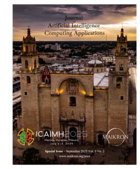
Journal of Artificial Intelligence and Computing Applications (2025) - **Special Issue** - 3(2): 18–24

Journal of Artificial Intelligence and Computing Applications



Expanded abstract

Detection of people with social phobia using pupillary position classification algorithms

Víctor Rangel-Fajardo[1], **David Cruz-Villavicencio**[1], **Erick Mpangi-Musungu**[1], **and Jesús Martínez-Castro**[1,*]

[1]Centro de Investigación en Computación, Instituto Politécnico Nacional

ABSTRACT

This study analyzes the visual attention of people with and without social phobia toward different images (photos of faces) depicting emotions such as normality, sadness, anger, and happiness. The analysis is divided into four main parts: The first part consists of identifying participants, both those with and without social phobia, using a psychometric instrument known as the BFNE-II. The second part focuses on the presentation of images composed of facial expressions, as well as the detection of the subjects' pupillary positions when observing these images. In the third part, the participants' pupillary positions are classified into characteristic points called fixations. Finally, in the fourth part of the study, these fixations are used as a measure to quantify the visual attention of people with and without social phobia, with the aim of identifying potential biases in both groups. Based on the analysis of visual attention to the different images, it was concluded that people with social phobia showed significantly greater visual attention to facial expressions of anger, compared to those without social phobia. On the other hand, non-phobic people tended to focus their attention primarily on images that depicted happy facial expressions.

Keywords: eye tracking, attention, social phobia

1. Introduction

Social phobia (SP) is defined as an unpleasant emotional state accompanied by somatic and psychological changes. It can present as an adaptive reaction or as a symptom associated with various medical and psychiatric conditions [1].

Social phobia is the second most common mental illness in Mexico, with a ratio of 3.6 women to every man. It is worth noting that only between 10% and 30% of all people affected by some type of phobic disorder seek help. Both social phobia and specific phobias manifest primarily during youth.

Various studies suggest that social phobia (social anxiety disorder) has a high comorbidity with other psychiatric disorders, especially depression, dysthymia, generalized anxiety disorder, and panic disorder. For clinical psychology professionals, it is essential to have valid and reliable instruments to detect this disorder. However, there is a lack of validated instruments in the clinical setting, due to the difficulty and financial resources involved in designing, evaluating, and validating them for different study populations.

For his part, Ressler [2] [3], in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) of the American Psychiatric Association (2013), established new standards for diagnosis and research on anxiety and related disorders. The DSM-5 classifies anxiety disor-

E-mail address: jamartinezc@ipn.mx
https://doi.org/10.5281/zenodo.17228364
 ISSN: 3061-8843

ders into eleven categories: 1) separation anxiety disorder, 2) selective mutism, 3) specific phobia, 4) social anxiety disorder (social phobia), 5) panic disorder, 6) agoraphobia, 7) generalized anxiety disorder, 8) substance- or medication-induced anxiety disorder, 9) anxiety disorders due to another medical condition, 10) other specified anxiety disorder, and 11) unspecified anxiety disorder.

Individuals with social anxiety show sustained attention biases towards socially threatening stimuli [4, 5, 6], so it is proposed that, by measuring the visual attention of individuals on different visual stimuli in the form of images, people with social phobia can be differentiated from those without social phobia.

This paper analyzes the visual attention of Mexican adults with and without social phobia by detecting their gaze when visual stimuli (in the form of an image) are presented to them on a monitor.

To identify individuals with social phobia, a psychometric instrument, the BFNE-II test, was used. This instrument consists of 12 questions in a 5-point Likert-type response format and measures a person's fear of being negatively evaluated by others.

2. Eye movements

This paper describes visual attention through two types of movements: voluntary movements of moving from one point to another (saccades) and gaze-holding movements (fixation). The movements that comprise a fixation are: tremor, drift, and microsaccade [7].

3. Pupillary positions, fixations and classification algorithms

The classification of pupillary positions into fixations is the grouping into n sets of the pupillary positions in such a way that they statistically describe the physiological behaviors of eye movements [8]. This process significantly reduces the size of the pupillary positions to representative tuples (fixations or characteristic points), where each tuple is composed of the arithmetic mean of the ocular positions in the group, the sum of the duration of each element, and the number of elements that make up the group.

Classification is a fundamental procedure in any analysis of eye movements, as poor classification can result in tuples that do not represent gaze behavior [9, 10].

The algorithms for classifying pupillary positions into fixations must evaluate statistical characteristics and physical magnitudes of the pupillary position samples, such as point-to-point velocity and/or acceleration of the samples, as well as data dispersion, among others. On the other hand, the physiological characteristics of eye movement behavior must be considered, such as the minimum duration time of a fixation, the minimum duration time of a saccade, and the maximum and minimum amplitude of a repositioning movement, among others. To achieve this, identification and classification algorithms for pupillary positions in fixations are used [11]. Some of the algorithms used are: dispersion-threshold identification algorithm (I-DT), velocity-threshold identification algorithm (I-VT), hidden Markov model identification algorithm (I-HMM), and the adaptive algorithm.

4. Materials and methods

There are different eye-tracking methods, which can be grouped into two major categories: invasive and non-invasive. The first category obtains the pupillary position directly by incorporating special devices into the eyeball, which implies physical interference with the subject and is therefore considered invasive. The second category estimates the pupil position indirectly, that is, it uses different techniques to estimate the position (one of them is through image analysis and known reference points), and thus does not physically interact with the subject (non-invasive). Invasive methods offer high precision and accuracy in measurement and are capable of detecting very small movements, such as tremors. However, they interfere with the natural characteristics of the subject's movements (for example, they reduce movement speed), in addition to causing discomfort. In contrast, indirect methods present lower precision and accuracy in measurement, but the movements obtained can be considered natural. Some non-invasive eye-tracking methods include the electrooculogram, the search coil, and infrared reflection. For this study, the infrared reflection method was chosen.

In [12], the twelve leading companies in the commercialization of eye-tracking products are presented, ordered by the number of publications from January 2016 to January 2017. For this work, the sensor *The Eye Tribe* was selected. For data capture, a communication module was developed between the sensor and the computer, capable of transmitting information at 30 and 60 Hz.

To calibrate the sensor, the subject is asked to look at nine points on the screen (see Figure 1), while the position of the pupil and the corneal reflection are determined at a given moment in time [13]. Current literature proposes various error metrics [14]. For a monitor with a resolution of 1280×1024 pixels, we use a value of ε between 30 and 60 pixels.

5. Procedure

The test begins with the *"Calibration"* operation, in which the subject is instructed to observe the reference points. The subject is asked to avoid moving their face during the test to prevent loss of sensitivity and high variance in the sample. Afterwards, they are informed that 10 images containing 4 faces each will be presented, and they should observe the face that draws their attention the most *"Visual Stimuli Presentation"*.

The ten *test images* are presented automatically, so the subject does not need to perform any additional action beyond viewing the stimuli. At the end of the presentation of the *test images*, a message will appear

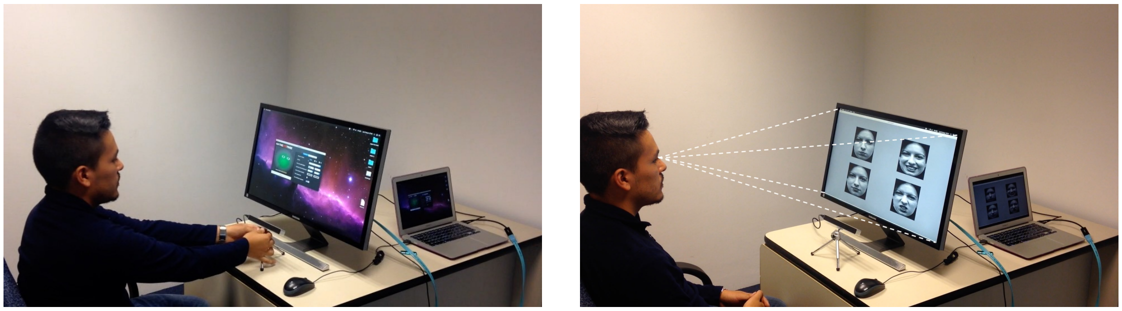


Figure 1. Experimental setup during its calibration stage. A point is randomly selected and viewed for one second; after this time, another point is selected. The process is repeated until all nine calibration points are displayed.

on the screen indicating the end of the test. At that moment, the subject may leave the testing room. The duration of the test ranges between four and ten minutes, depending on the calibration process.

The test images were taken from a sample of 145 undergraduate students, with an average age of 21.3 years. Of them, 62.7% were women. The image capture was carried out under controlled lighting and sound conditions.

The selected visual stimuli are based on work similar to the studies conducted by Paul Ekman on basic emotions [15, 16].

On the other hand, n_i images were presented, each composed of k different stimuli. The images were shown to individuals with social phobia (SP) and without social phobia (No-SP). During a time period t_i, the eye movements of the individuals with SP and No-SP were recorded as the n_i images were presented to them.

As part of the proposed hypothesis, it is suggested that individuals with social phobia exhibit sustained attention biases toward stimuli perceived as threatening. The stimuli that have been shown to be effective in tests with individuals with social phobia are faces with different facial expressions, as described in [4].

In this work, the image set of facial expressions "AU-Coded Facial Expression Database," developed by Cohn-Kanade [17], was used. This repository contains 2000 images of 200 individuals aged between 18 and 30 years. The image set is composed of different facial expressions (neutral, joy (happiness), sadness, anger, surprise, fear, and disgust) categorized according to the "FACS" standard [18]. From the Cohn-Kanade image set, ten faces (five women and five men) were used. An oval crop was applied to each face in order to suppress fixations centered on protruding elements of the stimulus (hair, earrings, ears, etc.). The goal was to ensure that the patients' attention was focused on the facial expression and not on distracting elements.

With the selected stimuli, ten images were composed and presented to the subjects. Each test image was composed of four stimuli (neutral, joy, sadness, and anger) arranged in a 2×2 grid (see Figure 2), where each cell is randomly assigned a stimulus.

6. Results and discussion

The expression of anger was the most focused on by individuals with SP, while the expression of joy was the most focused on by individuals without SP. The remaining expressions (neutral and sadness) are included to obtain better discrimination in attention levels between individuals with SP and No-SP.

Cutoff points (total score on the Likert-type test from which a subject is considered to present symptoms of social phobia) on fear scales vary depending on the population and social environment. In Mexico, in the state of Michoacán de Ocampo, Marcelo A. et al. *et al.* [19] validated the instrument (BFNE-II test) for the adult population of Michoacán, finding a cutoff point of 42 with a sensitivity of 94% for having social phobia. During this research, a version of the original BFNE-II test was applied and evaluated based on the cutoff point published by Marcelo A. et al. The results are described in [12].

The conventional analysis of eye movements was carried out through the study of fixations (pauses in regions of interest) or saccadic movements (rapid movements between fixations). To this end, the four algorithms described in Section 3 were evaluated with the aim of determining the optimal one for our objective. The four algorithms consider velocity as the main discriminative feature, so the calculation of velocities between samples must be unified for each one. Likewise, it is necessary to calculate noise-free velocities; for this purpose, the equations proposed by *Engbert and Kliegl* [20] are used, where the calculation is performed on the x and y coordinates for the left and right eyes independently.

One parameter of the pupillary position algorithms is the velocity threshold, from which it is determined whether a position $(x_i,\ y_i)$ with velocity $\dot{v}_i$ belongs to a fixation group or not. This threshold varies depending on the type of test and the sampling rate. The specialized literature proposes a threshold of 1.5 px/ms [8], but does not specify the type of experiment. In this work, a velocity threshold is calculated for the actual experimental conditions at a sampling frequency of 60 Hz. For this, the mean velocities per test image were ana-

Figure 2. The images represent four stimuli (neutral, joy, sadness, and anger), among which the subject had to primarily select one.

lyzed for each person. That is, each person contributes ten mean velocity values to the analysis, where each mean velocity describes the subject's general behavior during the presentation of a test image.

In Figure 3, the behavior of mean velocities for 145 subjects is presented. The data show a positively skewed distribution with a heavy tail, suggesting a high concentration of velocities close to zero. According to theory, low velocities are due to fixation movements (microsaccades), while high velocities correspond to repositioning movements (saccades). Once the velocity threshold parameter ε is determined, the classification algorithms for fixations are tested and evaluated.

The decision criteria for determining the optimal algorithm among I-VT, I-DT, I-HMM, and the Adaptive algorithm are described in [12].

6.1 Attention by stimulus

Using the attention tuples T_{At}, the attention time for each stimulus was measured, distinguishing between individuals with SP and without SP. The attention times per stimulus of the subjects with SP and without SP are compared using quartiles to generate the boxplots shown in Figure 4.

While individuals with SP focus their attention on the threatening stimulus, individuals with No-SP direct their attention to the happiness (joy) stimulus.

Boxplots are an efficient comparison method that allows subtle trends in the data to be observed; however, they do not allow a trend to be validated conclusively. For this purpose, a normality analysis was conducted (see [12]).

Finally, quantile-quantile (Q-Q) plots are presented with the aim of comparing the actual distribution with a theoretical normal distribution. In each plot, two distributions are compared, where each corresponds to the attention time of a group of individuals (SP and No-SP) for the same stimulus k_i.

From Figure 5, it is possible to state that a significant difference does exist: in the case of the Joy (happiness) stimulus, individuals with No-SP exhibit longer attention times compared to individuals with SP (see [12]), which suggests that the hypothesis of the study was confirmed for the study population.

7. Conclusions

This research presents results that could be useful for the eventual confirmation of the hypothesis proposed at the beginning of this work for the individuals examined.

It was verified through statistical methods that the level of attention in individuals with SP is higher for stimuli considered threatening (Anger), where the mean attention time was 1.6s, compared to individuals without SP, whose mean attention time for the threatening stimulus was 1.05s. A similar trend was observed when comparing the opposite stimulus (Happiness), where individuals without SP recorded a mean of 1.88s, while individuals with SP recorded a mean attention time of 1.26s.

The Neutral and Sadness stimuli did not show significant changes; the same level of attention was maintained for both types of individuals analyzed. A possible cause of this is the similarity that exists in the facial expressions of these stimuli.

It is important to mention an issue regarding the experimental design. Initially, the test was designed to display 30 test images for 10 seconds each. After the first trials, we received feedback indicating how exhausting this procedure was, since the observer had to avoid moving their face or body during the entire test (5 minutes). As a result of this overexposure to the images, the quality of the fixations detected in the final test images was poor. The main reason was that most individuals changed their posture after a certain number of images had been shown. This change in posture affected sensor detection and, consequently, the measurement. The final experiment was designed to use 10 test images, which equates to a stimulus exposure of 1.66 minutes. With this number of stimuli, uniform detection was achieved throughout the entire test.

In this research, visual attention was quantified as the time a subject devotes to an area of interest. An-

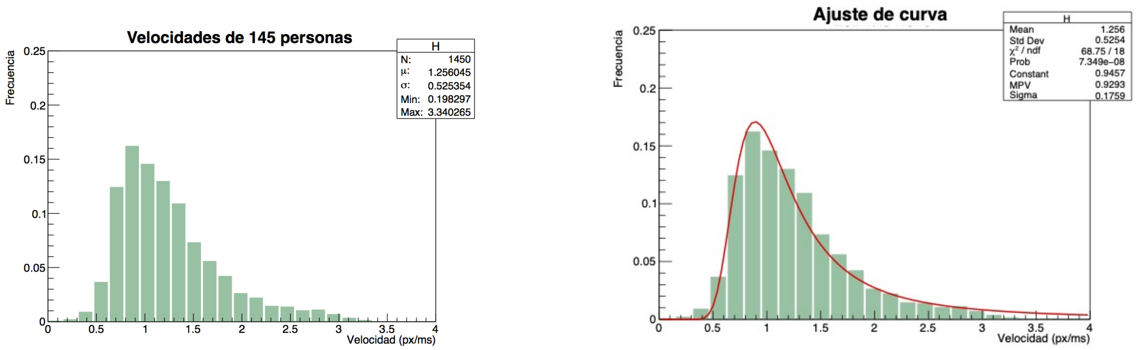


Figure 3. Velocity distribution and its fitted curve.

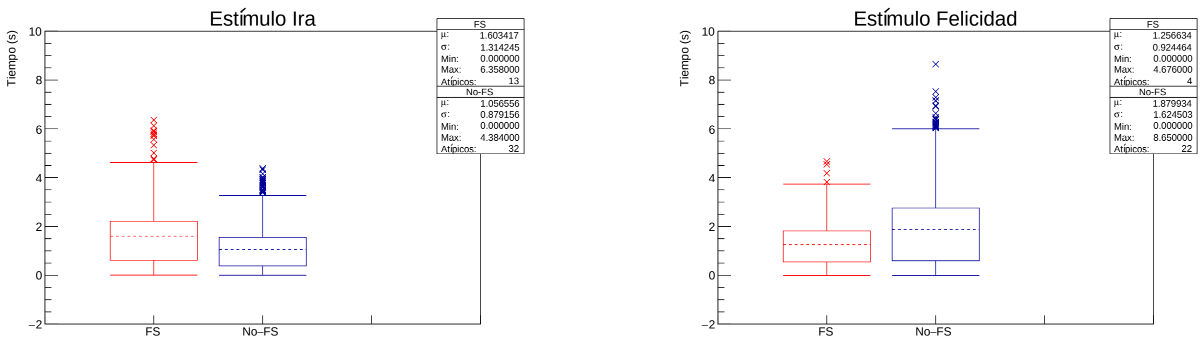


Figure 4. Boxplot for the anger and happiness stimuli.

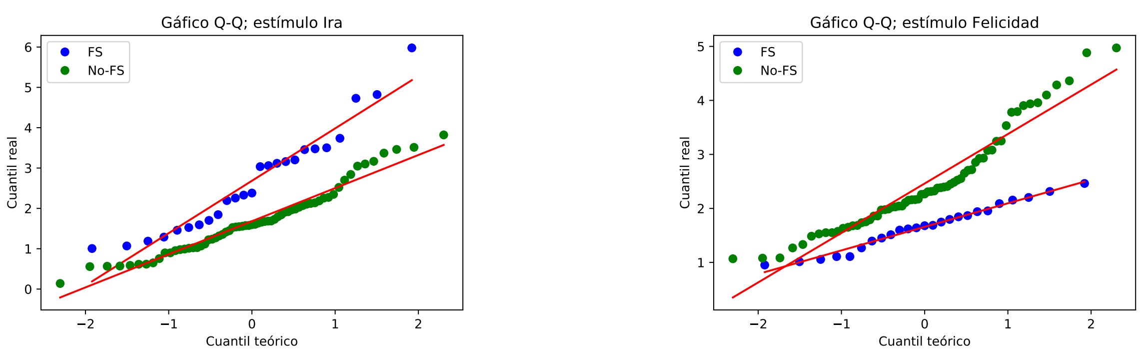


Figure 5. Q-Q plot for the anger and happiness stimuli.

other contribution of this work was the evaluation of the quality of the fixations detected by each evaluated algorithm, in the absence of a reference standard. The main proposed measure was the calculation of the nearest neighbor index, which assesses the dispersion of the data within a fixation.

Ethics Statement

This study used anonymized data. No personally identifiable information was collected. Ethical approval was deemed unnecessary according to the guidelines of Instituto Politénico Nacional.

CRediT authorship contribution statement

Víctor Rangel-Fajardo: Conceptualization, Methodology, Software, Investigation, Validation, Resources, Formal analysis, Project administration, Data curation. **David Cruz-Villavicencio:** Software, Writing - review & Validation. **Erick Mpangi-Musungu:** Writing - review & editing. **Jesús Martínez-Castro:** Conceptualization, Methodology, Formal analysis, Investigation, Writing original draft, Writing - review & editing, Visualization, Supervision, Project administration.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors utilized Grammarly and ChatGPT to refine sentence structure and enhance readability. No content was generated by AI; all scientific insights and original ideas are the authors' own.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

The authors thank the IPN and Secihti for their support in completing this work.

References

[1] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR*. Washington, DC: American Psychiatric Association Publishing, 2022. doi: 10.1176/appi.books.9780890425787 . [Online]. Available: https://www.psychiatry.org/psychiatrists/practice/dsm

[2] K. Ressler, *Anxiety Disorders*. Oxford: Oxford University Press, Incorporated, 2015. [Online]. Available: http://ebookcentral.proquest.com/lib/anahuac-ebooks/detail.action?docID=2012685

[3] A. Moukheiber, G. Rautureau, F. Perez-Diaz, R. Soussignan, S. Dubal, R. Jouvent, and A. Pelissolo, "Gaze avoidance in social phobia: Objective measure and correlates," *Behaviour Research and Therapy*, vol. 48, no. 2, pp. 147–151, 2010. doi: 10.1016/j.brat.2009.09.012 . [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19863948/

[4] T. Bantin, S. Stevens, A. L. Gerlach *et al.*, "What does the facial dot-probe task tell us about attentional processes in social anxiety? a systematic review," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 50, pp. 40–51, 2016. doi: 10.1016/j.jbtep.2015.04.009 .

[5] C. A. Schofield, A. L. Johnson, A. W. Inhoff, and M. E. Coles, "Social anxiety and difficulty disengaging threat: Evidence from eye-tracking," *Cognition and Emotion*, vol. 26, no. 2, pp. 300–311, 2012. doi: 10.1080/02699931.2011.602050 . [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/02699931.2011.602050

[6] N. K. Sahu, N. S. Harshit, R. Uikey, and H. R. Lone, "Beyond questionnaires: Video analysis for social anxiety detection," 2024. [Online]. Available: https://arxiv.org/abs/2501.05461

[7] S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, "The role of fixational eye movements in visual perception," *Nature Reviews Neuroscience*, vol. 5, no. 3, pp. 229–240, 2004. doi: 10.1038/nrn1348 .

[8] D. Salvucci and J. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the Eye Tracking Research and Applications Symposium*, Jan 2000, pp. 71–78. doi: 10.1145/355017.355028 .

[9] A. Lazarov, D. Basel, S. Dolan, D. G. Dillon, D. A. Pizzagalli, and F. R. Schneier, "Increased attention allocation to socially threatening faces in social anxiety disorder: A replication study," *Journal of Affective Disorders*, vol. 290, pp. 169–177, 2021. doi: 10.1016/j.jad.2021.04.063 . [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165032721003955

[10] L. R. Lidle and J. Schmitz, "Assessing visual avoidance of faces during real-life social stress in children with social anxiety disorder: A mobile eye-tracking study," *Child Psychiatry & Human Development*, vol. 55, no. 1, pp. 24–35, Feb 2024. doi: 10.1007/s10578-022-01383-y . [Online]. Available: https://doi.org/10.1007/s10578-022-01383-y

[11] R. Andersson, L. Larsson, K. Holmqvist *et al.*, "One algorithm to rule them all? an evaluation and discussion of ten eye movement event-detection algorithms," *Behavior Research Methods*, vol. 49, no. 2, pp. 616–637, Apr 2017. doi: 10.3758/s13428-016-0738-9 . [Online]. Available: https://doi.org/10.3758/s13428-016-0738-9

[12] V. Rangel-Fajardo, "Análisis de la fobia social mediante la clasificación de las posiciones pupilares ante estímulos visuales," Tesis de Maestría, Centro de Investigación en Computación, Ciudad de México, México, Diciembre 2018.

[13] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, 3rd ed. London: Springer, 2007.

[14] S. A. Johansen, J. San Agustin, H. Skovsgaard, J. P. Hansen, and M. Tall, “Low cost vs. high-end eye tracking for usability testing,” in *CHI’11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2011, pp. 1177–1182. doi: 10.1145/1979742.1979744.

[15] P. Ekman, “An argument for basic emotions,” *Cognition & Emotion*, vol. 6, no. 3-4, pp. 169–200, 1992. doi: 10.1080/02699939208411068.

[16] ——, “Basic emotions,” *Handbook of Cognition and Emotion*, pp. 45–60, 1999. doi: 10.1002/0470013494.ch3.

[17] J. F. Cohn, T. Kanade, and Y. Tian, “Au-coded facial expression database,” Digital Repository, 2000, obtenido de http://www.pitt.edu/˜emotion/ck-spread.htm.

[18] P. Ekman, “Facial action coding system (facs),” *A Human Face*, 2002.

[19] M. A. Bravo, F. González Betanzos, Castillo Navarro, and F. Padrós Blázquez, “Escala de miedo a la evaluación negativa versión breve (bfne): propiedades psicométricas en muestras clínicas y de universitarios en méxico,” *Acta Colombiana de Psicología*, vol. 18, no. 1, pp. 69–77, 2015. doi: 10.14718/ACP.2015.18.1.4.

[20] R. Engbert and R. Kliegl, “Microsaccades uncover the orientation of covert attention,” *Vision Research*, vol. 43, no. 9, pp. 1035–1045, 2003. doi: 10.1016/S0042-6989(03)00084-1.