

# SEEG-BUAP: Un corpus de señales electro-encefalográficas con emociones inducidas

Illiana Morales, Darnes Vilariño, Cristina Aguilar, Josefa Somodevilla

Benémerita Universidad Autónoma de Puebla,  
Facultad de Ciencias de la Computación, Puebla, México

{illiana.mrls.t,dvilarinoayala,mariajsomodevilla}@gmail.com  
crisaguilarc@hotmail.com  
<http://www.lke.buap.mx/>

**Resumen.** En este artículo se presenta un corpus de señales electro-encefalo-gráficas etiquetado manualmente con emociones inducidas a través de videos y encabezados de noticias. El corpus, al que se le ha denominado SEEG-BUAP, fue construido en base a una población de 40 personas a quienes se le mostró una serie de videos que detonaban algún tipo de emoción del estilo “alegría”, “tristeza”, etc. Se espera que la colección de datos construida y puesta a disposición pública sirva como base para el desarrollo de modelos de clasificación automática basadas en emociones básicas detectadas a través de señales electro-encefalográficas.

**Palabras clave:** Corpus etiquetado, señales electro-encefalográficas, clasificación automática.

## SEEG-BUAP: A Corpus of Electroencephalographic Signals with Induced Emotions

**Abstract.** In this paper we present a manually annotated corpus of electroencephalographic signals with emotions induced through videos and news headlines. The corpus, named SEEG-BUAP, has been constructed using 40 persons to whom a video has been shown, expecting to trigger some kind of emotion such as “happiness”, “sadness”, etc. We expect that the dataset constructed, will be freely available for further use in the construction of classification models for automatic detection of basic human emotions employing electroencephalographic signals.

**Keywords.** Manually annotated corpus, electroencephalographic signals, machine learning.

### 1. Introducción

El electroencefalograma (EEG) permite sondear la actividad neuronal, ya sea en entornos clínicos o de investigación. Médicamente, es una prueba estándar

para el diagnóstico de epilepsia, así como una serie de otras condiciones relacionadas con traumatismo y patología [10,13].

En el ámbito de la investigación, un EEG se utiliza para estudiar las respuestas neuronales ante ciertos estímulos externos como en [2], donde al desarrollar actividades como bailar y tocar guitarra, obtuvieron los datos de manera manual para un grupo de alumnos de Salsa en pareja. En otro trabajo [11], se interpretan las señales electroencefalográficas en la pronunciación del habla imaginada, en particular un vocabulario de 5 palabras, donde cada una de las palabras fue repetida 33 veces en 27 diferentes sujetos de prueba, sin embargo, en dicho trabajo no se menciona un corpus de prueba y/o entrenamiento disponible públicamente.

Para poder realizar tareas importantes relacionadas con la clasificación automática es necesario contar con conjuntos de datos de entrenamiento adecuados. En particular, cuando se diseña un corpus deben tomarse en cuenta distintos aspectos tales como la finalidad, los límites y el tipo de corpus [12]. En la actualidad es posible encontrar algunos corpora disponibles y relacionados con EEG, por ejemplo, conjunto de datos llamado “EEG Motor Movement/Imagery” cuenta con mas de 1500 registros de EEG de 1 y dos minutos, obtenidos de 109 voluntarios a los cuales se les pidió realizar ciertas tareas como: abrir y cerrar los puños con ojos abiertos, o abrir y cerrar los puños con los ojos cerrados [4]. Estos datos fueron capturados por 64 canales de registro EEG utilizando el sistema BCI200 [8]. Por otro lado, la base de datos CHB-MIT Scalp EEG Database fue obtenida en el Hospital de Niños de Boston, y se compone de registros EEG de personas pediátricas con convulsiones. Esta base de datos contiene información de 22 personas (5 hombres de edades entre 3-22, y 17 mujeres de edades entre 1.5-19) [9]. La base de datos EPILEPSIAE (una colección europea de información sobre epilepsia) contiene los registros EEG de 275 pacientes de los centros de epilepsia del Hospital Universitario de Friburgo, Alemania, del Hospital de la Universidad de Coimbra, Portugal, y del Hospital de la Pitier-Salpêtrière París, Francia [7]. Finalmente, el corpus HUT-EEG liberó recientemente 14 años de registros EEG clínicos recogidos en el Hospital de la Universidad de Temple. La colección consta de 16,986 sesiones de 10,874 sujetos únicos; los registros han sido organizados y se combinan con informes clínicos que describen los pacientes y las exploraciones [6].

Relativamente poca de esta información está disponible públicamente a la comunidad de investigación en una forma que sea útil para el aprendizaje automático [6], y la que se encuentra disponible está dirigida a estudios médicos. Hasta donde tenemos conocimiento, en la literatura no se cuenta con un corpus disponible públicamente de señales EEG con etiquetas de emociones. Es por eso, que en este artículo se describe la creación de un corpus de señales EEG con emociones inducidas mediante contenido multimedia.

El contenido restante de este artículo está estructurado como sigue. En la Sección 2. se presentan conceptos relacionados con la electro-encefalografía, en particular, los métodos de obtención de las señales EEG. En la sección 3. se presenta el proceso de adquisición de datos, describiendo no solamente como

se registran los datos sino también las emociones seleccionadas. Los resultados obtenidos se muestran y discuten en la Sección 4.. Finalmente, en la sección 5. se dan las conclusiones y trabajo a futuro.

## **2. Electroencefalografía**

El electro-encefalograma es el registro de la actividad eléctrica que se genera en el cerebro por las neuronas en el interior del mismo; siendo ésta obtenida mediante electrodos superficiales en la base del cráneo [5]. A medida que el registro de la actividad eléctrica del cerebro se realiza mas cerca del cerebro, se puede obtener una señal mas limpia.

En este trabajo, la adquisición de las señales EEG se realiza mediante la diadema de Emotiv EPOC+ [3], la cual ha sido principalmente diseñada para aplicaciones de investigación contextualizadas prácticas. Emotiv EPOC+ ofrece acceso datos de señales EEG con una calidad relativamente alta, utilizando software propietario con un SDK avanzado. Aunque también es posible acceder a esta información mediante software libre o lenguajes de programación como Python. Este dispositivo utiliza sensores superficiales que se ubican en el cuero cabelludo para leer la información electro-encefalográfica; así, el dispositivo cuenta con 14 canales de lectura, mas dos de referencia ubicados detrás de las orejas (ver Figura 1). Los sensores con los que cuenta la diadema de Emotiv Systems se colocan en la parte occipital, parietal y frontal de la cabeza; su nomenclatura indica la región de la misma donde están ubicados: Frontal (F), Central (C), Parietal (P), Occipital (O), Temporal (T) y Fronto-Parietal (FP). La diadema se complementa con un receptor USB, el cual permite que las señales adquiridas mediante los sensores sean enviadas a una computadora. Las señales electroencefalográficas se encuentran en el rango de frecuencias de 0 a 50 Hz y se pueden clasificar en cuatro bandas de frecuencia: Delta, Theta, Alpha y Beta.

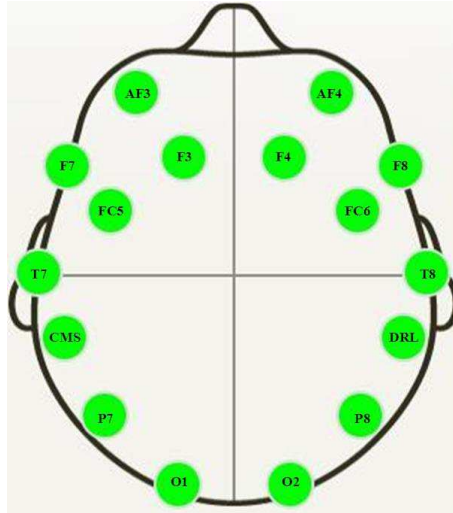
En la siguiente sección se discute la metodología utilizada para la adquisición de los datos para la construcción del corpus SEEG-BUAP.

## **3. Adquisición de los datos**

La construcción de un corpus de información etiquetada manualmente requiere indudablemente de una metodología que permita garantizar la calidad de los datos. Así, en esta sección se describe la manera en que el corpus fue creado.

### **3.1. Emociones seleccionadas**

Para la confección del corpus SEEG-BUAP se tomaron en cuenta las seis emociones básicas propuestas por Ekman [1], puesto que estas emociones se encuentran presentes en todos los individuos y pueden ser detonadas a través de estímulos. A continuación se enumeran cada una de ellas.



**Fig. 1.** Sensores de la diadema Emotiv Epoc+.

1. Alegría,
2. Tristeza,
3. Ira,
4. Miedo,
5. Asco,
6. Sorpresa.

Disparar dichas emociones en los seres humanos no es una tarea sencilla, sin embargo, nuestra metodología plantea inducir la emoción a través de videos.

### 3.2. Etiquetas de los videos

En una primera captura de datos se seleccionaron 3 videos que detonaran la emoción de “Alegría”. Se escogieron videos de comediantes nacionales e internacionales con una duración entre 3 y 5 minutos.

Los videos fueron manualmente etiquetados como se muestra en la Tabla 1. En aquellos intervalos de segundos donde presumiblemente se detonaba cierta emoción, se utilizó la etiqueta correspondiente. En el caso de no detonarse ninguna de las seis emociones básicas, la etiqueta para ese intervalo de segundos es “Ninguna”. Con ayuda de un sistema computacional desarrollado durante este proyecto fue posible realizar captura de imágenes del rostro del sujeto de prueba para comprobar que la emoción haya estado presente, quedando la captura de datos como se muestra en la Figura 2.

Una vez etiquetado cada video con el lapso de tiempo en donde una posible emoción debía ser inducida a una persona. El siguiente paso es adquirir las señales electro-encefalográficas y construir el corpus. La siguiente sección describe este proceso.

**Tabla 1.** Ejemplo de etiquetas en los videos.

Miedo		
Segundo inicial	Segundo final	Etiqueta
1	52	“Ninguna”
53	73	“Miedo”
74	79	“Ninguna”
80	106	“Miedo”
107	127	“Ninguna”



**Fig. 2.** Reproducción de video y adquisición de datos.

### 3.3. Sistema de adquisición de señales

Para la adquisición automática de los datos se desarrolló un sistema computacional que permite, utilizando dos procesos concurrentes, reproducir los videos y capturar los datos etiquetados automáticamente. Este sistema se desarrolló en Python, utilizando la API que Emotiv Systems proporciona a los desarrolladores. Al utilizar la API, el sistema permite recuperar las señales en bruto de los 16 sensores de la diadema, con una frecuencia de muestreo de 128 muestras por segundo.

Los datos obtenidos son almacenados en un formato *csv* (como se puede ver en la Tabla 2). Cada fila contiene el segundo en que fue capturado, el número

de muestra que representa (de un máximo de 128), el valor asociado al sensor y la etiqueta de la emoción.

**Tabla 2.** Ejemplo de datos recuperados (datos redondeados a cero decimales).

Segundo	Muestra	AF3	F7	F3	FC5	T7	P7	...	AF4	Emoción
37	126	4162	4152	4160	4182	4189	4173	...	4166	Alegría
37	127	4161	4150	4162	4180	4187	4170	...	4164	Alegría
37	128	4168	4153	4173	4183	4186	4175	...	4171	Alegría
38	1	4174	4175	4207	4198	4187	3329	...	4178	Ninguna
38	...	...	...	...	...	...	...	...	...	Ninguna
38	128	4237	4293	4242	4264	4179	4182	...	4185	Ninguna

El proceso de adquisición de datos tuvo que ser afinado constantemente ya que, por ejemplo, cuando inicialmente se realizó una captura de datos en un salón de clases a 11 sujetos de prueba, se pudo concluir que es necesario disminuir distracciones. En la Figura 3 se puede observar a un sujeto de prueba con distracciones que evidentemente alteran el resultado del experimento.



**Fig. 3.** Sujeto de prueba distraído.

En la captura final se citó a cada sujeto de prueba a acudir a un laboratorio y se le situó dentro de una cabina blanca, colocándole además audífonos para no generar distracciones. Se aumentó la lista reproducción de videos a 6, considerando cada una de las emociones básicas, obteniendo un total de 36 videos que fueron previamente etiquetados por anotadores humanos. De estos 36 videos se seleccionaron 1 de cada emoción para formar la lista de reproducción que vería el sujeto de prueba.

## 4. Constitución del corpus SEEG-BUAP

En total se realizaron los experimento con un grupo de 26 personas (16 hombres y 10 mujeres), con un rango de edad de entre 20 a 23 años, estudiantes universitarios con perfil de Computación en las siguientes condiciones: descansados y sin presión.

### 4.1. Filtrado de datos

En total se obtuvieron más de 1,100,000 muestras a partir de las 26 pruebas realizadas. Se descartaron 8 pruebas debido a inconvenientes con el hardware utilizado (dispositivo sin batería suficiente, problemas de conexión, etc.), y a que en algunas ocasiones la emoción no se detonó como se esperaba, lo cual pudo ser comprobado con las imágenes que se sacaron como apoyo a la creación de este corpus. Las pruebas utilizables son las que forman parte del corpus SEEG-BUAP. El corpus final cuenta con más de 650,000 muestras etiquetadas manualmente con la emoción correspondiente<sup>1</sup>.

## 5. Conclusiones y trabajo a futuro

En este trabajo se presenta un corpus de señales EEG con etiquetado manual de emociones inducidas a través de videos. Para la correcta creación del corpus SEEG-BUAP, fueron necesarias varias pruebas, en cada una de ellas se pudo mejorar la calidad de los datos. El corpus puesto a disposición pública cuenta con mas de 650 mil muestras etiquetadas de acuerdo a una de seis posibles emociones básicas.

Como trabajo a futuro se busca utilizar el corpus creado para proponer un modelo de clasificación de emociones con señales EEG. El procesamiento de las señales EEG debe ser llevado a cabo cuidadosamente debido a que hasta el momento se cuenta únicamente con los datos planos (*raw*), es decir, tal cual se ha leído desde el dispositivo. Dado que los datos obtenidos son valores reales, se considera que un clasificador basado en una mezcla de gaussianas sería el que obtendría los mejores resultados, sin embargo, no se descarta el uso de otro tipo de clasificadores, como por ejemplo, máquinas de soporte vectorial. También se espera utilizar el corpus de imágenes para la creación de una base de conocimientos para retroalimentar facialmente un robot humanoide.

## Referencias

1. Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W.A., Pitcairn, T., Ricci-Bitti, P.E., Scherer, K., Tomita, M., Tzavaras, A.: Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology* 53(4), 712–717 (1987)

<sup>1</sup> El corpus puede ser descargado gratuitamente de <http://lke.cs.buap.mx/seeg-buap>

2. Elizondo, J.J.E., Beristain, L.J., Martínez, R.A.R., Perales, A.C., Quevedo, E.M., Sandoval, J.A.R.: Metodología para el análisis de señales encefalográficas en actividades lúdicas. *Academia Journal* 6(5), 1348–1353 (2014)
3. Emotiv E.P.O.C.: Software Development Kit (2010)
4. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: Physiobank, physio-toolkit, and physionet. *Circulation* 101(23), e215–e220 (2000)
5. Guevara Mosquera, S.D.: Adquisición de señales electroencefalográficas para el movimiento de un prototipo de silla de ruedas en un sistema BCI. Master's thesis, Universidad Politécnica Salesiana, España (2012)
6. Harati, A., Choi, S., Tabrizi, M., Obeid, I., Picone, J., Jacobson, M.P.: The temple university hospital eeg corpus. In: *Global Conference on Signal and Information Processing (GlobalSIP)*, 2013 IEEE. pp. 29–32 (Dec 2013)
7. Ihle, M., Feldwisch-Drentrup, H., Teixeira, C.A., Witon, A., Schelter, B., Timmer, J., Schulze-Bonhage, A.: Epilepsiae – a European epilepsy database. *Comput. Methods Prog. Biomed.* 106(3), 127–138 (Jun 2012)
8. Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R.: Bci2000: A general-purpose brain-computer interface (bci) system. *IEEE Transaction on Biomedical Engineering* 51(6), 2004 (2004)
9. Shoeb, A.H.: Application of machine learning to epileptic seizure onset detection and treatment. Ph.D. thesis, Massachusetts Institute of Technology (2009)
10. Tatum, W.: *Handbook of EEG Interpretation*. Springer Demos Medic Series, Springer Publishing Company (2007)
11. Torres-García, A., Reyes-García, C., Villaseñor Pineda, L., Ramírez-Cortés, J.: Análisis de señales electroencefalográficas para la clasificación de habla imaginada. *Revista mexicana de ingeniería biomédica* 34, 23–39 (2013)
12. Torruella, J., Llisterri, J.: Diseño de corpus textuales y orales, chap. 2, pp. 45–77. *Filología e informática. Nuevas tecnologías en los estudios filológicos* (1999)
13. Yamada, T., Meng, E.: *Practical Guide for Clinical Neurophysiologic Testing: EEG*. M - Medicine Series, Lippincott Williams & Wilkins (2009)