Universidad Autónoma de San Luis Potosí

Doctorado Institucional en Ingeniería y Ciencia de Materiales (DICIM)

"Seguimiento de la mirada utilizado como instrumento para la evaluación del campo visual"

"Eye tracking used as an instrument for visual field evaluation"

Tesis que para obtener el grado de

Doctor en Ingeniería y Ciencia de Materiales

Presenta:

Eduardo Arturo Martínez González

Directores de tesis:

Dr. Víctor Hugo Méndez García
Dr. Francisco Alfonso Alba Cadena

San Luis Potosí, S.L.P.   Junio de 2024.

To the first teachers in my life: my parents.

Abstract

Perimetry is a fundamental test used for the monitoring and detection of glaucoma as well as some retinal diseases and brain injuries. The objective of this test is to evaluate the extension and condition of the visual field (VF) using visual stimuli such as cards, spheres, lights, fingers, etc. For years, Standard Automated Perimetry (SAP) has been the most widely used perimetry method by ophthalmologists for the VF evaluation. However, due to the subjectivity of the test as the main drawback faced by SAP, new paradigms has been proposed for the evaluation of the VF, which seek to maintain a balance between the evaluation time and the precision of the results. Eye Movement Perimetry (EMP) is a proposed paradigm that has shown advantages over SAP by using the patient's gaze paths as responses during the evaluation. However, the reliability of the EMP results depends on the correct acquisition, manipulation and interpretation of gaze paths. Nowadays, most of research in EMP is focused on showing the clinical feasibility of the paradigm and the advantages of the paradigm over SAP, while very little research is focused on the processing and analysis of gaze paths.

In this thesis we propose a paradigm for the VF evaluation based on EMP and machine learning techniques. The proposed paradigm is designed to evaluate the central thirty degrees of VF using the staircase strategy and the visual patterns C30-2 or C24-2. Using machine learning techniques for trajectory analysis, the paradigm is able to determine which of the visual stimuli presented during the assessment were seen and which were not seen, allowing reconstruction of the subject's monocular VF. To validate the paradigm, a series of monocular evaluations were performed on 20 young participants. The results obtained were compared with a ground truth, obtained an average accuracy index > 96.5%.

Resumen

La perimetría es una prueba esencial para la detección y seguimiento del glaucoma así como también algunas enfermedades de la retina y lesiones cerebrales. El objetivo de esta prueba es evaluar la extensión y la condición del campo visual (VF, por sus siglas en inglés) utilizando estímulos visuales como tarjetas, esferas, luces, dedos, etc. Por años, la perimetría automática estandar (SAP, por sus siglas en inglés) ha sido el método más utilizado por los oftalmólogos para la evaluación del VF. Sin embargo, debido a la subjetividad de la prueba como el principal problema que enfrenta SAP, se ha generado la necesidad de desarrollar nuevos métodos para la evaluación del VF que permitan mantener un balance entre el tiempo de evaluación y la precisión de los resultados. La perimetría de movimiento ocular (EMP, por sus siglas en inglés) es uno de los métodos propuestos para la evaluación del VF que ha mostrado ventajas sobre SAP al utilizar los movimientos oculares del paciente como respuestas durante la evaluación. Sin embargo, la confiabilidad de los resultados de EMP depende de una correcta adquisición, manipulación e interpretación de las trayectorias visuales. Actualmente, la mayoría de las investigaciones referentes a EMP están enfocadas en resaltar la viabilidad clínica del método y sus ventajas sobre SAP, mientras que muy pocas investigaciones se centran en el procesamiento y el análisis de trayectorias.

En esta tesis presentamos un método para la evaluación del VF basado en EMP y técnicas de aprendizaje de máquinas. El método propuesto está diseñado para evaluar los treinta grados centrales del VF utilizando la estrategia de umbral en escalera y los patrones visuales C30-2 o C24-2. Utilizando técnicas de aprendizaje de máquinas, nuestro método es capaz de determinar cuáles de los estímulos utilizados durante la evaluación fueron o no vistos, permitiendo así realizar una reconstrucción del VF monocular del sujeto. Para validar el método, se realizó una serie de evaluaciones monoculares con 20 participantes jóvenes. Los resultados obtenidos fueron comparados con los resultados proporcionados por un experto, obteniendo así índices de precisión $> 96.5\,\%$.

Contents

Chapter 1

Introduction

The eyesight is a complex system that allows some living beings to acquire information from the environment through images. Thanks to this sophisticated and complex sense, humans and some animals have developed abilities increasing their opportunity to survive. Perhaps for humans, the eyesight is the most dominant sense in their lives. It makes most activities easier, such as driving, walking, reading, cooking, working, etc. than using only the touch or a combination of other senses. However, rarely people care about their visual system until a damage (often irreparable) in their visual field (VF) begins to appear. The VF is the total area than a person can see when stares in a fixed point. According to the last statistics from the World Health Organization (WHO), at least 2.2 billion people have a vision impairment, where at least half of these cases could have been prevented [1]. The main causes of vision impairment until 2023 were age-related macular degeneration, cataract, diabetic retinopathy, uncorrected refractive errors and glaucoma [1], where the majority of these can generate total or partial irreversible blindness.

For years, different ophthalmologic tests and devices have been developed to evaluate, monitor or diagnose different visual impairments. One of these tests used for more than 150 years, is the perimetry, whose goal is to evaluate the people's VF using different techniques and stimuli such as spheres, fingers, lights color cards or cards with patterns. Nowadays, Standard Automated Perimetry (SAP) is considered as the gold standard for VF assessment and as a useful test for diagnosing and monitoring certain neurological and ophthalmological diseases [3] [4] and has remained without major changes for approximately 50 years [5]. In this test, whose average evaluation time per eye is between 4 to 6.5 minutes, a device called campimeter is used to strategically project light stimuli over a hemispherical screen with a uniform luminous background. During the VF evaluation, the patient must simultaneously perform two tasks: 1) stare at a reference stimulus at all times while peripheral stimuli are presented in different positions and with different intensities for short times, and 2) press a button when a peripheral stimulus is perceived. Once the test is finished, the campimeter shows a reconstruction of the patient's VF with the values of the sensitivities perceived during the evaluation.

Collaboration between patient and technician is essential for reliable results in SAP. However, SAP is limited by large test-retest variability and exposed to different factors that may affect its reliability, such as patient motivation, patient motricity, patient

fatigue, and technician performance [6, 7]. Perhaps, the most difficult and tiring task in SAP is to maintain the gaze on the reference stimulus for long periods of time, as it is an unnatural task for people. Studies by Kosnik et al. [8] and Demirel et al. [9] have supported this fact, concluding that fixing the gaze on a reference point for long periods affects the patient's performance during the test and the reliability of his results.

In order to solve some problems faced by SAP, methods based on Visual Evoked Potential (VEP) [10–13] or Eye Movement Perimetry (EMP) **(see the section 2.3)** have been proposed as new alternatives to evaluate the people's VF. Nevertheless, the difficulty of obtaining and/or processing the acquired signals in a robust way are weaknesses of these methods that should be improved.

In this thesis we proposed a methodology based on EMP and machine learning in order to evaluate the VF. This methodology attempts to reduce the subjective responses taking decisions according to the gaze trajectories obtained during the evaluation.

1.1 Thesis goals

To develop threshold test for the evaluation of the VF based on EMP and machine learning methods in order to produce reliable results from a low-cost portable setup.

1.1.1 Particular goals

1. To implement a VF evaluation base on FASTPAC threshold strategy for the 30° central using the standard clinical patterns C24-2 and C30-2.

2. To detect the blind spot during the test.

3. To generate a reconstruction map of the VF from the gaze path analysis.

4. To validate the prototype system.

1.2 Thesis organization

This thesis is organized as follows:

- **Chapter 2** presents the basic concepts and techniques used in this thesis.

- **Chapter 3** presents a description of the methodology implemented for the VF evaluation.

- **Chapter 4** presents the results, conclusions of the paradigm implemented.

- **Chapter 5** present the discussion of the results as well as some ideas for future work.

- **Appendix A** presents a description of the machine learning techniques implemented in Chapter 3.

- **Apendix B** shows the visual maps obtained for each participant during the validation of the methodology.

- **Apendix C** shows a list of publications and participation in conferences about our work.

Chapter 2

Preliminary concepts

2.1 The visual field

The VF is a fundamental concept to understand the perimetry's goal. According to [14, 15] the VF corresponds to the total area than a person can see when focuses the vision on a static reference point. When this area corresponds only to one eye, the VF is said to be monocular. On the other hand, when this area corresponds to the combination of both left and right monocular VF, it is said to be binocular or total. Figure 2.1 shows the monocular and binocular VF.
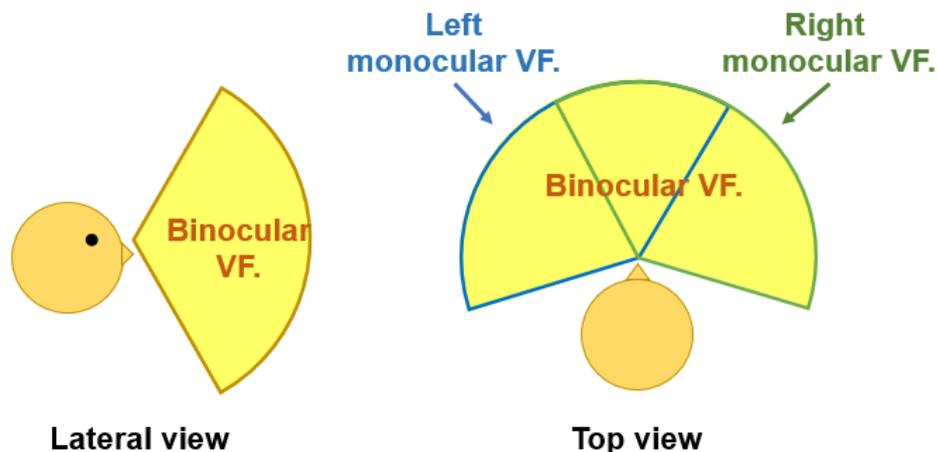
Figure 2.1: A 2D representation of the VF. The area under the green shape corresponds to the right monocular VF, whereas the area under the blue shape corresponds to the left monocular VF. The total VF is represented in yellow. (b) Average limits of the VF and blind spot location for the right eye.

The VF has been studied for many centuries. Perhaps the pioneer in quantifying its extension was Ptolemy around 150 BC who noticed that it was a roughly circle shape [5, 16]. However, in 1510 Leonardo Da Vinci contributed to this quantification by reporting that the temporal field was 90° [16]. Later in the early 1800s, Thomas Young reported the monocular VF size as 50° upwards, 70° downwards, 60° in the nasal direction and 90° to temporal direction, where Johannes Purkinje around the 1820s refined those limits reported by Young [5, 16]. At present, the extension of the monocular VF reported in the literature taking as its origin the center of the macula has

on average the following measurements: 60° upwards, 70° downwards, 60° in the nasal meridian and 90° to temporal meridian as shown in Figure 2.2; however these limits may vary slightly from person to person [14,15]. In the monocular VF a particular zone called the blind spot, which corresponds to the retinal area without photoreceptors, is the unique natural blind zone. This elliptic-shape zone with horizontal width of $6° - 10°$ and a height of $10° - 20°$ is positioned between $12° - 20°$ and slightly below the temporal meridian as shown in Figure 2.2 [17,20].
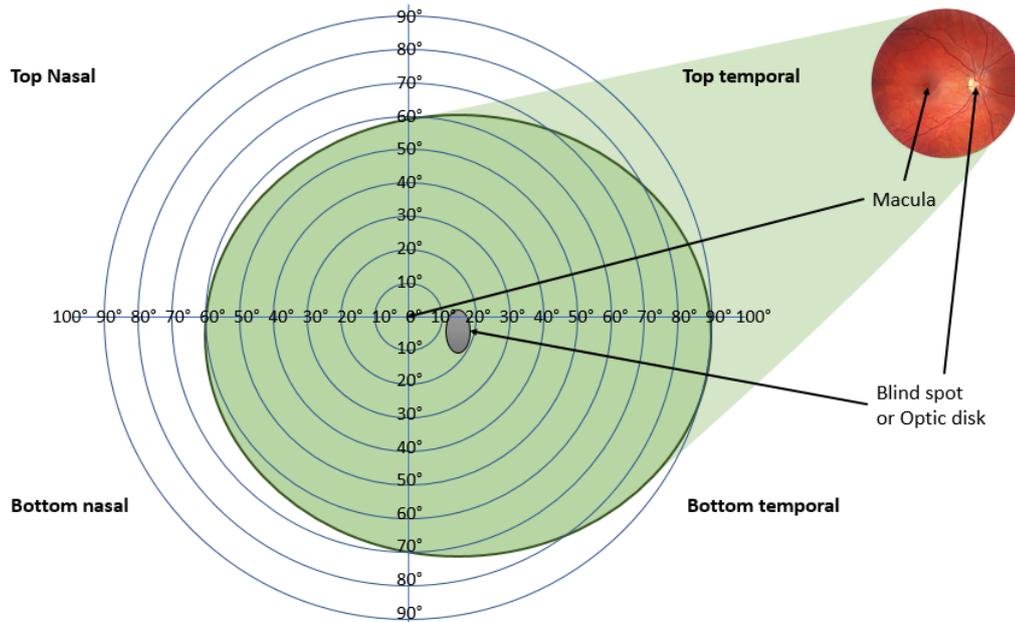

Figure 2.2: Average limits of the VF and blind spot location for the right eye.

The visual sensitivity across the VF is variable across the field and also is different for each person. This non-linear sensitivity is high in the first 10° and it decreases beyond this. The reason for this is related to the distribution of rods (photoreceptors responsible for vision at low light levels) and cones (photoreceptors responsible for high spatial acuity and the color vision) in the retina as shown in Figure 2.3. A useful analogy to understand the VF behaviour is the Traquair's island of vision in a sea of darkness: "Like an island arising out of the ocean, there is a shore that marks the absolute limit of vision. Just as an island rises from the shore to some inland prominence or hill, visual sensitivity increases as one proceeds from the peripheral boundary of vision to a peak at the fovea" [17]. According to this comparison, the blind spot could be seen as a deep pit in the island where the sensitivity in its border is low. Studies performed by Jaffe G. et al. [21] and Johnson C. et al. [22] have shown that sensitivity in the VF decreases with the age, where changes in the ocular media, linear reduction in pupil diameter and neural losses in the retino-geniculostriate pathway are some of the main causes [17].
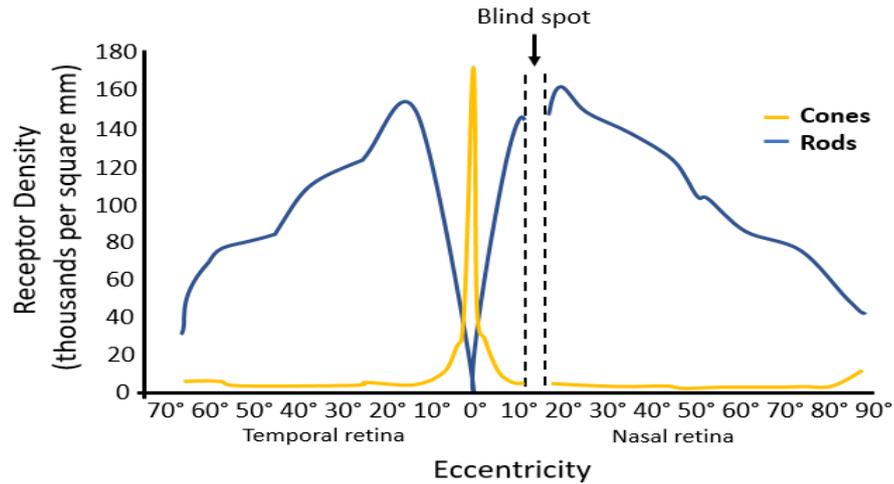
Figure 2.3: The distribution of rods and cones.

2.2 Manual and automated perimetry

The perimetry's goal is to detect anomalies in the VF by measuring the visual sensitivity in different positions. These anomalies are detected with the help of visual stimuli such as lights, color cards, fingers or spheres, which can be static or kinetic during the evaluation. To detect these anomalies, two types of perimetry can be used: manual and automated; each with its advantages and disadvantages listed below.

-Manual perimetry

Manual perimetry requires an operator to present the stimuli, monitor the patient's fixation and record his/her response if the latter is necessary. The Goldmann perimetry is the most common device used in manual perimetry; nevertheless, the VF evaluation can be done without it (e.g. by confrontation or tangent screen). During the evaluation, the patient must stare at a static reference while a stimulus is moved slowly from a blind zone in the periphery to the center of the VF. When the patient sees the stimulus, the stimulus' position is recorded by pressing a button and the process is repeated again in a new peripheral position. This interaction between patient and specialist is performed as many times as necessary evaluating the area of interest of the VF. At the end of the test, as shown in Figure 2.4, the recorded positions are interpolated forming *isopters* which represent the limit of the area where a specific stimulus can be seen.

- Advantages
  1. Given that the specialist monitors the patient's fixation, the responses given by the patient can be discarded if him/her is distracted, re-evaluating those discarded positions.
  2. The test can be customized according to the interest area to be evaluated by the specialist.
  3. If a suspicious visual defect is detected during the evaluation, the specialist can go back and evaluate it in detail.
  4. It can better detect visual defects outside the 30° central and some residual defects than the automatic perimetry.

6

- Disadvantages

  1. The results of the evaluation and its execution depend from the specialist's skills.

  2. Knowledge of anatomy and visual diseases are required for the specialist to perform the test.

  3. It requires more time and attention from the patient to perform the VF's evaluation.
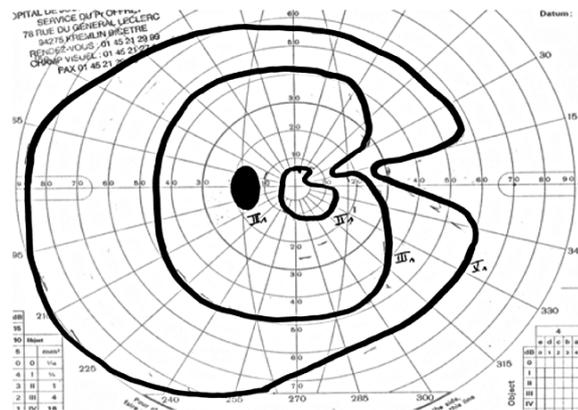
  4. It is a subjective test.

Figure 2.4: Real results of the evaluation of the VF using manual perimetry. Note how the isopters indicate an approximation of the patient´s healthy visual area. Source: An Atypical Case of Foster Kennedy Syndrome - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Goldman-visual-field-of-the-left-eye-showing-a-nasal-deficit-with-an-excluded-blind-spot_fig2_49687733.

-Automated perimetry

Automatic or computerized perimetry is performed by a computerized device that uses standard techniques and strategical fixed positions to evaluate the VF. These techniques and positions help reduce the patient-operator interaction, but demand additional concentration from the patient. SAP or white on white perimetry (because white stimuli are projected on a white background) is considered as the gold standard for the VF evaluation. SAP is performed by a device called campimeter, where the Humphrey field analyzer is the most used device in the clinical field; however, other devices such as Octopus, Topcon and Dicon can perform the automated perimetry.

Like the manual perimetry, the patient must stare at a static reference at all times. However, unlike manual perimetry, in which a stimulus is moved from a non-seeing area to a seeing area, SAP uses static light stimuli (sometimes changing its light intensity) that flicker for a few milliseconds. Every time the patient detects a stimulus, he/she presses a button to record his/her response. At the end of the test, the device gives different results such as numerical map of sensitivities, grayscale map and deviation maps as shown in Figure 2.5.
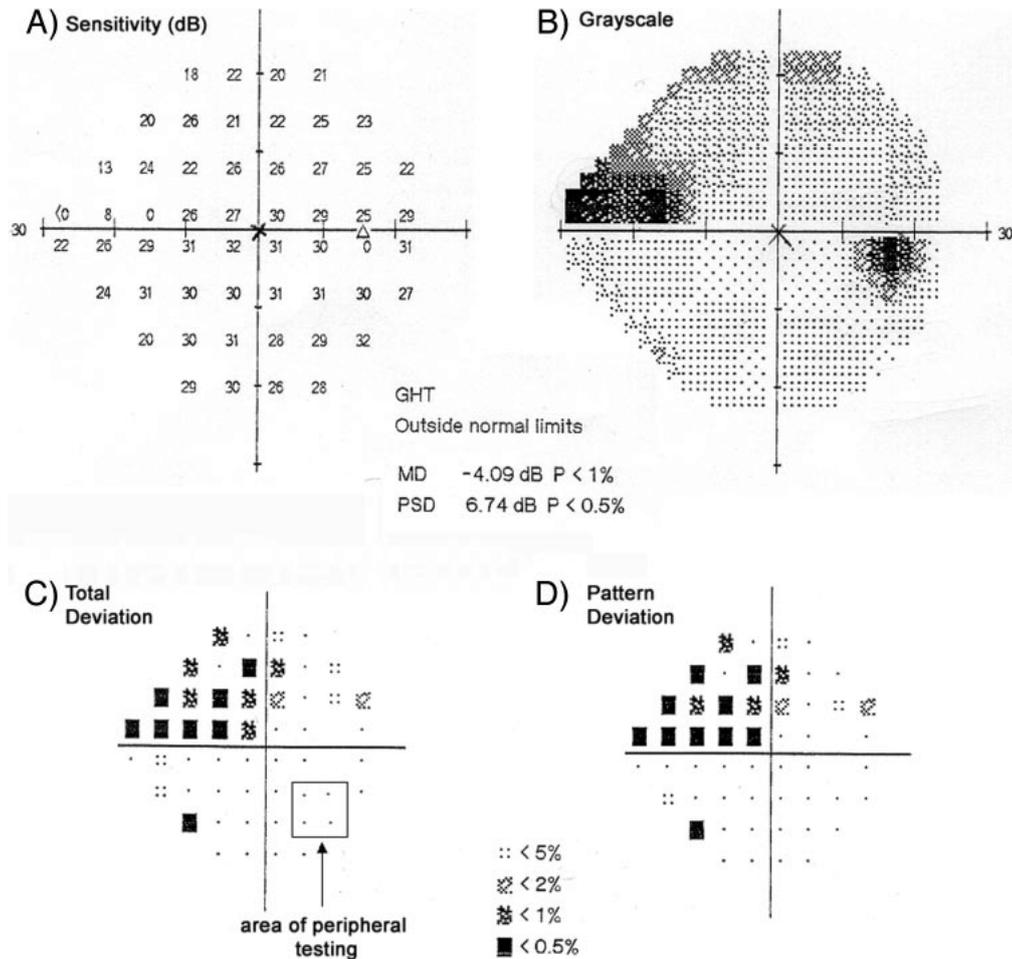
Figure 2.5: Sample right eye visual field (HFA, SITA Standard 24-2) from one glaucomatous participant (aged 66) illustrating placement of the peripheral stimuli in an area of relatively normal performance (within the normative 95% confidence limits for the total deviation plot). Source: Psychophysical Measurement of Neural Adaptation Abnormalities in Magnocellular and Parvocellular Pathways in Glaucoma - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Sample-right-eye-visual-field-HFA-SITA-Standard-24-2-from-one-glaucomatous-participant_fig1_8544315.

- Advantages

  1. Anyone with a basic knowledge of the device can run the test.
  2. It is standardized and provides quantitative results.
  3. It gives more precision in the 30° and 10° central than manual perimetry.
  4. It uses standardized strategies and patterns that reduce the patient-operator interaction.

- Disadvantages

  1. It is difficult to perform in children or adults with impaired cognitive skills.
  2. It does not have flexibility to evaluate specific defects or areas.

3. It requires greater concentration on the part of the patient, and in some cases a longer evaluation time.

4. It is a subjective test.

5. A specialized device is required to perform the test.

2.2.1 Suprathreshold and threshold strategies

The *visual threshold* at a specific location on the retina is the luminance value at which a stimulus has a 50% probability of being detected [16, 17]. As mentioned in the previous section, the goal of perimetry is to measure visual sensitivity in VF to detect visual abnormalities. In perimetry, there are two strategies to locate and/or monitor the visual anomalies: suprathreshold strategy and threshold strategy.
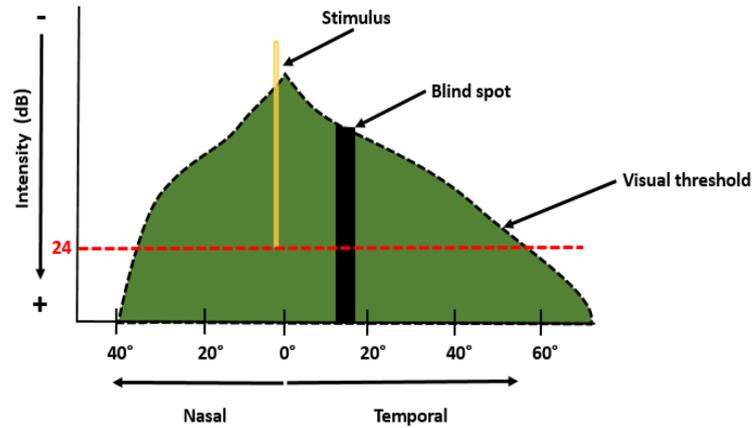
-Suprathreshold strategies

Suprathreshold strategies, used as a screening test, make it possible to locate decreases in visual sensitivity within the VF by means of static light stimuli. These stimuli have an intensity value above the visual threshold, so that a person with a healthy VF should see them without any problem. The intensity values of the stimuli are automatically selected by the campimeter from an internal database; however, if the specialist considers it necessary to modify the intensity value given by default, he/she can do it.

There are three suprathreshold strategies to evaluate the VF: 1) using a single intensity value, 2) threshold related, and 3) two-level suprathreshold or three zone.

- **Using a single level:** This strategy, as shown in Figure 2.6 (a), implements stimuli with an unique intensity value during the VF evaluation (a common intensity value is 24 dB, but it can be changed by the specialist). At the end of the test, the reconstruction of the VF indicates which stimuli were seen and which were not seen. Nevertheless, when implementing this strategy, some central visual defects may be unnoticed. The reason for this is due to the use of stimuli with light intensity values too large for the central visual threshold, so that the stimuli used in this zone have a high probability of being seen by the patient due to a higher visual sensitivity in the central area.

- **Threshold related:** Unlike the previous strategy, in which a constant intensity is used to evaluate VF, the threshold related technique uses light stimuli that exceeds the visual threshold by 2 dB to 6 dB as shown in Figure 2.6 (b). Similar to the results of previous strategy, the results of threshold related strategy indicate whether the stimulus was seen or not seen. However, it does not always detect early abnormalities in the VF.

- **Two-level suprathreshold or three zone:** This technique is based on the previous technique to evaluate the VF. Nevertheless, it takes the technique a step further by classifying the patient's responses as: normal, relative defect, and absolute defect. A normal classification corresponds to a stimulus seen by the subject during the test. On the other hand, if a stimulus is not seen with the initial light intensity, then a stimulus is presented in the same position with the maximum light intensity that the perimeter can provide. If this new stimulus

is seen, then it is classified as a relative error, otherwise, it is classified as an absolute error. Figure 2.7 shows a representation of this technique where the green area exemplifies a patient's VF and the dash line represents the healthy visual threshold.

(a)
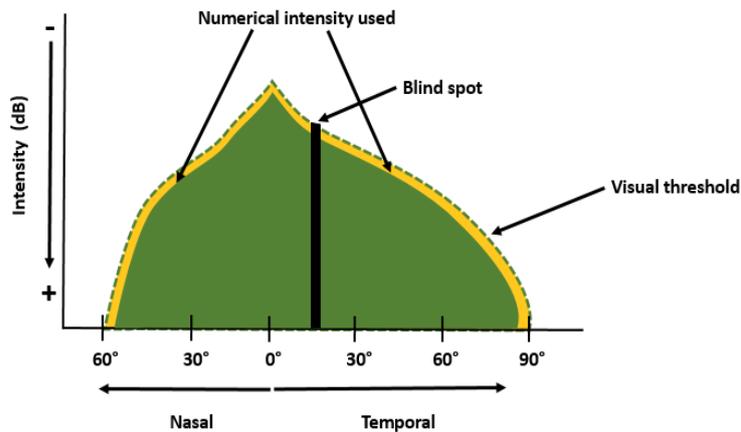
(b)

Figure 2.6: Some suprathreshold techniques. (a) Single level technique. The VF is evaluated with the same intensity (red dash line) in all the strategical positions. Notice how the intensity of the light used in the central areas far exceeds the vision threshold in the center of the VF. (b) One-level suprathreshold strategy. The intensity of the stimuli used during the evaluation (yellow area) is between 2-6 dB above the visual threshold (green dash line).
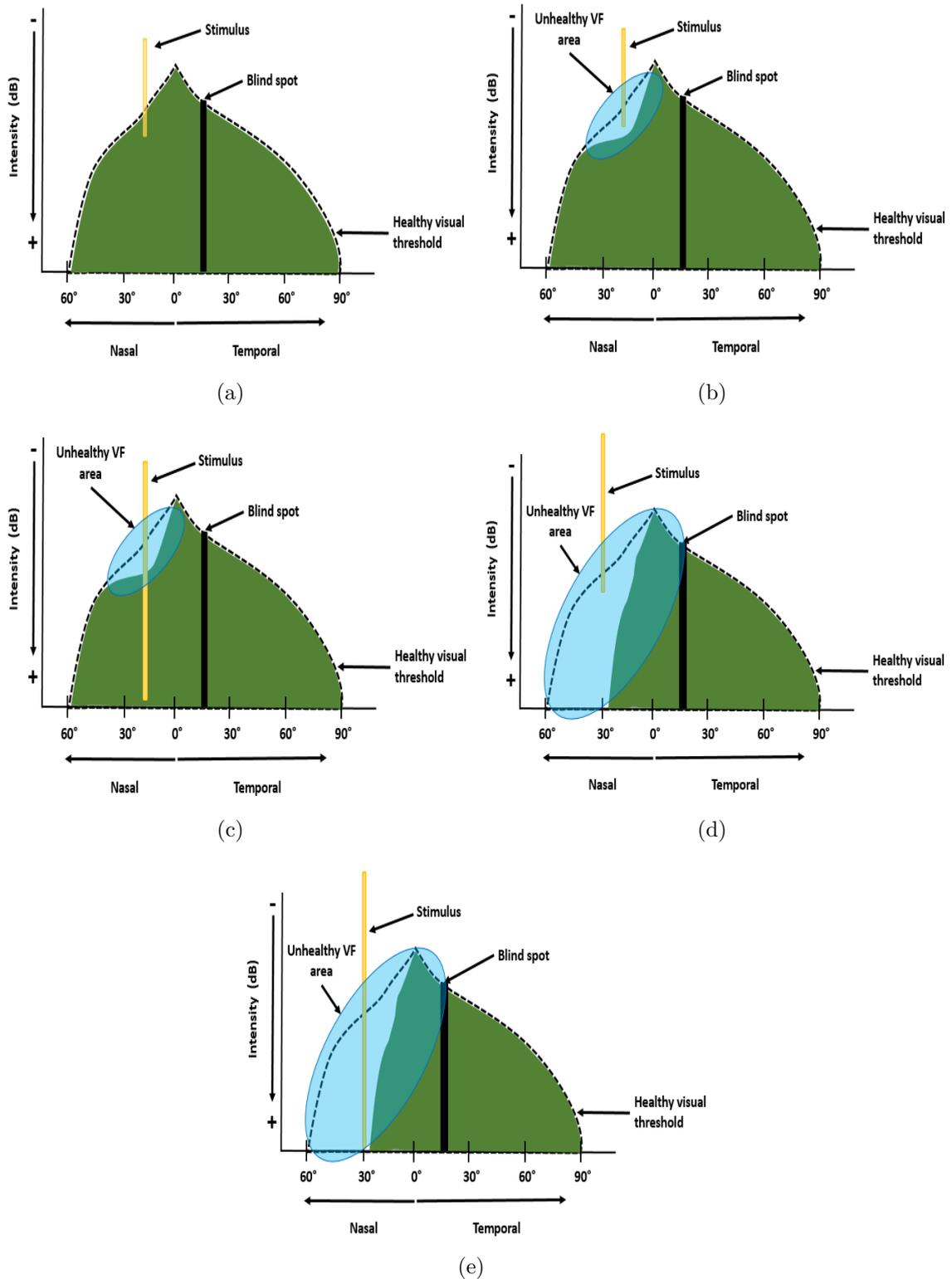
Figure 2.7: Two-level threshold technique classifications. (a) Stimulus seen so it is classified as normal threshold. (b) and (d) Stimulus not seen the first time. (c) Relative defect (stimulus seen with the maximum luminous intensity). (c) Absolute defect (stimulus not seen with the maximum luminous intensity).

-Threshold strategies

The threshold strategies use static stimuli with varying light intensity to determine the patient's visual threshold. These techniques allow the detection of early scotomas, even within the central 10° of the VF. However, threshold strategies require more time to perform the evaluation than the suprathreshold strategies.

- **Bracketing strategy or double crossing:** This technique, considered as a standard technique in static threshold perimetry and used in most glaucoma clinical trials [23], has a simple algorithm but takes about 10-15 minutes of evaluation per eye, whose evaluation time is longer than the other threshold strategies.

  As shown in Figure 2.8, the campimeter begins with a luminous intensity below to the visual threshold (infrathreshold stimulus), and increases its intensity in steps of 4 dB until the patient can see the stimulus (here the first crossing). The stimulus then decreases in intensity in 2 dB steps until the patient cannot see it (here the second crossover). The last value recorded as "seen" by the patient is considered the visual threshold at that position; however, campimeters like Octopus consider the visual threshold as the average value between the last seen and not see luminous intensity (the intensity value of the second crossing).

- **Staircase strategy:** This strategy, also know like FASTPAC strategy (Carl Zeiss Meditec, Inc., Dublin, CA), reduces approximately the evaluation time by 36% compared to the bracketing strategy. As shown in Figure 2.9, the visual threshold approximation begins with an unseen stimulus (infrathreshold) that increases its luminous intensity in 3 dB steps until the patient can see it. This luminous intensity that the patient sees is considered the visual threshold. However, the process can also be reversed; that is, to start with a suprathreshold stimulus and decrease its luminous intensity in 3 dB steps until the patient cannot see it. In this case, the visual threshold is considered to be the last sensitivity to light perceived by the patient.

- **SITA Algorithms:** The Swedish Iterative Threshold Algorithms (SITA), which are based on Bayesian models, were developed to reduce evaluation time without compromising the reliability of the results. At present, SITA algorithms are the most used by specialists, displacing the classic bracketing strategy and FASTPAC strategy. There are two SITA algorithms: SITA standard and SITA fast. The former is analogous to bracketing strategy and the latter is analogous to FAST-PAC strategy. The SITA standard algorithm, which is based on the bracketing strategy, uses 4 dB increments / decrements for the first threshold crossing and 2 dB decrements / increments for the second threshold crossing. On the other hand, SITA-fast uses 4 dB steps to estimate the visual threshold instead of 3 dB steps as the FASTPAC strategy does. Both algorithms reduce the evaluation time by about a half, compared to their analog strategies, which in turn reduces the patient's fatigue during the evaluation.
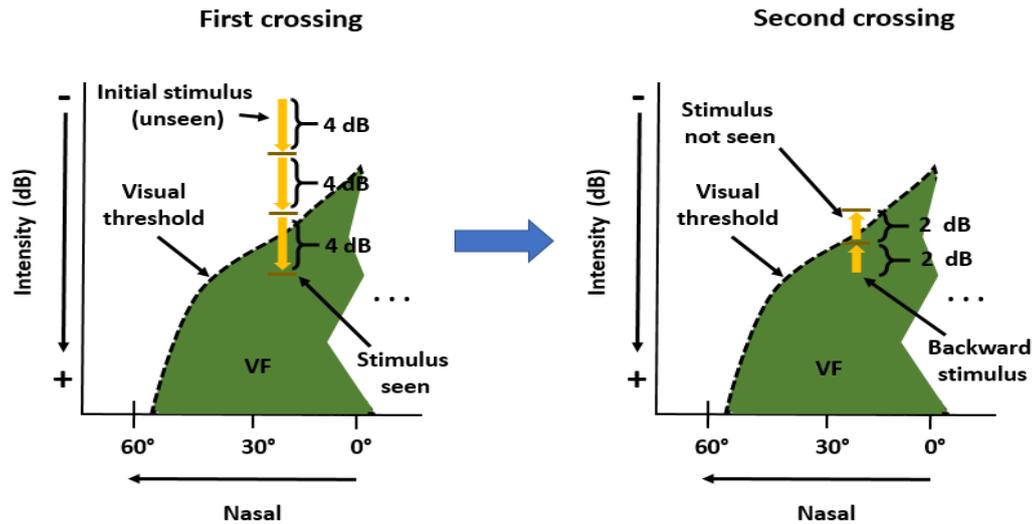
Figure 2.8: Bracketing strategy. The strategy begins with an initially invisible stimulus that increases its intensity until the patient sees it (image on the left). Subsequently, the stimulus is shown by slowly reducing its intensity until the patient cannot see it (right image). The last light intensity seen by the patient is considered the visual threshold.
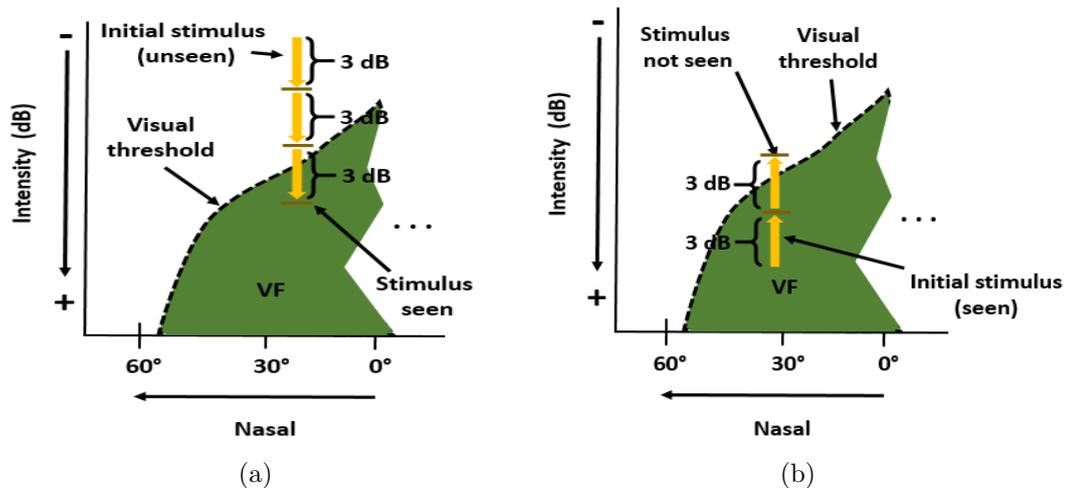
Figure 2.9: Staircase strategy. (a) Visual threshold computed from a infrathreshold stimulus. (b) Visual threshold computed from a suprathreshold.

2.2.2 Visual patterns used in SAP

In theory, visual sensitivity can be measured across the entire VF, however this is impractical and time consuming. In practice, different position grids or patterns are used to evaluate different areas of the VF. Patterns such as C30-2, C24-2, C10-2 and the macular pattern, are used to evaluate the central 30° using the threshold strategy; out of these, the patterns C30-2 and C24-2 are the most used because they allow better detection of glaucoma defects and some neurological diseases. Figure 2.10 shows the distribution of the positions for the C30-2 and C24-2 patterns, where the first uses 76 positions with a distance between positions of 6°, whereas the second uses 54 positions

to evaluate the VF. On the other hand, patterns such as C-40, C-64, C-76 and C-Armaly are used with suprathreshold strategy to explore the central 30°; out of these, C-76 is perhaps the most widely used due to its similar distribution to the C30-2 pattern. Patterns like 60-4 and the nasal step pattern are used to evaluate beyond the central 30°, whereas patterns such as CC-60, CC-81, CC-120, CC-135, CC-246 and CC-Armaly are available to explore the complete VF if it is required by the specialist.

(a)   (b)

Figure 2.10: The most used patterns to evaluate the central 30°. (a) Pattern C24-2 used to evaluate the right eye. (b) Pattern C30-2. The separation between positions in both patterns is 6°.

2.3 Eye movement perimetry

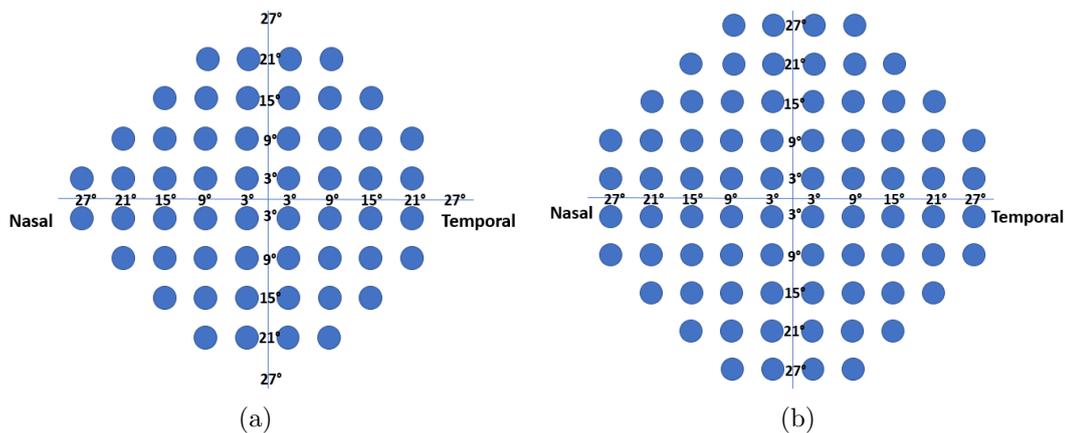EMP is a paradigm proposed to solve some problems faced by SAP such as loss of fixations or the excess of subjective responses. Unlike traditional VF evaluation used in SAP, where the eye movements are restricted and a button is used to record the person's responses, EMP gives eye movement freedom in order to obtain saccadic eye movements (SEM) during the evaluation, where these ocular movements are used as a response during the VF evaluation. Perhaps the first research in SEM dates back to the 1970s, where eye movements began to be analyzed under certain visual stimuli and conditions [31, 32]. However, in the early 1980s, the EMP paradigm began to appear when Jernigan [33] combined SEM and perimetry by evaluating oculomotor responses as an index of perception using a tangent screen and eye tracking technology (biometric model 200 eye movement monitor).

Thanks to technological advancement in eye tracking, devices like video-based Eye Trackers (ET) have proven to be a popular choice for detecting and recording eye movements. These portable and non-invasive devices, depicted in Figure 2.11, contain an infra-red light source used to generate a corneal reflection while a speed camera records the person's eyes or face to be processed in real time. Using computer vision algorithms to detect the center of the pupil and the corneal reflection from each frame, the gaze's position is estimated over a limited area. This technological advantage in eye tracking has allowed to contribute to the EMP, where this paradigm was forgotten for a time (possibly due to the difficulty of data acquisition and the inaccessibility of the devices for it).
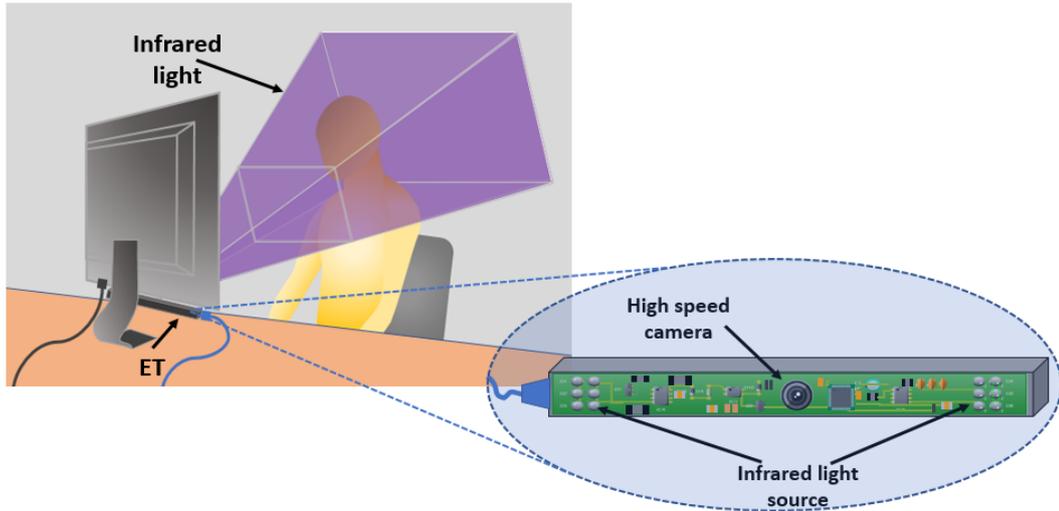
Figure 2.11: Eye tracker

In 1995, Kim et al. [24] collected gaze trajectories from healthy participants and glaucoma patients incorporating an ET to the perimeter. In general, the study results showed similarities with the SAP results, with differences mostly shown in the sensitivity threshold of 4 dB or less. Fourteen years later, Murray et al. [25] proposed a method called Saccadic Vector Optokinetic Perimetry (SVOP) to evaluate the VF in children. SVOP implements a suprathreshold strategy using the C-40 pattern, where the system was validated using data from both healthy and unhealthy children and adults. The results of SVOP were compared with the equivalent Humphrey Field Analyzer C-40 suprathreshold test, obtaining an average percentage of points agreeing with a normal visual field of 99.2% and 99.1% for healthy adults and children, respectively, and 98.8% of points in agreeing with Humphrey results for adults with unhealthy VF. In 2016 a new validation of SVOP was presented [26], where indices such as sensitivity, specificity and accuracy were computed from: 1) analysis of the overall VF pattern, and 2) analysis of the VF points. For the first analysis, the results were 90.9%, 88.5%, and 89.0% for sensitivity, specificity and accuracy respectively; while for the second analysis, the results were 69.1%, 96.9%, and 95.0% for sensitivity, specificity and accuracy, respectively. On average, SVOP reported shorter execution times than SAP for those participants without visual defects (1.37 minutes for adults and 1.5 minutes for children) but for those participants with visual defects, SVOP required longer times than SAP with a mean of 2.4 minutes. In 2018 Murray et al. [27] presented a new version of SVOP using a threshold strategy and a C24-2 pattern. SVOP was compared with SAP C24-2 SITA fast, obtaining similar mean threshold values with a correlation coefficient of 0.95; as well as sensitivity and specificity values of 97.7 % and 77.9 % respectively to identify normal versus abnormal visual fields with SVOP [28]. In 2019 Jones et al. [29] proposed a portable system called Eyecatcher, which implements a suprathreshold strategy based on a modified C24-2 pattern. The results of Eyecatcher were compared with SITA standard C24-2 using the Humphrey Field Analyzer. Eyecatcher reported consistent results and better evaluation times than SAP. A year later, Jones [30] presented a new Eyecatcher version based on C24-2 pattern and a similar threshold algorithm used by the Humphrey Field Analyzer. In this new Eyecatcher version, Jones et. al. reported similar results to SITA standard detecting spatial

differences in sensitivity across the normal VF.

These studies have demonstrated the advantages of EMP over SAP, which are: 1) EMP guarantees the concentration of the patient during the test by allowing to detect when fixation on the reference stimulus is lost, and implementing actions to counter these events. 2) EMP can be implemented as an economic and portable system because the minimum system requirements are a CPU, an screen (i.e a computer screen or a tablet), and a low-cost ET, and 3) the method is implemented by software, so it can be easily configurable according to the requirements. Nevertheless, EMP weaknesses such as evaluation time or robust trajectory analysis need to be improved.

Chapter 3

Visual Field evaluation

In this chapter, we describe the equipment and the methodology used for the VF's evaluation: 1) test paradigm, 2) data analysis, and 3) reconstruction of the VF.

3.1 Prototype system

The prototype system developed for the VF evaluation, which is shown in Figure 3.1, consist of a laptop (windows 7 under 64-bits and intel core i3), an ET (Tobii 4C; Tobii Technology, Stockholm, Sweden) with a operation range between 50 cm - 90 cm and an image sampling rate of 90 Hz, a monitor (23" with 16:9 aspect ratio and a resolution of 1920×1080@60 Hz), an additional keyboard used for the interaction of the participants with the system during the evaluation, a chin rest, and a dark cabin. In order to cover the largest possible evaluation area during the test, the distance between the screen and subject was 51 ±1 cm, which was constantly monitored with the ET.
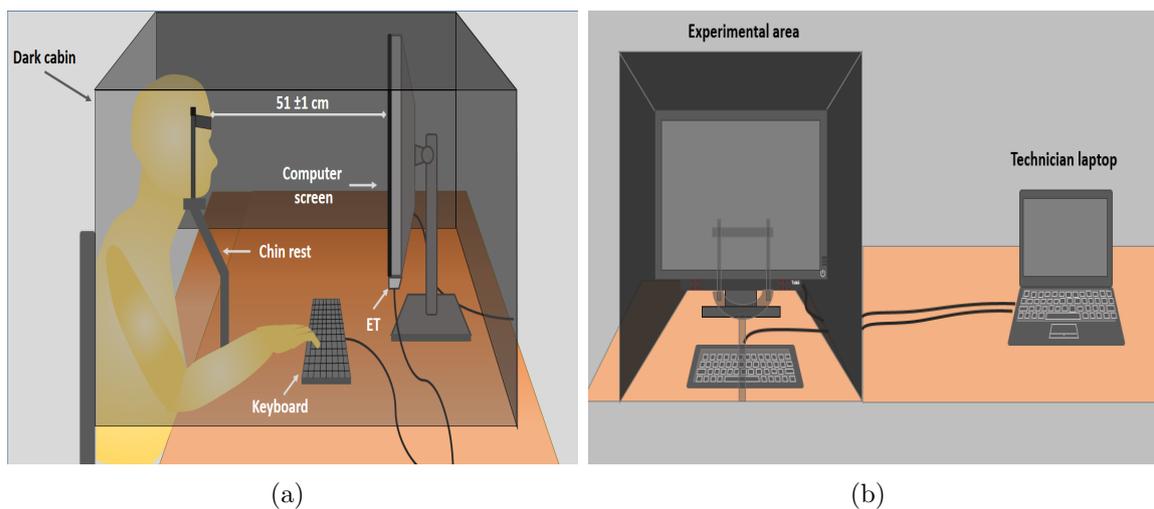

(a)   (b)

Figure 3.1: Prototype setup. (a) Lateral view. (b) Frontal view.

3.1.1 Software used

Tobii Eye Tracking Core software v2.13.4.7864 x86, Processing and R languages were used to develop the test. The first software was used to configure the ET's parameters and calibrate the device. The second one was used to create the graphical environment for the test and for data manipulation during the evaluation, and the last one was used for data processing and analysis. To perform the communication between Processing and the ET, an external library (GazeTrack: Eye-tracking for Processing (Tobii EyeX and 4C) 2.0.4 developed by Augusto Esteves) was used. Additional to the graphical test development, a graphical user interface was developed. The interface, shown in Figure 3.2, collects the basic personal data allows the configuration of the test, also provides the technician with certain graphical information that allows monitoring of patient performance during the test.
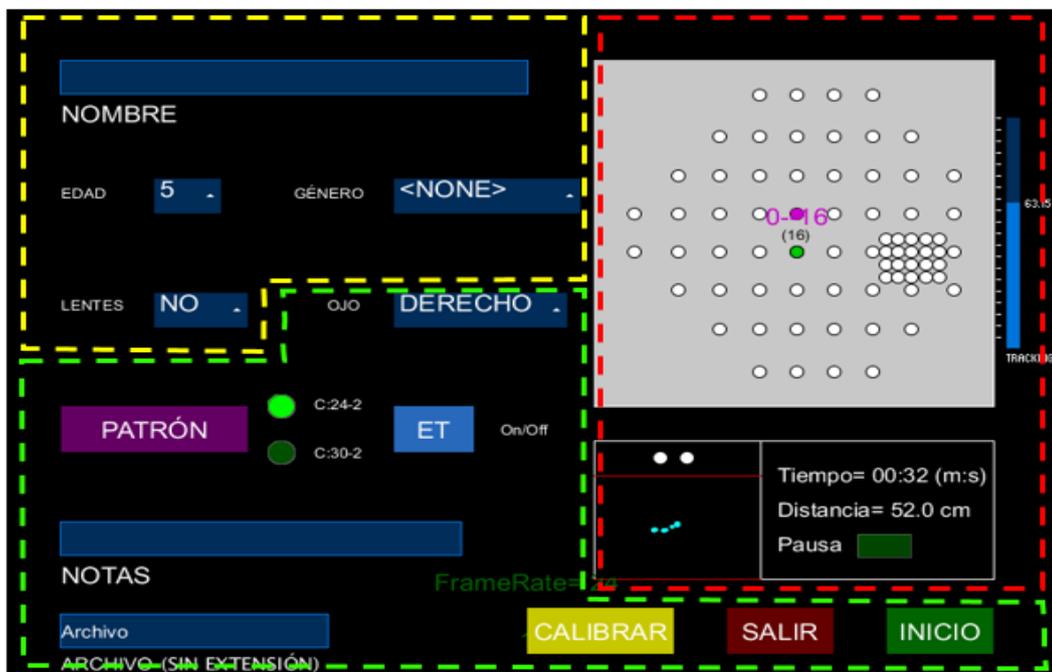
Figure 3.2: Graphical user interface. Basic personal information (yellow zone). Basic configuration area (green zone). Monitor area (red zone).

3.2 Test paradigm

Similar to SAP, the proposed EMP paradigm displays peripheral stimuli to calculate the intensity threshold over the different positions in the VF while the person is looking at a static reference stimulus, located on the center of the screen. The paradigm is based on the FASTPAC strategy and the popular C24-2 and C30-2 patters. During the evaluation, an ET is used to record and monitor eye movements at all times. However, the only ET data stored for analysis corresponds to the set of gaze positions (i.e., the gaze trajectory) acquired during the time period in which each peripheral stimulus is displayed.
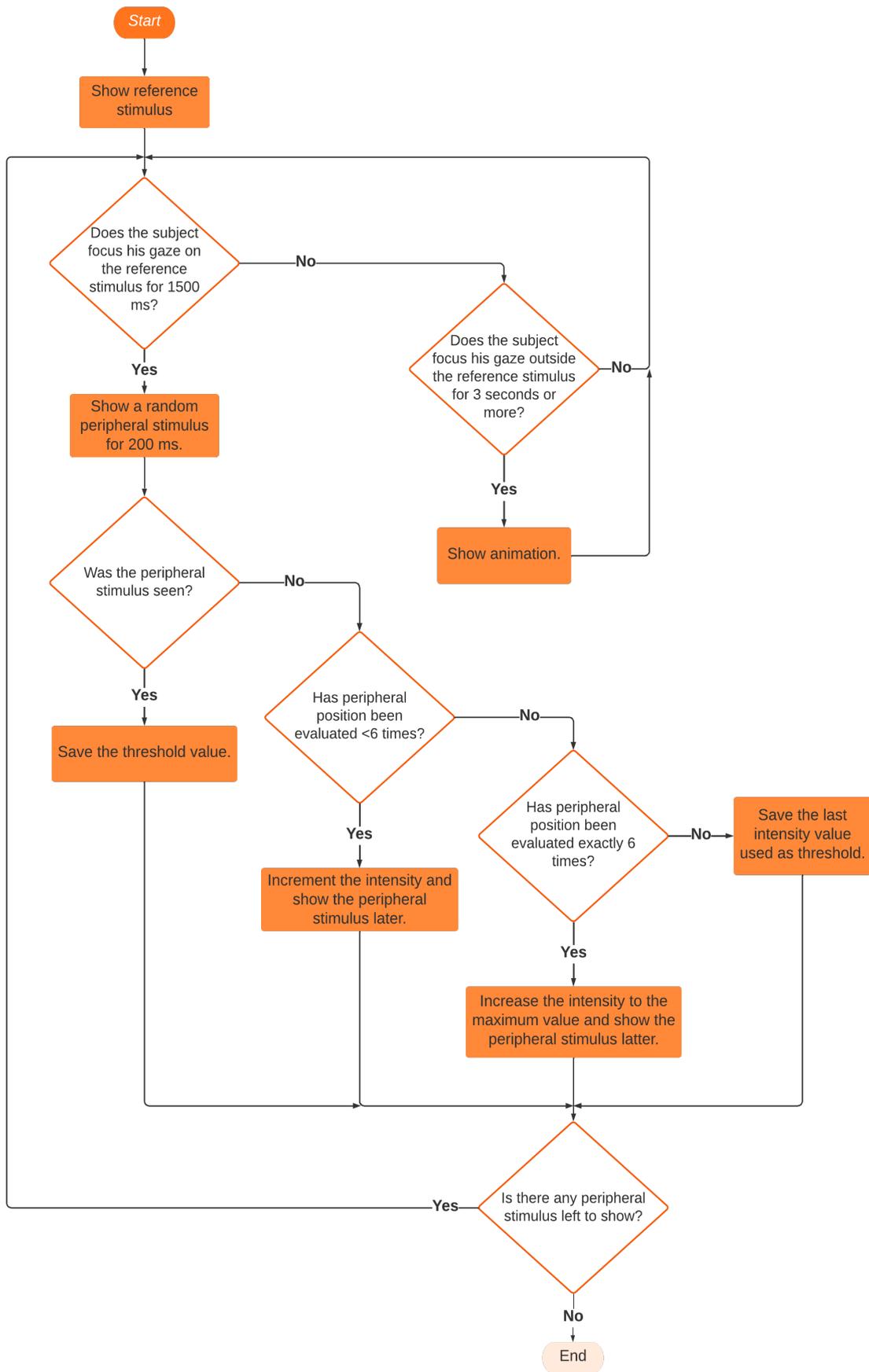
Figure 3.3: Algorithm used to evaluate the VF.

Additionally, and as reference to compare the ocular responses under the same experimental conditions, an emulation of SAP test is simultaneously carry out, so the person is instructed to perform the following actions: *"stare at the reference stimulus at all times and if you perceive a peripheral stimulus, just follow it with the gaze and record your response pressing a button[1]"*.

The test begins displaying a static orange reference stimulus (a circle with a similar size used in SAP, Goldman size III) at the center of the computer screen over a black background. If the person maintains the gaze at the reference stimulus for a period of 1.5 seconds, a peripheral gray stimulus with a given intensity is displayed for 200 milliseconds. During this time, the subject can record his/her responses and at the end of it, the peripheral stimulus disappears and a new peripheral stimulus with other intensity value is ready to be shown once the person looks at the reference stimulus again for the established time (1.5 seconds). This ensures that the person is always looking at the reference stimulus when the peripheral stimulus is presented. The procedure, shown in Figure 3.3, is repeated with each of the positions of the VF pattern until the visual threshold has been determined for all positions using the button or when a maximum number of seven repetitions is performed at each position.

Unlike SAP, where the technician must monitor the patient's fixation on the reference stimulus at all times and alert the patient when this instruction is not followed, our paradigm automatically detects fixation losses on the reference stimulus and pauses the evaluation waiting until the subject redirects his gaze towards the reference stimulus. In order to motivate this, an animation is used if a loss of fixation on the reference stimulus is detected for more than 3 seconds. The animation, which is depicted in Figure 3.4, consists of a large colorful circle placed on the center of the screen, which gradually decreasing in diameter until it matches with the reference stimulus.

3.2.1 Visual field evaluation area and sequence

During the evaluation, the patterns C24-2 and C30-2 are available to be used. However, the size of screen and the distance between participant and the screen determine the total area to be evaluated. According to the set-up used, in average, the maximum extension of the VF to be evaluated corresponds to $\pm21°$ along the horizontal direction and $\pm15°$ along the vertical direction, which makes it impossible to evaluate some peripheral stimuli in the VF patterns used[2].

Blind spot location is routinely measured during the VF evaluation to accurately interpret test results and reduce diagnostic errors [34]. In addition to the above, SAP uses the blind spot location to monitor fixation losses during the evaluation (Heijl-Krakau method [35]) or as a reference stimulus during the evaluation. In order to accurately locate the blind spot, a grid of 20 peripheral stimuli with a separation of 2 degrees between them and covering a range of 11° to 19° on the temporal and 1° to 7° over inferior temporal was used. Figure 3.5(a) shows, for the right eye, the maximum extension of the VF evaluated with a total of 63 peripheral stimuli.

Threshold strategies, such as FASTPAC, require starting with an infra-threshold or supra-threshold intensity (called the *initial intensity*) and slowly varying this intensity

---

[1]For practicality, the button corresponds to the space key on the participant's keyboard.

[2]Using the C30-2 and C24-2 patterns in our paradigm, practically the same positions within the VF are evaluated due to the limited evaluation area. For this reason, from now on we will only refer to the C24-2 pattern for reference.

Figure 3.4: Screenshots of the animation used when a loss of fixation on the reference stimulus is detected. (a) Detection of a loss of fixation. (b) Start of animation after a loss of fixation for 3 seconds from its start. (c) Animation after 1000 milliseconds. (d) Animation after 1500 milliseconds from its start. (e) Animation after 2000 from its start. (f) End of the animation after 2500 milliseconds from its start.

(increasing or decreasing it, depending on the approach) to find the visual thresholds during the evaluation. If the initial intensity is far enough from the visual threshold, the evaluation time will increase. During the test, each of the 63 positions are simulated in random order with a maximum number of seven repetitions to decrease the evaluation time required to find the visual thresholds. However, seven repetitions may not be enough in some peripheral positions to find the visual threshold; especially if the initial intensity used in that positions is far from this. Based on the knowledge about the behavior of visual intensities in a healthy VF (high visual sensitivity in the center and a non-linear decrease in the periphery), a simple strategy was implemented. This strategy consists in evaluating the VF sequentially by zones from the center to the periphery. To do this, the visual pattern used was divided into four zones as shown in Figure 3.5(b). For each zone $z = 1, 2, 3$ and 4, the initial intensity used to calculate the visual thresholds is defined by the following expression:
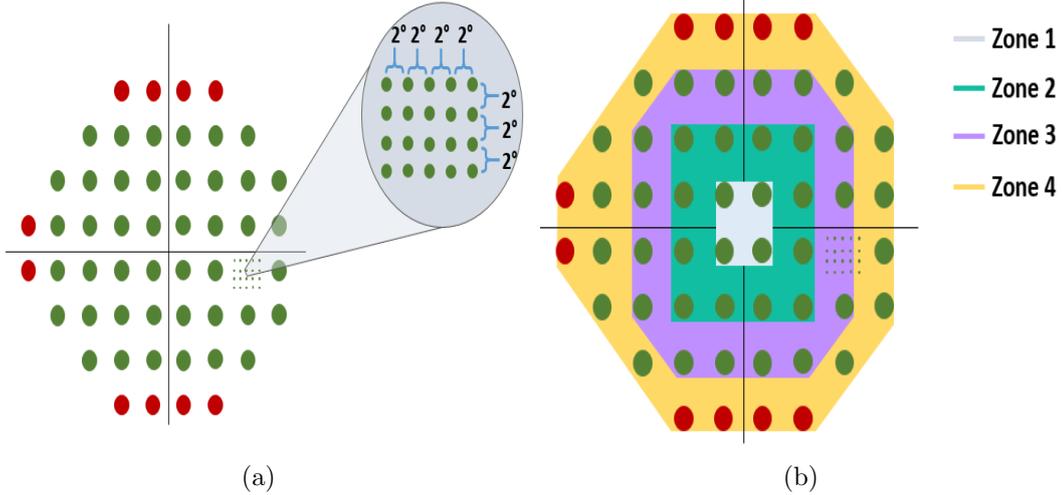
Figure 3.5: Evaluation area and classification. (a) Positions evaluated in the VF. (b) Zones of the pattern C24-2. In both panels, (a) and (b), the green circles represent the evaluated positions while the red circles represent the non-evaluated positions.

$$initial\ intensity_z = \min(VT_{z-1}) - 3\ dB, \tag{3.1}$$

where VT_z={*visual thresholds registered in the zone z*} and $VT_0 \equiv \{approximation$ *of the minimum visual threshold*}. That is, the initial intensity used in each zone corresponds to an intensity 3 dB below the minimum visual threshold recorded in the previous zone.

3.2.2 Initial settings

Before each VF evaluation, two necessary tasks were performed for each subject: 1) carry out a calibration process of the ET to guarantee its correct operation during the test, and 2) compute the *approximation of the minimum visual threshold*.

The ET calibration is a quick and easy process that involves looking at seven evenly spaced small circles on the monitor until they disappear. The seven circles are separated into three groups: 1) one circle on the center of the screen, 2) three circles located at the top and bottom of the screen (2 near the top corners and one on the bottom center), and 3) two circles placed near the bottom corners of the screen and one on the center of the top of the screen. Each group is used in order, and none of the items in the following groups are used until all the items in the current group are used. Figure 3.6 shows the position of the seven circles used in the calibration process as well as the elements of each of the groups. After this process, the ET calibration is verified using nine equispaced circles on the screen (3 at the top, 3 at half the height of the screen, and 3 at the bottom), where the software verifies that the gaze position obtained by the ET is as close as possible to the 9 circles. If there is a large error between the estimated position and the circles, the calibration process is repeated. Once the calibration process is complete, the *approximation of the minimum visual threshold* is computed using an animation. The animation, depicted in Figure 3.7, shows a gray ring positioned at the gaze coordinates calculated by the ET on a

black background. The ring starts with an infra-threshold intensity value and slowly increases its intensity until the subject can perceive it. As soon as the person perceives the ring, he/she must register the value by pressing the space key. This value registered corresponds to the *approximation of the minimum visual threshold.*
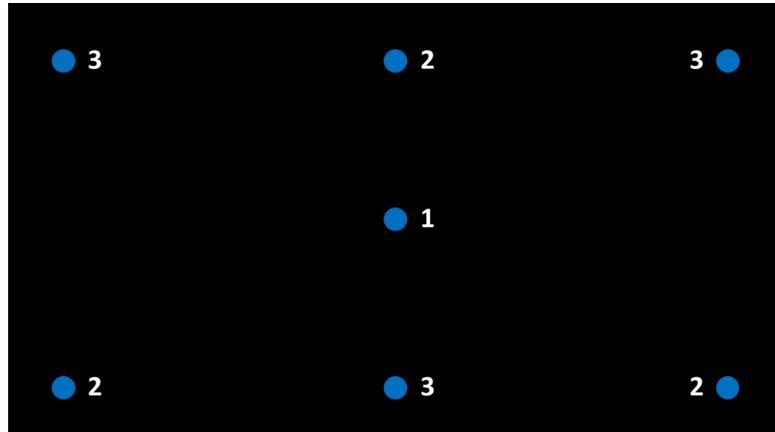

Figure 3.6: Positions of the circles used in the calibration process. The number next to the circle indicates the group to which it belongs.
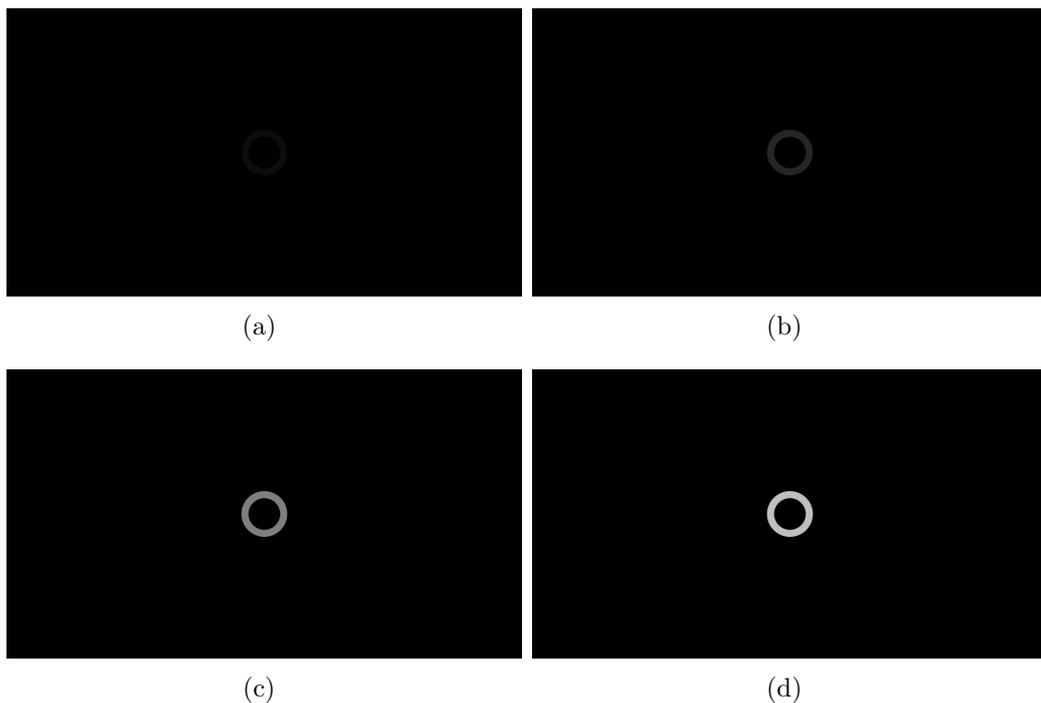

Figure 3.7: Some frames of the animation used to calculate the *minimum visual threshold.* (a) Frame in time= 1000 milliseconds. (b) Frame in time= 2500 milliseconds. (c) Frame in time= 4000 milliseconds. (d) Frame in time= 5500 milliseconds.

3.2.3 Participants and experimental protocol

To validate the system, a cohort of twenty participants (13 women and 7 men), aged 19 - 43 years ($\mu = 26 \pm 5$ years) were recruited at Facultad de Ciencias de la Universidad Autónoma de San Luis Potosí. For three weeks, all the participants voluntary conducted three sessions (one session per week on the same day and time), where in each session both VFs were evaluated. None of the participants had known VF deficits like glaucoma, cataracts, or any other similar ocular or neurological impairments. However, participants with low myopia and/or astigmatism were included (60% of them), so the use of corrective glasses was allowed during the test.

The protocol used during the sessions was the following:

- **First session.** Before performing the test, the goal of the experiment and the use of the information obtained from the test were explained to the participants. All the participants that accepted to take part in the experiment voluntarily signed a consent form and all their questions were answered. Next, all the test instructions were given using an illustrative video and a training period was provided to help them become familiarized with the test. At the end of the training period, a short break of 5 minutes was given. After the break, one of the participant's eyes was randomly selected and covered with a patch, to then sit down and place the participant inside the black cabin at a distance of 50 ± 1 cm from the screen and start the VF evaluation process. At the end of the evaluation of the first eye, the subject took a 5 minute break and then the other eye was evaluated.

- **Second and third sessions.** Before performing the test, the instructions were given again and a practice test was given if the participant considered it necessary. As in the first session, the order of the eyes to be evaluated was randomly selected and short breaks of 5 minutes were given before each monocular evaluation.

During the VF evaluation process, short breaks were made every 2 minutes, pausing the test and restarting it after this time without the need to recalibrate the ET. However, the participants had the flexibility to pause the test at any time if they deemed it necessary. Figure 3.8 summarizes the protocol implemented during the VF evaluation.

3.3 Data analysis

The gaze path analysis in EMP is an important process that ensures the reliability of the results. In some EMP prototypes, the decision rule to determine which stimuli were observed during the evaluation is based on a static threshold; that is, a stimulus is considered seen if a given feature of the gaze path (i.e., the distance to the stimulus) exceeds a static threshold. However, this decision rule is inefficient since it does not consider the behavior of the gaze path, which contains the information necessary to correctly classify which of the stimuli were seen. Here the gaze path analysis based on two machine learning techniques is described, as well as their performance.
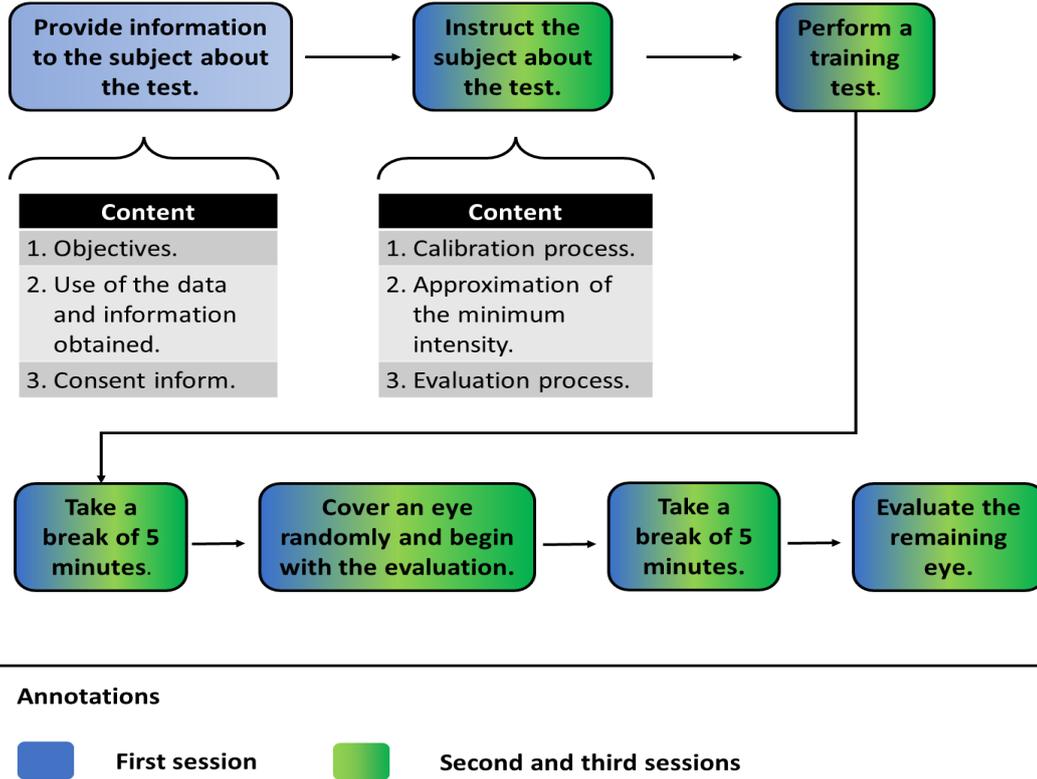
Figure 3.8: Protocol used for the VF evaluation.

3.3.1 Feature extraction

The extraction and correct selection of features are the most important processes in machine learning since from these, the model used will be able to classify the data correctly. The fixation either near or on the peripheral stimulus alone doest not guarantee that the intensity used to showing that stimulus corresponds to the visual threshold. For example, Figure 3.9 shows the minimum distance of each trajectory and its associated peripheral stimulus. Under a threshold-based classification, those stimuli below the established threshold (i.e., the dashed line) would be classified as seen stimuli. However, a person who cannot initially sees the peripheral stimulus may be tempted to search for it on the screen and possibly find it, thus misclassifying the stimulus as seen due to the short distance between it and the gaze position. To prevent the above, it is necessary to obtain more information about the gaze path behaviour in order to guarantee a correct classification. A useful feature corresponds to the direction of the gaze trajectory, which will remain stable in the case of the stimuli being seen, or highly variable if the person scans the screen for some stimulus. Based on the above, a peripheral stimulus is considered as seen if its associated gaze path meets two requirements: 1) the direction of the gaze path is close to the direction of the peripheral stimulus, and 2) the person performs a fixation on or sufficiently near the peripheral stimulus.

For the n peripheral stimuli used in the test (Figure 3.5(a)), each of them can be defined as a vector $\vec{PS_i} \in \Re^2$ with its respective magnitude and direction $|\vec{PS_i}|$, θ_{PS_i}, where the subscript $i = 1, 2, 3, ..., n$ indicates the number of the peripheral stimulus. On the other hand, a gaze path $GP_{i,j} = \{\vec{P_k} = (x_k, y_k) \,|\, k = 1, 2, 3, ..., 26\}$ is a finite set of twenty six gaze positions as shown in figures 3.10(a) or 3.10(b), where the subscripts
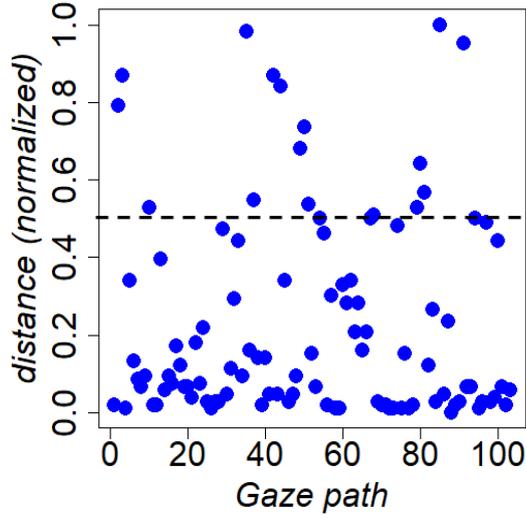
Figure 3.9: Scatter plot of the minimum distance between the gaze position and the peripheral stimuli.

i, j refer to the gaze path j of the stimulus i with $j = 1, 2, 3, ..., \leq 7$ (remember that a stimulus i is shown a maximum of 7 times, so that each peripheral stimulus would have a maximum of 7 associated gaze paths). Applying least squares regression on the 2D positions of a path, as shown in figures 3.10(c) or 3.10(d), we calculate the straight line that best fits them. This straight line can be characterized by a direction vector $\vec{LR}_{i,j}$, whose angle denoted as $\theta_{LR_{i,j}}$, corresponds to the angle[3] of the straight line computed using least squares regression. This direction vector provides an approximation of the gaze path direction from which one can compute the angular difference between the gaze path and the peripheral stimulus as

$$\triangle \theta_{i,j} = abs(\theta_{LS_{i,j}} - \theta_{PS_i}), \tag{3.2}$$

which is the first feature to be estimated.

In order to compute the second feature, the minimum distance between each element of $\vec{GP}_{i,j}$ and $\vec{PS}_i$, computed as

$$d_{i,j} = \min_{k}\{|\vec{PS}_i - \vec{P}_k|\}, \tag{3.3}$$

provides information about how close is the person's gaze to the peripheral stimulus. That is, a small value of $d_{i,j}$ indicates that the person fixed the gaze on or near the peripheral stimulus at some point across the trajectory.

Figure 3.11 shows a scatter plot of the two features computed from the evaluation of the right eye of a subject. Notice how both features, $d_{i,j}$ and $\Delta\theta_{i,j}$, help to better identify the seen stimulus (swarm of points in the scatter plot near the origin) from not seen stimuli (swarm of points in the scatter plot far from the origin), this compared with the single feature shown in Figure 3.9.

---

[3]To maintain the same reference units with respect to the visual pattern, the directions of the vectors will be expressed in degrees.
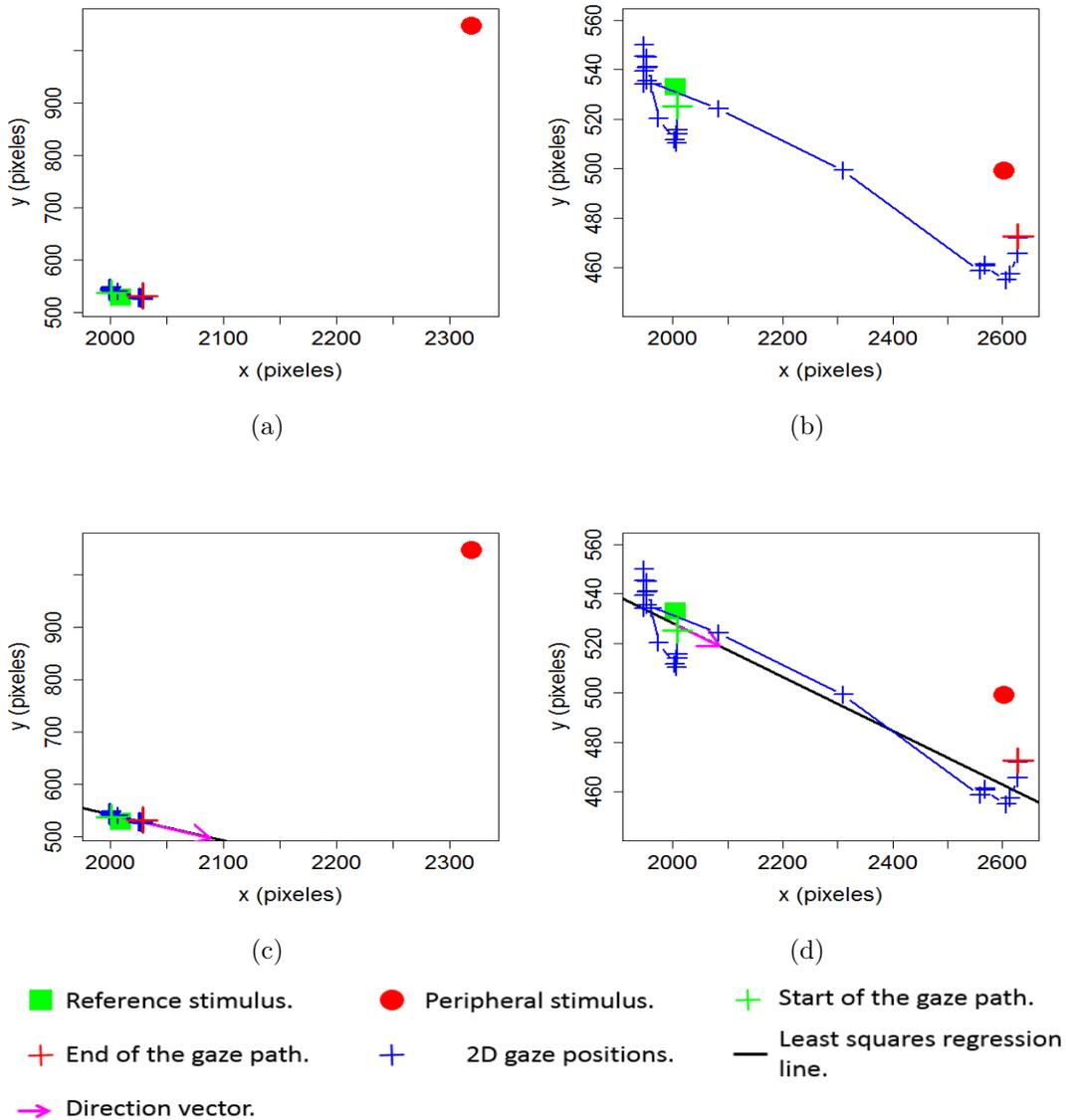
Figure 3.10: Feature extraction. (a),(b) Examples of gaze paths.(c),(d) Estimation of the gaze path direction using least squares regression.

3.3.2 Gaze paths classification using machine learning

There are different machine learning techniques to classify data such as Nearest Neighbour, Support Vector Machines, Discriminant Analysis and others, each with its advantages and complexity over the others. In order to classify the peripheral stimuli in two classes according to the behaviour of the associated gaze path: *Class 1={stimuli seen}* and *Class 2={stimuli not seen}*, two techniques were implemented. This section briefly describes the two implemented classifiers; for more details, please consult Appendix A. The first technique, is the combination of two classifiers: k-means and Bayes classifier [18] (from here on we will refer to this combination of classifiers as KBC) whose structure is shown in Figure 3.12.

The idea of this classifier is to first perform a pre-classification of the data into the two desired classes using the k-means classifier, and then regroup the data based on

Figure 3.11: Scatter plot of minimum absolute distance vs. angle difference.
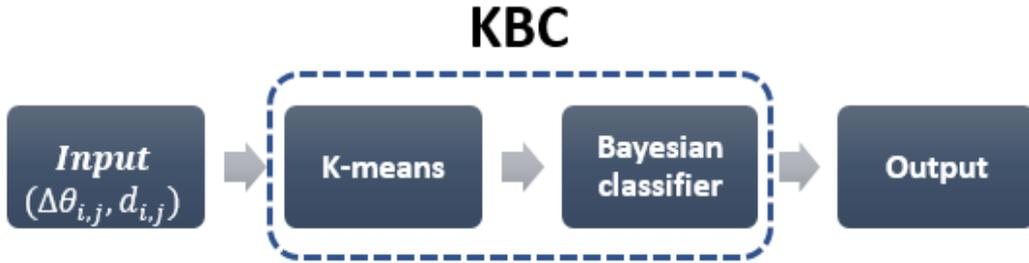
Figure 3.12: KBC diagram.

Bayesian decision theory. That is, once that k-means separated the data in the two classes desired as shown in Figure 3.13(a), each class is assumed to have a Gaussian distribution with a given mean vector and covariance matrix ($Class\ 1 \sim \mathcal{N}_1(\mu_1, \Sigma_1)$ and $Class\ 2 \sim \mathcal{N}_2(\mu_2, \Sigma_2)$). Under this consideration, the elements of the class are iteratively reclassified making use 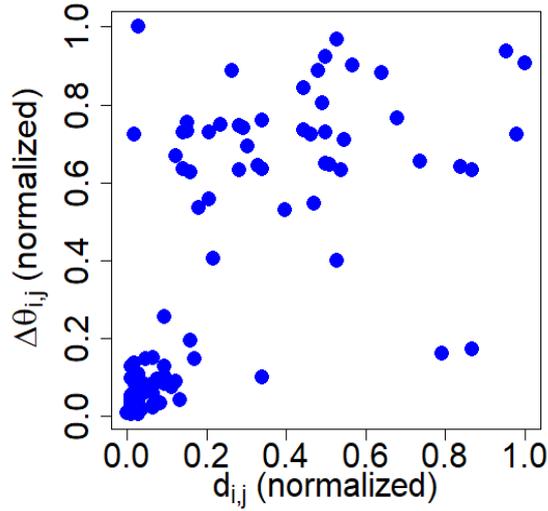of the Mahalanobis' distance, whose metric considers the mean and covariance matrix, both of which are computed at each iteration until convergence. Figure 3.14 shows the process diagram for KBC, while Figure 3.13(b) shows the data classified using iteratively the Bayes classifier. Notice how the darkgreen data, initially classified as seen by K-means, were adjusted using the Bayes classifier. The advantage of KBC is that no prior database is necessary to perform the data classification; instead, an ad-hoc model is trained for each subject and each test, as a result, the classifier can adapt to different ages of the population. On the other hand, under this configuration the classifier has the disadvantage that it is not possible to classify the information in "real time", which could increase the evaluation time as a consequence of over-evaluated positions.

The second technique applied to classify the gaze path corresponds to an Artificial Neural Network (ANN) [19]. The architecture of the ANN, as shown in Figure 3.15(a), consists of two inputs, two hidden layers with four neurons each, and one output, all with a sigmoid activation function. The ANN was trained with 1000 epochs using the

cross entropy loss function and gradient descent method with a learning rate of 0.01 and momentum of 0.8 in the back propagation process. The database used to train the ANN consist of 1484 gaze paths that were obtained from four additional subjects who did not complete the three experimental sessions.

(a)   (b)
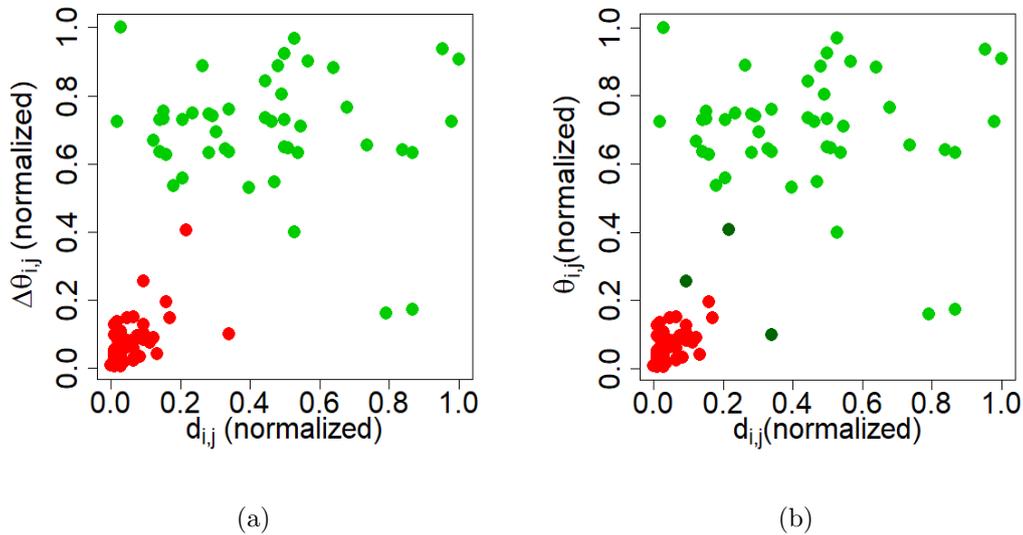
Figure 3.13: Data classification using KBC. (a) Initial classification using K-means. (b) Final classification using Bayes classifier. The red points correspond to the *Class 1* and the green points correspond to the *Class 2*.
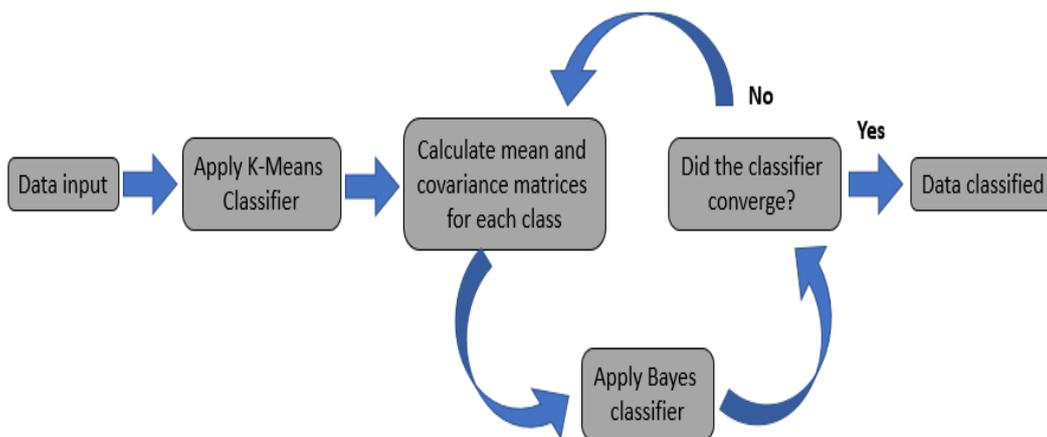
Figure 3.14: Diagram of the proposed KBC process.

Figure 3.15(b) shows the classification of the data used for the KBC example in Figure 3.13 into the two desired classes using the ANN. For this example, notice the slight difference in the results between ANN and KBC, where ANN separates better the two classes than KBC.

29
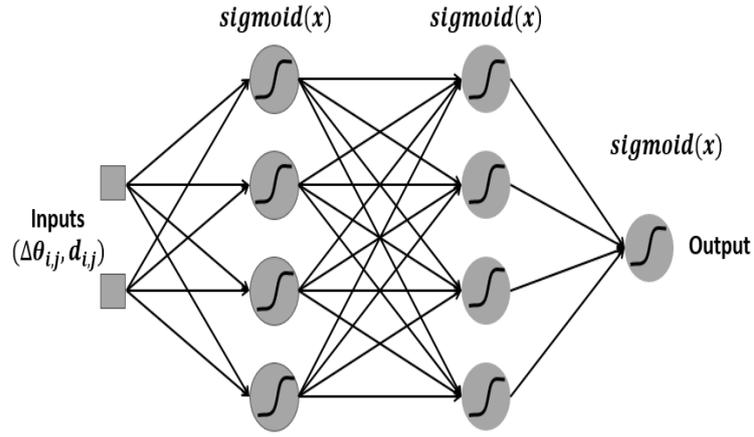
(a)
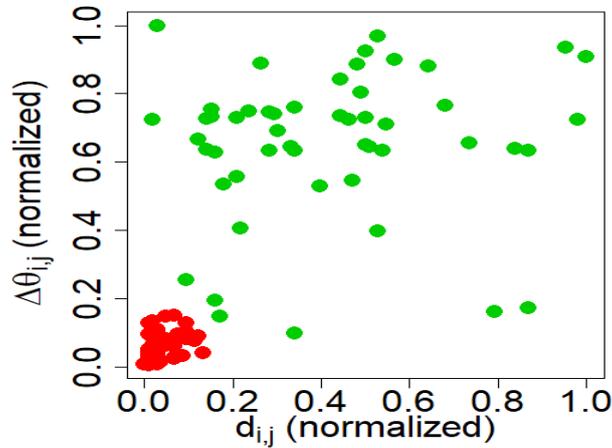
(b)

Figure 3.15: Classification using ANN. (a) ANN structure. (b) Data classified.

3.3.3 The ground truth

All the gaze paths obtained from the participants during the three session were visually analyzed and manually classified one by one, having a total of 16179 gaze paths ($\mu_{per\ test} = 143.82 \pm 25.83$). According to the behaviour of the gaze paths from the reference stimulus to the peripheral stimulus, each of them was assigned to one of the following three sets: *Type 1={direct gaze paths between stimuli}*, *Type 2={fixations on the reference stimulus or gaze paths where the peripheral stimuli were not seen}*, and *Type 3={gaze paths with dubious behaviour }*. This classification, considered as *the ground truth*, labels the peripheral stimulus in one of the following classes: *Class 1* (stimuli seen), *Class 2* (stimuli not seen) and a new class, *Class 3={Stimuli whose classification is uncertain}*.

Analysis of Type 3 gaze paths

The particular behavior of the trajectories included in the Type 3 set makes it difficult to determine whether the peripheral stimulus was seen or not due to the subjectivity

of its interpretation. In order to provide a better understanding of the behavior of the participants' gaze during the evaluation, an analysis of these trajectories in search of any relationship between them and the position of the peripheral stimuli was performed. To do the above, the different gaze paths that are included in this set were divided in two new sets: *Subtype 1={gaze paths with broken behavior}* and *Subtype 2={straight-line gaze paths}*. As shown in Figure 3.16, Subtype 1 trajectories are those in which the trajectory reaches or nearly reaches the peripheral stimulus; however, one or more abrupt changes in direction characterize this type of gaze trajectories. On the other hand, Subtype 2 gaze paths are characterized by having a direction similar or close to that of the peripheral stimulus, but the participant never performs a fixation on it. Instead, fixation is within the area of a circular's crown with origin in the peripheral stimulus and radii of approximately 30% to 50% of the magnitude of the vector representing the peripheral stimulus.
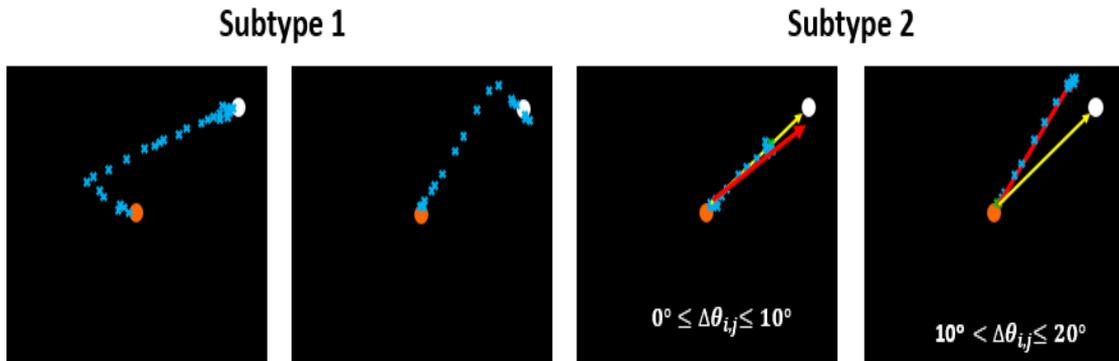
Figure 3.16: Examples of Type 3 gaze paths. The blue crosses represents the gaze path, the orange circle and the white circle represent the reference and the peripheral stimuli respectively, the yellow arrow corresponds to the expected direction, and finally, the red arrow corresponds to the estimated direction.

Subtype 1 and Subtype 2 trajectories were quantified according to VF's area where its associated peripheral stimulus appeared. To simplify this, the quantification was performed only in two areas: 1) the central area, which includes all the stimuli of the zones 1,2 and 3, and 2) the peripheral area that includes the remaining positions (see Figure 3.5 (b)).

3.3.4   Gaze path analysis versus the ground truth

In order to perform a comparison between the gaze path analysis and the ground truth, the common classification indices such as sensitivity, specificity, accuracy and error rate were computed. Taking into account Type 3 gaze paths, the quantification of these indices were performed under two scenarios: 1) Type 3 gaze paths are considered as not seen, and 2) Type 3 gaze paths are left out of the analysis, that is, approximately 6% of the data per test are not used to test the classifiers.

3.4 Reconstruction of the visual field

Unlike the SAP reconstruction map, where visual thresholds are represented in gray scale, visual thresholds computed from gaze path analysis are represented by a heat map that covering a color palette ranging from blue to red. These two colors, represent the minimum and maximum intensity values used during the test respectively. The heat map was selected because it is easier to visualize the different visual thresholds and their behavior in the VF map than with gray scale, where for the latter, it is necessary some experience to distinguish the different shades of gray. In addition to the color palette used, the positions not evaluated were represented in black dotted lines, while the dark-gray color was reserved to represent those VF positions where the participant could not see the intensity threshold used during the evaluation. These dark-gray positions should not be necessarily interpreted as blindness, rather that the intensity used in those positions was not sufficient to cross the visual threshold. In addition to the heat map, two additional maps are shown: 1) a numeric map with the visual threshold values used (in dBs) to render the heat map, and 2) a symbolic map that indicates whether the estimated visual thresholds are correct or incorrect based on the ground truth (false positives, false negatives, true positives and true negatives). For the numerical map, the single values correspond to the visual thresholds computed, the values in parentheses correspond to the intensity thresholds that the participant did not see (or in the case of machine learning techniques, these intensities were classified as not seen), and the acronym N.E (not explored) corresponds to the C24-2 positions that were not evaluated. On the other hand, the symbolic map uses dots for represent the correct visual thresholds estimated (true positives and true negatives) according to the ground truth, the symbol plus represent the false positives, while minus sign is used to represent the false negatives. Like the numeric map, unexplored positions are represented as N.E. However, there are some positions where it was not possible to verify the visual threshold since the trajectories belong to the Type 3 set; these positions are represented as N.C (not computed). Figure 3.17 summarizes the symbology used in the reconstruction maps.

38 dB.	37 dB.	36 dB.	35 dB.	34 dB.	33 dB.	32 dB.	31 dB.
30 dB.	29 dB.	28 dB.	27 dB.	26 dB.	25 dB.	24 dB.	23 dB.
22 dB.	21 dB.	20 dB.	19 dB.	18 dB.	17 dB.	16 dB.	

Not explored. Unseen visual intensity.

(a)

N.C	Not computed.	
( )	Unseen visual intensity value.	
-	False negative.	

•	Correct visual threshold computed.	
N.E	Not explored.	
+	False positive.	
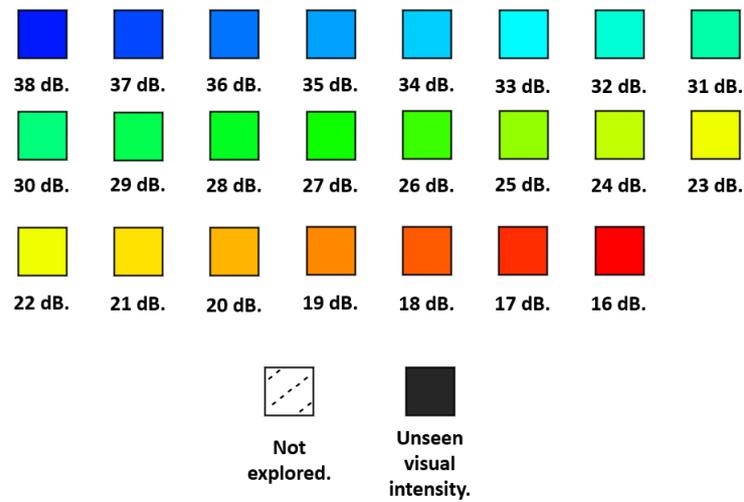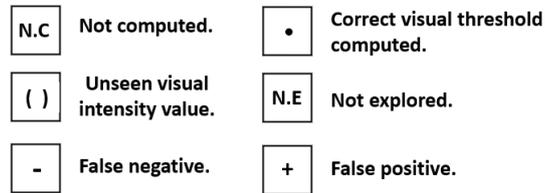
(b)

Figure 3.17: Symbology used in the reconstruction maps. (a) Symbology used in the heat map. (b) Symbology used in the numerical and/or symbolic maps.

Chapter 4

Results and discussion

4.1 Results

The results in this chapter show summarize statistical values of the gaze paths analysis. If you need details about a participant's reconstruction maps, see Appendix B.

4.1.1 Visual maps

Figures 4.1 to 4.5 show VF reconstruction maps of the right eye of a male participant. These reconstructions, which were obtained from the implemented machine learning techniques and also from SAP emulation, correspond to the three sessions performed by the participant. The order of the VF reconstruction maps is as a follows: 1) heat maps, 2) numerical map without the blind spot grid, 3) numerical map for the blind spot grid, 4) symbolic map without the blind spot grid, and 5) symbolic map for the blind spot grid.

From the heat maps or numerical maps is easy to see 1) the great similarity between the visual thresholds computed using the machine learning techniques and the manual responses, 2) the expected behavior of a healthy VF: greater visual sensitivity on the center of the VF and a decreased visual sensitivity on the periphery, and 3) the correct blind spot location. However, slight variations on the visual thresholds computed were found in each session. Such variations may be due to several factors such as: 1) classification errors obtained from the implemented machine learning techniques, 2) missed responses from the participant, 3) willingness and cooperation of the participant during the evaluation, and/or 4) the participant's experience in becoming familiar with the test.

Based on the threshold values from the ground truth, the figures 4.4 and 4.5 show the classification errors between the visual thresholds computed from the gaze path analysis and visual thresholds values of the ground truth. Here, it is easy to observe the correct value of the visual thresholds in the center of the VF (positions with a dot) and slight classification errors in some peripheral positions: false positives and/or false negatives represent by the plus and minus symbols respectively. Compared to SAP emulation, note how KBC and ANN correctly provide lower nasal visual thresholds during the first session. Similarly, note how ANN correctly detects visual thresholds in the third session. The above shows the control that the paradigm has during the evaluation by reducing the subjective responses of the SAP emulation.
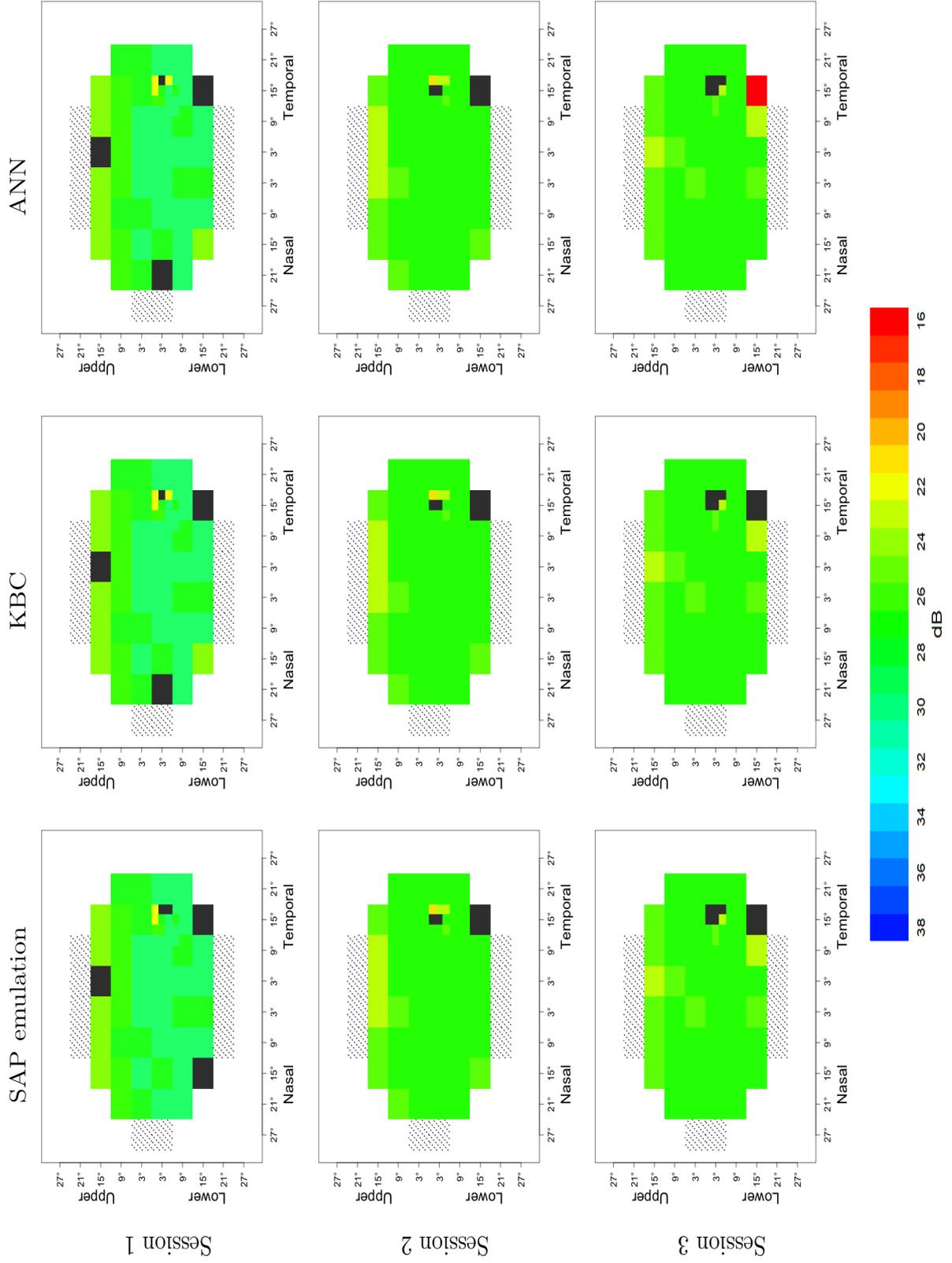
Figure 4.1: Reconstructions of the VF of the right eye: heat maps.

SAP emulation

KBC

ANN

Session 1

Session 2

Session 3
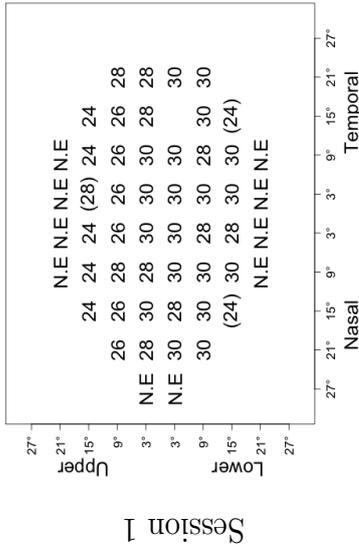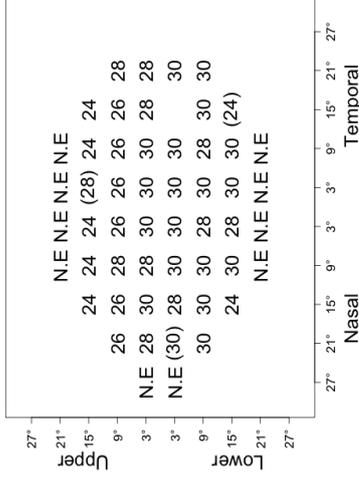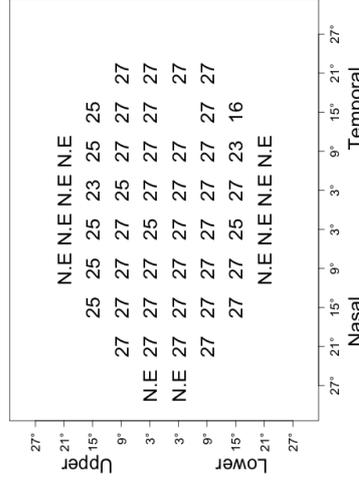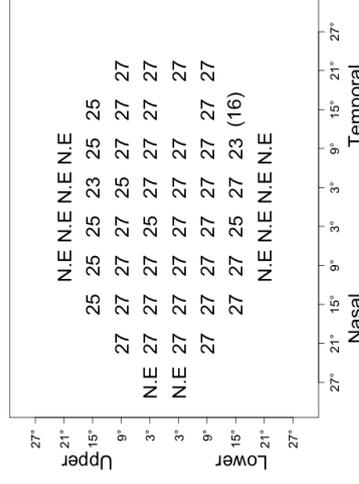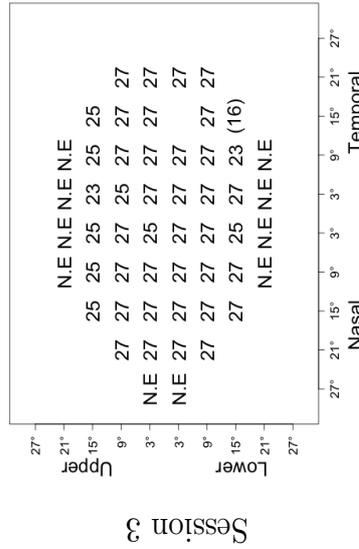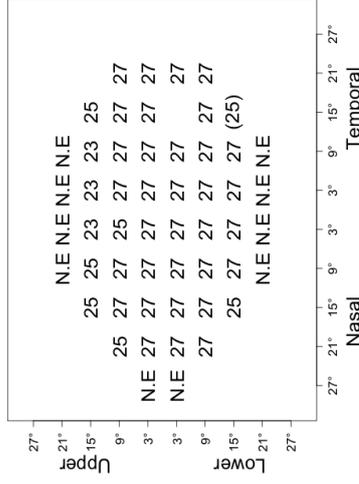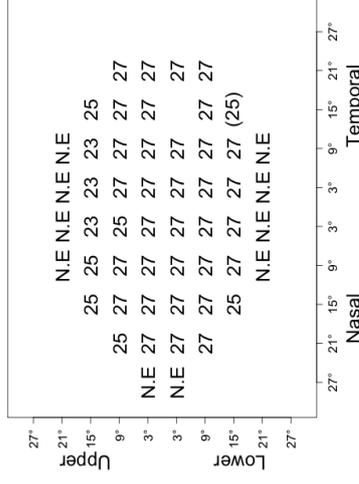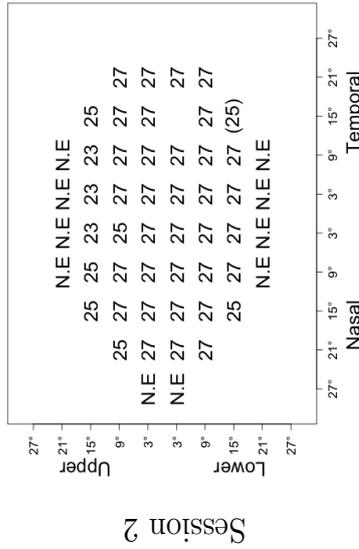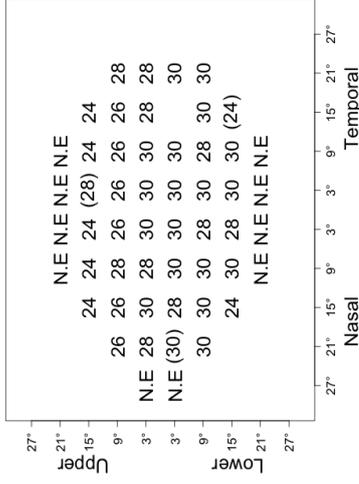
Figure 4.2: Reconstructions of the VF of the right eye: numerical maps.
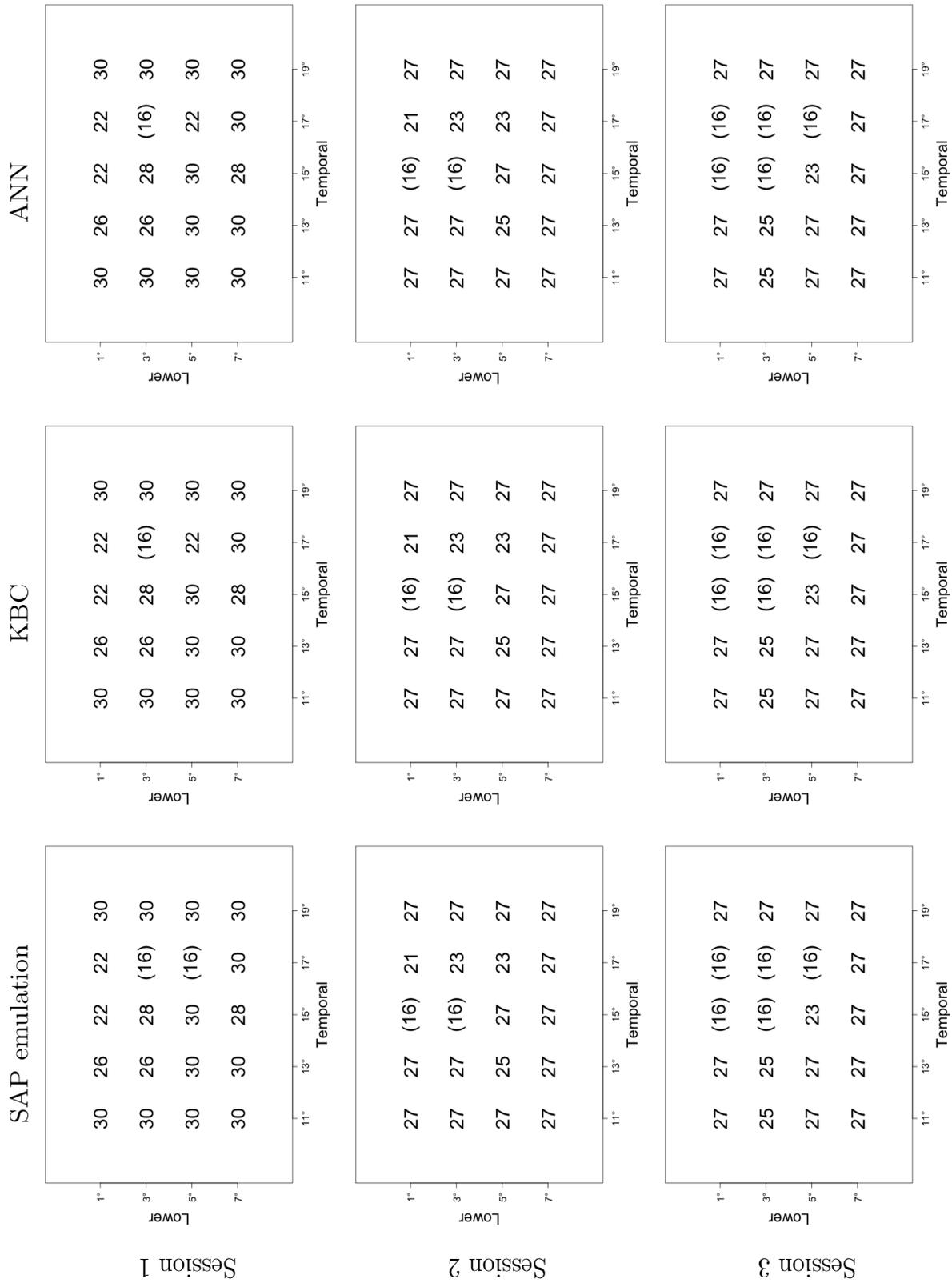
Figure 4.3: Reconstructions of the VF of the right eye: numerical maps of the blind spot.
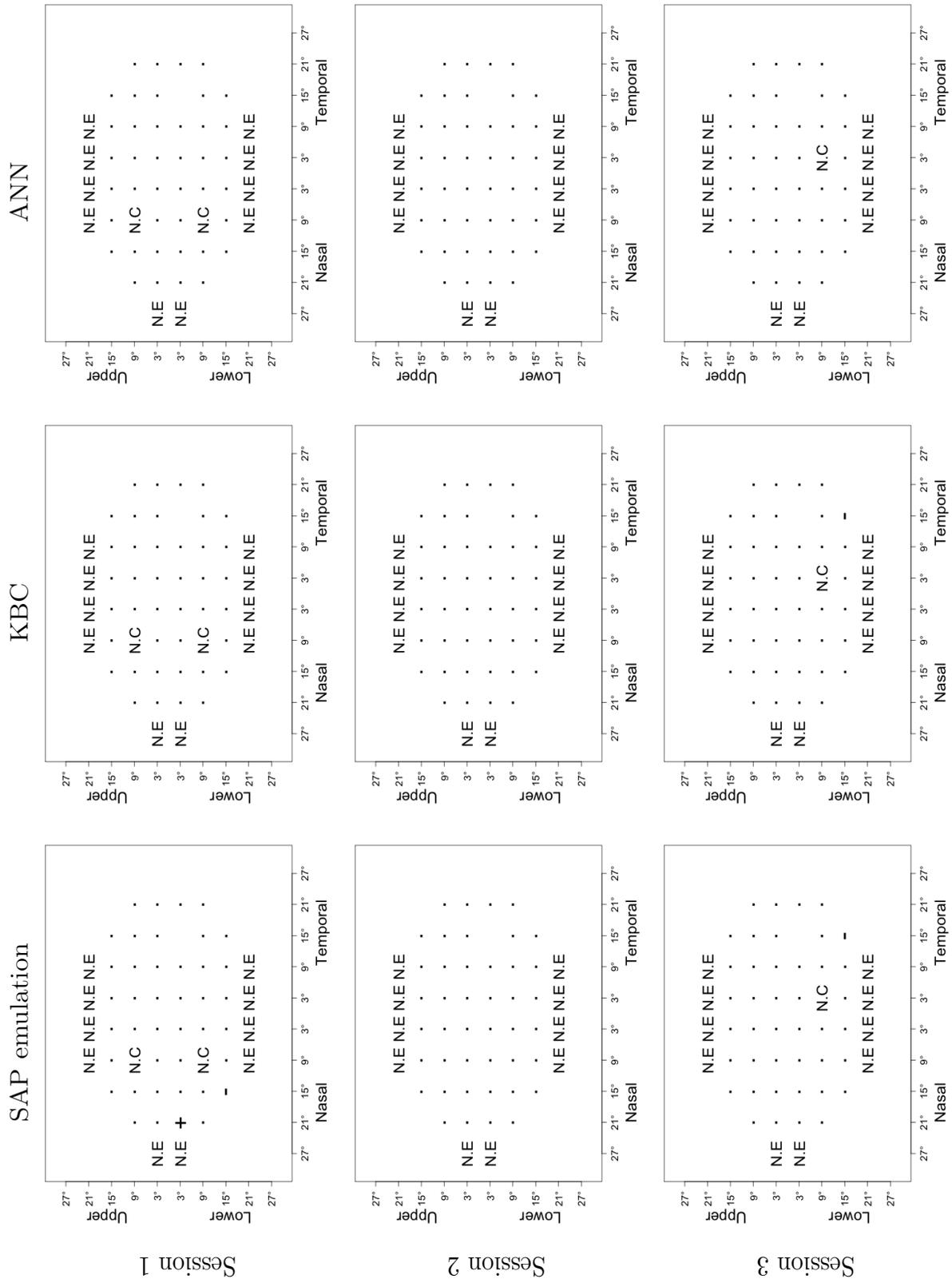
Figure 4.4: Reconstructions of the VF of the right eye: difference against ground truth.
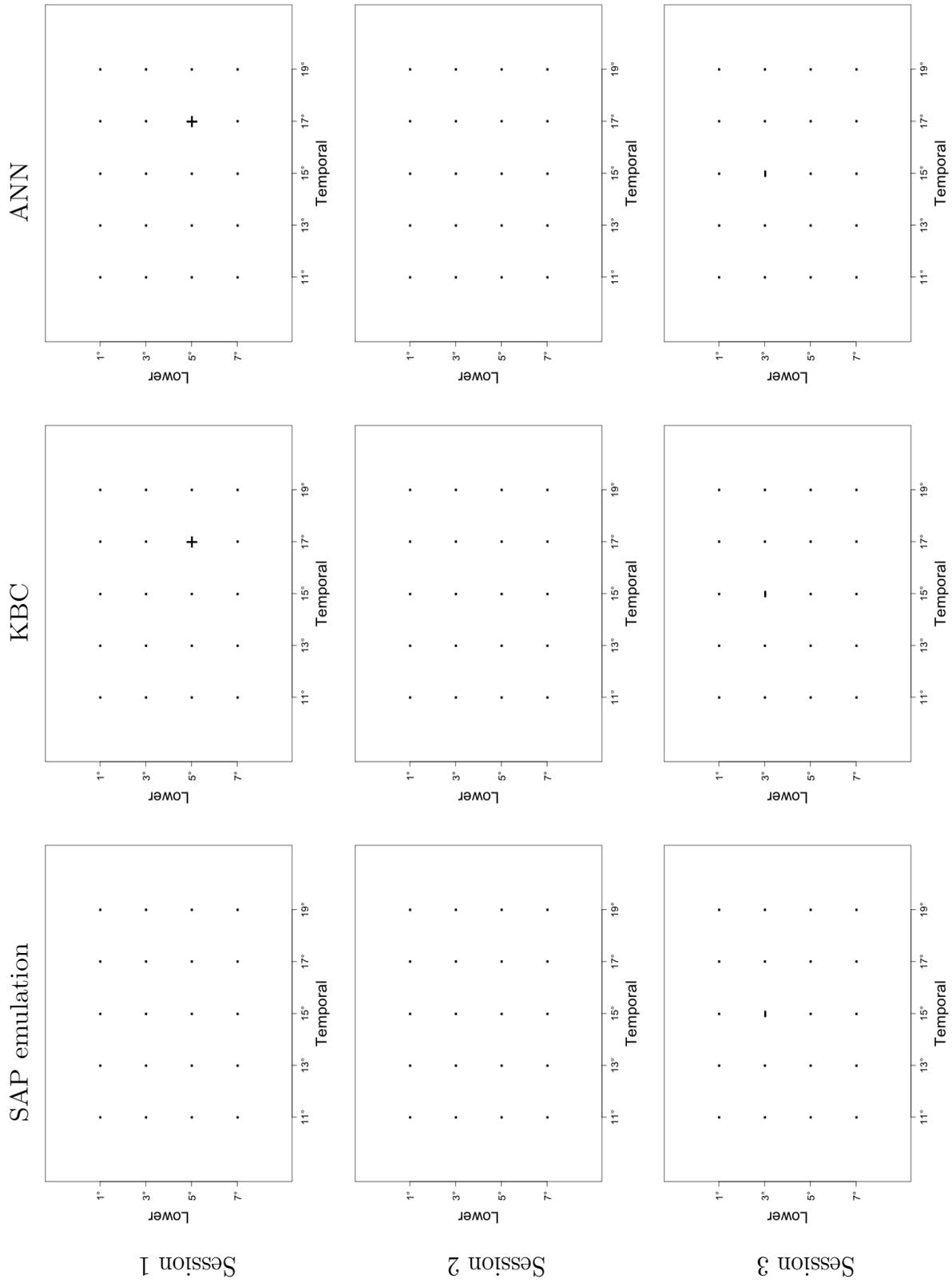
Figure 4.5: Reconstructions of the blind spot for the right eye: difference against ground truth.

However, for the first and last session we identified some positions within 9° whose visual threshold was impossible to calculate because Type 3 gaze trajectories were found in these positions (positions with N.C). For this particular participant, a total of 3 positions were identified.

4.2   Type 3 gaze path analysis

Figure 4.6 shows, for each session and eye, the average percentage of Type 3 gaze paths detected in the ground truth. For these average percentages, the proportion of subtype 1 gaze paths are present in dark colors, while for subtype 2 gaze paths, their representation is given in light colors. For all sessions, the average percentage of type 3 gaze paths was less than 2.80% for the left eye and 3.20% for the right eye. Nevertheless, it should be noted that the percentage decreases for both eyes throughout the sessions. This decrease may be the result of experience gained by the participant in each session.

(a) Left eye   (b) Right eye
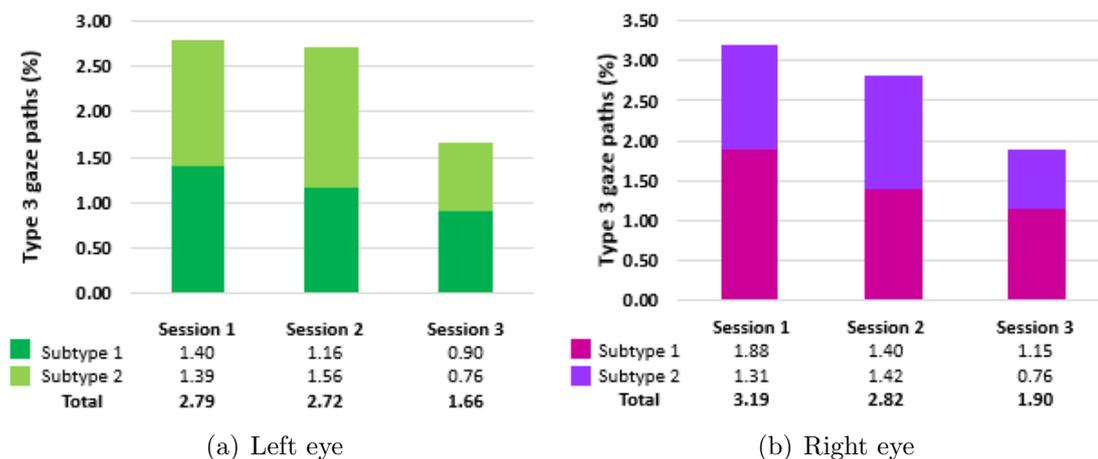
Figure 4.6: Average percentage of Type 3 gaze trajectories found in the ground truth, for each eye and each session.

According to the analysis performed for this type of trajectories, Figure 4.7 shows, for each session and eye, the average percentage of stimuli with Type 3 gaze paths reported as seen in SAP emulation. We consider this quantification necessary because by comparing two paradigms such as SAP and ours, we can identify classification errors and correctly quantify the accuracy of the results in our paradigm. Results shown in the Figure 4.7, indicate that, on average, between 60% and 80% of the stimuli with Type 3 gaze paths were reported as seen by the participants in each session. This average percentage suggests that the participant perceived most of these stimuli, but it is not clear whether the stimuli were perceived when the participant was staring at the reference stimulus or at some unrelated point in the screen. For instance, Subtype 1 trajectories can be explained by the user scanning the screen trying to find the stimulus, and then changing the scanning direction when stimulus is found. For this reason, we perform the subsequent evaluation considering the two scenarios described in Section

3.3.4: one scenario where Type 3 stimuli are considered not seen, and another scenario where they are considered ambiguous and are removed from the testing data sets.
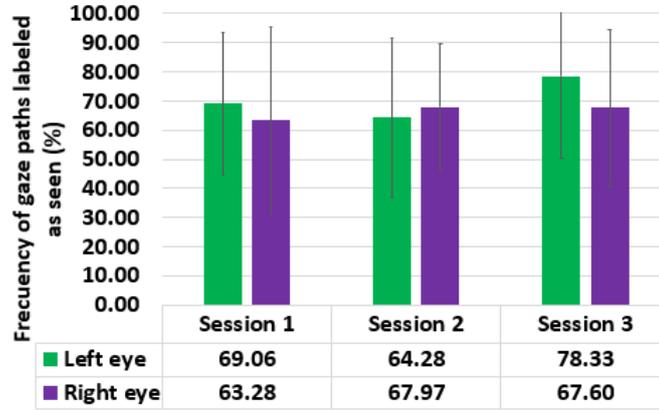
Figure 4.7: Percentage of Type 3 gaze path reported as seen in SAP emulation.

Table 4.1 shows, for each session, the mean percentage of Subtype 1 and Subtype 2 gaze paths according to the area in which they were found (for details about this quantification see the section 3.3.3). The quantification shows that, throughout the sessions carried out by the participants, Area 2 reported an increase of Subtype 1 gaze paths. For the left eye, an increase of 36.48% was obtained between the first session and the last session ; while for the right eye the increase was 13.66%. On the other hand, Subtype 2 gaze paths do not shown a progression; instead, they show fluctuating values in both eyes across the sessions.

Subtype 1 gaze paths

	Right eye		Left eye	
	Area 1 (%)	Area 2 (%)	Area 1 (%)	Area 2 (%)
Session 1	63.39	36.61	58.24	41.75
Session 2	51.19	48.81	47.30	52.79
Session 3	26.92	73.07	44.58	55.41

Subtype 2 gaze paths

	Left eye		Right eye	
	Area 1 (%)	Area 2 (%)	Area 1 (%)	Area 2 (%)
Session 1	39.89	60.11	38.54	61.46
Session 2	61.73	38.27	26.85	73.15
Session 3	16.67	83.33	51.67	48.83

Table 4.1: Mean percentage of Subtype 1 and Subtype 2 gaze paths according to the area in which they were found.

Left eye

	Sensitivity (%)			Specificity (%)			Accuracy (%)		
	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN
Session 1	82.15	96.87	99.28	93.98	96.18	94.33	88.03	96.38	96.65
Session 2	87.85	97.44	99.43	94.11	96.38	94.81	90.32	96.68	96.97
Session 3	91.17	99.17	99.47	95.05	97.68	96.33	93.33	98.37	97.83
μ	**87.06**	**97.83**	**99.39**	**94.38**	**96.74**	**95.16**	**90.56**	**97.14**	**97.15**

Right eye

	Sensitivity (%)			Specificity (%)			Accuracy (%)		
	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN
Session 1	80.81	97.95	99.00	93.83	94.20	92.73	86.99	95.97	95.74
Session 2	87.03	98.01	99.59	94.76	95.53	95.34	90.99	96.52	97.18
Session 3	91.85	98.39	99.68	95.91	97.92	96.80	93.96	98.10	97.99
μ	**86.56**	**98.12**	**99.42**	**94.83**	**95.88**	**94.96**	**90.65**	**96.87**	**96.97**

Table 4.2: Sensitivity, specificity and accuracy considering Type 3 gaze paths as not seen.

Left eye

	Sensitivity (%)			Specificity (%)			Accuracy (%)		
	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN
Session 1	82.15	96.87	99.28	97.10	98.40	97.58	89.53	97.56	98.44
Session 2	87.85	97.44	99.43	96.74	98.94	98.53	91.70	98.08	98.99
Session 3	91.17	99.17	99.47	96.59	99.28	98.81	94.08	99.20	99.09
μ	**87.06**	**97.83**	**99.39**	**96.81**	**98.87**	**98.30**	**91.77**	**98.28**	**98.84**

Right eye

	Sensitivity (%)			Specificity (%)			Accuracy (%)		
	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN
Session 1	80.81	97.95	99.00	96.88	97.50	97.02	88.39	97.76	98.04
Session 2	87.03	98.01	99.59	97.29	98.04	98.49	92.30	97.98	98.95
Session 3	91.85	98.39	99.68	97.99	99.75	99.60	95.10	99.11	99.60
μ	**86.56**	**98.12**	**99.42**	**97.38**	**98.43**	**98.37**	**91.93**	**98.28**	**98.86**

Table 4.3: Sensitivity, specificity and accuracy leaving out Type 3 gaze paths from the analysis.

4.3 Comparison between the gaze path analysis and the ground truth

Tables 4.2 and 4.3 show the mean percentage of sensitivity, specificity and accuracy of the results obtained from the SAP emulation, KBC and the ANN under the two scenarios described in section 3.3.4. For the first scenario, the results yielded accuracy values between 90.5% and 97.5%, whereas for the second scenario, accuracy results slightly improved with values between 91.5% and 99.0%. For both eyes, ANN was the method with the highest sensitivity and accuracy values in both scenarios, while for specificity values, KBC was better than the other methods. Figure 4.8 shows the percentage error rates for SAP emulation, KBC and ANN. From this figure it is easy to see a clear advantage of the machine learning techniques with respect to SAP emulation, where ANN was the best-performing method. Taking the SAP emulation error rate as reference, ANN decreased this error rate between a 67.5% and 70% in both eyes for the first scenario, whereas for the second, ANN decreased approximately an 85.6% the error rate in both eyes.
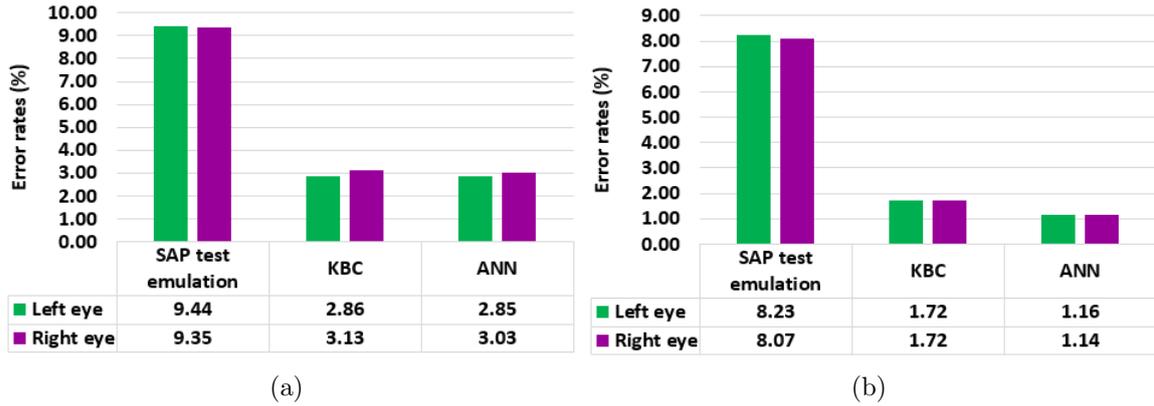
Figure 4.8: Percentaje of error rates for space key, KBC and ANN. a) Scenario 1 results. b) Scenario 2 results.

4.4 Test time

Figure 4.9 shows the mean percentage of test time per session. For the right eye, results show a slight increment of 0.5 minutes from the first to the last session, whereas for the left eye, the time practically was the same during the three sessions. That is, from the first session to the third the increment only was 0.09 minutes. However, neither increment is significant, as the standard derivation is much lager than the differences. On average, the test time for the three sessions in both eyes is 7.2 minutes. Some investigations such as that Roggen et al. [37] or Bengtsson et al. [36] reported average evaluation times of 6.1 and 7.05 minutes respectively when implementing SITA Standard test with healthy people using the C30-2 pattern. Although the current evaluation time of the paradigm proposed in this thesis is close to the evaluation time of the mentioned investigations, one should take into account two factors: on one hand, due to the limitations of the eye-tracker system, the proposed system evaluates a smaller area than a clinical campimeter, with a smaller number of positions (e.g., 64 positions vs. 76 positions in the C30-2 pattern); because this, the proposed system could take 1 or 2 minutes longer than SITA Standard test to evaluate the VF. On the other hand, the proposed system allows the person to take small rests (e.g., close the eyes or look elsewhere) during the test, which could artificiality increase the evaluation time, but would result in a less tiring test overall.
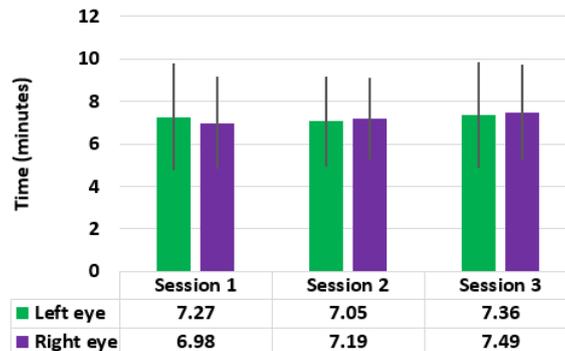
Figure 4.9: Average test time per session corresponding to the left and right eyes.

4.5 Discussion

4.5.1 Reconstruction of the VF

We present a threshold paradigm for the VF's evaluation based in EMP and machine learning techniques. The paradigm, which was implemented and validated with twenty young healthy participants, yielded similar results to those described for a healthy VF, that is, 1) a decrease of visual acuity from the center of the VF towards the periphery, and 2) a correct approximation of the position of the blind spot. Despite the fact that most of the reconstructions maps fulfil the above description, all of them had slight variations in the visual thresholds computed between sessions and paradigms. Some studies performed by Junoi et al. [6], Stewart et al. [38], Phillip et al. [39] or Gardiner et al. [40] have proven that the results of the SAP test are different for the same person regardless of whether these results correspond to evaluations with a difference of weeks, hours or even minutes. These studies have attributed the visual thresholds differences to different factors involved in the test such as the conditions in which the test was carried out (i.e. the technician's experience conducting the test, lighting conditions, strategy used, time of day the assessment was performed or the time of the year), the physical and health conditions of the patient (i.e., age, the pupil size of the patient or illnesses like diabetes or ocular hypertension) and/or the patient's experience with the test.

In some VF reconstruction maps obtained from KBC and ANN, we detected some positions in the three central degrees where the calculated visual thresholds were different than expected, that is, positions classified as unseen. Based on the ground truth, we found that most gaze paths from these central positions correspond to fixations on the reference stimulus. Nevertheless, the participants recorded the visual threshold values at those positions in SAP emulation. Figure **??** shows the VF reconstruction for a participant where the positions (3° nasal, 3° upper) and (3° nasal, 3° bottom) shows the behavior mentioned. Note that despite not having estimated the expected visual thresholds in the positions mentioned (a visual threshold represented by shades of green), KBC and ANN classifiers correctly classify these positions as unseen due to the participant's fixation on the reference stimulus. Studies in EMP performed by Murray et al. [25], [26] or Jones et al. [29] have reported similar behaviors when the three central degrees are evaluated. These behaviors are due to the difficulty that some people have in generating a gaze path from the reference stimulus to the peripheral stimuli due to the short distance between them; another factor could be the low temporal and spatial resolution of the ET used in this study. In our case, this phenomenon occurred for 18 participants in at least one session. Studies performed by Jones et al. [29], Brandao et al. [41] or Rumelhart et al. [42] have proposed new VF patterns or modifications to the current patterns by eliminating or relocating the stimuli used to evaluate the three central degrees. However, these new patterns may not be able to identify visual defects in this area in a timely manner.
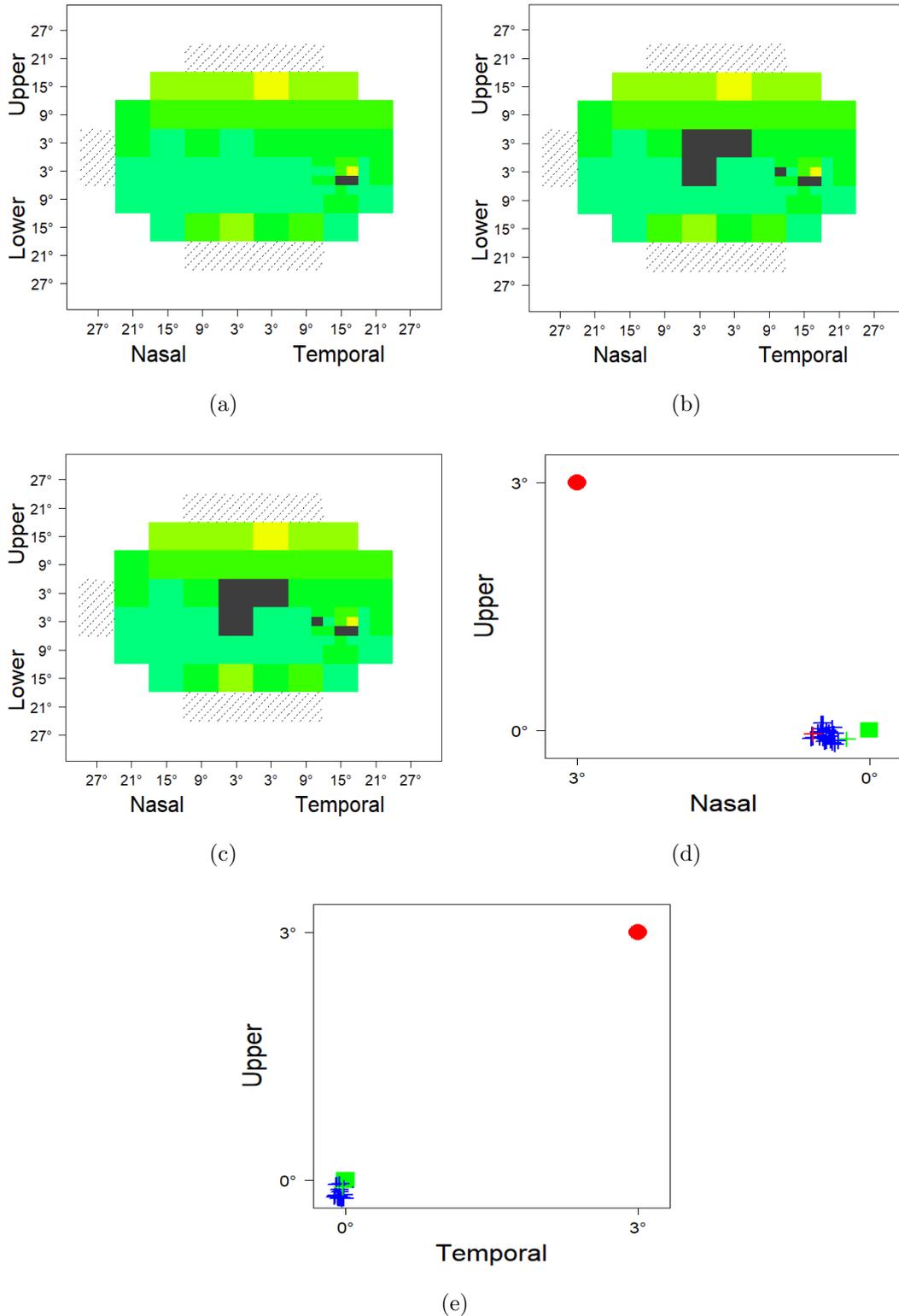
Figure 4.10: Examples of VF reconstructions where there are differences in visual thresholds within the central three degrees. (a) Reconstruction of VF from SAP emulation. (b) Reconstruction of VF from KBC. (c) Reconstruction of VF from ANN. (d) Gaze path for central stimulus at position 3° nasal and 3° upper, (e) Gaze path for central stimulus at position 3° temporal and 3° upper.

During the gaze trajectory analysis, the blind spot was not fully found in four participants in at least one session. This behavior was identified for both eyes; however the phenomenon was more common in the left eye than the right eye. For the left eye, the analysis shows that neither SAP nor machine learning techniques identified the blind spot of three participants during the first session, while for the right eye, this phenomenon appeared in one participant. Table 4.4 shows the participants and the sessions in which it was not possible to locate the blind spot (for more details see Appendix B). Most of the missed blind spot locations in these participants are due to errors in the classification according to the ground truth. However, for the first session of subjects 3 and 11, we found that all stimuli used to find the blind spot were seen. We believe that the reason for the above is due to an incorrect alignment between the subject and the screen during the evaluation, since in the following sessions it is possible to estimate the position of the blind spot in these subjects. On the other hand, subjects 7 and 9 generated uncertainty in the first session when detecting the blind spot. For subject # 7, the only invisible position (temporal 18, inferior 1) was labeled as a false positive, while the participant fully saw the stimuli in the neighboring positions. On the other hand, subject # 9 generated Type 3 gaze paths in some of the positions of the grid used to locate the blind spot, generating uncertainty in its location. Although there is documentation indicating that the blind spot presents absolute and relative scotomas, as well as slight variations in the size and position of the blind spot [20], [34] or [44] , adjustments in the methodology are necessary to guarantee the correct estimation of the blind spot in people (for example, incorporating a fast process with few stimuli to estimate the position of the blind spot before the test and thus modify the position of the grid used in the evaluation).

Left eye									
	First session			Second session			Third session		
Participant	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN
3	×	×	×	✓	✓	✓	✓	✓	×
7	✓	✓	✓	✓	✓	✓	✓	✓	✓
9	×	×	×	✓	✓	✓	✓	✓	✓
11	×	×	×	✓	✓	✓	✓	✓	✓

Right eye									
	First session			Second session			Third session		
Participant	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN	SAP emulation	KBC	ANN
3	✓	✓	✓	✓	✓	✓	✓	✓	✓
7	×	×	×	✓	✓	×	×	✓	✓
9	✓	✓	✓	×	✓	✓	✓	✓	✓
11	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 4.4: Participants for whom the blind spot was not located. The red cells with an '×'mark indicate the reconstruction map where the blind spot was not located. The white cells with a check mark indicate the reconstruction maps where the blind spot was located.

Detection of suspicious abnormal positions in the VF was found in four participants: participant # 6, participant # 8, participant # 12 and participant # 17. Some visual thresholds in SAP emulation, KBC and ANN reconstruction maps, positioned in 15° upper and between the 15° nasal and temporal were impossible to compute either one or both eyes during the three sessions. Reconstructions maps in Figure 4.10 shows, for the participant #12, the upper area (positions in dark gray) where the values of the

visual thresholds were not computed. Note that for sessions 2 and 3, visual thresholds in 15° upper and between 9° nasal and 9° temporal were not computed. Although the positions in (3° nasal, 15° superior), and (3°temporal, 15° superior) in the first session were not evaluated, the rest of the positions without estimated visual threshold in the 15° upper are the same as in sessions 2 and 3. Given this situation, we consider necessary to clinically evaluate the participants to rule out or validate said anomalies in the VF. On the other hand, the participant # 10 generated contrasting results in the three sessions due to the notable differences between the reconstruction maps for both eyes. Figure 4.11 shows the VF reconstruction maps of the right eye for this particular participant. Note how KBC and ANN reconstruction maps contain several dark-grey areas outside of the blind spot compared to reconstruction maps for the SAP emulation. On the surface, these contrasting differences could be interpreted as a concerning visual impairment given the large number of dark-gray positions. Based on our analysis for this atypical participant, Table 4.5 summarizes the total of positions evaluated in both VFs during the three sessions, as well as the total of positions with no visual threshold computed during the evaluations. In average, we found that 61 positions of the right VF and 62.33 positions of the left VF were evaluated. Of these values we found that 32.02% of the positions evaluated for the right VF, and 23.57% for the left VF were impossible to determine the visual threshold value using KBC; whereas for ANN, the right and left VF percentages were 36.73% and 24.11% respectively. According to the ground truth, the stimuli in these positions were classified correctly as not seen, where it was quantified that at least 48.67% of them the participant fixated his gaze on the reference stimulus instead of generate a gaze path to peripheral stimuli. This behavior of the participant may be caused by different situations: difficulty in performing both task simultaneously during the evaluation, ocular mobility deficiencies, ET detection problems caused by the corrective glasses used by the participant, and others. However, these assumptions need to be tested to determine the causes that yielded these results.
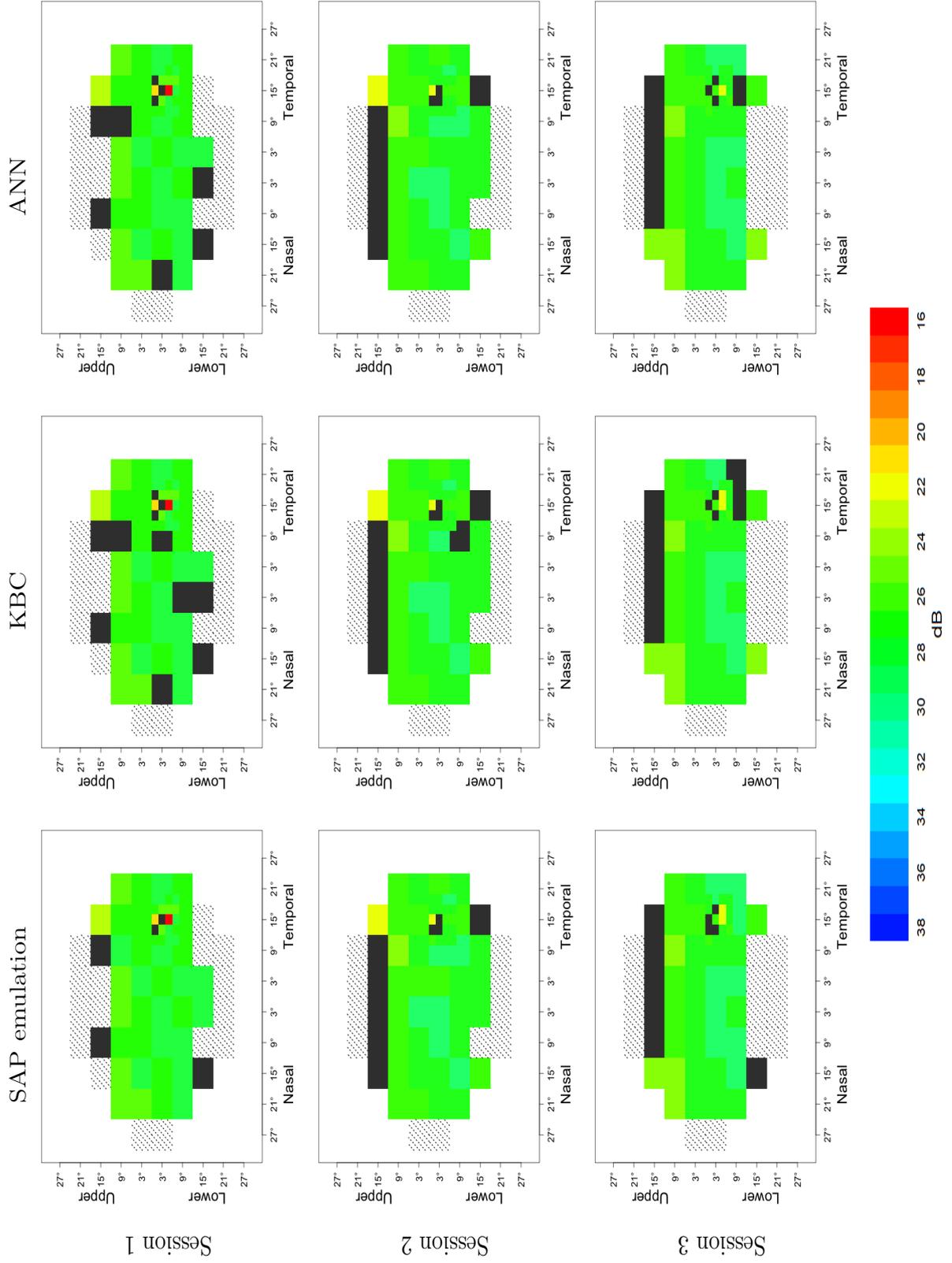
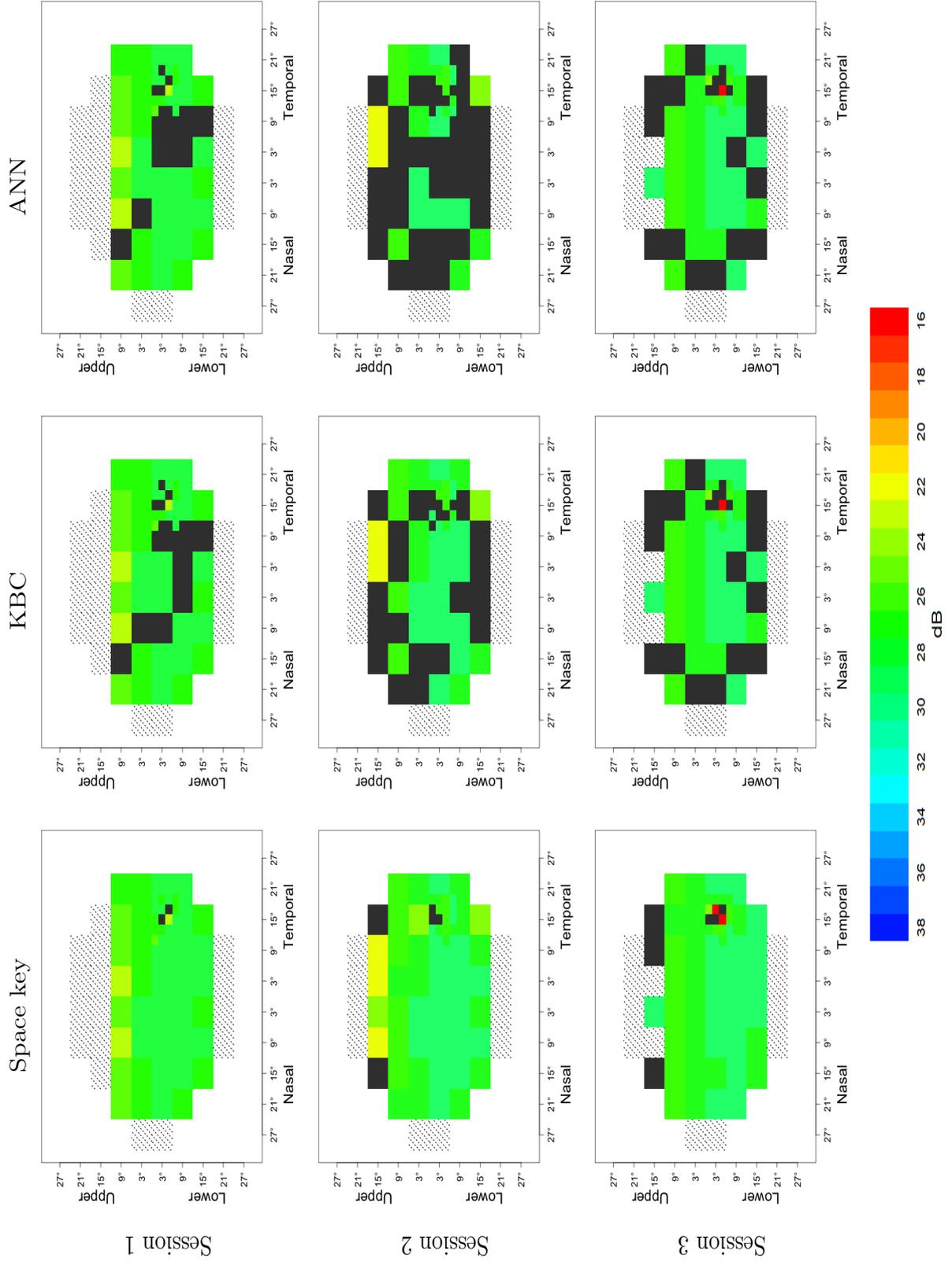Figure 4.11: Reconstructions of the VF of the right eye (heat map) of participant # 12

Figure 4.12: Reconstructions of the VF of the right eye for a participant with unsuccessful evaluations.

		Left eye				

Session	Total of positions evaluated	KBC Visual thresholds not computed		ANN Visual thresholds not computed	
		Positions classified as unseen (%)	Fixations performed on reference stimulus (%)	Positions classified as unseen (%)	Fixations performed on reference stimulus (%)
1	62	22.58	42.85	22.58	42.85
2	62	33.87	47.61	35.48	45.45
3	63	14.28	55.55	14.28	66.66
μ	**62.33**	**23.57**	**48.67**	**24.11**	**51.65**

		Right eye				

Session	Total of positions evaluated	KBC Visual threshold not computed		ANN Visual threshold not computed	
		Positions classified as unseen (%)	Fixations performed on reference stimulus (%)	Positions classified as unseen (%)	Fixations performed on reference stimulus (%)
1	58	24.13	100	22.41	100
2	63	39.68	48	55.55	65.71
3	62	32.25	55	32.25	55
μ	**61**	**32.02**	**67.66**	**36.73**	**73.57**

Table 4.5: Quantification of VF positions with no computed visual threshold for an atypical participant.

4.5.2 Analysis of gaze paths

The Type 3 gaze paths identified in the ground truth provided interesting information about the behavior of the visual system during the evaluation. This type of gaze paths occurs in less than 4% of the total number of stimuli evaluated in each session, where broken gaze paths were the most common in both eyes. As the participants gained experience in conducting the evaluation, it was observed that Subtype 1 gaze trajectories usually occurred in areas such as the periphery or near the blind spot where the stimuli require higher visual intensities to be detected. On the other hand, for Subtype 2 gaze paths the behavior was different. For the left eye it was observed that these gaze paths tend to be found more frequently in peripheral areas of the VF, while for the right eye the behavior is the opposite; that is, Subtype 2 gaze paths are usually found more in central areas of the VF. We hypothesize that the concentration of this type of gaze paths towards one or several specific areas should be the same in both eyes; however, to verify this hypothesis, it is necessary to carry out further tests that involve more than three sessions, so this verification will remain pending as future work.

According to the results shown in Figure 4.7 at least 60% of stimuli with Type 3 gaze paths were recorded as seen by the participants, where based on the results in Figure 4.6, these responses are likely related to stimuli seeking and also related to the evaluation of the central three degrees of VF as mentioned above. In order to reduce this percentage, it is necessary to detect this type of trajectories in real time, so that the corresponding positions can be reassessed in order to compute the correct visual threshold. However, the implementation of this decision rule may increase the evaluation time in our paradigm, so it may be necessary to investigate different alternatives that allow a balance between evaluation time and accuracy results. A possible solution based on the results shown in Table 4.3 is to discard the peripheral positions with Type 3 gaze paths and estimate the visual threshold values using information from

their neighbours. Going a step further, we could apply the same idea to reduce evaluation time; that is, during the evaluation we could discard some peripheral positions and estimate their visual thresholds values using the information from the neighbours.

4.5.3 Comparison between machine learning techniques and SAP emulation

The common indices used to measure the classifiers performance showed that KBC and ANN were better than SAP emulation with accuracy percentages between 96% and 99% for KBC and ANN, while the accuracy values for SAP emulation were between 90% and 92%. From the two machine learning techniques implemented, ANN was slightly better than KCB, showing accuracy differences for the fist and second scenarios of 0.01% and 0.56% for the left eye, and differences of 0.10% and 0.58% for the right eye. The superiority of machine learning techniques represent an decrease in the error rate with respect to SAP emulation between 66% and 70% if Type 3 gaze paths are included in the analysis, and between 78% and 86% if Type 3 gaze paths are not included. Murray et al. [26] have reported indices of sensitivity, specificity and accuracy of 69.1%, 96,9% and 95.0% respectively; however, their methodology is different from ours. Specifically, they perform SAP and EMP tests separately using different devices, which modifies the experimental conditions. Therefore, it is difficult to make a comparison between our paradigm and their paradigm.

4.5.4 Limitations

In EMP paradigms where a computer screen or any other display device is used to display peripheral stimuli, the VF area to be evaluated and the estimation of visual thresholds are conditioned by the technical specifications of these devices. For example, the size of the screen used and the distance between the subject and the monitor determine the evaluation area, while the technologies in screens such as LCD, LED and their variations, determine the range of luminance that will be used during the evaluation. Similar to some EMP studies such as [25–30], our paradigm was unable to fully evaluate the thirty central degrees and use the same range of sensitivities that SAP uses. However, these limitations did not prevent us from evaluating the central fifteen degrees and detecting the blind spot.

Compared against SAP using the SITA Standard C24-2 test, which is one of the most common test used by the ophthalmologists, the mean evaluation time for our paradigm is larger than of the SAP test. On average, the test time for SAP is 6.5 minutes per eye, while for our paradigm, the evaluation time was 7.2 minutes. In spite of this difference of time, a quick survey performed by the participants showed that none of them found the test exhausting, thanks to the short breaks allowed during the evaluation.

Chapter 5

Conclusions

At present, EMP is a developing paradigm whose qualities make it very attractive for VF evaluation, enabling portable, configurable and low-cost implementations capable of providing better control during the evaluation.

In this thesis, a paradigm based on EMP and machine learning for the VF evaluation was presented and validated. The paradigm was designed to estimate visual thresholds at each evaluated position within the central thirty degrees using the C30-2 or C24-2 patterns, as well as to identify the blind spot within this area. Nevertheless, due to the size of the screen used during the experimental stage, the evaluation area was reduced, evaluating in average the central fifteen degrees of the VF. Despite the above, we were able to locate the blind spot and estimate most of the visual thresholds within this area. The reconstruction maps obtained from each participant were satisfactory and showed the behavior described for a healthy VF; that is, a gradual increase of the visual sensitivity from the periphery to the VF center, as shown in figures 4.1 y 4.11.

An important contribution of the paradigm proposed was the gaze paths analysis using machine learning techniques, yielding mean accuracy results greater than 96.5%. This percentage represents an error rate reduction between 67% and 86% compared to methodologies that depend on a subjective patient's response, such as SAP .

Thanks to the visual analysis of the gaze paths for the creation of the ground truth, we were be able to find visual behaviors that were not initially considered. These behaviors, which correspond to the Type 3 gaze paths and the fixations performed by the participants on the reference stimulus when the central three degrees were evaluated, are of great importance in the visual analysis of the gaze paths due to the impact they can have on the results. Although KBC and ANN were able to correctly classify the most stimuli in the central three degrees, Type 3 gaze path detection will have to be considered in future versions of the proposed system.

The evaluation time is one of the disadvantages of our paradigm with respect to SAP. In our case, the average time is 7.2 minutes while the average test time for SAP is 6.5 minutes. Although, the participants who supported the validation of our system did not report inconveniences or negative comments about the evaluation time (perhaps due to the short breaks given during the evaluation), it is necessary to validate our paradigm with children and elderly people. The foregoing, with the aim to identify the necessary factors that may help to improve our paradigm by an adequate balance between precision and evaluation time.

Despite some problems that our paradigm faces, in this thesis we show a low-cost alternative for the VF evaluation with satisfactory results. Although EMP is currently

far from replacing SAP, the constant improvements in Eye Tracking technology as well as new machine learning techniques will allow the development of more robust and accurate tests based on EMP which, together with SAP, may allow the timely identification of visual impairments.

5.1 Future work

Based on the experience obtained in this work, this section presents some ideas that could be developed in the future.

5.1.1 Methodology modifications

There are various modifications in the methodology that could be performed in order to improve or solve some limitations identified in this work.

1. **Increasing the evaluation area.** By positioning the reference stimulus on the screen corners, the VF can be evaluated by zones, in this way the evaluation area can be expanded without the need to modify the size of the screen size or incorporate two or more screens into the system. Some studies in EMP [25–28] constantly move the reference stimulus during the evaluation. While it may be true that these methodologies increase the evaluation area, it is not clear if these studies consider factors such as the participant's head position, or the angle and position of the eye during the evaluation, both of which have a influence in the test and need be considered. To solve the problem, it could be convenient to adjust the position of the screen during the evaluation through a mechanical system, taking the position of the subject as a reference. This could greatly simplify the system by avoiding the incorporation and synchronization of new sensors used to determine the person's position, as well as the reduction of calculations to adjust the position of the peripheral stimuli in real time during the evaluation.

   In the validation stage, a high percentage of people who could not perform a saccadic movement in the central three degrees was identified due to the short distance between stimuli. A possible way to solve this problem could be to estimate an approximation of the visual thresholds in the three central degrees based on the visual threshold values of the neighboring locations and the initial threshold (similar to a labelling problem in image analysis). The approximation of these values can not only be used in the central three degrees, we could also implement the same idea in some positions in the central thirty degrees to reduce the number of positions to be evaluated, and therefore reduce the time during the evaluation.

2. **Gaze path analysis.** The extraction, selection and preprocessing of features from a database is one of the most important processes in machine learning, as it often determinates whether or not the prediction model will provide reliable results. In order to improve our system, we need find new features that help identify the three types of gaze paths mentioned in this work. For example, one could experiment with the latency and the number of saccadic movements under the following hypothesis: if a stimulus is seen, one would expect to observe a fast response time of the subject (perhaps in the fist half of the total time

of the peripheral stimulus), as well as a single saccadic movements towards the peripheral stimulus. Conversely, if some stimulus is not seen, one would expect to observe longer reaction times (or no reaction time in case of a fixation) with one or more saccades in direction to other than the peripheral stimulus, whereas for Type 3 gaze paths, the last saccadical movement would be in the direction of the peripheral stimulus.

The incorporation of real-time gaze paths analysis is a necessary and important step to perform decisions during the evaluation, for example, when a Type 3 gaze path is detected. To do this using machine learning, it is necessary to implement supervised techniques, so it becomes necessary to improve the database for this purpose. The aim is to create a database that covers a wide sector of people; that is, from youth to older adults with and without visual impairments. Among the different supervised machine learning techniques that one could implement are Naive Bayes, Support Vector Machine (SVM), Random Forest and others. According to the results, ANN could be a good option, however it is necessary to modify the architecture in order to detect the Type 3 gaze paths. An interesting technique that should be considered is Convolutional Neural Network (CNN). This deep learning technique, which has become prominent in image processing, combines the feature extraction and the classification process by using a full connected network with convolution stages, subsampling stages and a fully connected ANN in its architecture. Figure 5.1 shows the basic architecture of a CNN. So, if one can manipulate the gaze path as an image, one could train a CNN for gaze path classification with the advantage that we would not have to worry about the extraction and selection of features for the prediction model, since the technique does this during the training.
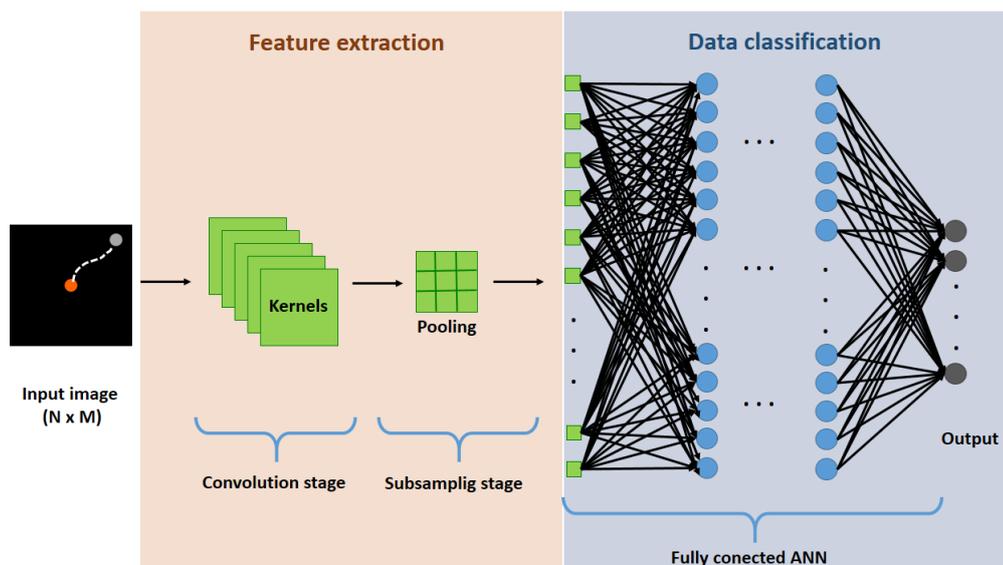
Figure 5.1: Basic architecture of a CNN. In practice, it is common to use two or more convolution and subsample stages to obtain better feature maps and to reduce computational load during CNN training.

5.1.2 Validation with children and elderly people

In order to improve the paradigm, it is necessary validate the system with children and elderly people with and without visual impairments. In this way, it is possible to identify and solve some problems that cannot be identified in a healthy young adult population, as well as to acquire information correlated with the study population, such as the evaluation time or the precision of the tests.

Appendix A

Machine learning techniques

This chapter describes the classifiers used during the gaze path analysis. For a better understanding of some techniques implemented, their descriptions are generalized for the two class case. However, these techniques can be adapted to n classes.

A.1   KBC classifier

Before explaining the KBC classifier, we define the following variables for a better understanding of the classifier.

- $\mathbf{D_q} = (\boldsymbol{\Delta}\theta_{\mathbf{i,j}}, \mathbf{d_{i,j}})$: corresponds to the peripheral stimulus q to be classified according to the computed features of its gaze path.

- $\mathbf{t}$: refers to the class number. In our particular case $\mathbf{t}$=1,2, where $1 = $ *stimulus seen*, and $2 = $ *stimulus not seen.*

- $\mathbf{C_t}$: corresponds to the centroid of class $\mathbf{t}$.

- $\mathbf{Dist_{q,Ce_t}}$: indicates the distance between $\mathbf{D_q}$ and $\mathbf{C_t}$.

- $\mathbf{Cl_t}$: indicates the class $\mathbf{t}$.

- $\mathbf{L_q}$ corresponds to the label of $\mathbf{D_q}$.

KBC consist of two classifiers: K-means and Bayesian classifier, whose description is as follows.

K-means is one of the most widely used clustering algorithms due to its good performance and easy implementation. This iterative and unsupervised algorithm groups data with similar features into r clusters using the minimum distance between the center of each cluster and the data as the decision rule; that is, the data is assigned to the closest cluster based on the metric used. Commonly, the used metric in the algorithm is the Euclidean metric; however, metrics such as Minkowski or the Manhattan distance can also be implemented. Once all the data is assigned to a cluster, the position of the center of each cluster is updated and the process of assigning data to the nearest cluster is repeated again. The algorithm stops until the position of the center of each cluster does not change significantly after its update. For the algorithm to work, the user must provide the total number of clusters into which the data will be separated, where each cluster is assigned a centroid that must be initialized at the start of the

56

algorithm. However, it is important to consider that the initial values of the centroids may have a significant impact in the classification of the data as shown in Figure A.1, where for the same data set (black circles), the two defined centroids represented as red and green diamonds generated clusters with different elements.

To classify m data points into one of the two classes using the K-means algorithm, the process can be summarized in the following steps:

1. Initialize the centroids as $C_t|_{t=1} = (0,0)$ and $C_t|_{t=2} = (0.8, 0.8)$. These centroids are chosen empirically, according to the expected behavior of the gaze paths: for a peripheral stimulus that was seen (class $\mathbf{t} = 1$), the gaze path should have a small distance to the stimulus and a small angular difference.

2. For each datum D_q, compute the distance between it and the centroids as

$$Dist_{q,C_t} = |C_t - D_q|. \tag{A.1}$$

3. Label D_q based on nearest centroid as

$$L_q = t \quad if \quad Dist_{q,C_t} < Dist_{q,C_u}, \quad for \quad t \neq u. \tag{A.2}$$

4. Update the positions of the centroids as

$$C_t = \frac{\sum_{q|L_q=t} D_q}{\sum_{q|L_q=t} 1}. \tag{A.3}$$

5. Repeat the steps 2-4 until the positions of the centroids are almost the same after the update.

On the other hand, a Bayesian classifier (also known as Naive Bayes classifier) is probabilistic classifier based on the Bayes theorem, where considering the features of the data as independent variables, the objective is compute the probability that D_q belongs to Cl_t; that is

$$P(Cl_t|D_q) = \frac{P(Cl_t)\, p(D_q|Cl_t)}{p(D_q)}, \tag{A.4}$$

From the above expression, $P(Cl_t)$ (known as *prior probability*) corresponds to a probabilistic weight assigned to each class t. In our particular case, these weights are considered equally likely; that is, $P(Cl_{t|t=1}) = P(Cl_{t|t=2}) = 0.5$. On the other hand, the denominator, $p(D_q)$, corresponds to a scalar value that guarantees that

$$\sum_{t=1}^{2} P(Cl_t|D_q) = 1. \tag{A.5}$$

With this in mind and for data classification purposes, such values can be omitted from the expression without affecting the decision rule. Therefore, the above expression can be rewritten as

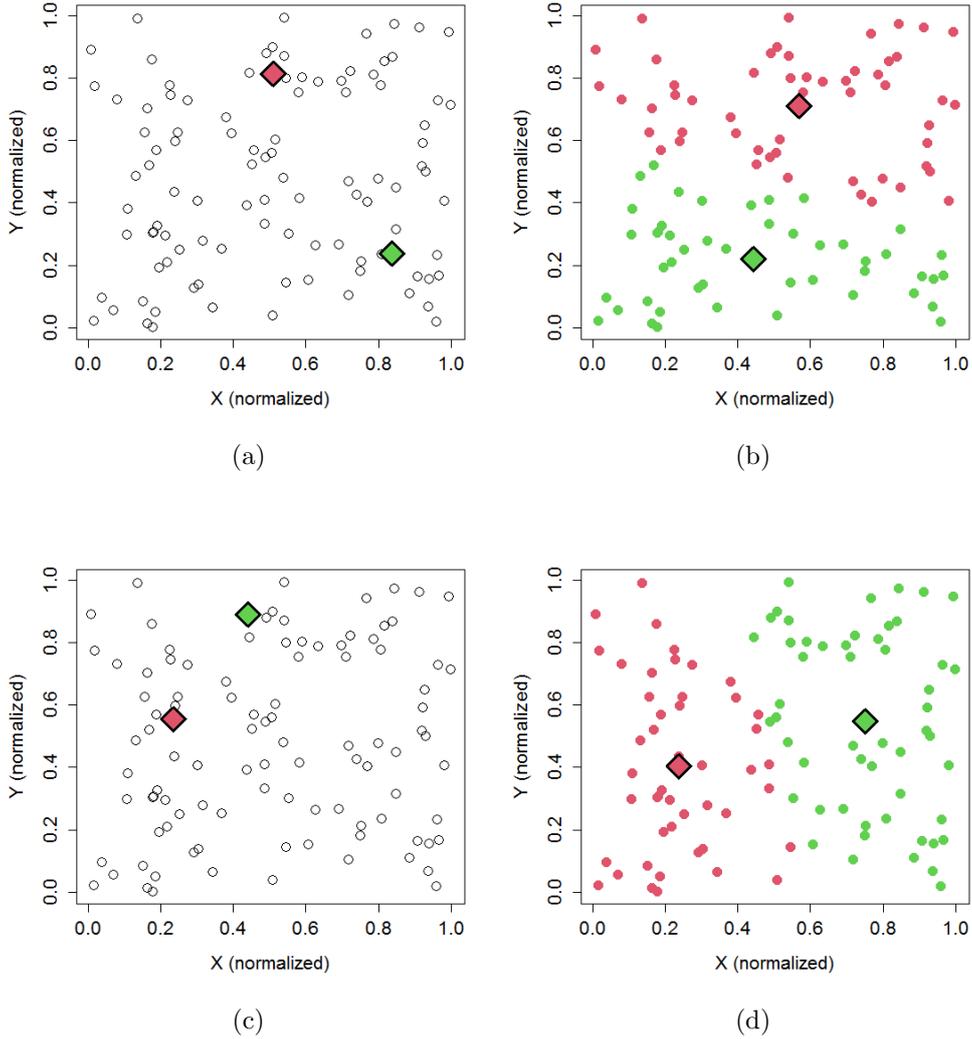$$P(Cl_t|D_q) \propto p(D_q|Cl_t). \tag{A.6}$$

Figure A.1: Different classification of data in two classes (red or green) according to the initialization of the centroids (diamonds). (a) Initialization of centroids as $centroid_1 = (0.51, 0.84)$ and $centroid_2 = (0.81, 0.24)$. (b) Classification of the data into two classes according to the initial position of the centroids shown in part (a). (c) Initialization of the centroids as $centroid_1 = (0.230.55)$ and $centroid_2 = (0.44, 0.89)$. (d) Classification of the data into two classes according to the initial position of the centroids shown in part (c).

An advantage of this classifier is that it can be represented as a set of discriminant functions $h_t(D_q)$, where considering that each class has a multivariate Gaussian distribution the equation A.6 can be rewritten as

$$P(Cl_t|D_q) \propto h_t(D_q) = \frac{e^{-\frac{1}{2}(D_q - \mu_t)^T \Sigma_t^{-1} (D_q - \mu_t)}}{(2\pi)^{\alpha/2} (\det \sum_t)^{1/2}}, \tag{A.7}$$

where $(D_q - \mu_t)^T \Sigma_t^{-1} (D_q - \mu_t)$ is known as the Mahalanobis distance squared. From the expression A.7, μ_t is the mean of the data belonging to Cl_t (in our particular case these values initially correspond to the values of the last update of the position of the clusters in K-means; that is, $\mu_t = C_t$), Σ_t is the covariance matrix of

Cl_t, Σ_t^{-1} and $\det \Sigma_t$ correspond to the inverse and the determinant of the covariance matrix, respectively, and finally $(D_q - \mu_t)^T$ is the transpose of $D_q - \mu_t$.

With this in mind, the decision rule of the Bayesian classifier is the following:

$$L_q = t \quad if \quad h_t(D_q) > h_u(D_q) \quad for \quad t \neq u. \tag{A.8}$$

Once all the data (D_q) have been labeled according to the decision rule expressed by A.8, the Bayesian classifier is applied iteratively until convergence, updating the values of μ_t and Σ_t in each iteration.

A.2 Artificial Neural Networks

An Artificial Neural Network (ANN) is a machine learning technique whose structure and name are inspired by the human brain. ANNs are designed to mimic the mechanisms of the brain, which contains millions of neurons designed to receive, process and transmit information. In the case of ANNs, the neuron's counterpart is called *perceptron* or *node*, which is the minimum expression for the ANN's development.

A perceptron, as shown in Figure A.2, is formed by n inputs, n weights, a bias, an activation function, and an output. The perceptron sums all the inputs multiplied by a weight plus the *bias*. This weighted output is passed through the activation function that provides a non-linear output on the perceptron, allowing discrete and non-linear problems to be solved. This process is described by the following expressions:

$$\alpha = WX + bias, \tag{A.9}$$

$$y = f(\alpha), \tag{A.10}$$

where y is the output of the perceptron, α corresponds to the weighted sum of the perceptron, $W = [w_1, w_2, ..., w_n]$ and $X = [x_1, x_2, ..., x_n]^T$ are vectors of weights and inputs respectively, *bias* is an adjustment parameter and it act as an input node that always produces a constant value, and $f(\alpha)$ is the activation function, where the functions most used are Sigmoid, Tanh and ReLU as shown in the equations A.11-A.13 ; however variations of ReLU such as LReLU, PReLU or GELU are other examples of activation functions.
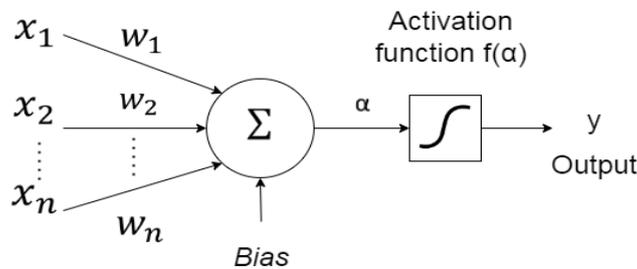

Figure A.2: Example of a perceptron.

$$f(\alpha) = \frac{1}{1 + e^{-\alpha}} \qquad \text{(Sigmoid function).} \qquad (A.11)$$

$$f(\alpha) = \frac{e^{\alpha} - e^{-\alpha}}{e^{\alpha} + e^{-\alpha}} \qquad \text{(Tanh function).} \qquad (A.12)$$

$$f(\alpha) = max(0, \alpha) \qquad \text{(ReLU function).} \qquad (A.13)$$

Therefore, an ANN consists of several perceptrons connected to each other and grouped in layers, where the outputs of a given layer are used as inputs of the next layer. In this way, introducing the input values to the ANN, a prediction is obtained in the output with an error value respect to the expected output value. Figure A.3 shows the general architecture of one of the most widely used ANNs, where an ANN with one input layer, one inner layer and one output layer is called as *a shallow neural network*, while an ANN with more than one inner layer it is called as *deep neural network*.
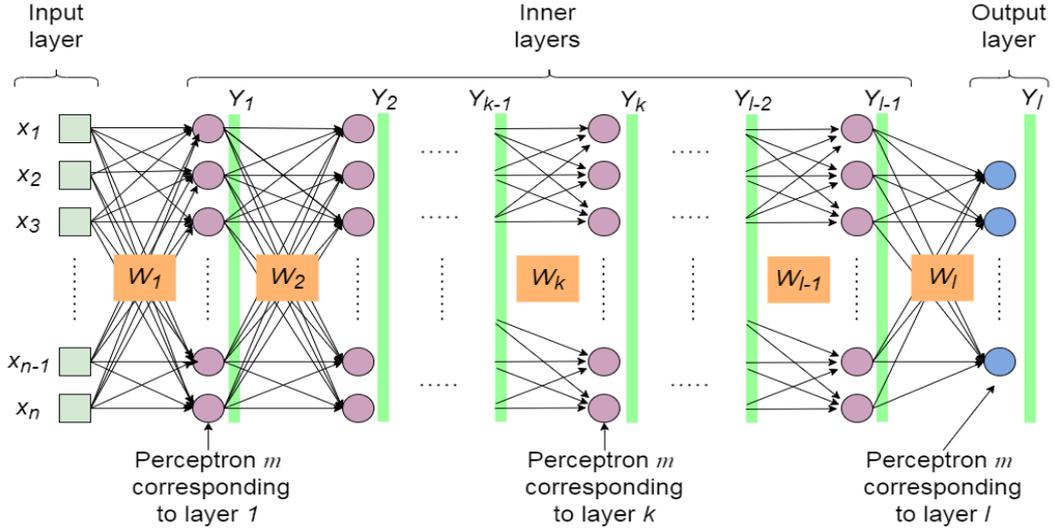
Figure A.3: Architecture of an ANN.

To compute the output in the ANN it is necessary to sequentially compute the outputs of its l layers [1]. Therefore, the equations A.9 and A.10 can slightly modified as

$$\alpha_k = W_k X_k + B_k, \qquad (A.14)$$
$$Y_k = f_k(\alpha_k). \qquad (A.15)$$

Where

---

[1]Since the input layer is not a layer composed of perceptrons, the expression for the calculation of the output layers applies only to the internal layers and the output layer of the ANN.

$$W_k = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} & \cdots & w_{1,n} \\ w_{2,1} & w_{2,2} & w_{2,3} & \cdots & w_{2,n} \\ w_{3,1} & w_{3,2} & w_{3,3} & \cdots & w_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{m,1} & w_{m,2} & w_{m,3} & \cdots & w_{m,n} \end{bmatrix},$$

$$X_k = [x_1, x_2, x_3, \cdots, x_n]^T, \quad \text{and} \quad \mathrm{B}_k = [bias_1, bias_2, bias_3, \cdots, bias_m]^T$$

correspond to the weight matrix, the input vector and the vector of bias respectively for the layer $k = 1, 2, 3, \cdots, l$ of the ANN. W_k is a weight matrix of size $m \times n$, where m is the total number of perceptrons in the layer k, and n is the total number of inputs for each perceptron; that is, the rows of the matrix correspond to the n weights of each perceptron. Depending on the number of perceptrons in the output layer, the output Y_k will either be a column vector of size m or a single value if there is only one perceptron in the output layer.

Similar to other supervised machine learning algorithms, ANN seeks to minimize a given error measure between the expected output and the model output during its training. To do this, ANN must adjust the values of the weights and biases of each perceptron. The way to adjust those values is done through the *backpropagation algorithm* [43] and *gradient descent algorithm*, where the first algorithm computes the gradient of the error with respect to the weights and biases, and the second algorithm uses those gradients to minimize the ANN output error.

Therefore, based on Figure A.3, the update of the weights and biases is as follows: Starting with the last layer we have the following partial derivatives:

$$\frac{\partial C}{\partial W_l} = \left(\frac{\partial C}{\partial Y_l} \right) \left(\frac{\partial Y_l}{\partial \alpha_l} \right) \left(\frac{\partial \alpha_l}{\partial W_l} \right), \tag{A.16}$$

$$\frac{\partial C}{\partial B_l} = \left(\frac{\partial C}{\partial Y_l} \right) \left(\frac{\partial Y_l}{\partial \alpha_l} \right) \left(\frac{\partial \alpha_l}{\partial B_l} \right). \tag{A.17}$$

Where C corresponds to the cost function used to compute the error in the ANN output, being the *mean square error* and the *cross entropy* some of the functions commonly implemented. By comparing the equations A.16 and A.17, it is easy to see that only the last partial derivative of the product is different. Therefore, both expressions can be rewritten as

$$\frac{\partial C}{\partial W_l} = \delta_l \left(\frac{\partial \alpha_l}{\partial W_l} \right) = (\delta_l) \left(X_l{}^T \right), \tag{A.18}$$

$$\frac{\partial C}{\partial B_l} = \delta_l \left(\frac{\partial \alpha_l}{\partial B_l} \right) = \delta_l. \tag{A.19}$$

Where $\delta_l = \left(\frac{\partial C}{\partial \alpha_l} \right) = \left(\frac{\partial C}{\partial Y_l} \right) \left(\frac{\partial Y_l}{\partial \alpha_l} \right)$ indicates how the error in the ANN output varies with respect to the weighted sum of each of the perceptrons in the layer l.

Now, continuing with the $l - 1$ layer we have the following partial derivatives:

$$\frac{\partial C}{\partial W_{l-1}} = \left(\frac{\partial C}{\partial Y_l}\right)\left(\frac{\partial Y_l}{\partial \alpha_l}\right)\left(\frac{\partial \alpha_l}{\partial Y_{l-1}}\right)\left(\frac{\partial Y_{l-1}}{\partial \alpha_{l-1}}\right)\left(\frac{\partial \alpha_{l-1}}{\partial W_{l-1}}\right), \tag{A.20}$$

$$\frac{\partial C}{\partial B_{l-1}} = \left(\frac{\partial C}{\partial Y_l}\right)\left(\frac{\partial Y_k}{\partial \alpha_l}\right)\left(\frac{\partial \alpha_l}{\partial Y_{l-1}}\right)\left(\frac{\partial Y_{l-1}}{\partial \alpha_{l-1}}\right)\left(\frac{\partial \alpha_{l-1}}{\partial B_{l-1}}\right). \tag{A.21}$$

Similar to what was done in the last layer, the expressions A.20 and A.21 can be rewritten as

$$\frac{\partial C}{\partial W_{l-1}} = \delta_{l-1}\left(\frac{\partial \alpha_{l-1}}{\partial W_{l-1}}\right) = \left(\delta_{l-1}\right)\left(X_{l-1}{}^T\right), \tag{A.22}$$

$$\frac{\partial C}{\partial B_{l-1}} = \delta_{l-1}\left(\frac{\partial \alpha_{l-1}}{\partial B_{l-1}}\right) = \delta_{l-1}. \tag{A.23}$$

Where $\delta_{l-1} = \left(\frac{\partial C}{\partial \alpha_{l-1}}\right) = W_l{}^T \delta_l \star f'(\alpha_{l-1})^2$ indicates how the error in the ANN output varies with respect to the weighted sum of the each perceptrons in the layer $l - 1$. Repeating the procedure described with the rest of the layers, it is observed that for any layer $k \neq l$ the partial derivatives of the error with respect to the weights and biases are expressed as

$$\frac{\partial C}{\partial W_k} = \delta_k\left(\frac{\partial \alpha_k}{\partial W_k}\right) = \left(\delta_k\right)\left(X_k{}^T\right), \tag{A.24}$$

$$\frac{\partial C}{\partial B_k} = \delta_k\left(\frac{\partial \alpha_k}{\partial B_k}\right) = \delta_k. \tag{A.25}$$

Where $\delta_k = W_{k+1}{}^T \delta_{k+1} \star f'(\alpha_k)$.

Once partial derivatives are computed for all layers, the weights and biases are updated as

$$W_{k(new)} = W_k + \beta\left(\frac{\partial C}{\partial W_k}\right) \tag{A.26}$$

$$W_{k(new)} = W_k + \beta\left(\frac{\partial C}{\partial B_k}\right) \tag{A.27}$$

where β is a numerical value between 0 and 1 indicating how fast the weights and biases are updated. The process described above (calculation of partial derivatives and update of values) is iteratively repeated until convergence is determined (e.g. when a given number of iterations is reached or when the norm of the derivatives is small).

The method for train an ANN can be summarized as follows:

---

[2]The symbol $\star$ will be used to indicate the element-wise matrix product.

1. Initialize randomly the weights and biases of the ANN.

2. Compute the output for the last ANN layer iteratively using the expressions A.14 and A.15.

3. Compute the partial derivatives for the last layer using the equations A.18 and A.19.

4. Compute the partial derivatives for the rest of the layers using the equations A.24 and A.25.

5. Update the weights and biases according to the equations A.26 and A.27.

6. Repeat the steps 2-5 until convergence criteria is met.

Appendix B

Experimental validation results: visual maps

This appendix compiles all the VF reconstruction maps of each participant who helped in the validation of the paradigm described in this thesis. For more details see the CD-ROM attached to this thesis.

Appendix C

Published works

The compilation of products generated from this work are listen as follows.

- Congress presentation:

  1. Martínez-González, E.A., Alba, A., Méndez, M.O. and Fernández- Wong J.. "Desarrollo de una prueba de concepto para la evaluación del campo visual mediante seguimiento de la mirada". 41° Congreso Nacional de Ingeniería Biomédica (CNIB). Oct, 2018. León, Gto.

  2. Martínez-González, E.A., Alba, A., Méndez, M.O. et al. "Eye tracking en la evaluación del campo visual". 2do Encuentro Nacional de Ingeniería Biomédica, Electrónica y Telecomunicaciones (ENIBET). Oct, 2018. Universidad Autónoma de San Luis Potosí.

  3. Martínez-González, E.A., Alba, A., Méndez, M.O. et al. "Eye tracking: una herramienta para la evaluación del campo visual". 3er Encuentro Nacional de Ingeniería Biomédica, Electrónica y Telecomunicaciones (ENIBET). 2019. Universidad Autónoma de San Luis Potosí.

- Research articles:

  1. Martínez-González, E.A., Alba, A., Méndez, M.O. et al. Developing a visual perimetry test based on eye-tracking: proof of concept. Health Technol. 10, 437–441 (2020). https://doi.org/10.1007/s12553-019-00366-9

  2. Martínez-González, E.A., Alba, A., Arce-Santana, E. et al. "A novel system for the automatic reconstruction of visual field based on eye tracking and machine learning". Multimed Tools Appl 82, 27193–27215 (2023). https://doi.org/10.1007/s11042-023-14464-4

ORIGINAL PAPER

Developing a visual perimetry test based on eye-tracking: proof of concept

Eduardo A. Martínez-González [1,2] · **Alfonso Alba** [1,2] · **Martín O. Méndez** [1,2] · **Jorge Fernández-Wong** [3]

Abstract
Computerized Perimetry (CP) is one of the clinical tests commonly used to evaluate peripheral vision and monitor the progress of eye diseases such as glaucoma. The aim of CP is to determine retinal sensitivity using luminous stimuli of variable intensity at different positions of the visual field. In modern campimetry devices, patients must respond to each perceived stimulus by pressing a button; however, this characteristic makes the test more susceptible to spurious and erroneous interpretations due to tiredness, lack of concentration, or device design flaws. This work presents an alternative paradigm for automatically assessing stimulus perception through a low-cost eye tracker and a computer monitor. We tested the preliminary version of the paradigm among eight subjects and obtained favorable results. In conclusion, our eye-tracking paradigm tool could help design more reliable visual field tests using low-cost portable equipment.

Keywords Computerized Perimetry · Eye tracker · Glaucoma · Visual field

1 Introduction

Human body deterioration is a natural and inevitable process, which negatively affects the health and lifestyle of most people. The sense of sight, which involves complex processing of signals in the nervous system, is one of the systems that are commonly affected by senescence [1]. According to 2017 statistics from the World Health Organization (WHO), approximately 253 million people suffer from visual disability, 36 million of which are completely blind. Adults aged 50 years or older account for a high percentage of the blind people, thus suffering mostly from cataracts (35%), refractive correction errors (21%) and glaucoma (8%) [2]. Similarly to many degenerative diseases, glaucoma is incurable and characterized by a gradual loss of the visual field, with no other symptoms

such as pain. However, multiple clinical tests are nowadays available to diagnose, evaluate, and monitor the progress of glaucoma and other diseases of the retina and the optic nerve. One of these tests is computerized campimetry (CC) or computerized perimetry (CP) [1].

CC employs luminous stimuli of variable size and intensity, located at different points within the visual field. The goal of the test is to evaluate the sensitivity of the retina to the different light intensities by asking the subject to respond to the stimuli by pressing a button each time a stimulus is perceived. Under this paradigm, CC demands a considerable amount of concentration from patients, who must also focus their sight toward a specific area (usually the centre of the screen) during the test. After a few minutes, patients often become easily tired and/or distracted, which may lead to spurious responses and thus unreliable results.

An eye-tracker (ET) is a device that registers gaze location within a certain area, often using an infrared light source and a high-speed camera. Currently, ETs are successfully used in different areas of research and entertainment [3]. In the medical field, they have been used for a wide range of purposes, including the assessment of visual adaptability of glaucoma patients in daily activities such as walking [4], determining the impact of cataracts in eye sensibility [5], or proposing new campimetry paradigms to overcome the difficulty that some people have in fixating their gaze in the central luminous

✉ Eduardo A. Martínez-González
arturomtz.2012@gmail.com

1 Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P, México

2 Laboratorio Nacional CI3M & CICSaB, Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P, México

3 Hospital de Especialidades Médicas de la Salud, San Luis Potosí, S.L.P, México

⚫ Springer

A novel system for the automatic reconstruction of visual field based on eye tracking and machine learning

Eduardo A. Martínez-González[1] · Alfonso Alba[1,2] · Edgar Arce-Santana[1,2] · Jorge Fernández-Wong[3] · Martin O. Mendez[1,2]

Abstract

Eye movement perimetry (EMP) is a paradigm developed to assess the visual field without the necessity of suppressing the natural eye movements during the test. Unlike the standard automated perimetry (SAP) where the patient's responses are recorded using a button, EMP uses the natural eye movements reflex as responses during the evaluation. The reliability of EMP depends on correctly determining whether a stimulus is seen or not which, in turn, depends on an adequate analysis of the eye movement data. However, many studies in EMP have focused on characterizing eye movements and only a few authors have documented their methods to determine whether a peripheral stimulus was seen during the test. Furthermore, many of them use static thresholds to perform the classification, but it is not clear how these threshold values were obtained. Based on the foregoing, we develop a threshold test based on FASTPAC C24-2 and EMP for the visual field assessment. Our method uses two machine learning techniques: (1) cascaded K-Means and Bayesian classifiers (KBC) and (2) an Artificial Neural Network (ANN) to classify whether a stimulus was seen or not. Our method was validated with twenty healthy participants (13 women and 7 men) aged 19–43 years (μ = 26 ± 5 years), where the participants performed both an EMP test and an SAP emulation test. Results were compared with gaze trajectories annotations performed by an expert, obtaining accuracy values between 96.8% and 98.9% for KBC and ANN, and values between 90.5% and 92% for SAP emulation.

✉ Martin O. Mendez
martin.mendez@uaslp.mx

[1] Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P, México

[2] Laboratorio Nacional CI3M, Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P, México

[3] Hospital de Especialidades Médicas de la Salud, San Luis Potosí, S.L.P, México

🖄 Springer

Bibliography

[1] World Health Organization (2023, August 10), "Blindness and vision impairment", . [Online]. Available: https://www.who.int/en/news-room/factsheets/detail/blindness-and-visual-impairment.

[2] World Health Organization (2019, October 8), "World report on vision",. Licence: CC BY-NC-SA 3.0 IGO. Available: https://www.who.int/publications/i/item/world-report-on-vision.

[3] L. M. Alencar, F. A. Medeiros, "The role of standard automated perimetry and newer functional methods for glaucoma diagnosis and follow-up",*Indian J Ophthalmol*, 2011 Jan, 59 Suppl(Suppl1):S53-8, doi: 10.4103/0301-4738.73694. PMID: 21150035; PMCID: PMC3038506

[4] G. Szatmáry, V. Biousse and N. J. Newman, "Can Swedish Interactive Thresholding Algorithm Fast Perimetry Be Used as an Alternative to Goldmann Perimetry in Neuro-ophthalmic Practice?",*Arch Ophthalmol*, 2002, 120(9):1162–1173, doi:10.1001/archopht.120.9.1162

[5] C. A. Johnson, M. Wall and H.S. Thompson, "A History of Perimetry and Visual Field Testing", *Optometry and vision science : official publication of the American Academy of Optometry*, vol. 88, pp E8-15, 01 2011, doi.org/10.1097/OPX.0b013e3182004c3b

[6] F. G. Junoy Montolio, C. Wesselink, M. Gordijn and N. M. Jansonius, "Factors That Influence Standard Automated Perimetry Test Results in Glaucoma: Test Reliability, Technician Experience, Time of Day, and Season", *Invest.Ophthalmol.Vis.Sci,*53 (11):7010-7017, 2012, doi: https://doi.org/10.1167/iovs.12-10268

[7] C. A. Johnson, C. W. Adams and R. A. Lewis, "Fatigue effects in automated perimetry ", *Appl Opt.* 198827(6):1030-1037. doi:10.1364/AO.27.001030

[8] W. Kosnik, J. Fikre and R. Sekuler, "Visual fixation stability in older adults", *Invest. Ophthalmol. Vis. Sci.*, 27(12):1720-1725, 1986.

[9] S. Demirel and A. J. Vingrys, "Eye Movements During Perimetry and the Effect that Fixational Instability Has on Perimetric Outcomes", *J. Glaucoma,*3(1) 28-35. PMID: 19920549, 1994.

[10] S. L. Graham, A. I. Klistorner, J. R. Grigg and F. A. Billson,"Objective VEP perimetry in glaucoma: asymmetry analysis to identify early deficits", *J. Glaucoma,*9:10–19, 2000, doi: 10.1097/00061198-200002000-00004

[11] I. Goldberg, S. L. Graham and A. I. Klistorner, "Multifocal objective perimetry in the detection of glaucomatous field loss", *Am. J. Ophthalmol*,133:29–39, 2002, doi: 10.1016/s0002-9394(01)01294-6

[12] B. Wangsupadilok, V. C. Greenstein, F. N. Kanadani, et al., "A method to detect progression of glaucoma using the multifocal visual evoked potential technique", *Doc. Ophthalmol* 118, 139–150, 2009, doi.org/10.1007/s10633-008-9149-2

[13] D. C. Hood, P. Thienprasiddhi, V. C. Greenstein, B. J. Winn, N. Ohri, et al., "Detecting Early to Mild Glaucomatous Damage: A Comparison of the Multifocal VEP and Automated Perimetry", *Invest. Ophthalmol, Vis. Sci.*, 45(2):492-498, 2004, doi: https://doi.org/10.1167/iovs.03-0602

[14] R. H. Spector Visual Fields. In: Walker HK, Hall WD, Hurst JW, editors. Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition. Boston: Butterworths; 1990. Chapter 116. Available from: https://www.ncbi.nlm.nih.gov/books/NBK220/

[15] R. Cubbidge, "Eye Essentials, Visual Fields", Elsevier, 2005.

[16] J. Walsh, "Visual Fields: Examination and interpretation", Third Edition, Oxford University Press, 2011

[17] J. J. S. Barton, and M. Benatar,"Field of Vision. A Manual and Atlas of Perimetry". Humana Press Inc., 2003.

[18] R. O. Duda, P. E. Hart and D. G. Stork, "Pattern Classification". 2nd ed. John Wiley & Sons, 2012.

[19] C. C. Aggarwal, "Neural Networks and Deep Learning: A Textbook". Springer International Publishing, 2018.

[20] M. F. Armaly. "The Size and Location of the Normal Blind Spot". *Archives of Ophthalmology*, 81(2), 192–201, 1969, doi:10.1001/archopht.1969.00990010194009

[21] G. J. Jaffe, J. A Alvarado and R. P. Juster, "Age-Related Changes of the Normal Visual Field", *Archives of Ophthalmology*, 104(7), 1021–1025, 1986, doi:10.1001/archopht.1986.01050190079043

[22] C. A. Johnson, A. J. Adams, J. D. Twelker and J. M. Quigg, "Age-related changes in the central visual field for short-wavelength-sensitive pathways", *Journal of the Optical Society of America A.*, 5(12), 2131, 1988, doi:10.1364/josaa.5.002131

[23] P. H. Artes, A. Iwase, Y. Ohno, Y. Kitazawa and B. C. Chauhan, "Properties of Perimetric Threshold Estimates from Full Threshold, SITA Standard, and SITA Fast Strategies", *Invest. Ophthalmol. Vis. Sci.*, 43(8):2654-2659, 2002, PMID:12147599

[24] D. E. Kim, M. Eizenman, G. E. Trope and C. Kranemann, "Eye movement perimetry", *Proceedings of 17th international Conference of the Engineering in Medicine and Biology Society*, Montreal, Quebec, Canada, IEEE, 1629–1630, 1995, doi: 10.1109/IEMBS.1995.579863.

[25] I. C. Murray, B. W. Fleck, H. M. Brash, M. E. MacRae, et al., "Feasibility of saccadic vector optokinetic perimetry: a method of automated static perimetry for children using eye tracking", *Ophthalmology*, 116:2017–2026, 2009, doi:10.1016/j.ophtha.2009.03.015

[26] I. C. Murray, L. A. Cameron, A. D. McTrusty, et al., "Feasibility, Accuracy, and Repeatability of Suprathreshold Saccadic Vector Optokinetic Perimetry", *Trans. Vis. Sci. Tech*, 5(4):15, 2016, doi: https://doi.org/10.1167/tvst.5.4.15.

[27] I. C. Murray, A. Perperidis, L. A. Cameron, et al., "Comparison of Saccadic Vector Optokinetic Perimetry and Standard Automated Perimetry in Glaucoma. Part I: Threshold Values and Repeatability", *Trans. Vis. Sci. Tech*, 6(5):3, 2017, doi: https://doi.org/10.1167/tvst.6.5.3.

[28] A. D. McTrusty, L.A Cameron, A. Perperidis, et al., "Comparison of Threshold Saccadic Vector Optokinetic Perimetry (SVOP) and Standard Automated Perimetry (SAP) in Glaucoma. Part II: Patterns of Visual Field Loss and Acceptability", *Trans. Vis. Sci. Tech*, 6(5):4, 2017, doi: https://doi.org/10.1167/tvst.6.5.4

[29] P. R. Jones, N. D. Smith, W. Bi and D. P. Crabb, "Portable Perimetry Using Eye-Tracking on a Tablet Computer—A feasibility Assessment", *Trans. Vis. Sci. Tech*, 8(1):17, 2019, doi: https://doi.org/10.1167/tvst.8.1.17

[30] P. R. Jones, "An open-source static threshold perimetry test using remote eye-tracking (Eyecatcher): description, validation and preliminary normative data", *Trans. Vis. Sci. Tech*, 9(8):18, 2020, doi: https://doi.org/10.1167/tvst.9.8.18

[31] R. B. Weber and R. B. Daroff. "Corrective movements following refixation saccades: Type and control system analysis", *Vision Research*, 12(3), 467–475, 1972, doi:10.1016/0042-6989(72)90090-9

[32] W. Becker and A. F. Fuchs, "Further properties of the human saccadic system: Eye movements and correction saccades with and without visual fixation points", *Vision Research*, 9(10), 1247–1258, 1969, doi:10.1016/0042-6989(69)90112-6

[33] M. E. Jernigan, "Structural analysis of eye movement response to visual field stimuli", *Computers in Biology and Medicine*, 10(1), 11–22, 1980, doi:10.1016/0010-4825(80)90003-7

[34] M. Wang, L. Q. Shen, M. V. Boland, et al., "Impact of Natural Blind Spot Location on Perimetry", *Scientific Reports*, 7(1), 2017, doi:10.1038/s41598-017-06580-7

[35] A. Heijl and C. E. T. Krakau, "An Automatic Static Perimeter Design and Pilot Study", *Acta Ophthalmologica*, 53(3), 293–310, 1975, doi:10.1111/j.1755-3768.1975.tb01161.x

[36] B. Bengtsson, A. Heijl and J. Olsson, "Evaluation of a new threshold visual field strategy, SITA, in normal subjects. Swedish Interactive Thresholding Algorithm", *Acta Ophthalmologica Scandinavica*, 76(2):165-169, 1998, doi: 10.1034/j.1600-0420.1998.760208.x

[37] X. Roggen, K. Herman, L. V. Malderen, M. Devos and W. Spileers, "Different strategies for Humphrey automated perimetry: FASTPAC, SITA standard and SITA fast in normal subjects and glaucoma patients", *Bulletin de la Société belge d'ophtalmologie*, 279, 2001, 23-33.

[38] C. W. Stewart and H. H. Hurshell, "Threshold variation in automated perimetry", *Survey of Ophthalmology*, Volume 37, Issue 5, 1993, Pages 353-361, ISSN 0039-6257, doi:10.1016/0039-6257(93)90065-F.

[39] B. Phillip, M. P. Selwyn, G. K. Yu Xiang, J. Algis J. Vingrys. "Scaling the size of perimetric stimuli reduces variability and returns constant thresholds across the visual field", *Journal of Vision*, 21(11):2, 2021, doi: 10.1167/jov.21.11.2

[40] S. K. Gardiner, "Differences in the Relation Between Perimetric Sensitivity and Variability Between Locations Across the Visual Field", *Invest. Ophthalmol. Vis. Sci.*, 59(8):3667-3674, 2018, doi: doi:10.1167/iovs.18-24303

[41] N. S. Kadavath Meethal, D. Mazumdar, R. Asokan,et al., "Development of a test grid using Eye Movement Perimetry for screening glaucomatous visual field defects". *Graefes Arch Clin Exp Ophthalmol*, 256, 371–379, 2018, doi:10.1007/s00417-017-3872-x

[42] L. M. Brandão, M. Monhart, A. Schötzau, A. A. Ledolter and A. M. Palmowski-Wolfe, "Applying a New Automated Perimetry Pattern Based on the Stimulus Distribution of the Multifocal ERG to Improve Structure-Function Investigation in Glaucoma", *J. Ophthalmol*, 2017:8780934, 2017, doi: 10.1155/2017/8780934

[43] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning representations by back-propagating errors", *Nature*,323(6088), 533–536, 1986, doi:10.1038/323533a0

[44] A. B. Safran, B. Mermillod, C. Mermoud, C. D. Weisse and D. Desangles, "Characteristic features of blind spot size and location, when evaluated with automated perimetry: Values obtained in normal subjects". *Neuro-Ophthalmology*, 13(6), 309–315, 1993, https://doi.org/10.3109/01658109309044579