

# PREMOC: Plataforma de reconocimiento multimodal de emociones

Ramón Zatarain-Cabada, María Lucia Barrón-Estrada, Gilberto Muñoz-Sandoval

Instituto Tecnológico de Culiacán, Culiacán, Sinaloa,  
México

{rzaratain, lbarron, m14170101}@itculiacan.edu.mx

**Resumen.** En años recientes la computación afectiva ha venido a mejorar la interacción humano-computadora, pues ayuda a la computadora a conocer el estado afectivo del usuario para mejorar la toma de decisiones. Este artículo presenta los avances en el proyecto PREMOC, una plataforma que brinda un servicio web para el reconocimiento de emociones en texto, imágenes de rostros, sonidos de voz y señales EEG de manera mono-modal y multimodal. PREMOC ayuda a los desarrolladores a integrar el reconocimiento de afecto a sus aplicaciones o sistemas de software. Cada uno de los reconocedores se implementó aplicando diferentes técnicas tanto para extraer características como para clasificar emociones; además para el reconocimiento multimodal se integraron las emociones mediante un sistema difuso. Esta plataforma ya está siendo utilizada por diferentes proyectos en el laboratorio de la Maestría en Ciencias de la Computación del Instituto Tecnológico de Culiacán.

**Palabras claves:** Computación afectiva, inteligencia artificial, reconocimiento de emociones.

## PREMOC: Multimodal Emotion Recognition Platform

**Abstract.** In recent years, affective computing has been improving human-computer interaction, because it helps computers to know the emotional state of the user to improve the decision-making process. This paper presents the advance in the PREMOC project, a platform that provides a web service for emotion recognition in text, facial images, speech and EEG signals in mono and multimodal mode. PREMOC assists developers to incorporate emotion recognition on their applications. Each recognizer was implemented with different techniques for feature extraction and emotion classification; for multimodal recognition, emotions were integrated through a fuzzy system. This platform is being used on different projects in the Computer Science Laboratory of Instituto Tecnológico de Culiacán.

**Keywords:** Affective computing, artificial intelligence, emotion recognition.

## 1. Introducción

Las computadoras están adquiriendo desde hace varios años la capacidad de detectar la emoción de sus usuarios con el uso de diferentes sensores y dispositivos. Esto facilita la interacción entre el usuario y la computadora para crear una línea de comunicación de dos canales, con el cual la computadora obtiene información adicional del usuario. A esta nueva área del conocimiento se le nombra comúnmente como Computación Afectiva [1]. Hoy en día se han desarrollado una gran cantidad de proyectos y aplicaciones en diferentes ámbitos como la mercadotecnia, la medicina, la educación, entre otras.

Uno de los dispositivos más usados para detectar emociones para la computación afectiva es la cámara, pues según Mehrabian [2] los gestos faciales transmiten el 58% de la comunicación humana. Otra razón por la cual es la más usada es la existencia de características universales en los rostros que reflejan ciertas emociones sin importar la cultura facial de la persona; estas emociones son conocidas como las emociones de Ekman [3].

Otro dispositivo utilizado en la computación afectiva es el micrófono, pues la evaluación de la voz humana también ha sido utilizada para definir emociones como por ejemplo en Kun Han [4], donde al igual que muchos trabajos relacionados con el sonido, dividen el audio en pequeños segmentos y con ellos entrenan una red neuronal. En este método existen dos variables muy importantes, la intensidad y la valencia.

También se puede detectar emociones a partir del texto o documentos, ya que con el uso del internet, redes sociales y blogs, la cantidad enorme de información en estos medios abre una ventana muy grande para encontrar nuevos elementos dentro de estos datos, como es en este caso las emociones o afecto. Existen distintos métodos en diferentes idiomas de como reconocer emociones en el texto. Uno de los métodos más utilizados es el de la técnica keywords [5], que consiste en tener una base de datos previamente clasificada que contiene palabras etiquetadas con una emoción y un grado de pertenencia.

Un dispositivo más que podemos encontrar en la literatura es la diadema que emite señales electroencefalográficas (EEG). Este tipo de dispositivos puede ser muy efectivo para encontrar señales en tiempo real como se hace en Liu [6] donde se identificó las regiones del cerebro más importantes para interpretar emociones en un usuario.

Cada uno de estos dispositivos alcanza una tasa de asertividad de la emoción dependiendo de su implementación. Hay trabajos que combinan dos o más dispositivos para mejorar esa tasa de asertividad o para considerar diferentes aspectos del usuario antes de determinar su emoción. A la técnica donde se usan dos o más dispositivos o sensores para reconocer emociones, se le nombra reconocedor multimodal de emociones o afecto.

PREMOC ofrece cuatro reconocedores de emociones en forma de servicio web, es decir, cualquier aplicación cliente puede consultar al servicio web ofrecido por PREMOC y obtener una emoción de una imagen de un rostro, de un texto, de un sonido de voz o de un archivo de señales electroencefalográficas. Una de las ventajas de PREMOC es su modo multimodal, donde el usuario puede enviar en la misma consulta, dos o más diferentes tipos de archivos para obtener una emoción o cada una por separado. Otra ventaja es su gran flexibilidad y facilidad para ser usado, lo cual se da por medio de un servicio Web. Finalmente, otra ventaja para usar esta herramienta es

la identificación de emociones en texto y voz para el lenguaje Español, lo cual no existe en otros reconocedores multimodales de emociones, en el estudio que llevamos a cabo.

Este artículo está organizado de la siguiente manera: en el capítulo 2 se describen trabajos relacionados a PREMOC, tanto reconocedores individuales, como multimodales y algunos ejemplos que ofrecen el servicio en línea. En el capítulo 3 se describe cada uno de los reconocedores de manera individual, así como también se habla de los métodos que tiene PREMOC para la integración multimodal de emociones y la estructura de la plataforma. En el capítulo 4 se mencionan los experimentos y en el capítulo 5 las conclusiones y trabajo futuro.

## **2. Trabajos relacionados**

Existen actualmente muchos trabajos relacionados a la obtención de emociones tanto monomodales como multimodales, y muchos de ellos trabajan con diferentes métodos de extracción y clasificación de características.

Entre los que detectan emociones por medio del texto, Seol [5] utiliza el algoritmo keywords para extraer características de las palabras y una red neuronal artificial basada en conocimiento (KBANN por sus siglas en inglés) para clasificarlas y obtener una emoción. Así como Shaheen [7] extrae las características del texto mediante reglas de reconocimiento de emociones y utiliza k-vecinos cercanos (KNN por sus siglas en inglés) para la clasificación de la emoción. También el método Random Forest fue usado por Zheng [8] para clasificación de emociones en el texto.

Por medio del rostro, Khandait [9] y Gangwar [10] utilizan una red neuronal para la clasificación de las características y obtención de emociones. Para la extracción de características utilizan SUSAN edge detection y Hue-Saturation-Value respectivamente. Mientras que De [11] usa distancia euclidiana para obtener la emoción y 2D Discrete Cosine Transform para la extracción de características.

En los trabajos relacionados a las emociones por medio de voz, Kostoulas [12] utiliza el algoritmo C4.5 de la librería Weka para la clasificación de la emoción. Vogt [13] con EmoVoice utiliza una máquina de soporte vectorial (SVM por sus siglas en inglés) por la rapidez para hacerlo en tiempo real, ya que EmoVoice es una aplicación en tiempo real. Caballero-Morales [14] utiliza Modelos Ocultos de Markov para la clasificación de la emoción (HMM por sus siglas en inglés) en su trabajo para el idioma español. Los 3 trabajos presentados utilizan Mel-Frequency Cepstral Coefficient (MFCC) como método de extracción de características.

En los trabajos donde se obtiene la emoción de una persona por medio de señales EEG los métodos son más variados. Por ejemplo, Liu [6] utiliza directamente los datos ofrecidos por Emotive Software Development Kit para la extracción de características y hace la clasificación de las emociones con un Modelo de Dimensión Fractal. Nie [15] obtiene las características EEG por medio de la Transformada Rápida de Fourier (FFT por sus siglas en inglés) y clasifica las emociones con SVM. Lee [16] utiliza NeuroScan para la extracción de características en las señales y las clasifica en emociones con Análisis Discriminante Cuadrático.

Hay trabajos bimodales como Alonso-Martin [17] que creó un sistema para la detección de emociones durante la interacción Humano-Robot, para que los robots sean más sociales, y utiliza como parámetros el rostro y la voz humana. También Busso [18]

creó un analizador bimodal de voz y rostros para demostrar que al usar ambos, la asertividad de la emoción dada es mejor que en modo individual. El autor utilizó SVM para clasificar emociones en el rostro y Praat para la voz. Chuang [19] trabajó con voz y texto, y usó SVM para la clasificación de emociones en la voz y keywords para el texto.

Existen también sistemas multimodales como el creado por D'Mello[20] para añadirle afecto al famoso sistema tutor inteligente AutoTutor donde además de utilizar voz y rostro, también tomaron en cuenta el lenguaje corporal. Caridakis [21] también utiliza estos 3 parámetros para la obtención multimodal de emociones y alcanza un 88,2% de asertividad, y clasifica 8 diferentes emociones con redes bayesianas. También hay investigación sobre aprovechar todo lo que se utiliza mientras se trabaja. Por ejemplo Paredes [22] trabaja en su investigación con mouse y teclado además de la cámara para la emoción a través del rostro.

En el trabajo realizado por Wagner [23] se introduce el nombre de Online Emotion Recognition System (OER) para referirse a los sistemas que ofrecen reconocimiento de emociones en línea. Wagner presenta Smart Sensor Integration que es un framework que ofrece reconocimiento de emociones de manera multimodal para voz, rostros y movimiento en línea. Otro framework multimodal es ABE [24] que consiste en una federación de agentes en línea donde cada agente se encarga de un dispositivo de entrada. ABE trabaja usando señales del cerebro (EEG), rastreo de ojos, expresiones faciales, conductividad en la piel, presión arterial y postura corporal.

### **3. PREMOC**

PREMOC es una plataforma en línea que ofrece reconocimiento de emociones en los dispositivos anteriormente mencionados, lo cual hace a través de un servicio web para que los desarrolladores agreguen reconocimiento de emociones a sus aplicaciones. La comunicación entre PREMOC y los usuarios es a través de formato JSON, en donde el usuario envía sus datos de entrada por medio de un POST en formato JSON, y recibe la emoción como respuesta.

#### **1.1. Reconocedores de emociones**

PREMOC tiene definidos cada uno de los reconocedores de forma autónoma, los cuáles reciben el archivo con datos de la imagen, voz, texto y/o información cerebral, lo procesan, y obtienen la emoción que regresan. Los reconocedores fueron creados por separado en el laboratorio de la Maestría en Ciencias de la Computación del Instituto Tecnológico de Culiacán, y algunos ya fueron utilizados en una aplicación [26, 27].

##### **1.1.1 Reconocedor de emociones en texto**

Para el reconocimiento de emociones en texto, PREMOC utiliza un algoritmo semántico denominado ASEM. ASEM es un algoritmo basado en keywords, pues contiene una colección de palabras previamente etiquetadas a una emoción.

El algoritmo ASEM cuenta con 5 diferentes corpus, con diferentes funciones cada uno de ellos: stop words, palabras impropias, corpus emocional, new words y corpus semántico.

Stop words. Es una colección de palabras que no influyen en la emoción de la oración, como artículos, preposiciones, entre otros. Las palabras que están aquí son eliminadas de la oración a evaluar.

Palabras impropias. Es una colección de, como su nombre lo dice, palabras impropias, que están calificadas a una emoción y su factor de probabilidad afectiva a dicha emoción (FPA) donde  $0 < FPA < 1$ .

Corpus emocional. Es una colección de palabras, que están calificadas a una emoción y su FPA a dicha emoción, donde  $0 < PFA < 1$ .

New words. Aquí se coleccionan las palabras nuevas que no se tienen en el corpus emocional. Esto es para el mantenimiento del algoritmo ya que las nuevas palabras encontradas aquí serán calificadas cada cierto tiempo por un experto. El experto determinará a que emoción pertenece esa palabra y su FPA.

Corpus semántico. Es una colección de palabras que potencializa o inhibe el valor FPA semánticamente. Por ejemplo, la palabra muy potencializa el valor FPA de la siguiente palabra.

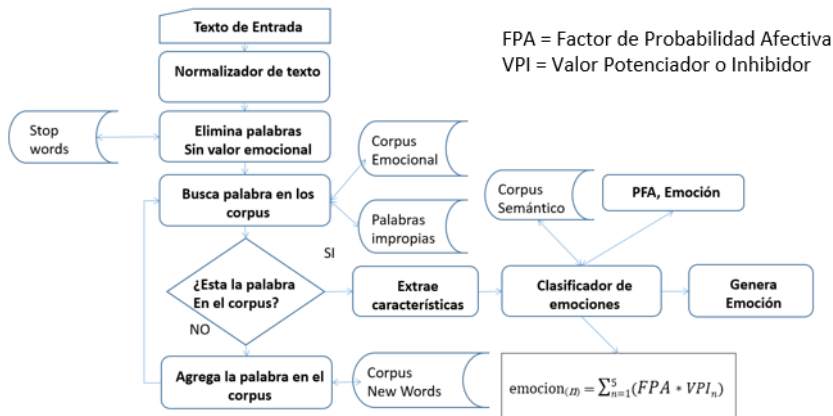


Fig. 1. Algoritmo ASEM.

Como se muestra en la Figura 1, el algoritmo ASEM funciona de la siguiente manera:

- a) Se recibe un texto de entrada.
- b) Se normaliza el texto quitando palabras que no agregan valor emocional usando el corpus Stop Words.
- c) Se busca cada palabra en el corpus emocional para obtener su emoción y su FPA. Si no está, se busca en el corpus palabras impropias. Si no está, se agrega al corpus new words para que sea más adelante clasificada por un experto. En la evaluación presente, dicha palabra es ignorada para la clasificación de la emoción en el texto.
- d) Se buscan palabras en el corpus semántico que puedan potencializar o inhibir los valores FPA ya obtenidos.

- e) Se calcula el valor de cada una de las 5 emociones (feliz, sorprendido, neutral, triste, enojado) en el texto con la fórmula:

$$\text{emocion}_{(n)} = \sum_{n=1}^5 (FPA * VPI_n).$$

- f) Se selecciona la emoción con el mayor valor obtenido y esa emoción se regresa al usuario.

### 1.1.2. Reconocedor de emociones en rostro

El reconocedor de expresiones faciales en PREMOC consiste de dos fases: entrenamiento de una red neuronal y uso del reconocedor.

La fase de entrenamiento consistió en entrenar la red neuronal encargada de la clasificación de rostros. Para esto se tomaron como datos de entrenamiento las imágenes del corpus RaFD [27]. Del corpus se tomaron 955 imágenes, 191 de cada una de las 5 emociones con las que se va a trabajar que son feliz, sorpresa, neutral, enojado y triste. De cada una de las imágenes se extrajo un vector de 10 características, donde cada una es una distancia euclidiana en diferentes puntos del rostro. Con esto se entrenó una red neuronal usando la librería Weka [28] para Java.

La fase de ejecución en línea funciona de la siguiente manera y como muestra la Figura 2.

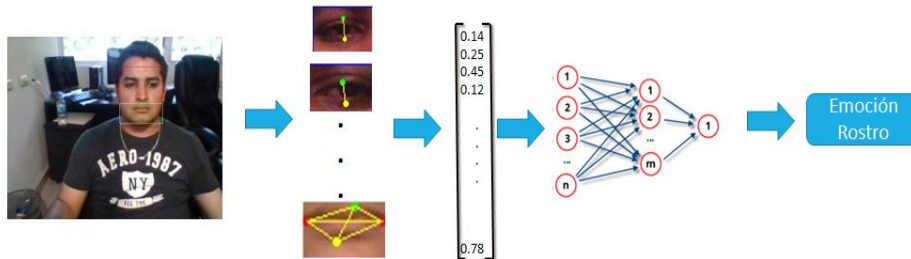


Fig. 2. Algoritmo del reconocedor de emociones en rostros.

- a) Recibe una imagen.
- b) Detecta el rostro en la imagen con OpenCV.
- c) Repite los siguientes pasos para boca, ojo izquierdo, ojo derecho, ceja izquierda, ceja derecha, lo llamaremos objetivo:
  - Ubicar dentro del rostro posible zona de ubicación del objetivo, se llamará región de interés.
  - Dentro de la región de interés aplicar un método de búsqueda en cascada de OpenCV para buscar en concreto el objetivo que estamos buscando.
  - Ya con el objetivo encontrado, se le aplican filtros de OpenCV [29] como se muestra en la Figura 3 para determinar los puntos clave en el rostro.
- d) Se crea un vector de características de 10 elementos, donde cada uno es una distancia euclidiana entre diferentes puntos clave ubicados en el rostro.
- e) Se ingresa el vector de características a la red neuronal ya entrenada para que lo clasifique a una de las emociones.

f) Regresa la emoción obtenida.

Este reconocedor regresa los valores de emociones para feliz, triste, enojado, sorpresa y neutral.

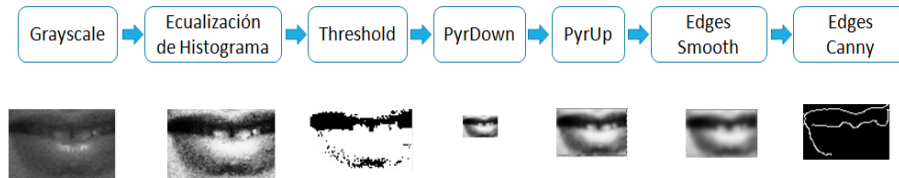


Fig. 3. Filtros de OpenCV en la boca.

### 1.1.3. Reconocedor de emociones en voz

El reconocedor de emociones en voz que utiliza PREMOC identifica la valencia del audio recibido (positiva o negativa) y también consiste en dos fases: entrenamiento y ejecución.

En la fase de entrenamiento se utilizaron un total de 45 audios de 9 sujetos diferentes, donde cada audio está clasificado en positivo o negativo. De los audios se extraen las características utilizando su espectrograma con la librería musicg [30] para obtener los parámetros para entrenar una SVM de la librería LibSVM [31]. La instancia de la SVM ya entrenada es guardada para utilizarla online en la clasificación de sonidos de PREMOC.

La fase de ejecución en línea funciona de la siguiente manera y como se muestra en la Figura 4.



Fig. 4. Algoritmo del reconocedor en voz.

- a) Se recibe un audio en formato *wav* serializado en Base64.
- b) Se obtienen el espectrograma del audio para obtener los valores de tono e intensidad a lo largo del audio.
- c) Se calculan los siguientes valores:
  - El valor de la intensidad más alta del audio.
  - El valor de la intensidad más baja del audio.
  - El valor del tono más alto del audio.
  - El valor del tono más bajo del audio.
- d) Estos 4 valores se introducen a la SVM para obtener la valencia de la voz.
- e) Se regresa la valencia obtenida.

Este reconocedor regresa la valencia de la voz (*positiva o negativa*).

#### 1.1.4. Reconocedor de emociones en señales EEG

El reconocedor de emociones por medio de señales EEG también consta de dos fases: entrenamiento e implementación en línea. Este reconocedor identifica la valencia de la emoción (positivo, negativo o neutral) del archivo recibido de señales EEG. El reconocedor fue creado y probado para la Interface Cerebro-Computadora (BCI) Emotiv Epoc [32]. Para seleccionar los canales apropiados que necesitamos del BCI, se consideró el trabajo de Mahajan [33] que usa los canales frontales AF3, AF4, F3, F4, F5 Y FC6. También el trabajo de Liu [6] que usa los canales AF3, F4 Y FC6. Al final se consideraron los canales AF3, F3, F4, FC5 y FC6 porque se concluyó que son los que más se aproximan a las emociones humanas.

La característica principal extraída de las señales EEG es el Exponente de Hurst, el cual es usado en análisis de series de tiempo para identificar un comportamiento inestable de las señales EEG para identificar tendencias en los datos. Se eligió el Exponente de Hurst por los buenos resultados mostrados en Wang [34](71.38 %).

La fase de entrenamiento esta descrita en [35]. Y la fase en línea funciona como muestra la Figura 5 y como se describe a continuación.



**Fig. 5.** Reconocedor de señales EEG.

- PREMOC recibirá un archivo edf con las señales EEG.
- Se extraerán la información solamente de los 5 canales mencionados anteriormente.
- Se calculará el Exponente de Hurst de cada canal para obtener un vector de tamaño 5.
- Alimentamos la SVM de la librería LibSVM [31] con el vector obtenido y se obtendrá la valencia de la emoción.

Este reconocedor regresa la valencia de la emoción: positiva, negativa o neutral.

#### 1.2. Integración de las salidas de los reconocedores

Para mayor control sobre lo que el usuario requiere, la plataforma ofrece 3 tipos de modo de respuesta, que son: *simple*, *multi* y *student*. Cada uno de ellos hace uso de los reconocedores individuales descritos en la sección anterior para determinar sus respectivas salidas.

##### 1.2.1. Modo simple

En el modo simple, la respuesta que ofrece PREMOC es la que da cada reconocedor por separado sin integrarlos; es decir, si un usuario requiere la emoción en un rostro y en un sonido de voz, PREMOC le da en respuesta la emoción del rostro y la emoción de la voz. Las emociones que regresa el modo simple dependen de cada reconocedor individual.



Como se muestra en la Figura 6, el integrador reúne las respuestas de los reconocedores individuales para crear el json de respuesta para el usuario.

### 1.2.2. Modo multi

En el modo multi, la respuesta que da el usuario es multimodal; es decir, regresa una sola emoción al usuario de los dispositivos que él requiera. En el mismo caso que el ejemplo anterior, si el usuario requiere la emoción en un rostro y en un sonido de voz y lo quiere en forma multimodal, PREMOC le da en respuesta una emoción obtenida después de analizar mediante reglas difusas la emoción que obtuvo desde el reconocedor de rostro y desde el reconocedor de voz. Las emociones que se generan con las reglas difusas son las mismas: feliz, sorpresa, neutral, enojado y triste.

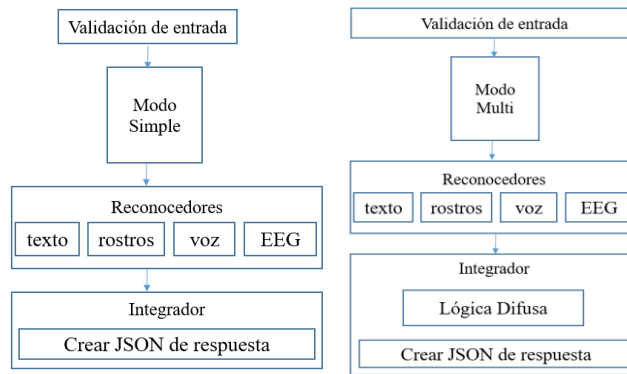


Fig. 6. Funcionamiento de Modo Simple y Modo Multi.

Como muestra la Fig. 6, en modo multi, una vez aplicados los reconocedores individuales, a diferencia del modo simple, utiliza un motor difuso que consta de una serie de reglas difusas para determinar una sola emoción considerando en cuenta dos o más.

Se creó un generador de reglas difusas para la creación de reglas en todas las combinaciones de dispositivos que pudieran darse. El generador de reglas está basado en la tabla de Whissell (Figura 7). Se ubica a cada emoción de los diferentes dispositivos en la tabla y se calcula un promedio para determinar la emoción final.

La tabla de Whissell está basada en activación y evaluación, y están distribuidas en ella las emociones.

Table 1. Ubicación de las emociones en la tabla de Whissell.

Emoción	Punto
Feliz	(2.2 , 1)
Sorpresa	(1.5 , 2)
Neutral	(0 , 0)
Enojado	(-0.5 , 1.1)
Triste	(-1.5 , -0.5)

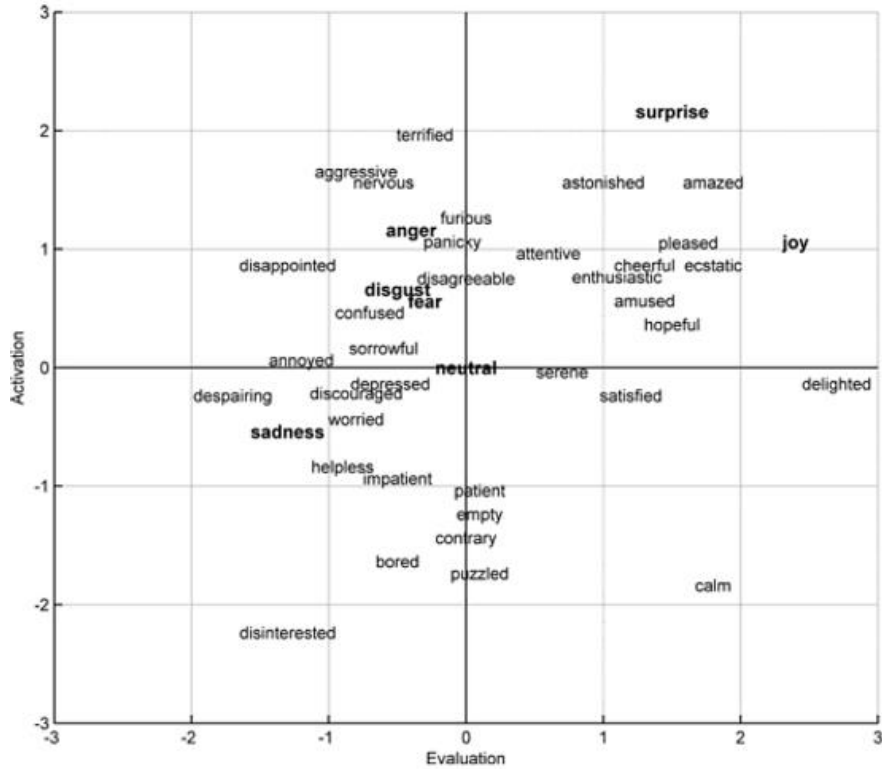


Fig. 7. Tabla de Whissell.

Los reconocedores de voz y de señales EEG solo regresan la valencia (emoción positiva y negativa), y también emoción neutral en el caso de señales EEG. Tomando en cuenta esto y para la integración de estas señales con el resto (expresiones faciales y texto), se realizaron diferentes pruebas para determinar una ubicación promedio de una valencia positiva o negativa en la tabla de Whissell y se determinó ubicarlas en los puntos que muestra la tabla 2.

Table 2. Ubicación de positivo y negativo de wav y EEG en la tabla de Whissell.

Emoción	Punto
Positivo	(1,0.8)
Negativo	(-0.6,0.1)

Con estos valores definidos, ya se pueden ubicar todos los resultados de los cuatro reconocedores dentro de la tabla de Whissell. El procedimiento para obtener una emoción multimodal en PREMOC es:

- Ubicar las emociones de los reconocedores individuales en la tabla de Whissell.
- Obtener un punto promedio entre las ubicaciones de las emociones.
- Calcular la distancia desde el punto promedio hasta la ubicación de las emociones de la tabla 1.

d) Seleccionar la emoción que tenga la distancia menor al punto promedio.

De esta manera se selecciona la emoción final para el modo multimodal en el generador de reglas; las reglas son creadas con el formato de la regla que se muestra en la Figura 8, y son procesadas por la librería jFuzzyLogic [36]. De esta manera se crean todas las reglas difusas necesarias para la integración de emociones en el modo multimodal. En total fueron creadas 470 reglas para las combinaciones de estos 4 reconocedores.



```
RULE 1: IF Txt IS Feliz AND Img IS Feliz AND Wav IS
Positiva THEN Emocion IS Feliz;
```

**Fig. 8.** Regla difusa.

### 1.2.3. Modo student

En modo student, el usuario obtendrá como respuesta el valor de una emoción tomando en cuenta las que normalmente son comunes en un ámbito de aprendizaje. Estas emociones son enganchado, aburrido, frustrado y neutral.

La creación de reglas difusas del modo student fue similar al modo multi, usando la tabla de Whissell. El modo student regresa al usuario las emociones educativas enganchado, frustrado, aburrido y neutral.

## 4. Experimentos

Actualmente PREMOC está siendo utilizado en varios proyectos dentro del laboratorio y entre ellos tenemos a EasyLogic y STAAM; ambos son Sistemas Tutores Inteligentes y Afectivos, donde el reconocimiento de emociones es realizado con PREMOC.

EasyLogic es un sistema web, que actúa como Sistema Tutor Inteligente Afectivo en el proceso de aprendizaje de programación de un lenguaje. EasyLogic utiliza el reconocimiento de emociones en rostro de PREMOC en modo student cada 15 segundos, para mostrar al estudiante dos tipos de ayudas mientras aprende lógica algorítmica: informativas y motivacionales (figura 10).

STAAM es un Sistema Tutor Inteligente Afectivo para matemáticas para niños de primer año de secundaria. STAAM utiliza el reconocedor de emociones en rostros y el reconocedor de emociones en texto de PREMOC para calificar los ejercicios resueltos por los alumnos. STAAM considera el estado afectivo del estudiante además del tiempo y la resolución del ejercicio y entrega un trofeo diferente al estudiante dependiendo del estado de ánimo con el que resolvió el ejercicio.

## 5. Conclusiones y trabajo futuro

Como conclusión, el reconocedor de emociones en imágenes según pruebas realizadas tiene una tasa de éxito del 80%. El reconocedor de emociones en texto tiene poco más de 80% según pruebas realizadas con estudiantes. El reconocedor de

emociones en voz obtuvo 64%. El reconocedor de emociones en señales EEG muestra su tasa de éxito en 70%. Las pruebas fueron realizadas con los mismos datos de entrenamiento.

Se puede concluir que tener un sistema de reconocimiento multimodal de emociones en línea, facilita la integración de afecto a los desarrolladores, ya que pueden crear su software sin realizar el entrenamiento de reconocedores de emociones.

Actualmente se está trabajando en mejorar el reconocedor de emociones en rostros con diferentes técnicas de extracción de características como: Action Units [37] y Binary Pattern [38]. Como trabajo futuro se considera la mejora del reconocedor de emociones en voz con más datos de entrenamiento y que no sea de solo valencia, sino, de las emociones *feliz*, *enojado*, *triste* y *neutral*. También se considera mejorar el reconocedor de texto y de señales EEG con otros tipos de extractores de características. También integrar el reconocimiento de otros dispositivos el rastreo de ojos.

## Referencias

1. Picard, R.W.: Affective Computing. No. 321, pp. 1–16 (1995)
2. Mehrabian, A.: Communication without words. *Psychol. Today*, Vol. 2, pp. 53–56 (1968)
3. Ekman, P.: Strong Evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique. *Psychol. Bull.*, Vol. 2, No. 1, pp. 268–287 (1994)
4. Han, K., Yu, D., Tashev, I.: Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine. pp. 223–227 (2014)
5. Seol, Y.S., Kim, D.J., Kim, H.W.: Emotion Recognition from Text Using Knowledge-based ANN. *Proceedings of 23rd International Technical Conference on Circuits/Systems, Computers and Communications*, pp. 1569–1572 (2008)
6. Liu, Y., Sourina, O., Nguyen, M.K.: Real-time EEG-based Human Emotion Recognition and Visualization. *Trans. Comput. Sci. XII*, Vol. 6670, Special Issue on Cyberworlds, pp. 256–277 (2011)
7. Shaheen, S., El-Hajj, W., Hajj, H., Elbassuoni, S.: Emotion Recognition from Text Based on Automatically Generated Rules. *IEEE Int. Conf. Data Min. Work.*, pp. 383–392 (2014)
8. Zheng, D., Tian, F., Liu, J., Zheng, Q., Qin, J.: Emotion Chat: A Web Chatroom with Emotion Regulation for E-Learners. *Phys. Procedia*, Vol. 25, pp. 763–770 (2012)
9. Khandait, S.P., Thool, R.C., Khandait, P.D.: Automatic Facial Feature Extraction and Expression Recognition based on Neural Network. *Int. J. Adv. Comput. Sci. Appl.*, Vol. 2, No. 1, pp. 113–118 (2011)
10. Gangwar, S., Shukla, S., Arora, D.: Human Emotion Recognition by Using Pattern Recognition Network. *Int. J. Eng. Res. Appl.*, Vol. 3, No. 5, pp. 535–539 (2013)
11. De, A., Saha, A., Pal, M.C.: A Human Facial Expression Recognition Model Based on Eigen Face Approach. *Procedia Comput. Sci.*, Vol. 45, pp. 282–289 (2015)
12. Kostoulas, T., Fakotakis, N.: A speaker dependent emotion recognition framework. *Proc. 5th Int. Symp. Commun. Syst. Networks Digit. Signal Process.*, University of Patras, pp. 305–309 (2006)
13. Vogt, T., André, E., Bee, N.: *EmoVoice — A Framework for Online Recognition of Emotions from Voice. Perception in Multimodal Dialogue Systems*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 188–199 (2008)
14. Caballero-Morales, S.O.: Recognition of emotions in Mexican Spanish speech: An approach based on acoustic modelling of emotion-specific vowels. *Sci. World J.*, Vol. 2013, p. 13 (2013)
15. Nie, D., Wang, X.W., Shi, L.C., Lu, B.L.: EEG-based emotion recognition during watching movies. *5th Int. IEEE/EMBS Conf. Neural Eng. NER*, pp. 667–670 (2011)

16. Lee, Y.Y., Hsieh, S.: Classifying different emotional states by means of eegbased functional connectivity patterns. *PLoS One*, Vol. 9, No. 4, pp. 1–13 (2014)
17. Alonso-Martín, F., Malfaz, M., Sequeira, J., Gorostiza, J.F., Salichs, M.A. A multimodal emotion detection system during human-robot interaction. *Sensors (Basel)*, Vol. 13, No. 11, pp. 15549–15581 (2013)
18. Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C.M., Kazemzadeh, A., Lee, S. Neumann, U., Narayanan, S.: Analysis of emotion recognition using facial expressions, speech and multimodal information. 6th Int. Conf. Multimodal interfaces, pp. 205–211, (2004)
19. Chuang, Z.J., Wu, C.: Multi-modal emotion recognition from speech and text. *J. Comput. Linguist. Chinese*, Vol. 9, No. 2, pp. 45–62 (2004)
20. D’Mello, S.K., Graesser, A.: Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Model. User-Adapted Interact.*, Vol. 20, No. 2, pp. 147–187 (2010)
21. Caridakis, G., Castellano, G., Kessous, L., Raouzaoui, A., Malatesta, L., Asteriadis, S., Karpouzis, K.: Multimodal emotion recognition from expressive faces, body gestures and speech. *IFIP Int. Fed. Inf. Process.*, Vol. 247, pp. 375–388 (2007)
22. Paredes, P., Berkeley, U.C., Sun, D.: Sensor-less Sensing for Affective Computing and Stress Management Technology. *Pervasive Comput. Technol. Healthc. Work.*, pp. 459–463 (2013)
23. Wagner, J., Andre, E., Jung, F.: Smart sensor integration: A framework for multimodal emotion recognition in real-time. 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, pp. 1–8 (2009)
24. Gonzalez-Sanchez, J., Chavez-Echeagaray, M.E., Atkinson, R., Bursleson, W.: ABE: An agent-based software architecture for a multimodal emotion recognition framework. Proc. - 9th Work. IEEE/IFIP Conf. Softw. Archit. WICSA, April 2016, pp. 187–193, (2011)
25. Zatarain Cabada, R., Barron Estrada, M.L., Hernandez, F.G., Bustillos, R.O.: An Affective Learning Environment for Java. IEEE 15th International Conference on Advanced Learning Technologies, pp. 350–354 (2015)
26. Zatarain-Cabada, R., Barrón-Estrada, M.L., García-Lizárraga, J., Muñoz-sandoval, G., Rios-Feliz, J.M.: Java Tutoring System with Facial and Text Emotion Recognition. *Res. Comput. Sci.*, Vol. 106, No. April 2016, pp. 49–58 (2015)
27. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., van Knippenberg, A.: Presentation and validation of the Radboud Faces Database. *Cogn. Emot.*, Vol. 24, No. 8, pp. 1377–1388, (2010)
28. Witten, I. H., Frank, E., Trigg, L., Hall, M., Holmes, G., Cunningham, S.J.: *Weka : Practical Machine Learning Tools and Techniques with Java Implementations*. Seminar, Vol. 99, pp. 192–196 (1999)
29. OpenCV Filtros. [Online]. Available: <http://docs.opencv.org/2.4/modules/imgproc/doc/imgproc.html>.
30. Musicg, Spectrogram. [Online]. Available: <http://deeplearning4j.org/canovadoc/musicg/extension/Spectrogram.html>.
31. Chang, C., Lin, C.: LIBSVM : A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.*, Vol. 2, pp. 1–39 (2011)
32. Inc, E., Epoc - Emotiv. [Online]. Available: <https://emotiv.com/epoc.php>.
33. Mahajan, R., Bansal, D., Singh, S.: A Real Time Set Up for Retrieval of Emotional States from Human Neural Responses. *World Acad. Sci. Eng. Technol.*, Vol. 8, No. 3, pp. 144–149 (2014)
34. Wang, X.-W., Nie, D., Lu, B.-L.: Emotional state classification from EEG data using machine learning approach. *Neurocomputing*, Vol. 129, pp. 94–106 (2014)
35. Barrón-Estrada, M.L., Zatarain-Cabada, R., Aispuro-Gallegos, C.L., Sosa-Ochoa, C.G. Lindor-Valdez, M.: Affective Environment for Java Programming Using Facial and EEG Recognition. *Res. Comput. Sci.*, Vol. 106, pp. 39–47 (2015)

36. Cingolani, P., Alcalá-Fdez, J.: jFuzzyLogic: a Java Library to Design Fuzzy Logic Controllers According to the Standard for Fuzzy Control Programming. *Int. J. Comput. Intell. Syst.*, Vol. 6, No. sup1, pp. 61–75 (2013)
37. Tian, Y.-L., Kanade, T., Cohn, J.F.: Recognizing upper face action units for facial expression analysis. *Proc. IEEE Conf. Comput. Vis. Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, Vol. 1, No. 2, pp. 1–19 (2000)
38. Feng, X., Pietikäinen, M., Hadid, A.: Facial Expression Recognition with Local Binary Patterns and Linear Programming. *Pattern Recognit. Image Anal.*, Vol. 15, No. 2, pp. 546–548 (2005)