Eye-to-Eye: Gaze detection as a proxy for medical doctor behaviour during appointments (June 2022)

Ricardo Antão, Master's Student

***Abstract*—The gaze direction of the physician during consultations plays an important role in patient satisfaction, with several studies pointing to the positive correlation between the amount of time the physician spends looking at the patient and patient satisfaction. Additionally, the effects of the virtual consultation environment on the gaze behaviour of the doctor have yet to be studied. Therefore, this study aims to assess the impact of the virtual consultation environment by comparing the amount of time the doctor spends looking at the patient between face-to-face and virtual consultations. The study population consisted of 14 doctors divided between 4 medical specialties: Gynaecology/Obstetrics (4), Neurology (3), Endocrinology (3) and General and Familiar Medicine (4). A desktop appearance-based gaze estimation pipeline was implemented in the clinical setup. Each doctor recorded 20 face-to-face consultations and 20 virtual consultations, which were then processed by the pipeline to obtain the percentage of time the doctor spent looking at the patient during the consultation. After obtaining the two distributions for each doctor the Mann-Whitney U test was used to compare both distributions. If a statistically significant difference ($p < 0.05$) was found, then the Cohen's *d* was used to calculate size effect. Overall, we found that all doctors, except for two, presented statistically significant differences or tendencies to look more at the patient in virtual consultations compared to face-to-face consultations. Within medical specialties, only one specialty presented doctors with no differences or tendencies between consultation environments, meaning that three out the four specialties involved presented clear tendencies to look more at the patient in virtual consultations when compared to face-to-face consultations.**

***Index Terms*—Gaze estimation, Physician-Patient Relationship,**

I. Introduction

The physician-patient relationship is a crucial component of the effectiveness of any health care system. Communication is one of the main components in a good physician-patient relationship where communication during medical interviews plays one of the most prominent roles [1].

The importance of non-verbal communication has gathered more and more attention from research studies that analyze the relationship between patient satisfaction and physician behaviour. A general practitioner can conduct between 120,000 and 160,000 interviews during a 40-year career [2]. Good communication skills are essential to reach improved management of chronic diseases like diabetes, hypertension, and others. In addition, patients are more adherent to medical recommendations and healthy behavioural changes when they are more informed and involved in the decision-making. Several studies indicate that the non-verbal cues of the physician are one of the most critical aspects of physician-patient communication [3, 4, 5, 6].

This work has been submitted for review on the 1^{st} of June 2022.

Communication in the physician-patient relationship is divided into verbal and non-verbal communication. Verbal communication is defined as communication behaviour with linguistic content [4]. This is classified according to the model described by Bird and Cohen-Cole model [7]. According to this model, we can classify verbal interactions into three key functions: data gathering to understand the patient (gathering information), development of a rapport and responding to the patient's emotions (developing therapeutic relationship), and patient education and behavioural management (decision making and management). Non-Verbal communication can be defined as communication behaviour without linguistic content and is typically distinguished by which part of the body is being used to express the behaviour. Face non-verbal behaviours include smiling, gazing, frowning, eyebrow-raising, and facial expressivity. Body non-verbal behaviour is expressed through posture or gestures. Vocal non-verbal behaviour includes loudness, voice pitch, monotony, and speech rate.

Among the non-verbal behaviour, the gaze is one of the most important cues to analyze in non-verbal communication. Studies have concluded that there is a positive correlation between patient satisfaction and the amount of eye contact between the physician and the patient [6]. The amount of time the physician is gazing at the patient and not at the patient's health records on the screen can improve the patient's perception and cognitive functioning [8].

The majority of the works on this topic use manual annotation systems to quantify non-verbal behaviours. This process is costly and laborious, which is not convenient or scalable. Recently, some methods for automated annotation systems have been proposed [9, 8]. However, they were designed for a very constrained environment which would not translate well to other consultation offices/setups. Nonetheless, they provide the first approaches to automatic classification of physician's gaze during medical appointments.

In this work we will focus on the automated analysis of the gaze direction of the physician during consultations, classifying it according to whether the physician is looking at the patient or not. It will aim to implement an appearance-based gaze estimation system in the clinical setup to support the analysis of the physician-patient relationship.

Additionally, the rise in the use of virtual consultations due to the COVID-19 pandemic provided an entirely different environment for the physician-patient relationship. Its impact in the physician-patient relationship is yet to be fully understood. A preliminary observation made by the Luz Saúde group during 2019 (in the pre-pandemic era) perceived a rise in the eye contact between the patient and the physician during video consultations compared to the face-to-face consultations. This

perception needs to be quantified and objectified clearly and scientifically. Therefore, in addition the first point, we will also aim to compare the gaze behaviour from the physician between the face-to-face and virtual consultation environment.

This work is organized into seven sections. Section II gives a very short overview of gaze estimation research explaining the method used in this study and it talks about the use of gaze estimation on the clinical setup. Section III explains the statistical study design, mentioning the study variable, the formulated hypotheses, the source population, the sample size and the statistical tests used. Section IV explains the methods used to record, extract and classify the gaze direction of the doctor during consultations. Section V presents the results of the study and provides the discussion and insights taken from them. Section X presents the discussion and main insights taken from the results presented. Section XI provides the overall conclusions from this study and the future works to be done to support this study.

II. Related Work

A. Gaze Estimation for images

Gaze estimation objective is to estimate a subject's gaze direction. The earliest attempts at gaze estimation consisted in the detection of eye movement patterns like fixation, saccades and smooth pursuits [10]. These early methods attached movement sensors around the eyes to detect the movement patterns mentioned before. However the evolution of computer vision technology enabled the creation of modern eye tracking software devices. Of the various modern gaze estimation methods developed in recent years, deep learning powered appearance-based gaze estimation systems are the methods that provide state-of-the-art accuracy in a robust manner (apart from proprietary commercial eye trackers).

Appearance-based gaze estimation methods learn functions that map directly from images to gaze direction by using image features like image pixels [11] or deep features [12] and are able to work with off-the-shelf web cameras. Regressing gaze directly from images is a difficult task due to the variety and complexity of eye appearance and several models were tested. However over recent years, deep learning methods have been found to be the most effective for the task with the majority of gaze estimation research today focusing on deep learning approaches [13, 14, 15, 16, 12, 17, 18, 19, 20, 21, 22].

The work in [15], *Gaze360*, is one of the state-of-the-art methods developed with deep learning techniques. It aimed to provide robust gaze estimation in unconstrained environments with a wide variety of head poses and illumination conditions. To achieve this, they propose both a dataset and a deep learning gaze estimation model, both denominated of *Gaze360*. The Gaze360 model is a video-based gaze tracking model using an RNN architecture, more specifically, a bidirectional Long-Short Term Memory (LSTM) [23], in conjunction with a backbone network and a fully connected layer. The model achieves very robust results, being able to provide state-of-the-art accuracy under unconstrained conditions like, extreme head poses or illumination changes.

B. Gaze Estimation on a clinical setup

The work in [8] proposed a gaze classification approach for doctor's gaze using CNNs. The CNN architecture used, as backbone, the VGG-16 architecture from [24] pre-trained on the ImageNet dataset, and added to it 1 Global Max Pooling layer, 1 Dropout layer and 5 fully connected layers. They then fine tuned the architecture with a dataset of raw videos from clinical interactions developed by the authors of the study. This dataset consisted of a set of 101 clinical interactions involving 10 doctors and 101 patients. Each interaction was comprised of 3 videos captured at the same time from 3 different cameras, a *Patient-Centered* camera focused on the patient, a *Doctor-Centered* camera focused on the doctor and a *Wide-frame* camera providing wide-view image of both the doctor and patient. The videos were annotated using the Noldus Observer XT Software [25]. In the end, the model proved to be very accurate with a 98.31% accuracy on the validation set and over 80% accuracy on the majority of independent hold out sets with unseen doctors and interactions. However, the need to use 3 different cameras with specific views of the consultation office makes this system unfeasible in many situations. In addition to this, it is unknown how the accuracy of the method translates to different consultation setups with different conditions, *e.g.* lighting conditions, patient positioning.

The work in [26] uses eye-tracking glasses to track the gaze of the doctor during face-to-face consultations. A total of 16 doctors seeing a total of 100 patients, each doctor seeing between 2 and 14 patients with the median being 6 patients. Doctor's gaze was measured with 3 different metrics: *face gaze duration*, *face gaze frequency* and *face gaze dwell time*. *Face gaze duration* corresponds to the total amount of time per minute the doctor spent looking at the patient, *face gaze frequency* corresponds to the amount of times per minute the doctor's gaze switched to the patient and *face gaze dwell time* is the time the doctor's gaze dwelled on the patients in each instance the doctor looked at them. The study concluded that there was a significant positive correlation between *face gaze dwell time* and *face gaze duration*, a significant negative correlation between *face gaze dwell time* and *face gaze frequency* and no correlation between *face gaze frequency* and *face gaze duration*. Additionally, the study also concluded that the amount of face gaze present in the beginning part of the consultation had positive and significant association to the amount of face gaze present in the beginning of the consultation. It also found that the amount of time the doctor spends looking at the patient during a consultation decreases in the final parts of the consultation compared to the beginning of the consultation.

III. Experimental Design

The goal of the study is to quantify and compare the amount of time the doctor spends looking at the patient between face-to-face and virtual consultations.

A. Hypotheses

The study variable chosen for this effect is the percentage of time the doctor spent looking at the patient during a consultation, denominated as *Patient%*. Two simple hypotheses were

formulated describing the possible outcomes of this study are the following:

- **Hypothesis 0 (H0)** - The doctor spends the same amount of time looking at the patient during face-to-face and virtual consultations.
- **Hypothesis 1 (H1)** - The doctor spends different amounts of time looking at the patient during face-to-face and virtual consultations.

B. Study Population

The source population for this study consists of 14 doctors divided among 4 different medical specialties. Choosing doctors among 4 medical specialties increases the scope of our study, at the expense of some robustness in the data. However, it was deemed more important understanding the impact of the virtual consultation environment on multiple types of medical specialties instead of focusing in just one, not knowing if the conclusions from this study could be applied in other medical fields. The 4 medical specialties chosen were:

- **Family and General Medicine** - 4 doctors
- **Endocrinology** - 3 doctors
- **Gynecology/Obstetrics** - 4 doctors
- **Neurology** - 3 doctors

The sample size for each doctor was defined as 20 face-to-face and 20 virtual consultations. Doctors from now on will be mentioned in the form of D*X* with *X* being the number assigned to the doctor in the beginning of the study.

C. Statistical test and Size Effect

To compare the *Patient%* distributions of each doctor the Mann-Whitney U test [27] was used. The Mann-Whitney U test is a non-parametric statistical test used to compare data between groups in different conditions and with different participants. The *Patient%* distributions were considered to be non-parametric since 20 results are not enough to check the normality of the distributions in most cases. Additionally, since face-to-face and virtual consultations have different participants, *i.e.* different patients, the consultation groups were considered to be groups in different conditions and with different participants. The Mann-Whitney U test outputs the *p-value* metric which allows us to see whether a statistical significant difference was found between distributions. The conventional *alpha* of $p < 0.05$ was chosen for this study, thus when $p < 0.05$ condition is met, a statistically significant difference exists between distributions.

When a statistically significant difference existed, the Cohen's *d* size effect [28] was used to calculate the size effect. The size effect will be used to assess how big of a difference did the virtual consultation environment cause on the amount of time spent looking at the patient during a consultation.

IV. Methodology

A. Consultation Recording

To record the consultations, an extra camera and computer were setup in the consultation rooms. The conventional consultation room for face-to-face consultations, illustrated in Figure 1a, possesses one computer screen, while the virtual consultation room for virtual consultations, illustrated in Figure 1b, possesses two computer screens. During face-to-face consultations, the patient is present in the room with the doctor while their health records are shown in the computer screen of the conventional consultation room. During virtual consultations, the patient's video feed is shown on one of the screens of the virtual consultation room while the health records are shown in the other screen.

(a) Conventional Consultation Room

(b) Virtual Consultation Room

Figure 1: Types of Consultation Rooms: Each room is was set up with an extra camera and computer to record the doctor's image during the consultation.

During both types of consultations doctors recorded the consultations with the extra computer by interacting with a very simple GUI, in which they just needed to press a button at the beginning and at the end of the consultation to start and stop the recording respectively. Additionally, due to privacy concerns, only the video feed and no audio was recorded. This way, only the doctor's image would be recorded.

B. Gaze Estimation Pipeline

The gaze estimation pipeline takes as input a video and the camera's extrinsic and intrinsic parameters and outputs the frame-by-frame 2D gaze estimates/points of gaze (*PoG*), which are points in the 2D screen plane, illustrated in Figure 3. The pipeline is composed of 3 main blocks illustrated in Figure 2.

The *Face/Landmark Detection* block serves as the pre-processing of the input video before the gaze estimation step. It takes the input recording and performs the face detection and 2D landmark annotation routines from *3DDFA_V2* [29]. The output of the block is the frame-by-frame face bounding box (locating the doctor's face on the frame) and the 2D landmark annotations of the doctor's face.

The *Gaze Estimation* block serves to extract the doctor's gaze direction during the consultation. It takes as input the consultation recording, the camera's intrinsic parameters and the output of the *Face/Landmark Detection* block. The gaze estimation is performed by the *Gaze360* model [15] and the output of the block is the frame-by-frame 3D gaze estimates/gaze directions.

The *Post-Processing* block converts the 3D gaze estimates/gaze directions into 2D gaze estimates/*PoG*s. It takes as input the 3D gaze estimates from the *Gaze Estimation* block and the camera's extrinsic parameters. With the camera's extrinsic parameters, *i.e.* the rotation and translation between the

Camera Coordinate System (CCS) and the Screen Coordinate System (SCS, the 2D screen plane), we can convert the 3D gaze estimates represented in the CCS to 2D gaze estimates in the SCS. Therefore the output of the block will be the frame-by-frame *PoG*s of the doctor during the consultation, which are going to be classified according to Figure 3.

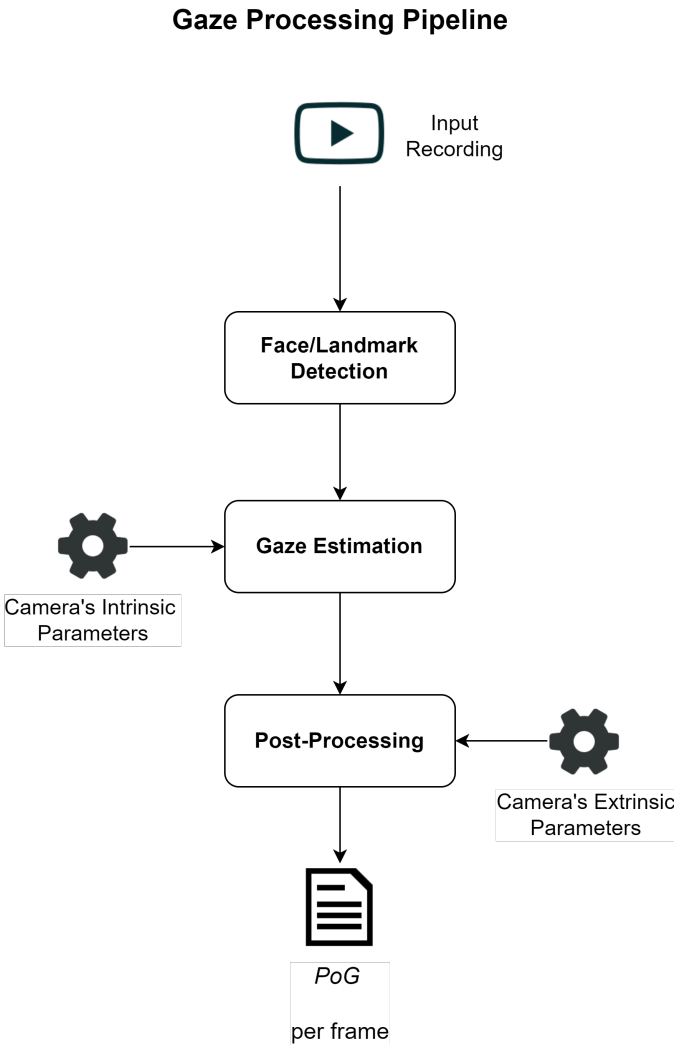

Figure 2: Gaze estimation pipeline composed of 3 main blocks: *Face/Landmark Detection*, *Gaze Estimation* and *Post-Processing*.

C. PoG classification

After obtaining the *PoG*s, we need to classify them in order to distinguish between the doctor looking at the patient and looking at screen/keyboard. Thus, we divided the screen plane into 5 zones as illustrated in Figure 3, *Above Screen*, *Screen*, *Keyboard*, *Right Of Screen* and *Left Of Screen*. In each consultation room the patient could be either on the left or on the right of the screen. Therefore, for each consultation room, either the *Right of Screen* zone or the *Left Of Screen* would be classified as the *Patient* zone. In the end, the classification allowed us to calculate how much time during the consultation the doctor spent looking at the patient by calculating the percentage of *PoG*s in the *Patient* zone. The percentage of *PoG*s classified as *Patient PoG*s was denominated as *Patient%* and it is the variable which is going to be compared in this study.

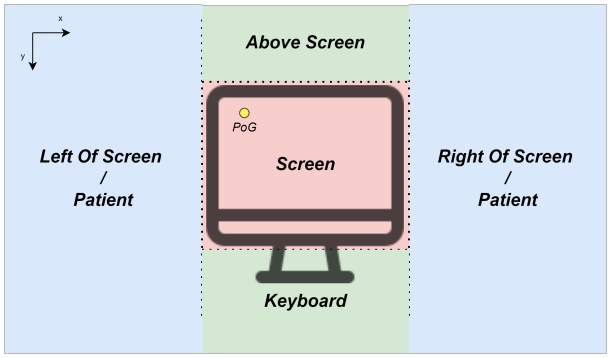

Figure 3: Gaze classification zones: *Above Screen*, *Screen*, *Keyboard*, *Right Of Screen* and *Left Of Screen*. For each consultation room, either *Right Of Screen* or *Left Of Screen* are considered to be the *Patient* zone depending on where the patient is positioned is relation to the screen.

V. Results

The results are divided by medical specialties. In each medical specialty, individual doctors are represented by D*X*, where *X* is the number assigned to each doctor in the beginning of the study. The individual doctor results are composed of three parts: the medians of the data distributions (Med_f and Med_v for the face-to-face and the virtual distributions respectively), the Mann-Whitney U test p-value plus size effect calculations (when a statistically significant difference is found) and the violin plot of the data distributions against each other. In this work, the violin plot shows, in addition to the distributions, all the data points of each doctor (represented with a black line). In addition to the individual doctors' results, we also present the results of the joint distributions of all doctors from that specialty, *i.e.* all face-to-face consultations and all virtual consultations in one specialty.

VI. Gynaecology/Obstetrics

In the Gynaecology/Obstetrics medical specialty, one doctor presents a statistically significant difference between *Patient%* distributions in face to face and virtual consultations, one doctor presents a tendency to have higher *Patient%* in virtual consultations and two doctors present no significant differences or tendencies between face-to-face and virtual consultations. The summary statistics of all doctors are shown in Table I and Figure 4 shows the violin plots of each doctor distributions.

Doctor D6's Mann-Whitney U test results in a *p-value* of 0.88, meaning no statistically significant difference exists between the face-to-face and the virtual distributions. Looking at Figure 4, we see that D6 face-to-face consultations' *Patient%* are more concentrated around the 50% mark, while the virtual consultations present a much wider range of values. This, combined with the face-to-face and virtual distribution medians (Med_f and Med_v) being almost identical, 51.8% and 51.9% respectively, leads us to conclude that D6 presents almost no differences between both consultation environments.

Table I: Gynaecology/Obstetrics results and summary statistics

Doctor	Med_f	Med_v	*p-value*	Cohen's *d*
D6	51.8%	51.9%	0.88	-
D9	52.3%	58.9%	0.06	-
D12	55.7%	75.1%	$1.3e-5$	0.69
D14	47.9%	43.4%	0.56	-
Joint	51.3%	57.5%	$0.7e-2$	-

Doctor D9's Mann-Whitney U test results in a *p-value* of 0.06, meaning no statistically significant difference exists between both distributions. However, looking at Figure 4, we see that D9's virtual distribution is slightly higher than the face-to-face distribution. Also confirmed by the distribution medians in Table I, where Med_v is higher, 58.9%, than Med_f, 52.3%. These results show that D9 presents a tendency to look more at the patient during virtual consultations than in face-to-face consultations even though it is not statistically significant.

Doctor D12's Mann-Whitney U test results in a *p-value* of 0.000013, meaning that there is a statistically significant difference. In addition, the Cohen's *d* effect size is equal to 0.69, meaning that there is a medium effect size when changing consultation environments. This significant difference is further confirmed when looking at the violin plot of Figure 4 and at the Med_f and Med_v medians, 55.7% and 75.1% respectively. Therefore, we can say that D12 does look more at the patient in virtual consultations as opposed to face-to-face consultations.

Doctor D14's Mann-Whitney U test results in a *p-value* of 0.56, meaning no statistically significant difference exists between the face-to-face and the virtual distributions. D14's results are very similar to D6's in the sense that the violin plots show the same characteristics in both doctors. Additionally, the face-to-face median is slightly higher than the virtual median, 47.9% and 43.4% respectively. Therefore, when looking at the *p-value* and at Figure 4 we conclude that D14, just like D6, has almost no differences when changing between face-to-face and virtual consultations.

Figure 4: Gynaecology/Obstetrics doctors violin plots

The Mann-Whitney U test on the joint data distribution of all doctors results in a *p-value* of 0.007, meaning there is a statistically significant difference between the face-to-face and the virtual distributions of the four doctors as a whole. Looking at Table I, the median of the face-to-face distribution, 51.3%, is lower than the virtual distribution's median, 57.5%. However, since half the doctors present no differences between and D9 only presented a tendency, this leads us to believe that the majority of this difference in the joint test comes from the D12 doctor which had around a 20% difference between distribution medians. Therefore, to reach a conclusion about the effect of the virtual consultation environment in the Gynaecology/Obstetrics as a whole we would need more data and participating doctors. Nonetheless, these are promising initial results.

VII. Endocrinology

In the Endocrinology medical specialty, two doctors present a statistically significant difference between *Patient%* distributions in face to face and virtual consultations and one doctor presents a tendency to have higher *Patient%* in virtual consultations. The summary statistics of all doctors are shown in Table II and Figure 5 shows the violin plots of each doctor distributions.

Table II: Endocrinology results and summary statistics

Doctor	Med_f	Med_v	*p-value*	Cohen's *d*
D3	37.7%	40.8%	0.11	-
D10	27.7%	42.2%	0.01	0.40
D15	44.6%	58.2%	$1.6e-5$	0.68
Joint	37.3%'s	50.9%	$0.7e-5$	-

Doctor D3's Mann-Whitney U test results in a *p-value* of 0.11, meaning no statistically significant difference exists between both distributions. However, we can note a tendency for higher *Patient%* in virtual consultations. When looking at Figure 5, we see that D3's virtual distribution values are slightly higher than the face-to-face distribution, further confirmed by the distribution medians in Table II, where virtual consultations have a higher median, 40.8%, than the face-to-face distribution, 37.7%. Therefore, taking all the results into account, we can say that D3 presents a tendency to look more at the patient during virtual consultations than in face-to-face consultations.

Doctor D10's Mann-Whitney U test results in a *p-value* of 0.01, meaning that there is a statistically significant difference between the face-to-face and virtual distributions. Additionally, the Cohen's *d* effect size of 0.40 indicates a medium sized effect when changing consultation environments. The violin plot of D10, shown in Figure 5 shows the clear difference in the *Patient%* metric between distributions, with the virtual distribution having its peak at a higher value than the face-to-face distribution. This significant difference is further confirmed by the medians in Table II, with the virtual distribution median being 42.2%, much higher than the face-to-face distribution median of 27.7%. Thus, we can safely say that D10 looks more at the patient during virtual consultations than in face-to-face consultations.

Doctor D15's Mann-Whitney U test results in a *p-value* of 0.000016, meaning that there is a statistically significant difference. In addition, the Cohen's d effect size is equal to 0.68, meaning that there is a medium effect size when changing consultation environments. This significant difference is further confirmed when looking at the violin plot of Figure 5 and at the Med_f and Med_v medians, 44.6% and 58.2% respectively, amounting to a difference of around 14%. Therefore, we can say that D15 looks significantly more at the patient in virtual consultations as opposed to face-to-face consultations.

Figure 5: Endocrinology doctors violin plots

The Mann-Whitney U test on the joint data distribution results in a *p-value* of 0.000007, meaning there is a statistically significant difference between the distributions of the three doctors as a whole. Looking at Table II, we can further confirm that the median of the face-to-face distribution is indeed lower than the virtual distribution median, 37.3% and 50.9% respectively. These initial results are very promising and indicate that the Endocrinology medical specialty has a tendency to look more at the patient during virtual consultations than in face-to-face consultations. However, generalizing these conclusions to the Endocrinology specialty as a whole in a robust manner needs more data to further corroborate these results.

VIII. Neurology

In the Neurology medical specialty, two doctors present a statistically significant difference between *Patient%* distributions in face to face and virtual consultations and one doctor presents a tendency to have higher *Patient%* in virtual consultations. The summary statistics of all doctors are shown in Table III and Figure 6 shows the violin plots of each doctor distributions.

Table III: Neurology results and summary statistics

Doctor	Med_f	Med_v	*p-value*	Cohen's d
D1	45.8%	48.3%	0.18	-
D2	54.9%	65.9%	$3.1e-3$	0.47
D8	32.8%	49.9%	$0.5e-5$	0.72
Joint	44.7%	57.7%	$2.1e-5$	-

Doctor D1's Mann-Whitney U test results in a *p-value* of 0.18, meaning no statistically significant difference exists between both distributions. However, the low *p-value* indicates a tendency to have a higher *Patient%* in virtual consultations. When looking at Figure 6, we see that D1's virtual distribution values are slightly higher than the face-to-face distribution, even though the medians have a negligible difference between them of around 3%. Nonetheless, when taking all the results into account, we can say that D1 presents a tendency to look more at the patient during virtual consultations when compared to face-to-face consultations.

Doctor D2's Mann-Whitney U test results in a *p-value* of 0.0031, meaning that there is a statistically significant difference between the face-to-face and virtual distributions with the Cohen's d effect size of 0.47 indicating a medium size effect when changing consultation environments. The violin plot of D2, shown in Figure 6 shows the clear difference between distributions with the virtual distribution being higher than the face-to-face distribution. This is also confirmed by the medians in Table III, with the virtual distribution median being 65.9%, a difference of around 10% when compared to the face-to-face distribution median of 54.9%. Therefore, doctor D2 looks more at the patient during virtual consultations than in face-to-face consultations.

Doctor D8's Mann-Whitney U test results in a *p-value* of 0.000005 and a Cohen's d effect size of 0.72, meaning that there is a statistically significant difference between the face-to-face and virtual distributions with a medium effect size. The violin plot of D8, shown in Figure 6 shows the clear difference between the face-to-face and virtual distributions, with the virtual distribution being clearly higher than the face-to-face distribution. Additionally, the medians difference of around a 17% further support this disparity, with the virtual distribution median being much higher than the face-to-face distribution, 49.9% and 32.8% respectively. Thus, doctor D8 looks significantly more at the patient during virtual consultations when compared to face-to-face consultations.

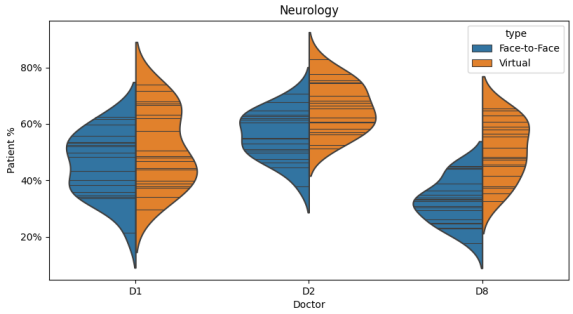

Figure 6: Neurology doctors violin plots

The Mann-Whitney U test on the joint data distribution results in a *p-value* of 0.000021, meaning that, generally, changing consultation environments has statistically significant effect for the three doctors. Looking at Table III, we can further confirm that the median of the face-to-face distribution is indeed lower than the virtual distribution median, 57.7% and 44.7% respectively. Therefore, in general, the Neurology doctors have a tendency to look more at the patient during virtual consultations than in face-to-face consultations.

IX. General and Family Medicine

In the General and Family Medicine medical specialty, two doctors present a statistically significant difference between *Patient%* distributions in face to face and virtual consultations and two doctors present a tendency to have higher *Patient%* in virtual consultations. The summary statistics of all doctors are shown in Table IV and Figure 7 shows the violin plots of each doctor distributions.

Table IV: General and Family Medicine results and summary statistics

Doctor	Med_f	Med_v	*p-value*	Cohen's *d*
D4	38.2%	43.2%	0.049	0.31
D5	62.4%	59.3%	0.19	-
D7	46.4%	55.7%	0.11	-
D16	43.6%	59.2%	0.0059	0.54
Joint	46.5%	57.1%	$2.9e-4$	-

Doctor D4's Mann-Whitney U test results in a *p-value* of 0.049 with a Cohen's *d* size effect of 0.31, meaning there is a statistically significant difference between both types of consultations with a medium size effect. When looking at Figure 7, we can clearly see that D4's virtual distribution is higher than the face-to-face distribution. Additionally, the medians, shown in Table IV, confirm this with a difference of around 5%, with the virtual median being higher than the face-to-face median. For this reasons, we can say D4 spends more time looking at the patient during virtual consultations than in face-to-face consultations.

Doctor D5's Mann-Whitney U test results in a *p-value* of 0.19, meaning that there is no statistically significant difference between the face-to-face and virtual distributions. however the low *p-value* indicates a tendency for D5 to have a higher *Patient%* in the virtual consultations when compared to face-to-face consultations. This tendency is supported by the violin plot of D5's distributions, where we can see that the virtual distribution extends further than the face-to-face distribution. However the medians of both distributions have a negligible difference between them of around 3%. Nonetheless, the *p-value* and the violin plots indicate a slight tendency for D5 to spends more time looking at the patient in virtual consultations.

Doctor D7's Mann-Whitney U test results in a *p-value* of 0.11, meaning that there is no statistically significant difference between the face-to-face and virtual distributions. Despite of that, both the violin plot and the medians indicate a tendency for D7 to have a higher *Patient%* in virtual consultations. The violin plot of D7 shows the peak of the virtual distribution being slightly higher than the face-to-face, in addition to the medians having a difference of around 9%. For this reasons, we can say that doctor D7 has a tendency to look more at the patient during virtual consultations than in face-to-face consultations.

Doctor D16's Mann-Whitney U test results in a *p-value* of 0.0059 with a Cohen's *d* size effect of 0.54, meaning there is a statistically significant difference between both types of consultations with a medium size size effect. When looking at the violin plot of D16's distributions, we can clearly see that the virtual distribution is higher than the face-to-face distribution. Additionally, the medians, shown in Table IV, confirm this with a difference of around 16%. Therefore, taking into consideration all results we can say D16 spends significantly more time looking at the patient during virtual consultations than in face-to-face consultations.

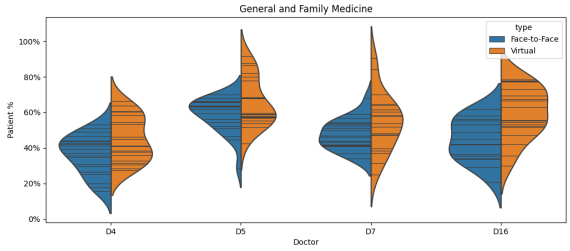

Figure 7: General and Family Medicine doctors violin plots

The Mann-Whitney U test on the joint data distribution results in a *p-value* of 0.00029, indicating the general tendency for doctors from this specialty to look more at the patient during virtual consultations. Looking at Table IV, we can further confirm that the median of the face-to-face distribution is indeed lower than the virtual distribution median, 46.5% and 57.1% respectively. Therefore, in general, the General and Family Medicine doctors have a clear tendency to look more at the patient during virtual consultations than in face-to-face consultations.

X. Discussion

Overall our findings provide another insight into the doctor-patient relationship. More importantly, they provide a look at the effect of virtual consultations on the doctor-patient relationship, specifically on their effect on the doctor's behaviour. In three of the four medical specialties analyzed, we found a pronounced general tendency for doctors to look more at the patient during virtual consultations. However, the other medical specialty (Gynaecology/Obstetrics) seems to be the least affected by the consultation environment change. Half of the doctors present no differences or tendencies between consultation environments, leading to the Gynaecology/Obstetrics specialty presenting very few tendencies between consultation environments. Even though these results are not enough to make robust generalizations about medical specialties as a whole, they still provide an insight into the effect of virtual consultations on doctor behaviour. Consequently, these promising initial results deserve further study and corroboration to better understand the doctor-patient relationship.

Concerning the individual doctors' results, these can be analyzed robustly through the Mann-Whitney U test and Cohen's *d* effect size results. Out of the fourteen participating doctors, seven showed statistically significant differences with medium-sized size effects, while another five showed tendencies to have higher *Patient%* in the virtual environment, with

only two showing no differences in gaze behaviour. Overall, these results show the virtual consultation environment's clear effect on the doctors from the study population, since all of the doctors, except for two, are affected by the change in consultation environments. Additionally, these results also show the virtual consultation medium does not degrade eye contact since no doctor looked at the patient less during virtual consultations.

Our study has some limitations. The first limitation relates to a caregiver being present and the patient in some face-to-face consultations recorded. The way we classify gaze estimates classified any gaze towards the caregiver as gaze towards the patient, which might lead to some overestimation. However, the goal of measuring the division of the doctor's attention between screen and patient is still accomplished. The second limitation is related to the way we classify gaze direction, which leads to the *Patient%* metric being a estimation of the real time percentage the doctor spends looking at the patient. The zone used to define the *Patient* zone was the area on the side of the screen where the patient is located. Thus, the *Patient* zone not only defines where the patient is located but also the surrounding area. The assumption used to make this division is that whenever the doctor looks to the side of the screen where the patient is, the doctor is looking at the patient. This is a strong assumption which leads to very good estimations of the real time the doctor spends looking at the patient through the *Patient%* metric. Another possible issue that was accounted for before the start of the study is related to the Hawthorne effect. The Hawthorne effect is a psychological effect that refers to the change in subjects' behaviours when they know they are being observed. To mitigate these effects, we kept the actions the doctor had to make that did not follow the regular consultation routine to a minimum, which is accomplished by only asking the doctor to click the *Record/Stop* button in the Recording interface in the beginning and at the end of the consultation.

Our study also has its strengths. The use of an automated gaze estimation pipeline led to the capture of a much larger number of interactions than previous studies of gaze in applied settings. While [26] provided 100 clinical interactions divided between 14 doctors, our study provided 260 face-to-face clinical interactions and 260 virtual clinical interactions. The use of an automated gaze estimation pipeline also allows us to study the gaze in more detail than with manually annotated gaze estimates. Additionally, using an appearance-based gaze estimation system like *Gaze360* provides a method that can be implemented in a variety of situations since it only needs an off-the-shelf webcam to be feasible.

In addition to assessing the virtual consultation environment's effect on doctor gaze behaviour, this study also provides, to our knowledge, the first implementation of an appearance-based gaze estimation pipeline in the clinical setup using an off-the-shelf webcam. This task came with many challenges, in addition to the usual robustness against head pose, glasses and environment changes, the gaze estimation system had to be able to deal with the obligatory use of masks. Since the use of masks was not common before the pandemic, all the gaze estimation datasets and systems were not tailored for the task of estimating the gaze of a mask-wearing subject. The solution, which consisted in using a different mask that did not occlude the mouth and nose region, worked remarkably well through the various tests performed and continued to work throughout the study. In the end, the gaze estimation pipeline implemented provides an easy way to record and analyze the doctor's gaze behaviour during consultations. It opens up many applications in the study of non-verbal communication in the doctor-patient relationship, some of which we will suggest in Section XI-A.

XI. Conclusions

In conclusion, this study successfully implemented a 3D appearance-based gaze estimation pipeline in the clinical setup. Additionally, the gaze estimation pipeline assessed the impact of the virtual consultations on the amount of time the doctors from the study population spend looking at the patient. It found that doctors from the study population spend significantly more time looking at the patient during virtual consultations, with a few exceptions. Also, when looking at the medical specialties analyzed, we found that three out of the four specialties present prominent tendencies to look more at the patient during virtual consultations when compared to face-to-face consultations. None of the specialties look less to the patients, suggesting that virtual consultations do not impact the eye contact between doctor and patient negatively. Conversely, these results show the sizable positive impact of virtual consultations on doctors' gaze behaviour.

The successful implementation of the gaze estimation pipeline allows us to analyze doctors' gaze behaviour in more detail and robustly than before. This ability is crucial to analyzing the impact of non-verbal communication in the physician-patient relationship to improve the quality and efficiency of the health care provided. The applications of gaze estimation in the clinical setup are immense. It could be used in the assessment of clinical setup changes. We assessed the impact of virtual consultations against face-to-face consultations. However, many other setup changes can also be analyzed. Software updates can be analyzed to increase doctors' efficiency when navigating the interfaces on the computer. Consultation room changes impact on doctor's gaze behaviours like lighting conditions changes and patient positioning changes can be analyzed thoroughly. More interestingly, gaze estimation can be used for digital biomarker analysis. During a consultation, the patient's attention can be analyzed to see how it evolves throughout the visit. Children's gazes during pediatric therapy sessions could be analyzed to assess if they are focusing on what the therapist wants and improve the overall therapy sessions in the future. Like pediatric therapy, patients with neurological/psychological disorders gaze could be studied during therapy sessions to assess the effectiveness of the therapy sessions and possibly improve them in the future.

In summary, doors opened by gaze estimation are immense, which, coupled with the study of other non-verbal cues, will enable the study of the relationship between people's non-verbal behaviours and actions.

A. Future Work

The impact of virtual consultations on doctors' gaze behaviour is apparent. However, it should continue to be analyzed to provide us with a better understanding of the consequences to the physician-patient relationship. Therefore, a large scale study involving more doctors with a higher number of consultations recorded should be performed. A study like this is essential to support and corroborate the initial results obtained to fully understand the impact of virtual consultations on doctors' gaze behaviour. In addition to a larger amount of samples collected, this future project could also include the extraction of more gaze features along with the *PoG*s, in the form of saccades, smooth pursuits or even face landmarks. This would allow for a deeper study of the doctor's gaze patterns.

Another work that could be done in support of this study is on the correlation between the *Patient%* metric and the actual percentage of time the doctor spent looking at the patient. To obtain the actual percentage, the doctor has to put on the eye-trackers that look like glasses, which provide images of the Field-Of-View (FOV) of the doctor and the gaze located in the FOV image (method used in [26]). Finding the exact relationship between the *Patient%* and the actual percentage of time is an important step on the support of these studies results.

Additionally, we need to study more than gaze to study the total impact of non-verbal communication in the physician-patient relationship. Therefore, coupling gaze estimation with other types of machine learning algorithms to analyze other non-verbal cues like voice pitch, monotony, facial expressions, or body posture is the next step in the studying the impacts of non-verbal communication in the physician-patient relationship.

References

[1] Jennifer Fong Ha and Nancy Longnecker. Doctor-patient communication: a review. *Ochsner Journal*, 2010.

[2] Rainer S Beck, Rebecca Daughtridge, and Philip D Sloane. Physician-patient communication in the primary care office: a systematic review. *The Journal of the American Board of Family Medicine*, 2002.

[3] Judith A. Hall, Jinni A. Harrigan, and Robert Rosenthal. Nonverbal behavior in clinician—patient interaction. *Applied and Preventive Psychology*, 1995.

[4] Marianne Schmid Mast and Gaëtan Cousin. The role of nonverbal communication in medical interactions: Empirical results, theoretical bases, and methodological issues. *The oxford handbook of health communication, behavior change and treatment adherence*, 2013.

[5] Yuval Hart, Efrat Czerniak, Orit Karnieli-Miller, Avraham E. Mayo, Amitai Ziv, Anat Biegon, Atay Citron, and Uri Alon. Automated video analysis of non-verbal communication in a medical setting. *Frontiers in Psychology*, 2016.

[6] Marianne Schmid Mast. On the importance of nonverbal communication in the physician–patient interaction. *Patient Education and Counseling*, 2007.

[7] J Bird and SA Cohen-Cole. The three-function model of the medical interview. an educational device. *Advances in psychosomatic medicine*, 1990.

[8] Tianyi Tan, Enid Montague, Jacob Furst, and Daniela Raicu. Robust physician gaze prediction using a deep learning approach. In *Proceedings of the IEEE International Conference on Bioinformatics and Bioengineering (BIBE)*, 2020.

[9] Daniel Gutstein, Enid Montague, Jacob Furst, and Daniela Raicu. Optical flow, positioning, and eye coordination: Automating the annotation of physician-patient interactions. In *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019.

[10] Laurence R. Young and Dr. Sheena. Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, 1975.

[11] Feng Lu, Yusuke Sugano, Takahiro Okabe, and Yoichi Sato. Adaptive linear regression for appearance-based gaze estimation. *IEEE transactions on pattern analysis and machine intelligence*, 2014.

[12] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. Appearance-based gaze estimation in the wild. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[13] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. Eye tracking for everyone. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[14] Tobias Fischer, Hyung Jin Chang, and Yiannis Demiris. RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[15] Petr Kellnhofer, Adria Recasens, Simon Stent, Wojciech Matusik, , and Antonio Torralba. Gaze360: Physically unconstrained gaze estimation in the wild. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019.

[16] Anjith George and Aurobinda Routray. Real-time eye gaze direction classification using convolutional neural network. *CoRR*, 2016.

[17] Xucong Zhang, Yusuke Sugano, and Andreas Bulling. Evaluation of appearance-based methods and implications for gaze-based applications. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2019.

[18] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. It's written all over your face: Full-face appearance-based gaze estimation. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, 2017.

[19] Qiong Huang, Ashok Veeraraghavan, and Ashutosh Sabharwal. Tabletgaze: dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets. *Machine Vision and Applications*, 2017.

[20] Seonwook Park, Shalini De Mello, Pavlo Molchanov, Umar Iqbal, Otmar Hilliges, and Jan Kautz. Few-

shot adaptive gaze estimation. In *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, 2019.

[21] Seonwook Park, Emre Aksan, Xucong Zhang, and Otmar Hilliges. Towards end-to-end video-based eye-tracking. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[22] Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato. Learning-by-synthesis for appearance-based 3d gaze estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[23] Alex Graves, Santiago Fernández, and Jürgen Schmidhuber. Bidirectional lstm networks for improved phoneme classification and recognition. In *International conference on artificial neural networks*, 2005.

[24] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, 2015.

[25] Patrick H Zimmerman, J Elizabeth Bolhuis, Albert Willemsen, Erik S Meyer, and Lucas PJJ Noldus. The observer xt: A tool for the integration and synchronization of multimodal signals. *Behavior research methods*, 2009.

[26] Chiara Jongerius, H. Boorn, Timothy Callemein, N. Boeske, Johannes Romijn, E. Smets, and Marij Hillen. Eye-tracking analyses of physician face gaze patterns in consultations. *Scientific Reports*, 2021.

[27] Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 1947.

[28] J Cohen. Statistical power analysis jbr the behavioral. *Sciences. Hillsdale (NJ): Lawrence Erlbaum Associates*, 1988.

[29] Jianzhu Guo, Xiangyu Zhu, Yang Yang, Fan Yang, Zhen Lei, and Stan Z Li. Towards fast, accurate and stable 3d dense face alignment. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.