD = 0.8) and information (average = 5.8, SD = 1.0) of the system were clear, but it needed to provide clearer instructions when students encountered errors (average = 4.4, SD = 1.0).
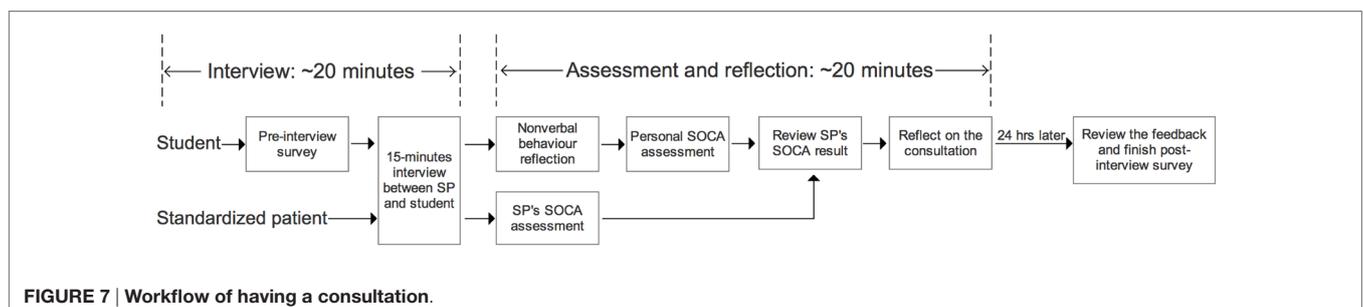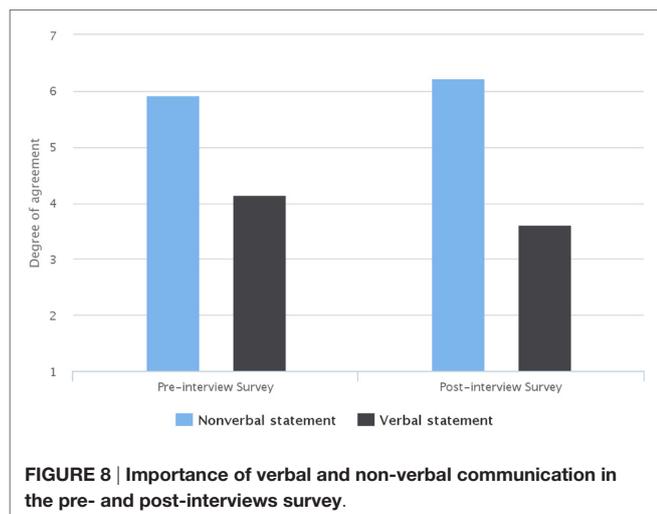

FIGURE 7 | Workflow of having a consultation.

FIGURE 8 | Importance of verbal and non-verbal communication in the pre- and post-interviews survey.

On average, students felt the feedback information was informative (average = 5.2, SD = 1.7) and clear (average = 4.8, SD = 2.0), and they would like to change their communication behavior with patients after reviewing the feedback information (average = 5.3, SD = 0.8). They also found the feedback on turn-taking, speaking ratio, smile intensity, frown intensity, and head nodding to be the most useful parts of the non-verbal behavior feedback, while the least useful feedback related to volume, sound pitch and body leaning.

Most of the students' comments on the non-verbal behavior feedback related to the question of whether it was possible to provide some suggestions to indicate to them what non-verbal behavior was good or bad. For example, some students mentioned that it was difficult to interpret the graph of certain non-verbal behaviors such as volume or sound pitch without providing to them the volume and sound pitch that is considered appropriate in medical consultations.

Accuracy

In Non-verbal behavior detection, we describe the algorithms we adopted for detecting non-verbal behaviors. The accuracies of most algorithms have been reported through testing on datasets made available by other research groups in the area of affective computing and computer vision. However, the accuracies of nodding, head shaking, and face-touching gestures are not reported. Thus, we randomly selected 5 video recordings of students from the 18 consultations and evaluated the accuracy of the algorithms.

The average length of the five selected video recordings was 17.5 min (SD = 1.15). In order to test the accuracy of the algorithms for detecting nodding and head shaking on our own data, a human annotator went through all the detected noddings and head shakings, and checked whether the reported video section contained a certain non-verbal behavior so that the false positive results could be generated. In total, 388 noddings (mean = 77.6, SD = 38.0) and 187 (mean = 37.4, SD = 16.3) head shakings were automatically detected from those five video

recordings by the algorithm. Fifty-one noddings (mean = 10.2, SD = 5.3) and 25 head shakings (mean = 5.0, SD = 3.0) were considered as false positives. In other words, the false positive rates of nodding and head shaking detection were 13.14 and 13.36%, respectively.

We used the same method to evaluate the accuracy of the algorithm for detecting face-touch gestures. However, unfortunately, the accuracy of this algorithm was not high. In total, 75 face-touching gestures were reported. As a result, the annotator found that 86.3% of these reported gestures were false positives. This result showed that this algorithm was not reliable. The main cause of this was the inaccuracy of the algorithm to sense the skin color pixels in images. By analyzing the video, we found the performance of this algorithm was significantly affected when the color of the video's background or students clothes were similar to the skin color. Another reason was that this algorithm would recognize face-touching gestures if students moved their hands around their facial area, even though their hands did not touch their face. In addition, the lighting conditions also affected the accuracy of this algorithm. Thus improving this algorithm will be an important job of our future works.

DISCUSSION

Based on our review of the literature, EQClinic provides a novel solution for medical students to practise their communication skills with automated non-verbal behavior feedback. With the increasing use of tele-health, EQClinic can be used by clinical doctors and nurses for helping them reflect on and learn communication skills in tele-health. Furthermore, its potential use is not limited to medical education but other professions were communication is important.

By comparing the Pre- and Post-interview surveys of the 13 students who completed the whole study, we found that students' attitude to verbal and non-verbal communication had been changed. The evaluation showed that students' conceptions about the importance of non-verbal behaviors increased, and the importance of verbal content was reduced with the EQClinic session. The results also showed that students paid attention to more kinds of non-verbal behavior, possible evidence of a more nuanced awareness. This pilot study clearly indicated that automated non-verbal behavior feedback enhanced the students' awareness of non-verbal communication. As non-verbal behavior is often performed unconsciously, increasing students' awareness of their non-verbal behaviors would be helpful for patients' perception of a clinical consultation.

The evaluation also showed that the user experience based on self-reports was satisfactory. The user experience could be improved with further participatory research exploring the type of features that students would like to see in the system.

Some free-text feedback from students indicated issues with communication skills training *via* video conferencing. Many students found it was difficult to maintain eye-contact with patients because they could not physically look at their eyes. Students were confused about where they should look during the consultation: the middle of the screen or the camera. In addition, some students

also mentioned that they were unable to observe some of the non-verbal cues of the SP as only the upper-body of the SP could be seen by the student.

The pilot also showed the importance of the infrastructure on the user experience. The stableness of the network in particular was a concern for the students. Eight (44%) students reported that they experienced different levels of network interruption during the consultations. Some students missed important parts of the conversation with the SP because of the interruption of the video conferencing. Video or sound lags also bothered the students, and contributed to inappropriate interruptions of the SP.

Another challenge we encountered during the study was that the lighting conditions of participants' environment had a significant impact on the results of the non-verbal behavior detection, especially on facial expressions and head movement recognition. In this study, we were asking the participants to have consultations in good lighting conditions. However, we did not clearly know the instructions for, and therefore could not accurately explain, these "conditions" specifically during the training sessions, so some of the video recordings were difficult to analyze. Thus, we believe that detailed instructions to help participants to select a suitable environment are needed in the future study. For example, in the instructions, we can state that a stable artificial lighting source is one of the preferred set-ups. In addition, in the future, we could also provide a testing page for participants to allow them to take a "sound check" video. By automatically analyzing the "sound check" video, the system can give some feedback to the participant about whether the lighting condition is appropriate for the system.

Currently, we are conducting a study with a larger cohort that investigates whether using EQClinic improves learning outcomes. If students using EQClinic improve their assessment scores, we will need to investigate if this learning also improves face-to-face consultations. Although students felt positive about the non-verbal behavior feedback, there were some limitations in EQClinic. The main limitations were: the accuracy of non-verbal behavior detection and providing meaningful suggestions for students.

Currently, EQClinic's non-verbal behavior feedback was that it could only identify non-verbal behavior, but it could not evaluate their quality or appropriateness to the specific type of tele-consultation. As we reported in the previous section, students would like to receive individualized suggestions about which specific types of non-verbal behavior they should pay attention to and how to improve them. However, due to the limited number of tele-consultations recorded in this pilot study, it was hard to develop a standard model of non-verbal behavior and provide valuable non-verbal behavior evaluation based on that model. With the increasing numbers of recorded tele-consultations, we think it is possible to develop models of non-verbal behavior in history taking tele-consultations with SPs using machine learning techniques.

AUTHOR CONTRIBUTIONS

CL and RC have contributed to the software development, data analysis, and writing, RL has contributed to the recruitment and data gathering, as well as to the writing.

FUNDING

REFERENCES

BioID. (2016). *BioID Face Database*. Available at: https://www.bioid.com

Buller, D., and Street, R. L. (1992). "Physician-patient relationships," In R. Feldman Editor. *Applications of Nonverbal Behavioral Theories and Research*. Hillsdale, NJ: Lawrence Erlbaum, 119–141.

Burgoon, J. K., Guerrero, L. K., and Floyd, K. (2016). *Nonverbal Communication*. Abingdon: Routledge.

Calvo, R. A. (2015). "Affect-aware reflective writing studios," in *Chapter 33. Handbook of Affective Computing*, eds R. A. Calvo, S. K. D'Mello, J. Gratch, and A. Kappas (New York, NY: Oxford University Press), 447–457.

Calvo, R. A., D'Mello, S., Gratch, J., and Kappas, A. (eds) (2015). *The Oxford Handbook of Affective Computing*. New York, NY: Oxford University Press.

Calvo, R. A., and D'Mello, S. K. (eds) (2011). *New Perspectives on Affect and Learning Technologies*, Vol. 3. Berlin: Springer Science & Business Media.

Chung, S.-C., Barma, S., Kuan, T.-W., and Lin, T.-W. (2014). Frowning expression detection based on SOBEL filter for negative emotion recognition. *Paper Presented at the Orange Technologies (ICOT), 2014 IEEE International Conference on*, Xi'an.

Collins, L. G., Schrimmer, A., Diamond, J., and Burke, J. (2011). Evaluating verbal and non-verbal communication skills, in an ethnogeriatric OSCE. *Patient Educ. Couns.* 83, 158–162. doi:10.1016/j.pec.2010.05.012

Dent, J., and Harden, R. M. (2013). "Clinical communication," in *A Practical Guide for Medical Teachers*, eds J. A. Dent and R. M. Harden (Sydney: Elsevier Health Sciences), 265–272.

DiMatteo, M. R., Taranta, A., Friedman, H. S., and Prince, L. M. (1980). Predicting patient satisfaction from physicians' nonverbal communication skills. *Med. Care* 18, 376–387. doi:10.1097/00005650-198004000-00003

Duggan, P., and Parrott, L. (2001). Physicians' nonverbal rapport building and patients' talk about the subjective component of illness. *Hum. Commun. Res.* 27, 299–311. doi:10.1093/hcr/27.2.299

Fung, M., Jin, Y., Zhao, R., and Hoque, M. E. (2015). ROC speak: semi-automated personalized feedback on nonverbal behavior from recorded videos. *Paper Presented at the Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Osaka.

GENKI-4K. (2016). *The MPLab GENKI Database, GENKI-4K Subset*. Available at: http://mplab.ucsd.edu

Gibbings-Isaac, D., Iqbal, M., Tahir, M. A., Kumarapeli, P., and de Lusignan, S. (2012). The pattern of silent time in the clinical consultation: an observational multichannel video study. *Fam. Pract.* 29, 616–621. doi:10.1093/fampra/cms001

Gorawara-Bhat, R., Cook, M. A., and Sachs, G. A. (2007). Nonverbal communication in doctor–elderly patient transactions (NDEPT): development of a tool. *Patient Educ. Couns.* 66, 223–234. doi:10.1016/j.pec.2006.12.005

Harrigan, J. A., Oxman, T. E., and Rosenthal, R. (1985). Rapport expressed through nonverbal behavior. *J. Nonverbal Behav.* 9, 95–110. doi:10.1007/BF00987141

Hartzler, A. L., Patel, R. A., Czerwinski, M., Pratt, W., Roseway, A., Chandrasekaran, N., et al. (2014). Real-time feedback on nonverbal clinical communication. *Methods Inf. Med.* 53, 389–405. doi:10.3414/ME13-02-0033

Hoque, M. E., Courgeon, M., Martin, J.-C., Mutlu, B., and Picard, R. W. (2013). Mach: my automated conversation coach. *Paper Presented at the Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Zurich.

Ishikawa, H., Hashimoto, H., Kinoshita, M., Fujimori, S., Shimizu, T., and Yano, E. (2006). Evaluating medical students' non-verbal communication during the objective structured clinical examination. *Med. Educ.* 40, 1180–1187. doi:10.1111/j.1365-2929.2006.02628.x

Jang, J. S. R. (2016). *Utility Toolbox*. Available at: http://mirlab.org/jang

Kawato, S., and Ohya, J. (2000). Real-time detection of nodding and head-shaking by directly detecting and tracking the "between-eyes". *Paper Presented at the Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, Grenoble.

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst.* 2, 26–41. doi:10.1145/357417.357420

Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall.

Larsen, K. M., and Smith, C. K. (1981). Assessment of nonverbal communication in the patient-physician interview. *J. Fam. Pract.* 12, 481–488.

Leppink, J., van Gog, T., Paas, F., and Sweller, J. (2015). Cognitive load theory: researching and planning teaching to maximise learning. In: J. Cleland, S. J. Durning Editors. *Res. Med. Edu.* pp. 207–218 Oxford, UK: Wiley Blackwell.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int. J. Hum. Comput. Interact.* 7, 57–78. doi:10.1080/10447319509526110

Liu, C., Scott, K., and Calvo, R. A. (2015). "Towards a reflective online clinic (ROC): a tool to support interviewing skills," in *CHI'15 Workshop on Developing Skills for Wellbeing*, Seoul.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression. *Paper Presented at the Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, San Francisco.

Mann, K. V. (2011). Theoretical perspectives in medical education: past experience and future possibilities. *Med. Educ.* 45, 60–68. doi:10.1111/j.1365-2923.2010.03757.x

Martin, L., and Friedman, H. S. (2005). "Nonverbal communication and health care," In R. E. Riggio, R. S. Feldman Editors. London: Lawrence Erlbaum *Applications of Nonverbal Communication*, 3–16.

Mast, M. S., Hall, J. A., Klöckner, C., and Choi, E. (2008). Physician gender affects how physician nonverbal behavior is related to patient satisfaction. *Med. Care* 46, 1212–1218. doi:10.1097/MLR.0b013e31817e1877

McLeod, S. A. (2013). *Kolb – Learning Styles*. Available at: www.simplypsychology.org/learning-kolb.html

Milborrow, S., and Nicolls, F. (2008). "Locating facial features with an extended active shape model," In F. David, T. Philip and Z. Andrew. *Computer Vision–ECCV 2008* (Springer), 504–513.

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica* 11, 23–27.

Pentland, A., and Heibeck, T. (2010). *Honest Signals: How They Shape Our World*. Cambridge, MA: MIT press.

Picard, R. W., (2000). *Affective Computing*, Cambridge: MIT Press.

Ross, M. J., Shaffer, H. L., Cohen, A., Freudberg, R., and Manley, H. J. (1974). Average magnitude difference function pitch extractor. *IEEE Trans. Acoust. Speech Signal Process.* 22, 353–362. doi:10.1109/TASSP.1974.1162598

Roter, D., and Larson, S. (2002). The Roter interaction analysis system (RIAS): utility and flexibility for analysis of medical interactions. *Patient Educ. Couns.* 46, 243–251. doi:10.1016/S0738-3991(02)00012-5

Simmenroth-Nayda, A., Heinemann, S., Nolte, C., Fischer, T., and Himmel, W. (2014). Psychometric properties of the Calgary Cambridge guides to assess communication skills of undergraduate medical students. *Int. J. Med. Educ.* 5, 212. doi:10.5116/ijme.5454.c665

Smith, L. I. (2002). *A Tutorial on Principal Components Analysis*, Vol. 51. Ithaca, NY: Cornell University, 65.

Stewart, M. A. (1995). Effective physician-patient communication and health outcomes: a review. *CMAJ* 152, 1423.

Tanveer, M. I., Zhao, R., Chen, K., Tiet, Z., and Hoque, M. E. (2016). *AutoManner: An Automated Interface for Making Public Speakers Aware of Their Mannerisms*.

Vinciarelli, A., Pantic, M., and Bourlard, H. (2009). Social signal processing: survey of an emerging domain. *Image Vision Comput.* 27, 1743–1759. doi:10.1016/j.imavis.2008.11.007

Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., et al. (2012). Bridging the gap between social animal and unsocial machine: a survey of social signal processing. *IEEE Trans Affective Comput.* 3, 69–87. doi:10.1109/T-AFFC.2011.27

Viola, P., and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Paper Presented at the Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, Kauai.

Wei, Y. (2009). *Research on Facial Expression Recognition and Synthesis*. Master Thesis, Department of Computer Science and Technology, Nanjing.

Weinberger, M., Greene, J. Y., and Mamlin, J. J. (1981). The impact of clinical encounter events on patient and physician satisfaction. *Soc. Sci. Med. E Med. Psychol.* 15, 239–244. doi:10.1016/0271-5384(81)90019-3

Young, J. Q., Van Merrienboer, J., Durning, S., and Ten Cate, O. (2014). Cognitive load theory: implications for medical education: AMEE guide no. 86. *Med. Teach.* 36, 371–384. doi:10.3109/0142159X.2014.889290