

MEDICIÓN DE LA EXPERIENCIA DE USUARIO MEDIANTE DATOS ENCEFALOGRAFICOS (EEG) PARA INTERFACES DE USUARIO

*Daniel Gaspar Figueiredo
Emilio Insfran
Silvia Abrahão
Jean Vanderdonckt*

12.1 INTRODUCCIÓN

Los diseñadores de software deben ser capaces de construir un software que satisfaga las necesidades de los usuarios y que proporcione una experiencia de usuario satisfactoria. Por un lado, para analizar la funcionalidad del software y para asegurar que no tenga errores existen distintas técnicas de aseguramiento de calidad que se pueden aplicar durante las fases de validación, verificación o pruebas. Por otro lado, los métodos de evaluación de usabilidad y experiencia de usuario (User Experience – UX) han ido evolucionando a lo largo de los años. Un método de evaluación de usabilidad/UX es un procedimiento sistemático compuesto por un conjunto de actividades bien definidas. Su propósito es medir una serie de características consideradas *a priori* como predictivas y/o informativas de la usabilidad/UX del software, para posteriormente interpretar los resultados, que permitan cuantificar el grado de usabilidad/UX alcanzado. Esta evaluación es

esencial para corregir los problemas detectados y mejorar la usabilidad o la experiencia de usuario del software.

El estándar ISO 9241-11 [9] define la usabilidad como “la medida en la que un producto se puede usar por determinados usuarios para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un contexto de uso especificado”. Sin embargo, en los últimos años, ha habido un cambio de énfasis hacia la experiencia del usuario, donde también las emociones y motivaciones son tenidas en cuenta. Como la UX subsume una gama de cualidades experienciales difusas como la felicidad, el disgusto, la sorpresa, por nombrar algunas, las controversias y dudas sobre la capacidad de medición de la UX son inevitables.

La búsqueda de la mejora de la medición de la UX ha ido cambiando a lo largo de los años. Por ejemplo, desde 2017, los cuestionarios UEQ (User Experience Questionnaire) han sobrepasado en número de usos a los cuestionarios del estilo SUS (System Usability Scale) o AttrakDiff, que son muy utilizados para medir la usabilidad y utilidad [6]. Por otro lado, los avances en electrónica, procesamiento por computador, la aplicación de la neurociencia, seguimiento ocular, entre otros, también nos permiten obtener métricas relevantes sobre la UX. Además, ya se están usando para evaluar diferentes campos de aplicación [19]. El acceso a estos datos fisiológicos nos permite evaluar de forma objetiva la reacción del usuario. Usando cuestionarios puede entrar en juego el recuerdo del usuario, una percepción tardía del software que se ha usado u otros factores como preferencias personales. Por este motivo, los datos obtenidos directamente a través de estas reacciones físicas son más fiables. Se puede obtener en qué instante de tiempo, o en qué fase del uso del software se ha realizado una mayor carga cognitiva, se ha tenido una menor distracción, o se ha experimentado mayor felicidad o angustia, entre otros. Existen varias formas de obtener estos datos fisiológicos, ya sea mediante seguimiento ocular, registro del pulso cardíaco, actividad cerebral, etc.

El hecho de que todos los procesos, físicos y mentales, estén controlados por nuestro cerebro sugiere que podemos obtener información mediante registros de ondas cerebrales. Por esto, el electroencefalograma (EEG) es una de las técnicas más fiables para medir la cognición humana, evaluar y monitorizar la salud mental de las personas [15]. Los datos fisiológicos toman especial valor cuando tenemos varios usuarios porque pueden reaccionar de forma distinta cuando usan un mismo software. Es decir, no existe un patrón universal que determine qué tipo de interacción con el software producirá una reacción positiva o negativa para todas las personas.

Uno de los elementos que genera distintas reacciones entre los usuarios cuando se mide la UX es el uso de interfaces de usuario (User Interfaces – UI). Una

interfaz de usuario inapropiada o que no se ajuste a las expectativas puede llevar a que los usuarios se frustren y dejen de utilizar un software [5]. Por tanto, un factor a tener en cuenta cuando se diseña el software es el punto de vista del usuario. Si únicamente se tiene en cuenta el punto de vista del desarrollador, dejando de lado el punto de vista del usuario, puede hacer que los usuarios no encuentren el propósito de usar la aplicación [1], [14].

En este capítulo presentamos una experiencia sobre la obtención de datos fisiológicos de la UX cuando distintos usuarios interactúan con una aplicación a través de su interfaz de usuario. Se presentan las métricas obtenidas, distintas alternativas de interfaz de usuario y los resultados de un experimento para comparar los datos fisiológicos recogidos mediante electroencefalogramas (EEG). Más concretamente, se va a comparar la carga cognitiva y la UX que supone el uso de distintos tipos de menú para distintos tipos de usuarios y poder así construir una base de conocimiento que permita conocer mejor al usuario y sus reacciones. El propósito del estudio es sentar las bases para el desarrollo de interfaces de usuario inteligentes que sean capaces de adaptarse no solo a las preferencias de los usuarios sino también a sus características y habilidades cognitivas y reacciones fisiológicas.

12.2 FUNDAMENTOS

12.2.1 Experiencia de usuario

La Experiencia de usuario (UX) ha tenido varias definiciones debido a su gran multidisciplinariedad. Según la norma ISO 9241-210, [8], la UX se define como “las percepciones y respuestas de una persona que resultan del uso o el uso anticipado de un producto, sistema o servicio”, incluidas las emociones, creencias, preferencias, percepciones, respuestas físicas y psicológicas, comportamientos y logros de los usuarios que se producen antes, durante y después del uso [7], [8]. Es decir, la experiencia de usuario es un factor clave cuando analizamos la calidad de un servicio o producto. Para evaluar la experiencia de usuario, se usan métodos e instrumentos que permiten obtener la percepción sobre el uso de un servicio o producto para poder así detectar fallos o aspectos a mejorar.

La experiencia de usuario se mide habitualmente por medio de entrevistas o cuestionarios sobre el sistema que se ha usado. Por ejemplo, para conocer la satisfacción del usuario, la facilidad de aprendizaje o el cumplimiento de expectativas, se usan cuestionarios con escalas que permiten expresar el agrado o el desagrado de un usuario tras usar la aplicación. Ahora bien, estas respuestas pueden estar condicionadas a un recuerdo inexacto de los usuarios u otros factores subjetivos, llevando así a mediciones imprecisas. Atendiendo a la definición que nos

ofrece la ISO y según Neumann & Westburym [13] también se pueden tener en cuenta las percepciones físicas y fisiológicas a la hora de medir o monitorizar la UX. Estas medidas son más objetivas que las medidas obtenidas por cuestionarios [13]. Además, la tarea de conseguir monitorizar las señales fisiológicas está siendo cada vez más viable gracias a la evolución de sensores y señales de procesamiento, y por tanto es natural su aplicación como herramienta para medir la UX.

Algunas de las medidas fisiológicas más utilizadas por la comunidad *Human Computer Interaction (HCI)*, según la recopilación realizada por Maia & Furtado [10] se pueden ver en la Tabla 12.2.

Medida fisiológica	Descripción
Respuesta galvánica de la piel (Galvanic Skin Response, GSR) o Actividad Electrodermica (Electrodermal Activity, EDA)	Se usa para cuantificar un potencial eléctrico entre dos puntos de la piel [11].
Ritmo cardíaco (Heart Rate, HR)	Se usa para medir el número de pulsaciones del corazón por minuto.
Variabilidad del ritmo cardíaco (Heart Rate Variability, HRV)	Es la variación del ritmo cardíaco (HR).
Ondas cerebrales medidas mediante electroencefalogramas (EEG)	Es la captura y evaluación de señales eléctricas provocadas por el cerebro [16].
Pulsación de volumen sanguíneo (Blood Volume Pulse, BVP),	Se usa para medir los cambios en el volumen sanguíneo en arterias y capilares.
Electromiograma (EMG)	Se usa para medir la respuesta muscular o la respuesta de la actividad eléctrica o la estimulación de un nervio del músculo [11].
Respiración	Mide el volumen de aire que sale de los pulmones de un individuo [11]. Por ejemplo, en estados relajados el ratio de respiración es menor.
Seguimiento ocular, (Eye Tracker)	Se usa para medir fijaciones y esfuerzo visual para encontrar elementos, entre otros [3].

Tabla 12.2: Medidas fisiológicas más utilizadas por la comunidad HCI

Las mediciones fisiológicas obtenidas por seguimiento ocular han sido las más habituales para medir la UX, sin embargo, las medidas basadas en EEG están siendo cada vez más utilizadas. De acuerdo con un estudio reciente sobre monitorización fisiológica [7], se destaca que el uso de mediciones EEG es cada vez mayor en los test de UX: “La amplia gama de estudios identificados en los que se utiliza esta herramienta indica que su aplicación es cada vez mayor en los test de UX y que tiene mucho potencial para extraer el conocimiento de las emociones. A pesar de ser una herramienta con protocolos de aplicación complejos, por el tiempo

necesario y la complejidad para interpretarlo, las posibilidades que ofrece para facilitar la detección de las emociones hacen que su uso vaya en aumento”.

Los avances tecnológicos permiten obtener cada vez con mayor facilidad y fiabilidad una monitorización de datos objetivos sobre la UX experimentada por los usuarios. Esto abre un nuevo campo de posibilidades haciendo que la obtención de datos precisos sea más accesible y por tanto pueda ser utilizada para desarrollar software que ofrezca una mejor experiencia de usuario.

12.2.2 Adaptación de interfaces de usuario

La adaptación de una interfaz de usuario consiste en modificar la interfaz de un software para que cumpla ciertos requisitos como las necesidades, deseos o preferencias de un usuario o grupo de usuarios. La personalización de la interfaz de usuario ya sea hecha de forma automática o por el propio usuario, son formas de adaptación de interfaz de usuario. Todos los elementos de la interfaz de usuario se pueden personalizar al gusto del usuario. Por ejemplo, podemos personalizar botones de acción, listas desplegables, selectores, etc.

En este capítulo vamos a tratar con la personalización de distintos tipos de menús de usuario. Para diferenciar los distintos tipos de menús, debemos diferenciar ciertas propiedades definidas en el espacio de diseño de los menús de usuario que se puede

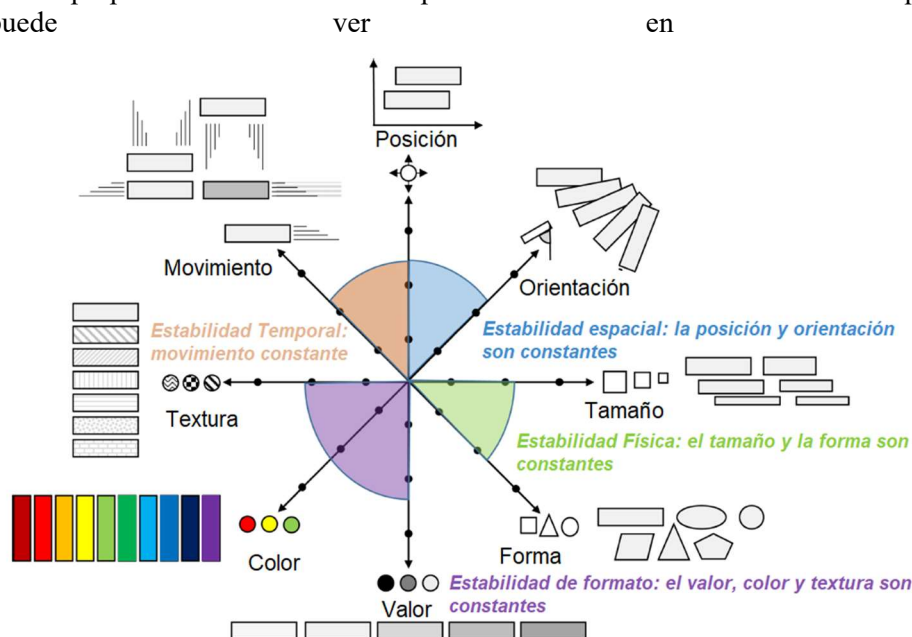


Figura 12.7. Las propiedades variables que podemos diferenciar en este espacio de diseño han sido definidas en [17] y se describen en la Tabla 12.3.

Propiedad	Descripción
Posición	Indica la localización en coordenadas x, y, z de los elementos del menú.
Valor	Indica aspectos de presentación como el uso del tema claro u oscuro, la saturación de color, etc.
Textura	Se usa para indicar patrones, degradados, etc. de los elementos del menú que permitan dar sensación de textura.
Color	Indica la gama de colores que permitan resaltar unos elementos del menú de otros.
Forma	Indica el área de activación de los menús y poder hacerlos más grandes o pequeños.
Tamaño	Indica el tamaño de los elementos de los menús haciendo que sean más o menos visuales y accesibles.
Orientación	Indica el ángulo de inclinación de los elementos para destacarlos sobre otros.
Movimiento	Indica si el menú tiene algún movimiento, variando la posición, orientación, forma o tamaño durante un tiempo determinado.

Tabla 12.3: Propiedades de los menús de usuario

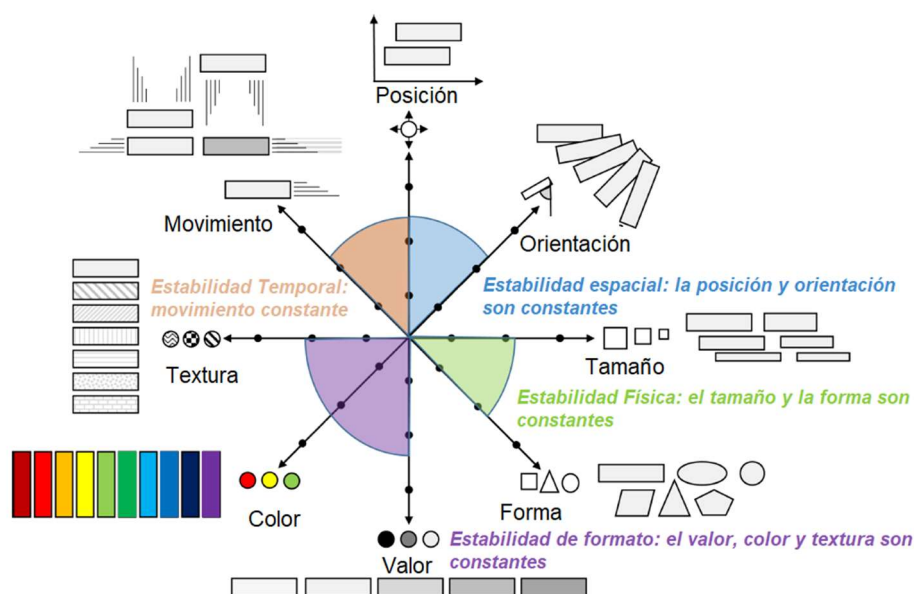


Figura 12.7: Espacio de diseño para los menús de usuario

La combinación de todas estas propiedades puede dar lugar a un gran número de tipos de menús de interfaz de usuario.

A continuación, presentamos 20 tipos de menús propuestos en [17] que usaremos en este trabajo, y que cubren las propiedades definidas en la

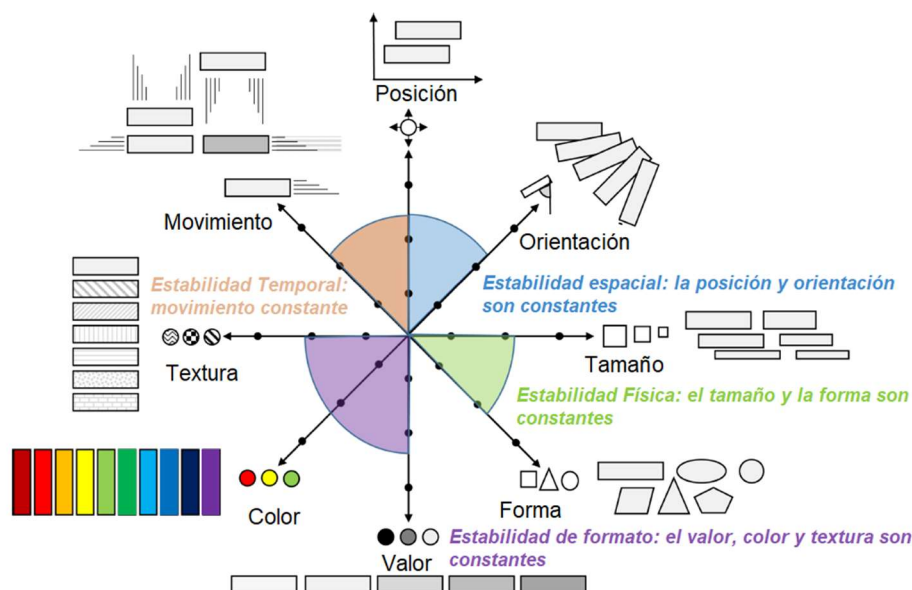
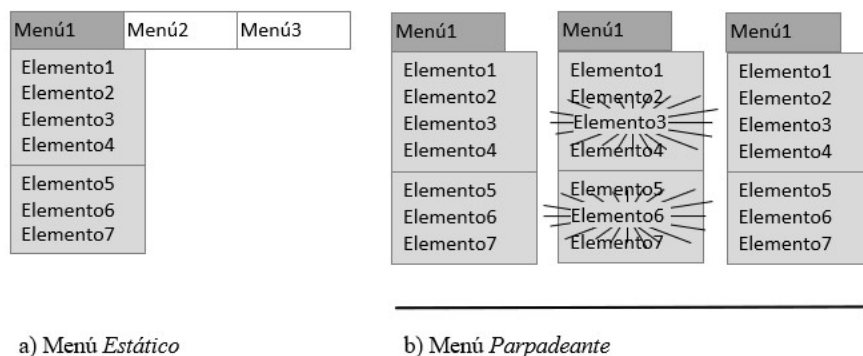


Figura 12.7.

El objetivo de variar las propiedades en los menús es resaltar alguno de sus elementos para que sea más atractivo y/o efectivo en su uso. En la presentación de los distintos tipos de menú se mostrarán ejemplos de estos tipos de menú resaltando 2 elementos de los 7 posibles, como se puede ver en las siguientes figuras.



a) Menú Estático

b) Menú Parpadeante

Figura 12.8: Menú Estático (a) y Menú Parpadeante (b)

- *Estático (Static)*. Esta primera versión del menú es la que no tiene propiedades variables. Como se puede ver en la Figura 12.8 (a) un menú

puede tener varios ítems principales (*Menú1*, *Menú2* y *Menú3*) y varios elementos dentro de cada ítem principal (*Elemento1*, *Elemento2*, *Elemento3*, ...). Los menús mostrados a continuación seguirán esta nomenclatura y formato.

- *Parpadeante (Blinking)*. Este tipo de menú resalta el elemento a resaltar mediante un parpadeo de ritmo lento. Este elemento forma parte de la categoría de “Menús con valores variables”. En la Figura 12.8 (b) se puede ver un ejemplo gráfico de este menú. La línea inferior horizontal representa la variable temporal, es decir, el parpadeo tiene una duración limitada.
- *Rotativo (Rotating)*. En este caso, el elemento gira 360° hasta que vuelve al estado inicial. Forma parte de la categoría de menús con orientación variable. En la Figura 12.9 (a) se puede ver un ejemplo gráfico de este menú.
- *Hoja (Leaf)*. Es un tipo de menú con una forma inusual. Consiste en representar los distintos elementos distribuidos en una estructura de árbol donde cada hoja sería un elemento del menú. Para este tipo de menú no se resaltan elementos en concreto, en cambio, se distribuyen de forma que se agrupa en cada rama un conjunto de elementos con funcionalidades parecidas. La vista gráfica de este menú se puede ver en la Figura 12.9 (b).

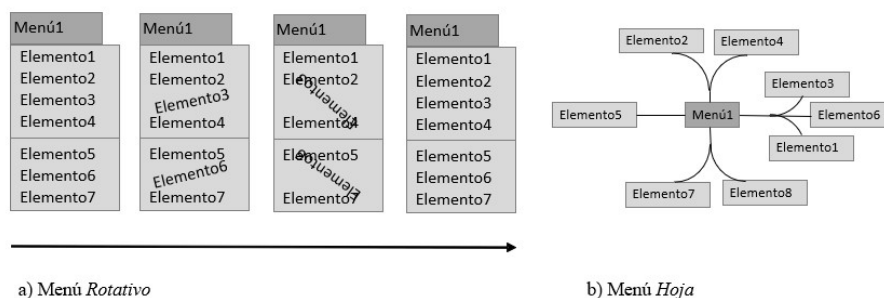


Figura 12.9: Menú rotativo (a) y Menú Hoja (b)

- *Dividido sin replicación (Split without replication)*. Muestra los elementos a resaltar en la cabeza del menú, siendo así los primeros elementos que el usuario ve si realiza un barrido de arriba abajo. Mueve los elementos de sitio, pues los elementos a resaltar dejan de estar en su lugar habitual para estar al inicio del menú. En la Figura 12.10 (a) se puede ver la representación visual de este tipo de menú.
- *Pulsante (Pulsing)*. Este tipo de menú hace un pulso hacia adelante en los elementos a resaltar. Cambiando su tamaño hasta que vuelven al estado inicial. Este tipo de menú se percibe como menos intrusivo que otros menús como el menú *Rotativo*. En la Figura 12.10 (b) se puede ver cómo se vería el efecto en el tiempo de este tipo de menú.

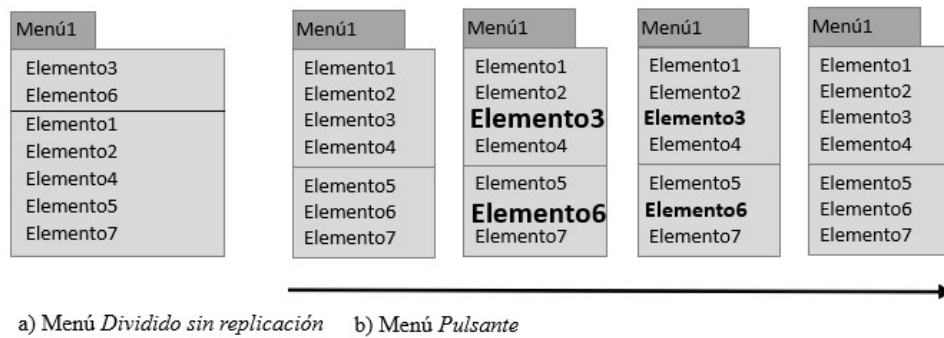


Figura 12.10: Menú Dividido sin replicación (a) y Menú Pulsante (b)

- *Temporal*. Es un tipo de menú que introduce el tiempo de mostrado de los elementos del menú como variable. Es decir, muestra los elementos del menú en dos fases: Al inicio solo se muestran los elementos a resaltar y tras un tiempo reducido (ej. 170 milisegundos), aparece el resto de los elementos. La Figura 12.11 (a) muestra la transición de este menú.
- *Efímero (Ephemeral)*. Este es un tipo de menú parecido al *Temporal*. En este caso, primero se muestra el elemento a resaltar y durante unos milisegundos (ej. 500) los elementos no predichos van apareciendo de forma gradual. En la Figura 12.11 (b) se pueden comparar los dos tipos de menú anteriores (*Temporal* y *Efímero*) para entender mejor la diferencia entre ellos.

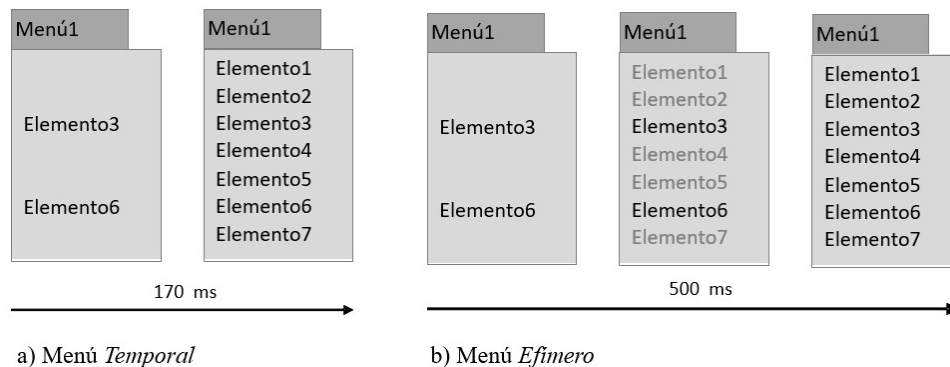


Figura 12.11: Menú Temporal (a) y Menú Efímero (b)

- *Desapareciendo fuera de contexto (Out-of-context Disappearing)*. Al inicio se muestran los elementos a resaltar, y durante unos milisegundos (ej. 500) se desvanecen. Al mismo tiempo, todo el menú aparece de forma gradual. En la Figura 12.12 (a) se puede ver con más detalle este menú.

- *Área de Activación Adaptable (Adaptive Activation-Area)*. Este tipo de menú ofrece un área de activación más grande para los elementos a resaltar. Es decir, el usuario tendrá más área sobre la que pulsar para realizar la navegación. En la Figura 12.12 (b) se representan las acciones del menú con distintas áreas de activación de modo que accionar el *Elemento3* o *Elemento6* sea más fácil.

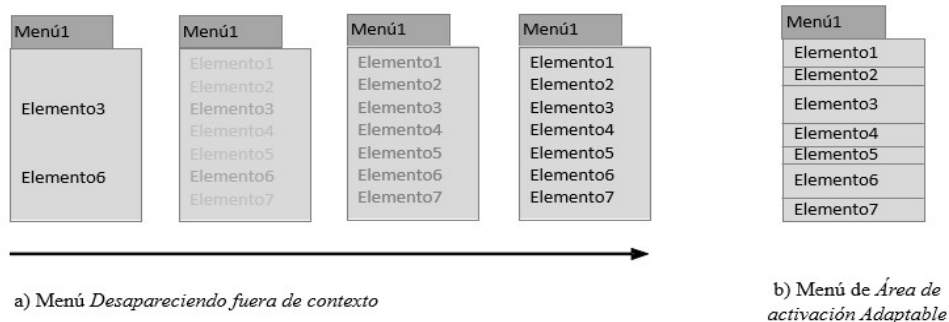


Figura 12.12: Menú Desapareciendo fuera de contexto (a) y Menú de Área de activación adaptable (b)

- *Negrita (Bolding)*. Enfatiza los elementos del menú usando una fuente en negrita para resaltar ciertos elementos. La Figura 12.13 (a) muestra un ejemplo del menú en *Negrita*.
- *Cursiva (Italizing)*. Enfatiza los elementos del menú usando una fuente en cursiva. Como ejemplo, podemos ver la Figura 12.13 (b).
- *Subrayado (Underlying)*. Enfatiza los elementos del menú usando el subrayado de la fuente. La Figura 12.13 (c) muestra este tipo de menú.
- *Escala de grises (Greyscale)*. Enfatiza los elementos usando una escala de grises. El elemento a resaltar se muestra en blanco (si el tema es oscuro) o negro (si se usa un tema claro) y el resto de los elementos en distintos tonos de gris dependiendo del grado de predicción que tengan. La Figura 12.13 (d) muestra como cada elemento tiene un color dentro de la escala de grises para resaltarlos más o menos.

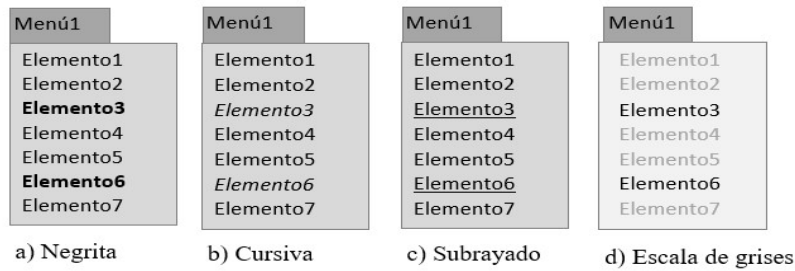


Figura 12.13: menús que usan la negrita (a), cursiva (b), subrayado (c) o escala de grises (d) para resaltar elementos

- *Tamaño de fuente variable (Morphing)*. Utiliza distintos tamaños de fuente para mostrar los elementos a resaltar más grandes y el resto más pequeños. En la Figura 12.14 (a), podemos ver como *Elemento3* y *Elemento6* tienen un tamaño de fuente superior al resto. A mayor tamaño de fuente, también tiene un área de activación mayor.
- *Tambaleante (Twisting)*. Cambia brevemente la orientación en el eje horizontal del elemento a resaltar. Podemos ver en la Figura 12.14 (b) una representación de cómo se moverían los elementos durante un breve periodo de tiempo.

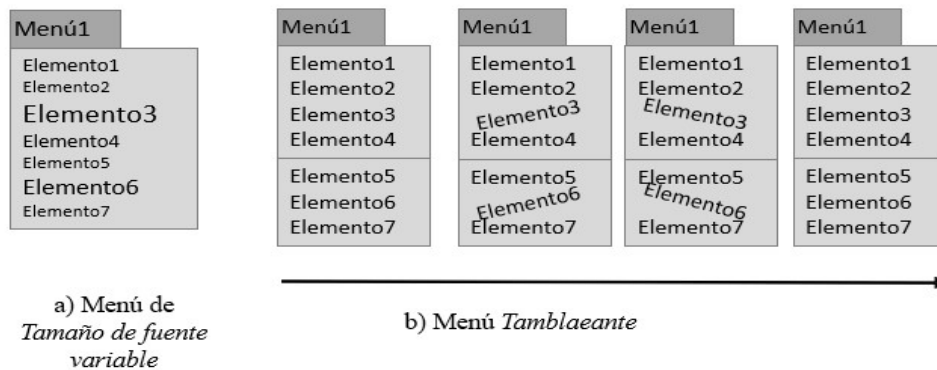


Figura 12.14: Menú de Tamaño de fuente variable (a) y Menú Tambaleante (b)

- *Cuadro (Boxing)*. El elemento a resaltar se encuadra en un marco blanco (si el tema es oscuro) o negro (temas claros) para resaltarlos. Este marco prevalece en el tiempo. Un ejemplo de este menú se puede ver en la Figura 12.15 (a).
- *Brillante (Glowing)*. Este tipo de menú es como el *menú de cuadro*, pero se usa un borde brillante para destacar más el elemento. Además, tras unos

breves instantes de tiempo el recuadro brillante desaparece. En la Figura 12.15 (b) se puede ver gráficamente el borde brillante que caracteriza este menú.



Figura 12.15: Menús de Cuadro (a), Menú Brillante (b), Menú de Resaltado (c) y Menú Rosa (d)

- *Resaltado (Highlight)*. En este tipo de menú se resalta el elemento mediante contrastes de colores sobre el menú estático. Los elementos predichos se enmarcan con un fondo que tiene contraste con el resto del menú para hacerlos más visibles. En la Figura 12.15 (c) podemos ver como se enmarcan los elementos 3 y 6.
- *Rosa (Pink)*. Una forma específica de usar el menú *Resaltado* es enmarcando los elementos con un fondo rosa. En la Figura 12.15 (d) vemos cómo se remarcan en rosa los elementos 3 y 6.

12.2.3 Electroencefalografía

Cualquier interacción de un usuario con un software produce en su cuerpo un conjunto de respuestas fisiológicas. Estas respuestas pueden ser interpretadas a partir de un conjunto de indicadores que reflejan el grado de percepción, positiva o negativa, que experimenta este usuario y que pueden ser utilizadas para medir de forma objetiva la experiencia de usuario (UX). En esta sección se va a profundizar en los conceptos de la electroencefalografía (EEG) para estudiar la UX como respuesta fisiológica provocada por la interacción de los usuarios con distintos tipos de menús de interfaz de usuario.

Para interpretar adecuadamente un electroencefalograma (EEG), debemos entender primero cómo funciona el cerebro. La actividad cerebral consiste en un conjunto de ráfagas de señales eléctricas que viajan a través de células, llamadas neuronas. Cuando una neurona se activa, transmite una corriente eléctrica. Ahora bien, si muchas neuronas se activan a la vez, se puede detectar esta señal a través del

cuero cabelludo utilizando un casco neuronal que registre esas leves señales eléctricas producidas entre las neuronas. Un casco neuronal consiste típicamente en un conjunto de sensores, llamados electrodos, que estando en contacto con el cuero cabelludo, monitorizan la señal eléctrica producida en toda la zona cerebral.

12.2.3.1 FRECUENCIA DE SEÑALES CEREBRALES

Los electrodos de los cascos neuronales, que monitorizan la actividad cerebral mediante las señales eléctricas, son capaces de captar señales distinguidas en cuatro clases según la frecuencia dominante: *ondas Beta* (frecuencia mayor a 13 Hz), *ondas Alfa* (frecuencia entre 8Hz y los 13Hz), *Ondas Theta* (frecuencia entre los 4 Hz y los 7 Hz) y *ondas Delta* (frecuencia inferior a 4 Hz). Cada frecuencia corresponde con el estado en el que se encuentra el sujeto que las emite. Por ejemplo, en la Figura 12.16 podemos detectar cuando un usuario parpadea o tiene los ojos cerrados gracias a la diferencia de potencial que se produce. Este es solo uno de los muchos estados o reacciones que se pueden obtener mediante estas señales.

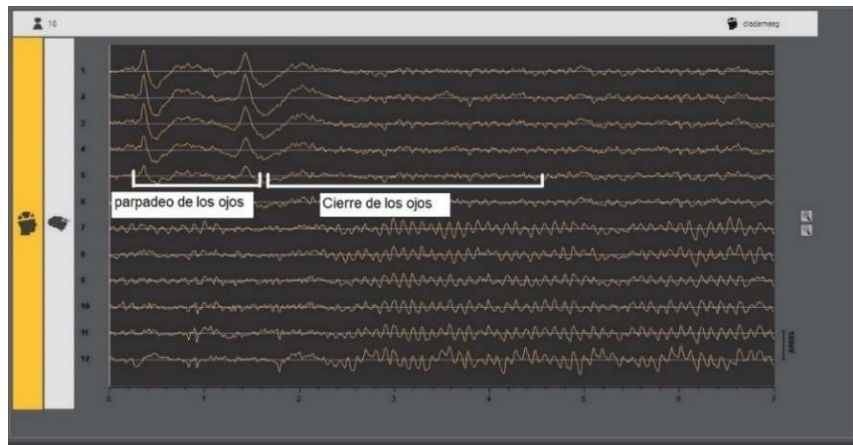


Figura 12.16: Cambios en la señal por tener los ojos cerrados o estar parpadeando

Estas frecuencias y las diferencias de potencial son distintas para cada usuario ya que no todos los humanos reaccionamos de igual forma ante todas las situaciones. Debido a esto, antes de empezar las mediciones se debe realizar una fase de calibración.

La fase de calibración se divide normalmente en las siguientes tareas:

- Tarea de familiarización. Se muestra una tarea de 30 segundos para que el usuario conozca qué debe hacer en la tarea de calibración. Para esta tarea no

registra ningún dato, solamente sirve para familiarizar al usuario con la tarea de calibración.

- **Lavado 1.** Se indica a los usuarios que cierran los ojos y entren en un estado de reposo. Esta fase de lavado sirve para desacoplar posibles respuestas emocionales entre una tarea a otra y obtener una respuesta emocional base para cuando el usuario entra en reposo. Esta fase se sitúa entre la familiarización y el resto de las tareas del estudio.
- **Ojos abiertos.** En esta fase se indica a los usuarios que entren en un estado de reposo con los ojos abiertos. Sirve para establecer una base de mediciones EEG para cada usuario.
- **Calibración.** Esta es una tarea con carga cognitiva que induce un estado de estrés leve controlado en el usuario. Para ello, se pide al usuario que cuente hacia atrás de 7 en 7 en voz alta, empezando por un número aleatorio. Esta tarea dura 1 minuto. Sirve para crear la base de conocimiento de cómo reacciona el participante ante situaciones de estrés y con carga cognitiva.
- **Lavado 2.** Igual que el Lavado 1 anterior, pero en este caso se sitúa entre la tarea de calibración y el posterior estudio que se quiera realizar.

12.2.3.2 MÉTRICAS UX OBTENIDAS MEDIANTE EEG

Como se ha mencionado anteriormente, mediante el análisis de las señales cerebrales podemos detectar el estado en el que se encuentra un usuario ante distintos estímulos. Con este objetivo, se definen un conjunto de métricas que pueden ser obtenidas por medio de dispositivos de EEG y que nos permitirá determinar la UX basada en la respuesta fisiológica de los usuarios:

- **Carga de trabajo (carga cognitiva).** Mide el enfoque neurológico o la concentración de un participante cuando se le presentan estímulos o durante experiencias. Es decir, representa el uso de recursos cognitivos para realizar una tarea o visualizar un estímulo. La carga de trabajo se expresa en porcentajes. Valores cercanos al 0% indican que el participante está muy distraído, mientras que un valor cercano al 100% indica que está muy atento al estímulo.
- **Implicación.** Mide el grado de conexión entre el participante y el estímulo o tarea. Este es un indicador más complejo que la carga cognitiva, ya que un participante puede estar atento a una tarea incluso si la información presentada no es de su interés. El compromiso se expresa como porcentaje. Un valor cercano al 0% indica que no hay conexión o vínculo con los estímulos. Un valor cercano al 100% indica un alto compromiso con los estímulos o la tarea.

- **Memorización.** Mide la intensidad de los procesos cognitivos relacionados con la formación de recuerdos futuros durante la presentación de estímulos o durante una experiencia. Captura el grado de almacenamiento, codificación y retención en la memoria. La memorización se expresa como porcentaje. Un valor de 0% indica que la probabilidad de que se recuerde el estímulo es baja. Un valor cercano al 100% indica una alta posibilidad de que el estímulo se retenga en la memoria de los participantes.
- **Valencia.** Mide el grado de atracción experimentado en respuesta a un estímulo o una situación, desde una reacción positiva o agradable hasta una reacción negativa o desagradable. La valencia se expresa como porcentaje. Un valor de 100% positivo o negativo equivale al valor medido en la fase de calibración. Por tanto, es posible obtener un nivel de valencia superior al 100% (positivo o negativo) si la reacción calculada en el experimento supera la medida durante la calibración.

Cabe destacar que todos los valores de estas variables dependen de cada participante y de su calibración. Cada individuo tiene una respuesta fisiológica diferente para cada situación. Para poder medir las respuestas fisiológicas como métricas usamos la fase de calibración. Durante esta fase obtenemos la respuesta base que el participante tiene para cada métrica. Es por esto por lo que, durante el análisis de datos podemos encontrar valores que superen el 100%. Estos valores por encima del 100% quieren decir que el valor medido para la métrica en cuestión es superior a la medida en la fase de calibración. En el caso de la métrica de Valencia, los datos pueden ser inferiores a -100% indicando que el valor medido en el experimento ha sido menor, una atracción más negativa que el medido en la calibración. Es por este motivo por el que es muy importante realizar una fase de calibración de calidad.

12.2.3.3 TOMA DE DATOS ELECTROENCEFALOGRÁFICOS

El electroencefalograma puede ser llevado a cabo mediante cascos neuronales que detectan las señales y variaciones de potencial creadas por la transmisión de información entre neuronas. Además, se debe usar un software especializado para la captura y el análisis de las señales eléctricas producidas por las neuronas para obtener un electroencefalograma que pueda ser utilizado para entender las distintas reacciones y estados en los que se encuentra el participante. A continuación, se presenta el casco neuronal se va a usar en este trabajo y también el software específico que nos permitirá la toma y análisis de datos.

Por un lado, usaremos el casco neuronal *Diadem* de la empresa *Bitbrain* [4]. Este casco neuronal de sensores secos dispone de 12 canales (y para ello usa 14

electrodos) para la captura de señales EEG. En la Figura 12.17 podemos ver cómo se distribuyen cada uno de estos sensores sobre el cuero cabelludo de un participante.

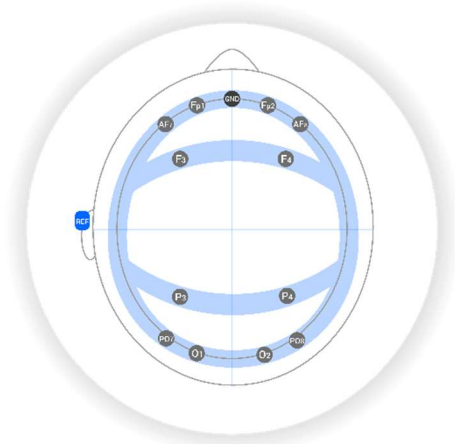


Figura 12.17: Distribución de los sensores en el casco Diadem. Tomado de [4]

Se ha elegido este dispositivo porque además de ser muy preciso y poco susceptible al ruido externo, posibilita la realización de experimentos de forma rápida gracias a su fácil colocación, número de sensores y rapidez de calibración sin perder la calidad de los datos obtenidos.

Para la toma de datos usaremos el software *SennsLab*. Este software nos permite conectar el casco neuronal al ordenador para leer y guardar los datos de las señales eléctricas de la actividad neuronal y las variaciones de potencial capturados. Para el análisis de las métricas obtenidas en función de las señales, usaremos *SennsMetrics*. Ambos programas son software propietario de *Bitbrain* [4].

12.3 DISEÑO DEL EXPERIMENTO

En esta sección se describen las principales características del experimento realizado para recoger información fisiológica de la interacción de usuarios con diferentes tipos de menús y usar esta información para evaluar la experiencia de usuario.

12.3.1 Objetivo

Según el modelo *Goal-Question-Metric* (GQM) [2], el objetivo de este experimento es *analizar* un conjunto de tipos de menús de usuario *con el propósito*

de compararlos con respecto a la carga cognitiva y experiencia de usuario que provocan desde el punto de vista de diseñadores de interfaces de usuario novatos en el contexto de un grupo de usuarios finales.

12.3.2 Pregunta de investigación

Dado que vamos a obtener métricas objetivas (carga cognitiva, implicación, memorización y valencia) mediante el uso de un dispositivo EEG de la respuesta fisiológica de los usuarios cuando interactúan con un interfaz de usuario, podemos formular la siguiente pregunta de investigación:

- RQ: ¿Existen diferencias significativas en la respuesta fisiológica de los participantes cuando utilizan distintos tipos de menú de usuario?

12.3.3 Variables

Hay una variable independiente definida por los 20 tipos de menú que se quieren comparar presentados en la *Sección 12.2.2: Adaptación de interfaces de usuario*. Las variables dependientes corresponden a las cuatro métricas obtenidas con el dispositivo EEG (carga cognitiva, implicación, memorización y valencia).

En la Tabla 12.4 se puede ver una descripción de cada una de las variables. Cabe destacar que las métricas obtenidas con el EEG son mediciones objetivas.

Tipo	Variable	Descripción
Independiente	Tipo de menú a comparar	Esta variable tomará como posible valor cada uno de los menús descritos en la <i>Sección 12.2.2: Adaptación de interfaces de usuario</i> .
Dependiente	Carga cognitiva	Nivel de esfuerzo mental realizado y atención ante un estímulo. Valores de 0 a 100.
	Implicación	Nivel de implicación entre el participante y un estímulo
	Memorización	Nivel de intensidad de procesos cognitivos para crear recuerdos futuros ante un estímulo
	Valencia	Nivel de atracción experimentado ante un estímulo

Tabla 12.4: Variables del estudio

12.3.4 Hipótesis

Para la pregunta de investigación formulada anteriormente, nos planteamos una serie de hipótesis para cada variable dependiente descrita en la Tabla 12.4. Las hipótesis nulas y su correspondiente hipótesis alternativa son:

- $H1_0$: No hay diferencias significativas en la carga cognitiva de los usuarios cuando utilizan distintos tipos de menús de usuario / $H1_a$: Hay diferencias significativas en la carga cognitiva de los usuarios cuando utilizan distintos tipos de menús de usuario.
- $H2_0$: No hay diferencias significativas en la implicación de los usuarios cuando utilizan distintos tipos de menús de usuario / $H2_a$: Hay diferencias significativas en la implicación de los usuarios cuando utilizan distintos tipos de menús de usuario.
- $H3_0$: No hay diferencias significativas en la memorización de los usuarios cuando utilizan distintos tipos de menús de usuario / $H3_a$: Hay diferencias significativas en la memorización de los usuarios cuando utilizan distintos tipos de menús de usuario.
- $H4_0$: No hay diferencias significativas en la valencia de los usuarios cuando utilizan distintos tipos de menús de usuario / $H4_a$: Hay diferencias significativas en la valencia de los usuarios cuando utilizan distintos tipos de menús de usuario.

12.3.5 Diseño

El experimento se ha diseñado para que los participantes realicen una única sesión en la que usarán de forma aleatoria los 20 tipos de menú que queremos comparar. Para cada participante se realizará una primera fase de calibración del casco neuronal, se le indicarán las tareas a realizar y hará uso, a modo de ejemplo para familiarizarse con la tarea experimental, de un menú distinto para seleccionar un elemento del menú. Después de esto, y con el tiempo que se necesite, harán uso de los distintos tipos de menú que forman parte del experimento.

12.3.6 Tarea experimental

Se realizarán sesiones individuales para cada participante. Cada uno usará los distintos tipos de menú en un orden distinto. Para realizar el experimento, seguirán los siguientes pasos:

1. Fase de calibración. Tras colocarse el casco neuronal, se realiza una fase de calibración para detectar los estados del participante. Estas fases están descritas en la *Sección 12.2.3.1: Frecuencia de señales cerebrales*.
2. Se mostrará al participante las instrucciones a seguir para realizar el experimento. Estas instrucciones indican al participante que deberá usar una serie de menús de usuario con un contexto concreto.

3. Se mostrará un ejemplo al participante de un tipo de menú que no se pretende analizar en un contexto distinto a los que usan los menús a analizar. Solo cuando el participante haya entendido la dinámica de la tarea podrá seguir adelante. Puede volver a usar el menú de ejemplo tantas veces como necesite. Esta fase se considera una fase de familiarización con los menús de usuario.
4. Una vez el participante está familiarizado con la tarea, se mostrarán los 20 tipos de menús, uno cada vez y de forma aleatoria:
 - a. Para cada tipo de menú se pide al usuario que seleccione un elemento del menú.
 - b. El usuario abre el menú y dependiendo del tipo se visualizará de una forma u otra.
 - c. El usuario seleccionará el elemento del menú que se le ha pedido.
 - i. En caso de seleccionar el que se le ha pedido, se contabilizará un acierto.
 - ii. En caso de no seleccionar el que se le ha pedido, se contabilizará un fallo.

12.3.7 Instrumentación

Para realizar el experimento, referente a la captura de datos electroencefalográficos, hemos usado el casco neuronal *Diadem*, de la empresa *Bitbrain*. Este casco neuronal de sensores secos dispone de 14 electrodos, como se puede ver en la Figura 12.17, y ofrece 12 canales de lectura de señales. Además, para mejorar estas señales se ha usado un amplificador con una resolución de 24 bits y una frecuencia de muestreo de 256 Hz. Para leer los datos e interpretarlos *a posteriori*, hemos usado *SennsLab* y *SennsMetrics*. Referente a la captura de datos demográficos, hemos usado un cuestionario entregado antes de empezar el experimento, además de un formulario de aceptación de participación en el estudio. Referente a la ejecución del experimento, hemos usado una presentación PowerPoint interactiva para simular el uso de los distintos tipos de menú. Adicionalmente, para comparar las métricas EEG con las reacciones expresivas del participante durante el análisis de datos, se ha usado una cámara web.

12.4 EJECUCIÓN DEL EXPERIMENTO

En el experimento han participado 21 personas seleccionadas por invitación entre conocidos y familiares de los autores de este estudio y usando un sistema de cita previa, de forma que los participantes pudiesen elegir un momento del día que les fuese más conveniente y sin prisas. A continuación, se muestran los distintos

perfiles de participantes, los contextos en los que se han aplicado los distintos tipos de menú y las amenazas a la validez.

12.4.1 Participantes

Los 21 participantes que ejecutaron el experimento son todos residentes en España, de edades entre los 15 y 63 años. Se han recogido también otros datos como el nivel social, estado marital, situación laboral y el nivel educativo de cada participante.

- Por géneros, han sido 14 hombres y 7 mujeres.
- Por edades, hay 12 participantes entre 15 y 39 años y 9 participantes entre 39 y 63 años.
- Por estado marital, la distribución ha sido: 13 casados y 8 solteros.
- Por situación laboral, 12 de los participantes están desempleados, 8 tienen empleo y solamente 1 está inactivo.
- En cuanto a estatus social, clasificado según la Tabla 12.5, definida por la *National Readership Survey (NRS)* [12] y cómo se puede ver en la Tabla 12.5, se ha obtenido la siguiente distribución:
 - 9 participantes pertenecen a la Clase C2.
 - 8 participantes pertenecen a la Clase D.
 - 2 participantes pertenecen a la Clase E.
 - 1 participante pertenece a la Clase B.
 - 1 participante pertenece a la Clase C.

Grado social	Descripción
A	Gerente, administrativo y profesional
B	Gerente intermedio, administrativo y profesional
C1	Supervisor, Administrativo profesional
C2	Trabajadores cualificados
D	Trabajadores semi-cualificados o no cualificados
E	Jubilados estatales, trabajadores ocasionales. Desempleados con únicamente beneficios estatales

Tabla 12.5: Definición de los grados sociales según la NRS [12]

- En cuanto al nivel educativo, la distribución ha sido la siguiente:
 - 7 participantes tienen el nivel de Bachillerato.
 - 5 participantes tienen el nivel de Educación Secundaria Obligatoria.

- 3 participantes tienen el nivel de Formación Profesional de Estudios Específicos.
- 3 participantes tienen el nivel de Primaria.
- 1 participante tiene el nivel de Graduado.
- 1 participante tiene el nivel de Máster.
- 1 participante tiene el nivel de Doctorado.

Dados estos datos y distribuciones, podemos concluir que la selección de los participantes ha sido satisfactoria porque tenemos variedad de perfiles y aleatoriedad. De esta forma, no tendremos datos sesgados por un perfil de participante dominante.

12.4.2 Objeto experimental

Para mostrar los 20 posibles tipos de menú se ha usado una presentación de PowerPoint interactiva simulando la interacción con los menús de aplicaciones convencionales en entornos de escritorio. Para dotar de contexto al uso de estos menús se han seleccionado dos aplicaciones de dominios típicos con menús conteniendo cerca de 20 acciones posibles. Estas aplicaciones son:

- La aplicación *Thunderbird*, un gestor de correo electrónico y en concreto el menú de *Mensajes* como se puede ver en la Figura 12.18 conjuntamente con su simulación en PowerPoint.
- El navegador web de *Microsoft Edge*, y en concreto el menú de *Configuración*, como se puede ver en la Figura 12.19 conjuntamente con su simulación en PowerPoint.

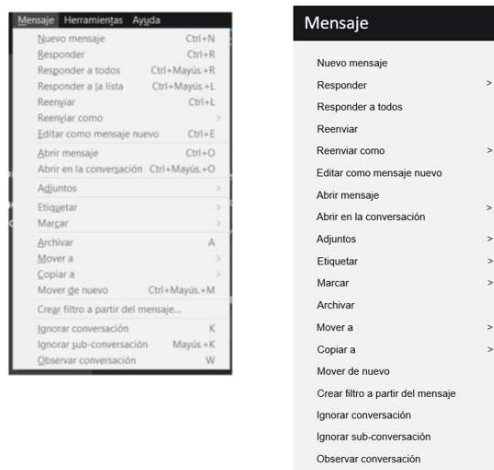


Figura 12.18: El menú de Mensajes de Thunderbird y el menú creado en PowerPoint

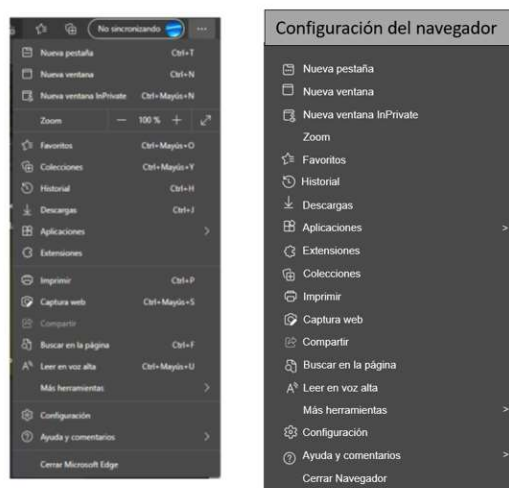


Figura 12.19: El menú de Configuración de Microsoft Edge y el menú creado en PowerPoint

Para cada menú creado en la presentación PowerPoint, se le ha añadido el efecto característico del tipo de menú a analizar (por ejemplo, parpadeo del elemento, subrayado, negrita). Cuando se pulsa en cualquier elemento del menú, se finaliza el análisis para ese menú y se pasa a analizar el siguiente tipo de menú. Si el usuario pulsa fuera de los ítems del menú, no se muestra ningún cambio.

12.4.3 Amenazas a la validez

Se deben considerar ciertos problemas que pueden suponer una amenaza a la validez del experimento.

- Con respecto a la validez interna, no todos los participantes reaccionan igual ante el mismo estímulo, por lo que debemos tener una muestra amplia. Para mitigar esto, usaremos una cantidad de participantes usual en este tipo de experimentos. Según [7], este tipo de estudios tiene una muestra de 19 participantes. Sin embargo, en un futuro se puede ampliar la cantidad de participantes y ampliar los datos del estudio para hacerlos todavía más fiables y completos.
- Con respecto a la validez externa, la mayor amenaza es la generalización de los resultados. Se va a analizar la carga cognitiva creada por distintos tipos de menú sobre 21 participantes distintos, y como se comenta en la validez interna, no todos reaccionan de igual forma. Generalizar los datos obtenidos puede llevar a errores en la conclusión. Para mitigar este factor, se tratará de obtener correlaciones mediante perfiles parecidos.
- Con respecto a la validez de constructo, la principal amenaza proviene del estado anímico del participante. Si al realizar el experimento no se encuentra bien, puede llevar a lecturas de estados con el casco neuronal correspondientes a este malestar y no al uso normal de los distintos tipos de menú. Para mitigar este factor, realizaremos las sesiones cuando mejor considere cada participante. Para ello se sigue un sistema de cita previa en el que el participante decide cuándo y dónde realizar el experimento. De esta forma no se provoca frustración a los participantes por irrumpir en el horario personal o por provocar desplazamientos no deseados.
- Con respecto a la validez de conclusión, las principales amenazas son la validez del agrupamiento de participantes en base a los datos obtenidos. Cabe la posibilidad de que haya algún dato anómalo y cree incongruencias en los grupos. Para mitigar la amenaza, previamente a analizar los datos y crear los grupos de usuarios se eliminarán los datos anómalos.

12.5 ANÁLISIS DE DATOS

Hemos recogido datos de medidas fisiológicas mediante EEG para los 20 tipos de menús de los 21 participantes. Para cada participante hemos obtenido en cada momento los valores de las métricas de carga cognitiva, implicación, memorización y valencia, para cada menú. Adicionalmente, hemos cronometrado el

tiempo desde que los participantes abren el menú hasta que seleccionan un elemento del menú y contabilizado si ha seleccionado el elemento que se le ha pedido o no. A continuación se responde a la pregunta de investigación.

12.5.1 Diferencias entre las métricas para cada menú

Para poder responder a esta pregunta de investigación, necesitamos realizar un análisis descriptivo y la prueba de hipótesis.

12.5.1.1 ANÁLISIS DESCRIPTIVO

Debemos analizar los datos de cada una de las variables dependientes descritas en la Tabla 12.4 que puedan dar respuesta a la pregunta de investigación. Es decir, realizaremos el análisis descriptivo para la carga cognitiva, implicación, memorización y valencia.

12.5.1.1.1 Carga cognitiva

Esta métrica se mide porcentualmente y tomará valores de 0 a 100. Como podemos ver en la Tabla 12.6, los valores obtenidos van desde el 6,59% de carga cognitiva en el menú *Escala de Grises* hasta el 95,97% en el menú *Pulsante*. Sin embargo, la media de todos los menús ronda los 28% al igual que la mediana. En cuanto a la desviación estándar hemos obtenido desviaciones de 9-21%.

Menú	Mínimo	Máximo	Media	Mediana	Desviación Estándar
<i>Área de activación Adaptable</i>	11,735532	57,27623	32,31535	26,8425	13,097996
<i>Brillante</i>	10,085575	65,12927	30,39607	29,15144	13,685651
<i>Cuadro</i>	9,671296	57,3097	26,00841	20,91558	15,56964
<i>Cursiva</i>	14,734321	48,02754	29,32603	27,89808	9,279312
<i>Desapareciendo fuera de contexto</i>	8,54884	47,19037	25,35378	26,18794	10,030357
<i>Dividido sin replicación</i>	9,805958	53,95066	25,77378	21,52243	13,078819
<i>Efímero</i>	9,889247	61,60098	27,09427	24,52382	14,504444
<i>Escala de grises</i>	6,592051	64,44095	32,24637	30,68132	15,445003
<i>Estático</i>	12,008472	53,59909	28,23478	23,89251	12,772637
<i>Hoja</i>	10,845278	50,06741	31,56597	28,74144	10,972955
<i>Negrita</i>	8,95466	56,32536	29,47165	30,37903	11,775788
<i>Parpadeante</i>	10,487956	67,25966	28,87284	27,24552	15,229945

<i>Pulsante</i>	12,500494	95,97756	31,41998	27,0529	21,765814
<i>Resaltado</i>	13,134918	50,47402	29,8352	30,22436	10,387769
<i>Rosa</i>	7,682273	58,20637	25,84976	22,57473	13,603914
<i>Rotativo</i>	12,11844	52,39843	27,98372	28,19865	11,887427
<i>Subrayado</i>	8,759597	58,32509	30,76031	29,69115	12,743563
<i>Tamaño de fuente variable</i>	8,214223	48,28281	26,75922	25,93932	13,846091
<i>Tambaleante</i>	7,78998	59,28485	31,28282	31,23442	14,269379
<i>Temporal</i>	10,469261	55,55197	27,06894	26,20469	10,08315

Tabla 12.6: Análisis descriptivo para la carga cognitiva registrada en cada menú

12.5.1.1.2 Implicación

Esta métrica se mide porcentualmente, es decir, toma valores de 0 a 100. Como podemos ver en la Tabla 12.7 los valores obtenidos van desde el 4,70% de implicación en el menú *Brillante* hasta el 85,32% en el menú *Parpadeante*. Sin embargo, la media de todos los menús ronda los 35% al igual que la mediana. En cuanto a la desviación estándar hemos obtenido desviaciones de 8-17%.

Menú	Mínimo	Máximo	Media	Mediana	Desviación estándar
<i>Área de activación Adaptable</i>	21,157716	51,89738	35,58588	32,28998	8,938768
<i>Brillante</i>	4,705916	82,91048	32,69038	33,87996	16,432457
<i>Cuadro</i>	12,465915	56,18234	38,45181	38,0653	11,334833
<i>Cursiva</i>	18,050013	74,0503	36,03912	33,99027	12,809782
<i>Desapareciendo fuera de contexto</i>	23,450631	59,30916	37,97024	35,21204	9,210745
<i>Dividido sin replicación</i>	15,289459	55,84784	34,75371	37,02327	10,921716
<i>Efímero</i>	18,819923	49,05656	34,6734	36,63376	9,099074
<i>Escala de grises</i>	24,733812	64,01579	36,87396	30,94793	12,291219
<i>Estático</i>	8,500359	54,09753	35,08706	36,5032	11,331032
<i>Hoja</i>	20,126054	62,94311	37,2901	38,76344	11,500415
<i>Negrita</i>	19,729196	63,19679	38,66456	35,23485	12,668005
<i>Parpadeante</i>	20,714005	85,32141	40,46771	35,74884	17,780458
<i>Pulsante</i>	9,727614	50,99411	31,64148	29,25857	10,353378
<i>Resaltado</i>	17,950384	65,10942	36,50434	36,50997	14,445706

<i>Rosa</i>	7,026976	69,03383	36,70837	36,29732	16,682378
<i>Rotativo</i>	24,516425	54,87948	37,84685	36,59092	9,019143
<i>Subrayado</i>	11,860113	61,7546	35,56418	35,65868	13,024379
<i>Tamaño de fuente variable</i>	9,389597	62,32713	33,85347	33,92697	13,571566
<i>Tambaleante</i>	15,839113	49,88178	32,129	34,39453	9,200989
<i>Temporal</i>	12,619598	53,91118	35,64959	36,73718	11,887638

Tabla 12.7: Análisis descriptivo para la implicación registrada en cada menú

12.5.1.1.3 Memorización

Esta métrica se mide porcentualmente, es decir, toma valores de 0 a 100. Como podemos ver en la Tabla 12.8, los valores obtenidos van desde el 3,77% de memorización en el menú *Parpadeante* hasta el 90,30% en el menú *Rosa*. Sin embargo, la media de todos los menús ronda los 25% al igual que la mediana. En cuanto a la desviación estándar hemos obtenido desviaciones de 7-19 %.

Menú	Mínimo	Máximo	Media	Mediana	Desviación estándar
<i>Área de activación Adaptable</i>	9,04746	56,25051	27,82702	28,77117	12,076042
<i>Brillante</i>	8,268952	60,00205	29,45793	24,42963	15,404636
<i>Cuadro</i>	6,084792	57,09333	25,56347	21,63427	15,138883
<i>Cursiva</i>	12,00897	52,96311	26,35469	26,62585	9,958247
<i>Desapareciendo fuera de contexto</i>	6,585114	45,07524	24,18505	22,56673	10,52963
<i>Dividido sin replicación</i>	5,644069	63,60312	23,51993	22,03408	11,672925
<i>Efímero</i>	11,982627	37,45341	23,97762	23,04074	7,801262
<i>Escala de grises</i>	4,804607	42,19185	26,63506	28,74626	9,236136
<i>Estático</i>	7,912212	44,9947	22,97196	22,28923	11,173381
<i>Hoja</i>	10,429301	74,84388	31,31497	30,81974	15,925372
<i>Negrita</i>	4,000788	39,80677	25,89922	27,40624	8,372582
<i>Parpadeante</i>	3,777807	46,87816	21,74701	21,63885	11,413668
<i>Pulsante</i>	5,831989	65,76197	26,54024	24,84127	14,517355
<i>Resaltado</i>	4,222922	57,11553	26,04337	23,14183	13,186086
<i>Rosa</i>	6,752716	90,30423	26,49149	23,81613	19,216676
<i>Rotativo</i>	6,152622	45,06925	24,57214	24,00927	11,929753

<i>Subrayado</i>	8,000352	43,21767	27,77193	29,16314	9,944795
<i>Tamaño de fuente variable</i>	7,219529	64,2873	27,87854	24,75616	15,860696
<i>Tambaleante</i>	5,684515	53,0405	28,87893	27,37847	12,327055
<i>Temporal</i>	13,758046	48,43856	28,35336	26,23172	11,20579

Tabla 12.8: Análisis descriptivo para la memorización registrada en cada menú

12.5.1.1.4 Valencia

Esta métrica se mide con valores dentro del rango de -100 a 100. El valor 100 indica agrado máximo y el valor -100 desagrado máximo. Se puede ver como un valor porcentual de desagrado (de -100 a 0) junto al valor porcentual de agrado (de 100 a 0). Como podemos ver en la Tabla 12.9, los valores obtenidos van desde -67,86 en el menú *Rotativo* hasta 73,17404 en el menú *Dividido sin replicación*. Sin embargo, la media de todos los menús ronda un valor de -0,16. Las medianas obtenidas están entre -11 y 6 en la escala de -100 a 100. Es decir, de media los menús no han sido ni agradables ni desagradables para los participantes, de media han provocado una reacción neutra en los participantes. En cuanto a la desviación estándar hemos obtenido desviaciones de entre 17 y 30.

Menú	Mínimo	Máximo	Media	Mediana	Desviación estándar
<i>Área de activación Adaptable</i>	-43,92719	36,59539	-6,9556729	-3,6359272	21,92717
<i>Brillante</i>	-40,43716	37,49988	-2,6483574	-11,6302325	23,56303
<i>Cuadro</i>	-53,62319	44,5511	-4,9694811	-6,8657606	27,19272
<i>Cursiva</i>	-41,78218	58,88779	0,2988773	-3,6480462	24,14058
<i>Desapareciendo fuera de contexto</i>	-48,85553	65,43911	1,8136379	1,7214565	27,83593
<i>Dividido sin replicación</i>	-55,00542	73,17404	1,5249248	6,3939171	30,56263
<i>Efímero</i>	-46,99454	55,47228	1,9348956	0,6788721	24,91897
<i>Escala de grises</i>	-53,37017	33,24656	-1,0887593	-1,2477364	21,62584
<i>Estático</i>	-56,87662	58,12731	0,5740602	0,9576762	26,75539
<i>Hoja</i>	-36,65959	32,56827	-2,3295205	-10,1411985	17,6323
<i>Negrita</i>	-55,94817	68,72531	3,6087988	6,852613	28,41458
<i>Parpadeante</i>	-53,52941	38,52346	-4,4875026	-5,0510378	25,17944
<i>Pulsante</i>	-49,6063	68,07741	6,6281509	5,6045066	25,85096
<i>Resaltado</i>	-52,31754	60,12035	-0,6860896	-0,3725372	30,85571

<i>Rosa</i>	-54,5	38,18029	-4,2408698	0,1852126	22,4961
<i>Rotativo</i>	-67,86472	58,29297	-1,9065796	-3,8703247	26,62181
<i>Subrayado</i>	-46,82883	35,91155	1,8360576	0,1859882	21,59421
<i>Tamaño de fuente variable</i>	-59,0998	46,29134	2,8046682	3,9225145	24,05646
<i>Tambaleante</i>	-30,94473	36,65461	-2,3553582	0,2927528	19,00239
<i>Temporal</i>	-58,16794	57,18127	3,3116201	6,2748754	27,72379

Tabla 12.9: Análisis descriptivo para la valencia registrada en cada menú

12.5.1.2 VALIDACIÓN DE LA HIPÓTESIS

Pasamos ahora a validar las cuatro hipótesis. Para ello debemos realizar las pruebas de homogeneidad mediante el test de Levene y posteriormente usar el test de Saphiro-Wilk para realizar las pruebas de normalidad. Los resultados de los test se pueden ver en la Tabla 12.10.

Para el test de Levene realizado, obtenemos que todas las métricas son homogéneas. Para la primera hipótesis, referente a las mediciones de la carga cognitiva, tras realizar el test de Saphiro-Wilk, obtuvimos que 6 de los 20 menús no seguían una distribución normal, por tanto, debemos usar el test de Kruskal-Wallis para poder validar la hipótesis. Para la segunda hipótesis, referente a las mediciones de la implicación, tras realizar el test de Saphiro-Wilk, obtuvimos que 3 de los 20 menús no seguían una distribución normal, por tanto, debemos usar el test de Kruskal-Wallis para poder validar la hipótesis. Para la tercera hipótesis, referente a las mediciones de la memorización, tras realizar el test de Saphiro-Wilk, obtuvimos que 5 de los 20 menús no seguían una distribución normal, por tanto, debemos usar el test de Kruskal-Wallis para poder validar la hipótesis. Para la cuarta hipótesis, referente a las mediciones de la valencia, tras realizar el test de Saphiro-Wilk, obtuvimos que los 20 menús seguían una distribución normal, por tanto, debemos usar el test de One-Way ANOVA para poder validar la hipótesis.

Métrica	Distribución normal	Homogénea	Método	p-valor	Diferencias entre menús
Carga cognitiva	No	Sí	Kruskal-Wallis	0,70758	No
Implicación	No	Sí	Kruskal-Wallis	0,90119	No
Memorización	No	Sí	Kruskal-Wallis	0,73043	No
Valencia	Sí	Sí	One-Way ANOVA	0,99158	No

Tabla 12.10: Validación de la hipótesis

Como podemos ver en la Tabla 12.10, tras aplicar One-Way ANOVA a las mediciones de valencia y Kruskal-Wallis al resto de métricas, no se han detectado

diferencias significativas entre los menús para ninguna de las cuatro métricas obtenidas con el EEG.

12.5.2 Discusión

Las cuatro hipótesis nulas se confirman y se descartan las hipótesis alternativas. Como se ha comprobado en la Tabla 12.10, no ha habido diferencias significativas entre los distintos tipos de menú para ninguna de las métricas. Dado el amplio espectro de posibilidades, con las múltiples propiedades de los menús, era esperable que se encontrasen diferencias significativas entre los tipos de menús planteados. Tras los resultados y los análisis hemos detectado algunas limitaciones que se deberán tener en cuenta para futuros experimentos para verificar que, efectivamente, no hay diferencias entre los menús o si hubiera diferencias qué factores lo determinan.

En primer lugar, debemos tener en cuenta el perfil de los participantes. Puede que el poco uso de un dispositivo informático influya en la percepción de los distintos tipos de menús. Para futuros experimentos se deben considerar participantes con un nivel mínimo de conocimientos y también con conocimientos avanzados.

En segundo lugar, se puede considerar realizar el estudio de una forma más dinámica. Detectamos que la sensación de falta de contexto puede favorecer a una mala percepción de los menús. Para mitigar este efecto, se usará una implementación concreta, o un prototipo web, que use estos mismos tipos de menú y que los participantes tengan que realizar tareas concretas y decidan qué acciones realizar para lo que tendrán que acceder a los menús correspondientes.

Otro factor a tener en cuenta es la complejidad de los menús. Típicamente, los menús clasifican acciones en apartados y en sub-apartados (sub-menús). En este experimento solamente se ha usado un menú normal sin sub-apartados. Creemos que usar estructuras más complejas ayudaría al impacto de los distintos tipos de menú de usuario analizados y además presentaría un contexto más realista.

En cuanto al diseño del experimento, creemos también que se debe dejar un tiempo de “reposo” entre el uso de varios menús. Basándonos en las fases de calibración, debería haber una fase de “Lavado” para eliminar reacciones de un menú que se producen durante la ejecución de los siguientes. En este experimento, se han encadenado los distintos tipos de menú, haciendo que, en un lapso muy corto de tiempo, los participantes usen varios menús, dificultado potencialmente la percepción de cada uno de ellos. Añadir un tiempo de reposo entre cada menú puede

hacer que las señales electroencefalográficas obtenidas por un menú no afecten al siguiente.

Por último, debemos tener en cuenta que los menús son el medio de transporte para encadenar acciones y realizar una tarea concreta. En este experimento la tarea concreta ha sido usar el propio menú. Se puede cambiar la dinámica del ejercicio, es decir, no pedir a los participantes que elijan un elemento concreto del menú, sino que realicen una tarea que, para ser realizada, se tenga que buscar y usar un elemento de menú. Con este cambio conseguiremos que el usuario no sienta que solo tiene la tarea de buscar y seleccionar elementos dentro de un mismo menú sino dentro de una aplicación.

12.6 CONCLUSIONES

En este capítulo se ha presentado un experimento para obtener información biométrica de varios usuarios con el objetivo de conseguir una base de conocimiento sobre qué tipos de menú favorecen más la experiencia de usuario. Se han monitorizado métricas como la carga cognitiva, la implicación, la memorización y la valencia durante el uso de varios tipos de menú. De la selección de los menús presentado podemos concluir que a los usuarios les ofrece una mejor experiencia de usuario usar menús que introducen la variable temporal en ellos. Es decir, menús como el *Efímero*, *Desapareciendo fuera del contexto*, *Temporal*. Además, también hemos detectado que los menús con estructura atípica, como pueda ser el menú de *Hoja*, suponen una carga cognitiva muy elevada, llevando a una mala UX. Como se ha comentado durante el capítulo, ofrecer una buena UX es un factor importante en la certificación de calidad de un software y desde el punto de vista de los diseñadores software la información obtenida en este experimento, aunque preliminar, puede ser muy valiosa para entender ciertas tendencias en las reacciones fisiológicas de los usuarios.

Por un lado, vemos que el uso de datos biométricos nos permite cuantificar la UX de cada uno de los menús individualmente. Esto nos permite centrarnos en la elección del mejor elemento de interfaz de usuario en lugar de abstraerse a las categorías o propiedades. Los datos biométricos que hemos obtenido pueden ser expandidos usando otro tipo de mediciones fisiológicas como las que se comentan en la Tabla 12.2, de hecho, como trabajo futuro buscamos replicar este experimento sincronizando dispositivos de seguimiento ocular y de respuesta galvánica de la piel (GSR) con el casco neuronal para obtener más métricas UX y datos más completos.

Por otro lado, la limitación de este estudio se refleja en la cantidad y variedad de elementos de interfaces de usuario comparados. Creemos que el dominio de elementos de interfaz de usuario es muy amplio y nos hemos centrado en un área

específica, los tipos de menús. Los datos obtenidos por tanto pueden ser potencialmente de utilidad, sin embargo, si usamos otros tipos de elementos de interfaz de usuario (por ejemplo, ventanas, campos de edición, botones de acción) podremos añadir más valor al conocimiento obtenido para el diseño de sistemas software. En esta línea, y como trabajo futuro, queremos expandir el estudio para comparar, por ejemplo, distintos tipos de selectores, tipos entrada de datos para el sistema, entre muchas otras opciones. A mayor espacio dentro del dominio de las interfaces de usuario que cubramos mejor conocimiento estaremos aportando al diseño de software desde el punto de vista de aseguramiento de calidad ofreciendo una buena UX.

Por lo tanto, la principal contribución de este trabajo es la creación de una base de conocimiento para diseñadores de software obtenida mediante datos biométricos de la interacción de usuarios con menús de interfaces de usuario. Esta base de conocimiento está basada en evidencias empíricas sobre cómo cada elemento del interfaz afecta a la UX y que podrá ser expandida con replicaciones y con otros experimentos que abarquen otros elementos del interfaz de usuario.

12.7 AGRADECIMIENTOS

Este trabajo ha sido financiado por los proyectos Adapt@Cloud (TIN2017-84550-R) y UX-Adapt (AICO/2020/113).

12.8 REFERENCIAS

- [1] Allanwood, G., & Beare, P. (2014). Basics interactive design: User experience design: Creating designs users really love. A&C Black.
- [2] Basili, V. R., & Rombach, H. D. (1988). The TAME project: Towards improvement-oriented software environments. IEEE Transactions on software engineering, 14(6), 758-773. <https://doi.org/10.1109/32.6156>
- [3] Bergstrom, J. R., & Schall, A. (2014). Eye tracking in user experience design. Elsevier.
- [4] Bitbrain, retrieved from <https://www.bitbrain.com/>
- [5] Cooper, A., Reimann, R., & Cronin, D. (2007). About face 3: the essentials of interaction design. John Wiley & Sons.

- [6] Díaz-Oreiro, I., López, G., Quesada, L., & Guerrero, L. A. (2019). Standardized questionnaires for user experience evaluation: A systematic literature review. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 31, No. 1, p. 14). <https://doi.org/10.3390/proceedings2019031014>
- [7] Iriarte, A. A., Erle, G. L., & Etxabe, M. M. (2021). Evaluación de la experiencia de usuario con monitorización fisiológica: revisión sistemática de la literatura. DYNA New Technologies, 8(1), 20-p.
- [8] ISO 9241-210, «9241-210.(2019)», Ergonomics of Human-System Interaction—Part 210: Human-Centred Design for Interactive Systems, 2019, retrieved from <https://www.iso.org/standard/77520.html>
- [9] ISO/IEC 9241-10, «9241-10.(1996)», Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs), Part 10: Dialogue Principles, 1996, retrieved from <https://www.iso.org/standard/16882.html>
- [10] Maia, C. L. B., & Furtado, E. S. (2016, October). A study about psychophysiological measures in user experience monitoring and evaluation. In Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems (pp. 1-9). <https://doi.org/10.1145/3033701.3033708>
- [11] Mandryk, R. L., Inkpen, K. M., & Calvert, T. W. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. Behaviour & information technology, 25(2), 141-158. <https://doi.org/10.1080/01449290500331156>
- [12] National Readership Survey, Social grade, retrieved from <http://www.nrs.co.uk/>
- [13] Neumann, D. L., & Westbury, H. R. (2011). The psychophysiological measurement of empathy. Psychology of empathy, 119-142.
- [14] Nielsen, J. (1989, September). Usability engineering at a discount. In Proceedings of the third international conference on human-computer interaction on Designing and using human-computer interfaces and knowledge based systems (2nd ed.) (pp. 394-401).

- [15] Radüntz, T. (2018). Signal quality evaluation of emerging EEG devices. *Frontiers in physiology*, 9, 98. <https://doi.org/10.3389/fphys.2018.00098>
- [16] Van Camp, M., De Boeck, M., Verwulgen, S., & De Bruyne, G. (2018, July). EEG technology for UX evaluation: A multisensory perspective. In *International Conference on Applied Human Factors and Ergonomics* (pp. 337-343). Springer, Cham. https://doi.org/10.1007/978-3-319-94866-9_34
- [17] Vanderdonckt, J., Bouzit, S., Calvary, G., & Chêne, D. (2019). Exploring a design space of graphical adaptive menus: Normal vs. Small screens. *ACM Transactions on Interactive Intelligent Systems*, 10(1), 1–40. <https://doi.org/10.1145/3237190>
- [18] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). Empirical Strategies. In *Experimentation in Software Engineering* (pp. 9- 36). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-29044-2_2
- [19] Zaki, T., Islam, M. N., & Zaki, T. (2021). Neurological and physiological measures to evaluate the usability and user-experience (UX) of information systems: A systematic literature review. *Computer Science Review*, 40, 100375. <https://doi.org/10.1016/J.COSREV.2021.100375>