



INAOE

Análisis de Usabilidad de Entornos Virtuales incorporando Surrogados de Actividad Cognitiva.

Por:

Arlem Aleida Castillo Avila

Tesis sometida como requisito parcial para obtener el grado de **Maestro en Ciencias en el Área de Ciencias Computacionales** en el Instituto Nacional de Astrofísica, Óptica y Electrónica

Supervisada por:

Dr. Felipe Orihuela-Espina, INAOE.

Dr. Luis Enrique Sucar Succar, INAOE.

Dr. Alberto L. Morán y Solares, UABC.

Agosto 2016

© Coordinación de Ciencias Computacionales
INAOE

Luis Enrique Erro 1
Sta. Ma. Tonantzintla,
72840, Puebla, México.



Índice

Lista de Acrónimos	VII
1 Introducción	2
1.1 Motivación y justificación	4
1.2 Problemática	5
1.3 Preguntas de investigación	7
1.4 Hipótesis	7
1.5 Objetivo General	7
1.6 Objetivos Específicos	8
1.7 Contribuciones	8
1.8 Resumen del capítulo	9
1.9 Estructura del documento	10
2 Marco Teórico	11
2.1 Experiencia de Usuario y Usabilidad	11
2.2 Entornos Virtuales Serios	16
2.3 Anatomía del ojo y del sistema oculomotor	17
2.4 Seguimiento del ojo y pupilometría	17
2.5 Neuroergonomía y Carga cognitiva	18
2.6 Modelos Gráficos Probabilistas	19
2.6.1 Redes Bayesianas	20
2.6.1.1 Inferencia	21

2.6.1.2	Aprendizaje	22
2.7	Distribuciones Multivariadas	24
2.7.1	Distribución de probabilidad conjunta	25
3	Trabajo Relacionado	26
3.1	Evaluación de usabilidad como parte de la experiencia del usuario	26
3.1.1	Métodos de evaluación de Usabilidad	27
3.1.2	Interferencia cognitiva en la evaluación de usabilidad	28
3.1.3	Evaluación de entornos virtuales	29
3.2	Medición de actividad cognitiva por métricas oculares	31
3.2.1	Seguimiento del ojo en IHC	32
4	Resultados	35
4.1	Implementación del algoritmo de seguimiento del ojo CECL+LSH	35
4.2	Carga cognitiva y entrenamiento virtual	36
4.3	Estudio de usabilidad de los entornos virtuales de entrenamiento	40
4.3.1	Protocolo de la sesión de entrenamiento	40
4.3.2	Cuestionario NASA-TLX	41
4.3.2.1	Obtención de pesos o importancia iniciales	42
4.3.2.2	Valoración	42
4.3.3	Cuestionario de usabilidad SUS	43
4.3.4	Cuestionario de perfil de usuario	44
4.4	Modelo de Red Bayesiana jerárquica del análisis de usabilidad	48
4.5	Aprendizaje automático de la estructura de la red Bayesiana	50
5	Conclusiones	53
5.1	Trabajo Futuro	55

Índice de Tablas

2.1	Principios de usabilidad en las pautas de diseño de software	13
2.2	Ciclo de vida de la Ingeniería de la Usabilidad de Mayhew	13
3.1	Clasificación de los Métodos de evaluación de usabilidad	29
4.1	Resultados de desempeño de CECL+LSH en comparación con la literatura.	36
4.2	Niveles de carga cognitiva de la primera y tercera sesión de entrenamiento cognitivo.	37
4.3	Niveles de carga cognitiva de la primera y tercera sesión de entrenamiento motor.	38
4.4	Puntajes del entrenamiento cognitivo.	38
4.5	Puntajes del entrenamiento motor.	39
4.6	Cuestionario de perfil general del Usuario	44
4.7	Resumen de los resultados del cuestionario SUS para todas las interfaces	44

Índice de Figuras

2.1	Modelo de Nielsen	12
2.2	Marco de definición de la norma ISO/IEC 9241	12
2.3	Componentes del proceso de UX	14
2.4	Relación entre los componentes del proceso de evaluación de usabilidad.	16
2.5	Ejemplo de red bayesiana	20
3.1	Relación entre el desempeño y la carga cognitiva	31
4.1	Comparación entre el desempeño por puntaje y por nivel de carga cognitiva del entrenamiento cognitivo	39
4.2	Comparación entre el desempeño por puntaje y por nivel de carga cognitiva del entrenamiento motor	39
4.3	Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento cognitivo (Interfaz 1 del nivel más bajo de abstracción)	45
4.4	Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento cognitivo (Interfaz 2 del nivel más bajo de abstracción)	46
4.5	Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento cognitivo (Interfaz del nivel más alto de abstracción)	46
4.6	Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento motor (Interfaz 1 del nivel más bajo de abstracción)	47

4.7	Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento motor (Interfaz 2 del nivel más bajo de abstracción)	47
4.8	Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento motor (Interfaz del nivel más alto de abstracción)	48
4.9	Estructura del perfil de usuario generada sobre el conjunto de datos con sobremuestreo (izquierda) y sobre el conjunto original (derecha)	49
4.10	Estructura del perfil del sistema generada sobre el conjunto de datos con sobremuestreo (izquierda) y sobre el conjunto original (derecha)	49
4.11	Estructura del modelo general sobre el conjunto de datos con sobremuestreo (izquierda) y sobre el conjunto original (derecha)	50
4.12	Estructura del conjunto original de datos WDBC por el algoritmo PC .	51
4.13	Estructura del conjunto reducido a 25 datos WDBC por el algoritmo PC	52
4.14	Estructura del conjunto sintético de datos WDBC por el algoritmo PC	52

AGPL: Affero General Public License.

Resumen

La evaluación de usabilidad es un componente clave en el análisis de la experiencia de usuario dentro del estudio de la Interacción Humano-Computadora. Los entornos virtuales tienen un análisis de usabilidad complejo con respecto a los sistemas tradicionales debido a varios factores entre los que se encuentra el estado cognitivo del usuario durante el momento interactivo. Este trabajo presenta la integración del nivel de carga cognitiva al proceso de evaluación de usabilidad a través de la obtención de un surrogado medido por técnicas de seguimiento ocular. El resultado final es una red bayesiana jerárquica que modela las relaciones de dependencia probabilística entre las características del usuario, las características del entorno virtual, el nivel de carga cognitiva y el resultado de una evaluación de usabilidad tradicional. Este modelo es capaz de lidiar con la limitación de tener un conjunto de datos reducido para los algoritmos de aprendizaje automático de la estructura de la red, aplicando una técnica de sobremuestreo por ajuste a distribuciones multivariadas.

Capítulo 1

Introducción

Aunque no existe un consenso sobre el alcance del concepto de experiencia de usuario (UX) [Lallemand et al., 2015], éste se puede entender como el conjunto de factores y elementos relativos a la interacción que una persona tiene con un entorno o dispositivo concretos, cuyo resultado es la generación de una percepción positiva o negativa del mismo. La evaluación de usabilidad es un paso importante en el desarrollo de sistemas interactivos, especialmente en aplicaciones de innovación como los entornos de realidad virtual, en los que se pretende que la interacción con el sistema proporcione una experiencia que haga sentir al usuario que se encuentra en una situación real. La evaluación de usabilidad de un sistema interactivo, es un componente clave en el análisis de la experiencia de usuario (UX).

Los entornos virtuales (VE, por sus siglas en inglés) son sistemas que permiten a una persona interactuar en un ambiente simulado y mostrar los efectos en tiempo real. Los VE tienen retos de evaluación de usabilidad únicos debido a su naturaleza. En los sistemas tradicionales orientados a la productividad, factores como la frustración y el estrés son considerados elementos negativos [Mendoza and Novick, 2005]. En un VE, y específicamente en un videojuego, pueden ser elementos positivos que aumenten el nivel de compromiso del usuario con el sistema y lo alienten a completar las tareas [Sánchez et al., 2008]. En el desarrollo de videojuegos se pretende llevar el concepto de “usabilidad” (*usability*) a un concepto más específico para medir la calidad de software llamado “jugabilidad” (*gameability*), en el que se toman en cuenta atributos como la motivación y la emoción como parte del conjunto que caracteriza al sistema interactivo. Esta extensión de la usabilidad tomando en cuenta el contexto del usuario es parte del

proceso de evaluación más cercano a la UX [Sánchez and Vela, 2014].

Para el objetivo de evaluar la usabilidad de VEs dentro del marco de la UX, los métodos tradicionales de evaluación suelen ser insuficientes [Moreno-Ger et al., 2012]. Se han desarrollado métodos de evaluación de usabilidad específicos para los VE basados en el uso de técnicas tradicionales, la mayoría en forma de cuestionarios de evaluación de usabilidad [Kalawsky, 1999][Gil-Gómez et al., 2013]. Aunque los cuestionarios son un método de indagación que provee información directa sobre la experiencia del usuario tienen varias limitaciones [Holzinger, 2005]. Aspectos como el número de preguntas, el tiempo de evaluación, etcétera, pueden desencadenar una interacción negativa con los participantes en el estudio de usabilidad y por lo tanto una evaluación sesgada.

Por otra parte, la interacción entre el usuario y un VE genera reacciones cognitivas y de comportamiento relacionadas a las necesidades y metas del individuo durante la realización de las tareas del VE. Se ha sugerido que ésto juega un papel importante en el desempeño de un usuario en un sistema computacional, lo que incluye a los VE [Norman, 2004] [Peter et al., 2007]. Conocer la actividad cognitiva durante el uso del sistema puede ayudar a determinar el nivel de satisfacción de un usuario [Picard, 2000], y en general, provee una gran cantidad de información relacionada a la UX.

La actividad cognitiva es el proceso mental a través del cual una persona recibe, integra, relaciona y modifica la información circundante, a través de los órganos sensoriales con el propósito de comprender la realidad [Ardila, 1995]. Esta actividad puede medirse con el uso de señales fisiológicas como indicadores de evaluación. El método más directo es con el uso de neuroimágenes funcionales como los son la electroencefalografía (EEG) o resonancia magnética funcional (fMRI), entre otras. Las desventajas del uso de estos métodos en la evaluación de UX es que se introduce un sesgo al requerir que la persona porte un dispositivo durante el episodio de interacción. El seguimiento del ojo es un método no intrusivo que evalúa el comportamiento ocular y a partir del cual se puede obtener información sobre el estado cognitivo, y específicamente sobre el nivel de carga cognitiva de una persona [Marshall, 2002]. Otra forma sencilla de obtener una aproximación al nivel de carga cognitiva es a través de cuestionarios. Un ejemplo es el

cuestionario NASA-TLX [Hart and Staveland, 1988], pero su evaluación no ocurre de forma concurrente con la ejecución de la tarea.

La usabilidad no es un concepto absoluto sino relativo a la naturaleza de la aplicación y dependiente del usuario. No obstante, incluso el mismo usuario podría evaluar la usabilidad de un sistema de forma substancialmente diferente si su estado cognitivo varía. En este sentido, la correcta interpretación de un resultado de una herramienta de usabilidad se ve favorecida por el complemento que supone la información adicional sobre dicho estado cognitivo.

Con la finalidad de fortalecer la evaluación de usabilidad de los VE, este trabajo busca modelar los componentes del proceso de evaluación integrando el nivel de carga cognitiva del usuario durante la interacción a través de la obtención de un surrogado de actividad cognitiva.

1.1 Motivación y justificación

En la actualidad, muchos sistemas computacionales funcionales no son utilizados por los usuarios finales debido a que carecen de elementos interactivos que alienten a la persona a completar tareas. Sin embargo, la evaluación de UX es compleja debido a que es dependiente del contexto del usuario e implica un gran número de variables a analizar. La usabilidad es un componente en la evaluación de la UX. En el caso de los VE, la evaluación de usabilidad presenta retos que no son abordados en métodos tradicionales. En la mayoría de los casos, el aplicar técnicas tradicionales para evaluar juegos serios lleva a estudios limitados debido a los requerimientos característicos de dichos sistemas. Realizar un estudio de usabilidad que se ajuste a las características particulares de un VE puede resultar costoso. Las pruebas con un número representativo de los usuarios finales puede conducir a estudios longitudinales que tomen mucho tiempo y requieran una gran cantidad de recursos financieros y humanos. Es conveniente desarrollar métodos de evaluación que reduzcan los costos de la evaluación de usabilidad. Una alternativa es

mediante el análisis de las relaciones que existen entre los componentes de la interacción (usuario y sistema) y el resultado de usabilidad, integrando el nivel de carga cognitiva que tiene el usuario como una evidencia contextual de la UX. La obtención de indicadores de carga cognitiva por técnicas de seguimiento ocular es conveniente debido a que no resultan intrusivos ni obstrusivos en el desempeño del usuario al interactuar con un VE. Más aún, el utilizar una cámara web de alta resolución reduce el costo de la necesidad de dispositivos especiales para este fin.

Una alternativa para modelar los componentes de la evaluación de usabilidad son los modelos gráficos probabilistas que representan las relaciones de dependencia entre variables mediante el uso de grafos. Esto se debe a la propiedad de dichos modelos para manejar la incertidumbre. La evaluación de usabilidad es un problema en el que la información está incompleta debido a la cantidad de variables que influyen en el problema. Otras ventajas de los modelos gráficos probabilistas son que permiten aprender sobre relaciones de dependencia y causalidad y permiten combinar conocimiento con datos. Debido a las dificultades de evaluación de VE, un estudio de usabilidad puede terminar en la adquisición de conjuntos de datos relativamente pequeños que limitan la automatización del modelado del flujo del proceso. Esto implica un problema para los algoritmos de aprendizaje automático en los modelos gráficos probabilistas, tanto globales como locales, ya que se requiere de una gran cantidad de datos para establecer las dependencias entre variables debido a que se basan en el resultado de pruebas estadísticas. Este trabajo explora los resultados de aplicar una técnica de sobremuestreo basada en el ajuste de datos a distribuciones multivariadas como una solución al problema del aprendizaje automático con bases de datos pequeñas.

1.2 Problemática

Debido a las características únicas de interacción de los VE, el proceso del análisis de UX en dichos entornos presenta dificultades en muchos aspectos.

Para mejorar el proceso de evaluación de UX a partir de la evaluación de usabilidad de los

VE es necesario complementar dicha evaluación con indicadores de la interacción entre un usuario y un VE. Cuando una persona está haciendo uso de un VE se desencadena una serie de procesos cognitivos directamente ligados a esa actividad. La carga cognitiva de la persona aumenta o disminuye en función de la dificultad que le represente la ejecución de dicha actividad. El nivel de carga cognitiva es un indicador que se puede utilizar para evaluar la interacción usuario - VE. La idea no es necesariamente disminuir el nivel de carga cognitiva, sino tener un equilibrio que mejore el desempeño del usuario para completar las tareas dentro del VE. Es necesario considerar otras características del perfil del usuario además del nivel de carga cognitiva para seleccionar una estrategia que lleve a la mejora del desempeño ya que la individualidad del usuario define su comportamiento en la interacción.

La integración del nivel de carga cognitiva a la evaluación de usabilidad puede hacerse con el uso de indicadores subjetivos. Sin embargo, para reducir el sesgo introducido por estos medios, se requiere de una estimación objetiva que es posible con la obtención de un sustituto (*surrogate*), que es la señal fisiológica acompañante del proceso cognitivo. La propuesta es el uso de indicadores basados en el seguimiento automatizado de actividad ocular a través del uso de una cámara web convencional como un método no intrusivo de alta disponibilidad y bajo costo.

Por otra parte, el proceso de evaluación de usabilidad, desde el enfoque de la UX, considera incluir un número considerable de variables propias del perfil del usuario, así como variables que caracterizan al entorno virtual, esto ligado al puntaje de usabilidad final. Para poder aprender las relaciones de dependencia entre las variables, se requiere un gran número de datos en atención a los requerimientos de los algoritmos de aprendizaje, ya que se basan en pruebas estadísticas para su objetivo. Al realizar una evaluación de usabilidad y de UX, muchas veces no se consiguen conjuntos de datos de grandes dimensiones debido a muchos factores. Por ejemplo, el hecho de que es difícil conseguir voluntarios para probar sistemas que requieran más de una sesión de prueba, como sistemas de entrenamiento; o en el caso de los entornos de rehabilitación virtual, donde es difícil conseguir una muestra representativa de los usuarios finales.

El problema de tener un bajo número de observaciones disponibles recae en la automatización del aprendizaje de modelos que puedan complementar el proceso de evaluación de usabilidad. Los algoritmos de aprendizaje que se utilizan en la generación de modelos gráficos probabilistas están basados en pruebas estadísticas que requieren una gran cantidad de datos. En este sentido, la propuesta es disponer de un gran número de datos representativos del conjunto de datos original mediante el ajuste de éstos a una distribución multivariada.

1.3 Preguntas de investigación

Ante la problemática descrita surgen las siguientes preguntas de investigación:

- ¿Se puede integrar el nivel de carga cognitiva a la evaluación de usabilidad con un surrogado basado en la actividad ocular medida por una cámara web?
- ¿Es posible modelar automáticamente el proceso de evaluación de usabilidad mediante un PGM cuando se tiene un bajo número de observaciones disponibles?

1.4 Hipótesis

Si se tiene conocimiento de las características contextuales del usuario, incluyendo el nivel de carga cognitiva durante el uso de un entorno virtual, y de las características de diseño de interacción del sistema, entonces es posible ofrecer una interpretación del resultado de una evaluación de usabilidad que esté menos influenciada por la subjetividad inherente al cuestionario.

1.5 Objetivo General

Establecer y validar un modelo computacional que muestre y explique la relación entre el nivel de carga cognitiva y el resultado de una evaluación de usabilidad; mediante la

integración de surrogados de actividad cognitiva, utilizando un método no obtrusivo, específicamente, el seguimiento del ojo en el espectro visible; para informar al diseñador del sistema y aliviar el sesgo que introduce la carga que un usuario tiene al interactuar con un VE.

1.6 Objetivos Específicos

Los objetivos específicos de la investigación son:

- Determinar el nivel de actividad cognitiva de un grupo de usuarios mediante el análisis de dilatación de la pupila utilizando un sistema de seguimiento remoto basado en una cámara de video.
- Establecer y validar un modelo de red bayesiana jerárquica que muestre las relaciones de dependencia entre las características del usuario que incluya la actividad cognitiva, las características del sistema y el resultado de una evaluación de usabilidad tradicional.
- Analizar cómo la reorganización de los elementos de interacción de un VE influyen en su apreciación de usabilidad.

1.7 Contribuciones

- La propuesta de un método biométrico de evaluación de usabilidad de entornos virtuales que incorpore el nivel de carga cognitiva presente en la interacción con el sistema.
- Un modelo computacional que explique las relaciones de dependencia entre el perfil de usuario, el perfil del sistema y el puntaje de una evaluación de usabilidad.

- Método para facilitar el aprendizaje automático de la estructura de redes bayesianas jerárquicas en conjuntos de datos pequeños mediante la explotación del sobremuestreo.
- Un algoritmo de extracción de indicadores pupilométricas a partir de una cámara web de bajo costo.
- Un análisis de usabilidad de la distribución de elementos interactivos en un VE.

1.8 Resumen del capítulo

La evaluación de usabilidad de un sistema interactivo es clave en el análisis de UX. Los VE tienen características particulares de interacción que hacen que el proceso de evaluación de usabilidad sea más complejo. Los métodos tradicionales para dicha evaluación suelen ser insuficientes. El uso de cuestionarios como única herramienta puede conducir a una evaluación sesgada. Por otra parte, durante la interacción del usuario y el VE se desencadena una serie de procesos cognitivos que pueden ser medidos por diversas técnicas, como el seguimiento del ojo, que tiene la ventaja de ser un método no intrusivo y no obstrusivo, además de reducir costos si se realiza con una cámara web convencional. Se propone construir un modelo computacional, específicamente un modelo de red bayesiana jerárquica, para representar el proceso de evaluación de usabilidad desde el enfoque de la UX, integrando el nivel de carga cognitiva que un usuario tiene como un indicador objetivo que reduzca el sesgo de la subjetividad del proceso evaluativo y que lidie con el problema de tener un bajo número de observaciones disponibles mediante la aplicación de una técnica de sobremuestreo en el proceso de aprendizaje automático de la estructura de la red.

1.9 Estructura del documento

La organización del documento es la siguiente. El capítulo 2 expone la teoría que fundamenta esta investigación, desde los conceptos de UX y usabilidad, los métodos de seguimiento ocular y la obtención del nivel de carga cognitiva, así como los conceptos al rededor de los modelos gráficos probabilistas y las técnicas de sobremuestreo utilizadas en este trabajo. El capítulo 3 da un panorama general de los trabajos más trascendentes relacionados con esta investigación y plantea la importancia del enfoque desde el que se abordan los problemas que se integran en este trabajo. El capítulo 4 describe la metodología propuesta en esta investigación, haciendo énfasis en las contribuciones que se derivan de la realización de este trabajo. El capítulo 5 expone y compara los resultados experimentales obtenidos a lo largo de este trabajo y finalmente en el capítulo 6 se dan las conclusiones finales y trabajo futuro que se desprende de esta investigación.

Capítulo 2

Marco Teórico

En este capítulo se revisan los conceptos claves para el desarrollo de esta investigación comenzando por definir el concepto de Usabilidad dentro del marco de la Experiencia de Usuario y dando algunas características particulares de los entornos virtuales serios. Posteriormente se explica el concepto de carga cognitiva y la relación que existe con los indicadores oculares. Finalmente la introducción a los Modelos Gráficos Probabilistas y específicamente a las redes bayesianas y los algoritmos de aprendizaje e inferencia, que son las bases de esta tesis.

2.1 Experiencia de Usuario y Usabilidad

El término *usabilidad* es un anglicismo que significa *facilidad de uso* [Bevan, 1995]. La ISO/IEC 9126 [ISO/IEC, 2001] define usabilidad como *la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario, en condiciones específicas de uso*. Se han construido un gran número de definiciones para el término, normalmente a través de la enumeración de los diferentes atributos o factores mediante los que puede ser evaluada, dependiendo finalmente cada definición del enfoque con el que pretende ser medida [Folmer and Bosch, 2004]. El modelo de Nielsen, que se aprecia en la figura 2.1, identifica 5 componentes: *facilidad de aprendizaje, eficiencia, retención, manejo de errores y satisfacción del usuario* [Nielsen, 1994]. De igual forma, la ISO/IEC 9241 [Bevan, 1998], cuyo marco de definición se muestra en la figura 2.2, asocia el término a 3 componentes: *eficiencia, eficacia y satisfacción*.

La usabilidad puede ser vista como un atributo de calidad. Desde este enfoque existen

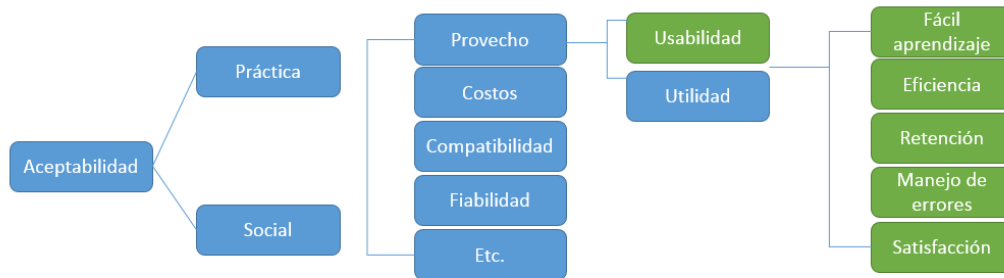


Figura 2.1: La figura muestra el lugar del concepto de usabilidad dentro del modelo general de aceptabilidad de software definido por Nielsen y los atributos que la definen. Se observa que está directamente asociado al provecho del producto de software que es uno de los aspectos determinantes de la aceptabilidad práctica y social del producto.

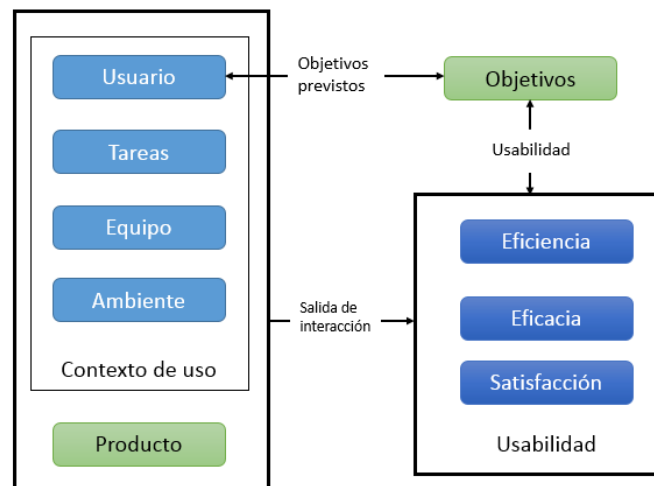


Figura 2.2: La figura muestra la interacción entre los componentes del marco de usabilidad de la norma ISO/IEC 9241. Las medidas de usabilidad se consideran en relación a los objetivos previstos por las especificaciones contextuales (contexto de uso) del producto. Los componentes de usabilidad son la eficiencia, eficacia y satisfacción. El contexto de uso incluye al usuario, las tareas el equipo y el ambiente.

Principio	Definición
Consistencia	Soluciones de la interfaz con los mismos principios sobre un conjunto de situaciones individuales.
Control interno	Sensación de participación y manipulación directa del usuario
Presentación visual	Información explícita y oportuna para el usuario en cada nivel de interacción
Manejo de errores	Gestión del error que contiene advertencias, información de acciones y estrategias de prevención y detección de errores
Reducción de trabajo	Facilitar el aprendizaje de las tareas que el usuario realiza.
Guía y apoyo	Técnicas con la intención de que el usuario perciba la existencia de soporte ante cualquier fallo.

Tabla 2.1: Principios de usabilidad en las pautas de diseño de software

Fase	Etapas	
Análisis de especificaciones	<ol style="list-style-type: none"> 1. Perfilado de usuario 2. Análisis de tareas 3. Definición de metas de usabilidad 4. Capacidades y limitaciones 	
Diseño	Nivel 1	<ol style="list-style-type: none"> 1. Trabajo de reingeniería 2. Modelado de diseño 3. Modelado de maquetas 4. Evaluación del modelado
	Nivel 2	<ol style="list-style-type: none"> 5. Definición de estándares de diseño 6. Establecimiento de estándares de prototipado 7. Evaluación de estándares
	Nivel 3	<ol style="list-style-type: none"> 8. Interfaz de diseño de usuario 9. Evaluación de la interfaz
Implementación	Evaluación postmortem	

Tabla 2.2: Ciclo de vida de la Ingeniería de la Usabilidad de Mayhew

principios que definen el concepto orientado a cumplir los objetivos del diseño de un producto de software [Keinonen, 2003]. La tabla 2.1 resume los principios a los que frecuentemente se hace referencia en las pautas de diseño.

Desde otro enfoque, la usabilidad se considera parte del proceso del desarrollo de un producto. Se suele hablar entonces de *ingeniería de usabilidad* y se refiere a un conjunto de fundamentos teóricos y metodológicos que aseguren el cumplimiento de los niveles de usabilidad requeridos para la aplicación [Montero et al., 2004]. Es decir, la especificación cuantitativa de la usabilidad de un producto. Las fases del ciclo de vida de la ingeniería de usabilidad son el análisis de especificaciones, el diseño y la implementación (tabla 2.2) [Mayhew, 1999].

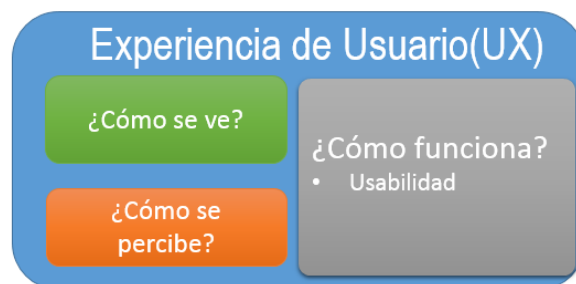


Figura 2.3: Componentes del proceso de UX

El concepto de UX se define como *"las percepciones y respuestas de un individuo a partir del uso y / o uso anticipado de un producto, servicio o sistema"* [Bevan, 1998]. La UX comprende la apariencia, la percepción y la funcionalidad de un sistema (figura 2.3). Se pueden identificar tres características claves de la UX [Albert and Tullis, 2013], (i) hay un usuario involucrado, (ii) el usuario interactúa con una interfaz y (iii) la experiencia es observable o medible. Sin embargo, evaluar dicha experiencia es complejo debido a la gran cantidad de posibles parámetros que involucra y de los que hasta ahora no hay un consenso final en la comunidad científica. Nielsen [Group, 2000] la define como un concepto integrador de todos los aspectos de la interacción entre el usuario final, los desarrolladores y el sistema. Se considera un cambio emergente del propio concepto de usabilidad [D'Hertefeldt, 2000], que extiende la evaluación de la funcionalidad de un sistema al considerar aspectos contextuales del usuario. Desde este enfoque se puede partir de la evaluación de usabilidad para obtener una gran cantidad de información sobre la UX [Nielsen, 1994].

En Interacción Humano-Computadora (IHC) la evaluación de usabilidad puede usarse para comprobar el efecto de la interfaz en el usuario, así por ejemplo, se pueden identificar atributos que promuevan estados y emociones positivas en el usuario, como satisfacción, relajamiento, etcétera; y/o eviten emociones negativas como: frustración, enojo o aburrimiento.

La evaluación de usabilidad no es una simple fase del proceso de diseño. De forma

ideal, la evaluación debe ocurrir a través de todo el ciclo de vida del proceso de diseño y desarrollo para tener una retroalimentación y hacer modificaciones al mismo. En la evaluación, hay un objeto que está siendo evaluado y un proceso a través del cual uno o más atributos son juzgados o se les da un valor [Karat, 1997]. La evaluación de la usabilidad, es una de las tareas más importantes que debe emprenderse cuando se desarrolla una interfaz de usuario [Woodward, 1998]. Tiene tres objetivos principales [Dix et al., 2004]:

1. Comprobar la extensión de la funcionalidad del sistema.
2. Comprobar el efecto de la interfaz en el usuario.
3. Identificar cualquier problema específico con el sistema.

El modelo básico de la evaluación de usabilidad comprende la selección de métodos, técnicas y herramientas apropiadas para evaluar los atributos de calidad que son de interés. Un método de evaluación de usabilidad es un procedimiento sistemático para grabar datos relacionados con la interacción del usuario final con un producto software o sistema [Fitzpatrick, 1999]. La clasificación de los métodos de evaluación de usabilidad puede hacerse a partir de muchos criterios, como puede ser el grado de intervención del usuario, el escenario, el uso de reglas y atributos definidos, etcétera. Las técnicas de evaluación de usabilidad definen un conjunto de actividades a ejecutar por los evaluadores. Pueden ser descriptivas (basadas en el usuario) o predictivas (basadas en la opinión de expertos) [Gediga et al., 2002]. Una herramienta es la adaptación de la técnica con base en el objetivo perseguido por el método, es decir, el atributo a evaluar. Otros componentes del modelo son el propio sistema y los requerimientos ergonómicos y funcionales del usuario, que desde el enfoque de UX se entienden como sus características contextuales. La figura 2.4 muestra la relación entre estos componentes.

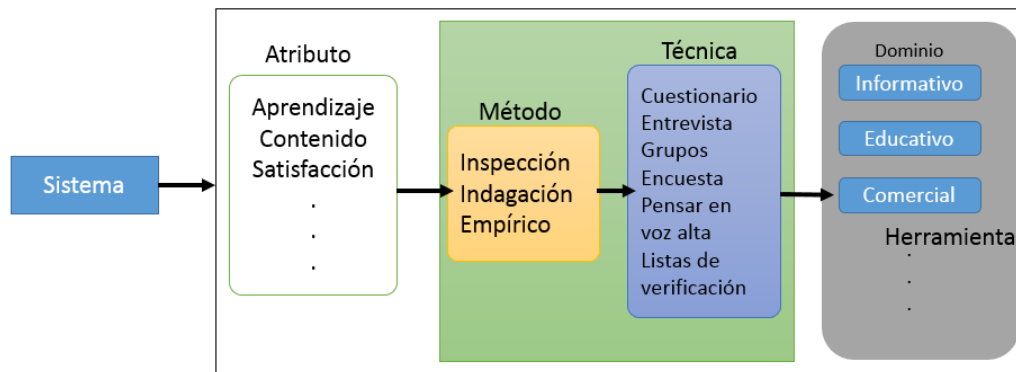


Figura 2.4: Relación entre los componentes del proceso de evaluación de usabilidad.

2.2 Entornos Virtuales Serios

La realidad virtual (RV) es una representación de cosas físicas a través de medios electrónicos que proporciona la sensación de estar en una situación real y en la que se puede interactuar con el entorno [Hilera et al., 1999]. La RV es una herramienta que ha encontrado aplicación entre otras áreas en la estimulación cognitivo afectiva. En esta área tiene el potencial de favorecer el aprendizaje, la autoestima, potenciar la creatividad y las habilidades digitales y al mismo tiempo generar motivación y entretenimiento. En particular, un entorno virtual serio (EVS) puede ser definido como una representación de la realidad con normas específicas de interacción que se utiliza para la educación, la salud, la política pública, y los objetivos estratégicos de comunicación [Zyda, 2005]. Es decir, aquellos entornos diseñados para propósitos principales distintos al entretenimiento.

Los entornos virtuales son frecuentemente, aunque no exclusivamente, animados en 3D y son sistemas interactivos cuya evolución depende de las acciones del usuario y con características especiales que los alejan de los sistemas tradicionales [Sánchez et al., 2008]. Un factor importante que marca esta diferencia es la motivación que existe para completar cada tarea del sistema. El flujo de interacción es diferente y lleva a requerir evaluaciones de UX a medida de estas particularidades.

2.3 Anatomía del ojo y del sistema oculomotor

El órgano de la visión está compuesto por los párpados, los globos oculares, el aparato lagrimal y los músculos oculares externos. El globo ocular mide unos 25 mm de diámetro y se mantiene en su posición gracias a los músculos extraoculares. La pared del ojo está formada por tres capas. La capa externa, que incluye la esclerótica (espesa, resistente y de color blanco) y en la parte anterior la córnea transparente. La capa media, incluye coroides, que contiene abundantes vasos sanguíneos, y el tejido conjuntivo del cuerpo ciliar y el iris. La capa interna se llama retina, en la que se encuentran las células sensibles a la luz (los bastones y los conos), recubiertas por una lámina externa de células epiteliales cúbicas que contienen melanina. Externamente, la retina descansa sobre la coroides; internamente, está en contacto con el humor vítreo [Ross et al., 2003]. El sistema oculomotor comprende, entre otras cosas, la relación biomecánica de los ojos con el espacio en el que se mueven: las órbitas [Pons, 2012]. Hay seis músculos extraoculares en cada ojo que mantienen el alineamiento del eje visual con movimientos coordinados: Rectos (superior, inferior, lateral y medial) y oblicuos (superior e inferior). Se ha demostrado que el movimiento del ojo está asociado con la activación de distintas parte del cerebro. El ojo está conectado al cerebro por un nervio (nervio oculomotor) y éste accede a diferentes departamentos del cerebro según tenga necesidad [Bandler et al., 1982].

2.4 Seguimiento del ojo y pupilometría

El seguimiento del ojo (*eye tracking*) es el proceso de evaluar, bien el punto donde se fija la mirada, o el movimiento del ojo en relación con la cabeza. Este proceso es utilizado en la investigación en los sistemas visuales, en psicología, en lingüística cognitiva y en diseño de productos. La extensa mayoría de diseños actuales son seguidores de ojos basados en vídeos. Una cámara enfoca uno o los dos ojos y graba sus movimientos mientras el sujeto mira una serie de estímulos. Se usan dos tipos generales de técnicas de seguimiento de

ojos: pupila brillante y pupila oscura. La diferencia entre ellas se basa en la localización de la fuente de iluminación con respecto a los ojos. La mayoría usan una frecuencia de muestreo de al menos 30 Hz. Aunque 50/60 Hz es lo más común, actualmente muchos seguimientos de ojos basados en video funcionan a 240, 350 o incluso 1000/1250 Hz, frecuencia que se necesita para captar en detalle los rápidos movimientos durante la lectura o durante los estudios de neurología. El movimiento de los ojos normalmente se divide en fijaciones y salidas, cuando la mirada se detiene en cierta posición y cuando se mueve hacia otra posición respectivamente. Las series resultantes de las fijaciones y las salidas se llama *scanpath*. El seguimiento de ojos en IHC normalmente investiga el *scanpath* con fines de usabilidad, o como método de entrada en muestras contingentes con la mirada, también llamadas interfaces basadas en la mirada.

2.5 Neuroergonomía y Carga cognitiva

Neuroergonomía es definida como el estudio del cerebro mientras trabaja [Parasuraman, 2003]. Su objetivo es determinar el funcionamiento del cerebro en el contexto de la cognición y la conducta del factor humano. Por ejemplo, el comportamiento de un trabajador en situaciones de peligro. Se apoya principalmente de técnicas de neuroimagen para monitorear la carga mental desde aspectos biológicos. Sin embargo, desde 1973 se tienen cimientos sobre el estudio de la carga mental a partir de índices basados en la actividad ocular [Kahneman, 1973].

Una herramienta útil de la neuroergonomía es la realidad virtual [Rizzo et al., 2007]. El uso de entornos virtuales facilita el entrenamiento de labores complejas, como el pilotaje o la telecirugía.

El concepto de *carga cognitiva*, en neuroergonomía, se relaciona con el uso más eficiente de las interacciones corticales y sus desviaciones. Según la Teoría de Carga Cognitiva, la información que entra al cerebro es procesada en tres diferentes estructuras:

a) la memoria sensorial, b) la memoria de trabajo, y c) la memoria de largo plazo. En psicología cognitiva, *carga cognitiva* se refiere al esfuerzo mental total en la memoria de trabajo.

El enfoque ergonómico busca una expresión neurofisiológica cuantitativa de la carga cognitiva que puede medirse usando instrumentos comunes, por ejemplo, usando el producto de presión de la velocidad de la sangre al corazón (RPP), como una medida de la carga de trabajo ocupacional cognitiva y física [Fredericks et al., 2005]. La respuesta pupilar es una medida que refleja directamente la carga cognitiva en la memoria de trabajo.

2.6 Modelos Gráficos Probabilistas

Los Modelos Gráficos Probabilistas (PGM por sus siglas en inglés) representan las relaciones entre variables aleatorias basadas en probabilidad. Proveen un marco de trabajo para manejar la incertidumbre de una manera computacionalmente eficiente [Sucar, 2015]. Un PGM es una representación compacta de una distribución de probabilidad conjunta, de la cual podemos obtener probabilidades condicionales y marginales. La probabilidad conjunta es obtenida por el producto de las funciones locales:

$$P(X_1, X_2, \dots, X_N) = K \prod_{i=1}^M f(Y_i)$$

donde K es una constante de normalización. Una forma natural de representar las relaciones de dependencia entre un conjunto de variables es usando grafos. De esta forma, un PGM se compone de: (i) un grafo, $G(V, E)$, que define la estructura del modelo; y (ii) un conjunto de funciones locales, $f(Y_i)$, que define los parámetros.

Esta representación en términos de un grafo y un conjunto de funciones locales es la base para la inferencia y aprendizaje de los PGMs. La *inferencia* es la obtención de las probabilidades condicionales o marginales de cualquier subconjunto de variables Z dado otro subconjunto Y . El *aprendizaje* se refiere a estimar la estructura y parámetros de

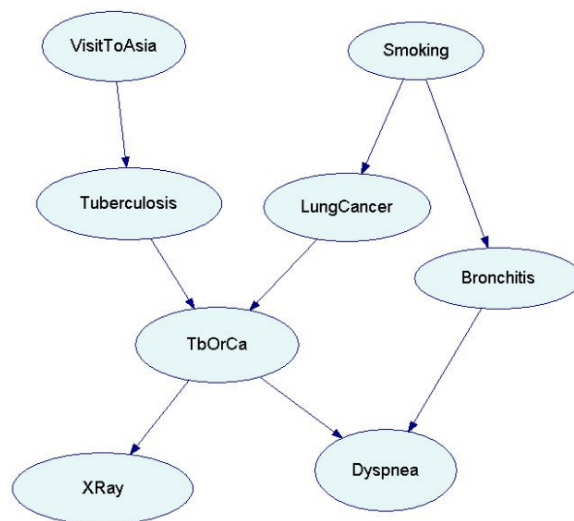


Figura 2.5: Ejemplo de red bayesiana. Cada nodo en la red corresponde a una condición de los pacientes en una clínica. Por ejemplo, "VisitToAsia" indica si el paciente recientemente visitó ese continente. Las flechas entre los nodos indican que hay relación probabilística conocida entre los estados de esos nodos; por ejemplo, fumar ("Smoking") incrementa la probabilidad de tener cáncer de pulmón ("LungCancer").

un modelo.

2.6.1 Redes Bayesianas

Una red bayesiana es un tipo de modelo gráfico probabilista, representado como un grafo dirigido y acíclico. Los nodos del grafo son los objetos y eventos del mundo real. Las aristas representan relaciones causales entre estos objetos, empezando en la variable causal y terminando en la variable efecto. Cada variable o nodo tiene asociada una función de probabilidad cuya dimensión depende de las aristas que llegan al nodo. La ausencia de una arista representa independencia condicional entre nodos.

Las relaciones propagan conocimiento del sistema una vez que se tiene evidencia sobre algún evento. Así se pueden aprender las probabilidades de los elementos a partir del conocimiento de alguno de ellos y las relaciones que tiene con los demás. En la figura 2.5 se muestra un ejemplo de una red bayesiana sencilla.

Además de la estructura, la red considera un conjunto de parámetros locales, que son las probabilidades condicionales para cada variable dados sus padres en el grafo. Dada la estructura y los parámetros podemos responder preguntas sobre los datos y sus relaciones probabilísticas, es decir, hacer inferencias sobre ellos.

2.6.1.1 Inferencia

La inferencia es la acción de calcular la probabilidad de cada estado de un nodo en la red cuando se conocen los valores que toman otras variables de la misma. La inferencia probabilística consiste en propagar los efectos de cierta evidencia en la red bayesiana para estimar su efecto en las variables desconocidas. Esto es, obtener las probabilidades posteriores. Si no hay variables desconocidas, se pueden obtener las probabilidades de todo el grafo.

Las variables de una red bayesiana, representadas por V , pueden dividirse en dos grupos dependiendo de su capacidad de ser observadas. Consideremos la partición $V = Q \cup O$. Sean $Q = \{q_0, q_1, \dots, q_{N-1}\}$, $O = \{o_0, o_1, \dots, o_{M-1}\}$. Denotamos los dos subconjuntos como los grupos de variables ocultas y observables respectivamente. Entonces, dado U subconjunto arbitrario de V , el objetivo de la inferencia es encontrar la función de distribución de probabilidad condicionada de que U tome el valor $\vec{u} = \{q_0, q_1, \dots, q_{N-1}\}$ dadas las variables observadas O . Esto se puede escribir como $P(U = \vec{u} | O)$.

Si $U \subseteq O$, se define la función de distribución de probabilidad como $P(U = \vec{u} | O) = \prod_{k=1}^K \delta(u_k - o_k)$, donde o_k es el valor observado para la variable k -ésima del conjunto U . Siendo $\delta(x) = 1$ si $x = 0$ y $\delta(x) = 0$ en cualquier otro caso. Si $U \subseteq Q$, se presenta un caso no trivial. Ahora se puede obtener la función deseada usando la regla de Bayes:

$$P(U = \vec{u} | O) = \frac{P(U = \vec{u} | O)}{P(O)} = \frac{P(U = \vec{u}, O)}{\sum_{\forall \vec{u}} P(U = \vec{u}, O)}$$

Se observa que es suficiente con calcular la función conjunta $P(U = \vec{u}, O)$ y marginalizar sobre U . Además, la probabilidad conjunta de U y O se obtiene marginalizando

$P(V)$ sobre el conjunto de las variables ocultas $Q \setminus U$, como se indica en la siguiente expresión:

$$P(U = \vec{u}, O) = \sum_{x \in Q \setminus U} P(x, U = \vec{u}, O)$$

La inferencia en redes bayesianas arbitrarias puede realizarse por:

1. Propagación exacta de probabilidades en una red simplemente conectada, que es aquella en la que sólo existe un camino entre dos variables de la red.
2. Inferencia aproximada.

2.6.1.2 Aprendizaje

La fase de aprendizaje de una red bayesiana incluye dos aspectos: aprendizaje de la estructura y de los parámetros. Cuando la estructura es conocida, el aprendizaje paramétrico consiste en estimar las tablas de probabilidad condicional desde los datos.

Aprendizaje paramétrico El aprendizaje paramétrico se fundamenta en descubrir los parámetros asociados a la estructura dada de una red bayesiana. Es decir, lo que se pretende es encontrar las probabilidades a priori de los nodos raíz y las probabilidades condicionales de las demás variables dados sus padres, de acuerdo con la información disponible.

Uno de los principales trabajos en el campo del aprendizaje de redes bayesianas es el de Herkovits y Copper [Herskovits, 1991]. Si se tienen suficientes datos para todas las variables, y se asume que la topología de la red es conocida, el aprendizaje es sencillo. La tabla de probabilidad condicional para cada variable puede ser estimada desde los datos basada en la frecuencia de cada valor obteniendo un estimador de probabilidad máxima.

Para actualizar las probabilidades de cada nodo observado, cuando se utiliza una estructura en árbol, se hace uso de las siguientes fórmulas:

Fórmula de probabilidades previas:

$$P(A_i) = \frac{a_i + 1}{s + 1}; i = k$$

$$P(A_i) = \frac{a_i}{s + 1}; i \neq k$$

Fórmula de probabilidades condicionales:

$$P(B_j | A_i) = \frac{(b_j + 1)}{[a_i + 1]}; i = k, j = 1$$

$$P(B_j | A_i) = \frac{(b_j)}{[a_i + 1]}; i = k, j \neq 1$$

$$P(B_j | A_i) = \frac{(b_j)}{[a_i]}; i \neq k$$

Donde:

$P(A_i)$: es la probabilidad asociada con el evento (nodo) de interés.

s , corresponde al número de casos totales.

i, j , los índices de las variables, k .

b_i, a_i es un conjunto de eventos.

Aprendizaje de estructura Consiste en obtener la estructura de la red bayesiana a partir de los datos. Es decir, las relaciones de dependencia y/o independencia entre las variables involucradas. Las técnicas de aprendizaje de la estructura dependen del tipo de organización de la red (árboles, poli árboles o redes multiconectadas).

El aprendizaje de la estructura es un problema complejo por: (i) el número de estructuras posibles es extenso aún con pocas variables y (ii) se requiere una gran base de datos para obtener buenos resultados de las medidas estadísticas.

Para el aprendizaje de la estructura hay dos principales tipos de métodos:

1. Métodos globales basados en búsqueda y puntaje
2. Métodos locales que utilizan test de independencia.

Métodos Globales: Realizan una búsqueda heurística sobre el espacio de estructuras de red, comenzando de una estructura inicial y generando una variación en cada paso [Heckerman, 2008] [Lam and Bacchus, 1994]. Se selecciona una estructura con base en un puntaje que mide que tan bien el modelo representa los datos.

Métodos Locales: Están basados en la evaluación de relaciones de dependencia o independencia entre subconjuntos de variables dados los datos, para secuencialmente obtener la estructura de la red. El más famoso en este enfoque es el algoritmo PC.

El algoritmo PC [Spirtes et al., 2000] comienza por recuperar un esqueleto de la red bayesiana y luego determinar la orientación de los arcos. Para determinar el esqueleto, comienza asumiendo un grafo completamente conectado y no dirigido y determina la independencia condicional para cada par de variables. Para esto, asume que hay un procedimiento que puede determinar si dos variables, X, Y , son independientes dado un subconjunto de variables S , $I(X, Y | S)$. Una alternativa es la medida de entropía cruzada.

Las técnicas de test de independencia dependen de tener suficientes datos para obtener buenos resultados. Aunque los algoritmos de búsqueda y puntaje son más robustos con respecto al tamaño del conjunto de datos, su desempeño también se ve afectado por el mismo.

2.7 Distribuciones Multivariadas

Sobre un dado espacio muestral podemos definir diferentes variables aleatorias. Por ejemplo, en un experimento binomial, X_1 podría ser la variable binomial (número de total de éxitos) y X_2 el número de éxitos en las k primeras pruebas. Si las variables aleatorias X_1, X_2, \dots, X_n están definidas sobre el mismo espacio muestral S se dice que están conjuntamente distribuidas.

2.7.1 Distribución de probabilidad conjunta

La función de distribución **conjunta** para las variables aleatorias X_1, X_2, \dots, X_n se define como

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) \equiv \text{Prob} X_1 < x_1, \dots, X_n < x_n$$

Donde $X_1 < x_1, \dots, X_n < x_n = X_1 < x_1 \cap \dots \cap X_n < x_n$ En otras palabras, $F_{X_1, \dots, X_n}(x_1, \dots, x_n)$ es la probabilidad de que las variables aleatorias X_i tomen simultáneamente valores en los intervalos $-\infty < X_i < x_i$ con $i = 1, \dots, n$.

Capítulo 3

Trabajo Relacionado

3.1 Evaluación de usabilidad como parte de la experiencia del usuario

La importancia que ha tomado el concepto de Experiencia del usuario (UX) dentro del diseño y desarrollo de productos de software ha crecido en los últimos años. La dificultad de evaluar la experiencia de un usuario radica en que no sólo se evalúa el desempeño, sino además la experiencia vivida [Kaye, 2007]. El objetivo principal de una evaluación de experiencia del usuario es conocer cómo se siente la persona ante el uso de una interfaz.

Alrededor de este concepto se han desarrollado un gran número de métodos que intentan obtener la mayor cantidad de información para medir la experiencia del usuario. Inicialmente se tomaban instrumentos desarrollados en las ciencias sociales para llevar a cabo la evaluación [Poels and Dewitte, 2006]. Es común que se desarrollen métodos a medida de un sistema específico y que no se pruebe su efectividad para aplicarse en casos posteriores. Parte del problema es que no hay un consenso sobre la definición y alcance del concepto de Experiencia de Usuario [Lallemand et al., 2015]. Aunque es frecuente utilizar cuestionarios para obtener datos cualitativos y cuantitativos acerca del estado emocional y afectivo de las personas [Bargas-Avila and Hornbæk, 2011] [Tonetto and Desmet, 2016], recientes enfoques toman ventaja de varios sensores que pueden colocarse en el cuerpo humano [Branco et al., 2005] [Tsai et al., 2009]. En [Chae et al., 2011] Chae et al. relacionan la distribución de la presión del cuerpo en el asiento de un automóvil con los puntajes de comodidad que da un conductor. Carrillo [Carrillo et al., 2015], utiliza

un dispositivo de EEG para obtener información sobre el estado emocional del usuario. En [Vermeeren et al., 2010], los autores listan una serie de características de los métodos de evaluación, tales como: el tipo de dato obtenido (cualitativo o cuantitativo), tipo de aplicación (genéricos, específicos), fuente de información (usuario, grupos, expertos), localización (laboratorio, campo, internet), periodo de la experiencia (momentánea, previa, longitudinal) y fases de desarrollo del sistema. Como parte de sus conclusiones mencionan la necesidad de métodos que faciliten el análisis en relación a los recursos requeridos para realizar la evaluación.

Considerando el concepto de usabilidad como parte de lo que constituye la experiencia del usuario, es posible decir que se puede partir de los métodos de evaluación de usabilidad y aportar información sobre esta experiencia, ya que uno de los componentes generales de la usabilidad es la satisfacción del usuario. Ejemplo de ello es el trabajo de Finstad [Finstad, 2010], quien correlaciona el cuestionario SUS [Brooke, 1996] con un equivalente de preguntas orientadas a evaluar directamente la experiencia del usuario. Por otra parte, Alegria et al. [Alegria et al., 2015] presentan un modelo que incorpora la identificación de emociones durante el uso del sistema al proceso de evaluación de usabilidad de los sistemas interactivos. Finalmente en [Chorianopoulos and Spinellis, 2004], Chorianopoulos introduce el concepto de *usabilidad afectiva*.

3.1.1 Métodos de evaluación de Usabilidad

La evaluación de usabilidad es considerada una de las tareas más importantes que debe emprenderse cuando se desarrolla una interfaz de usuario [Woodward, 1998]. Tiene tres objetivos principales [Alan et al., 2004]: (i) Comprobar la extensión de la funcionalidad del sistema, (ii) Comprobar el efecto de la interfaz en el usuario y (iii) Identificar cualquier problema específico con el sistema. De forma ideal, la evaluación de usabilidad provee una retroalimentación sobre el sistema y la experiencia del usuario.

Existen muchos métodos de evaluación de usabilidad que pueden clasificarse por el lugar de realización, el tipo de técnica, la automatización y el tipo de participantes

[Granollers i Saltiveri et al., 2004]. La tabla 3.1 resume la clasificación de los métodos en referencia a las técnicas que utilizan.

Dentro de esta clasificación se encuentran los métodos biométricos, aquellos que a partir de la observación directa de respuestas biológicas del cuerpo determinen el estado de la persona mientras utiliza la interfaz [Pike et al., 2012][Kwon et al., 2003][Carrillo et al., 2015].

Si hacemos una clasificación con respecto a la automatización de los métodos, la gran mayoría son clasificados como métodos manuales. La automatización de la evaluación se ha visto en diferentes investigaciones [Möller et al., 2006] [Tarta and Moldovan, 2006][John and Kieras, 1996]. La base de la automatización es la identificación de las tareas generales del proceso de evaluación [Ivory and Hearst, 2001] (i) Capturar datos de usabilidad, que pueden ser cualitativos o cuantitativos, (ii) Analizar los datos para generar métricas de usabilidad y (iii) Interpretar y sugerir soluciones a los problemas identificados.

3.1.2 Interferencia cognitiva en la evaluación de usabilidad

El papel de los procesos cognitivos y emociones como componentes integrales de la experiencia del usuario ha sido recientemente estudiado en el campo de la IHC [Alegria et al., 2015]. Es posible inferir una relación entre sobrecarga cognitiva y la facilidad de aprendizaje que tiene un usuario ante un sistema. La carga cognitiva puede estimarse usando cuestionarios como el NASATLX [Hart and Staveland, 1988]. Instrumentos como Emocards [Zenk et al., 2008] han sido desarrollados para integrar las emociones a la evaluación de sistemas interactivos en general. Emocards consiste en tarjetas animadas que representan diferentes respuestas emocionales. También han surgido metodologías para la integración de elementos afectivos [Alegria et al., 2015]. Sin embargo, el trabajo alrededor de integrar procesos cognitivos a la evaluación de sistemas interactivos y específicamente a entornos de realidad virtual es muy reciente y con margen de mejora. La medición de emociones mediante surrogados de actividad cognitiva se ha abordado

Tipo de método	Descripción	Técnicas	Usuarios finales
Inspección	Evaluadores examinan la interfaz [Wharton et al., 1994]	Caminata cognitiva, Evaluación heurística Inspección formal	No
Indagación	Retroalimentación del usuario [Brooke, 1996]	Cuestionarios Entrevistas Encuestas	Sí
Observacional	Evaluadores observan la interacción [Sutcliffe and Gault, 2004] del usuario con la interfaz	Pensamiento en voz alta Observación directa Observación indirecta	Sí
Biométrico	El estado del usuario es determinado por mediciones biológicas [Pike et al., 2012]	Evaluación cognitiva Evaluación ergonómica	Sí

Tabla 3.1: Clasificación de los Métodos de evaluación de usabilidad

por diferentes métodos. Algunos involucran señales electrofisiológicas, como el uso de EEG o ECG [Petranonakis and Hadjileontiadis, 2010] [Agrafioti et al., 2012]. Pike [Pike et al., 2012] propone una evaluación cognitiva de usabilidad (CUES) a través de la integración de la señal electroencefalográfica en conjunto con cuestionarios y técnicas tradicionales. La limitación principal de CUES es que únicamente presenta la información para que sea analizada e interpretada posteriormente por los evaluadores, sin establecer relación con los puntajes de usabilidad. Otros hacen uso de señales no electrofisiológicas como el análisis del habla [Kwon et al., 2003] o el reconocimiento de rostros [Maglogiannis et al., 2009], o señales fisiológicas tales como la conductancia de la piel, el ritmo cardiaco, el cortisol presente en la saliva, etcétera. Para los fines de la evaluación de usabilidad objetiva, son preferibles aquellos métodos no obstrusivos, es decir, que no interfieren directamente con el desempeño del usuario.

3.1.3 Evaluación de entornos virtuales

En la última década han surgido diferentes enfoques que abordan los problemas específicos de los entornos virtuales en relación a la evaluación de usabilidad [Bowman et al., 2002]. Muchos de ellos tienen conceptos y metas similares a los tradicionales aplicados a cualquier sistema interactivo, pero el uso en el contexto del entorno virtual es distinto. En [Bowman et al., 2002], Bowman, Gabbard y Hix organizan los aspectos

que diferencian la evaluación de usabilidad de los entornos virtuales en categorías: tipos de ambiente físico, tipos de evaluadores y tipos de usuarios. Se han propuesto heurísticas, es decir, pautas generales y no reglas específicas de diseño, que abordan la usabilidad en la evaluación de entornos virtuales tomando en cuenta principios de diseño propios de este tipo de interfaces como son la necesidad de ajuste de la tarea, la representación natural de objetos y la facilidad de navegación, entre otros [Sutcliffe and Gault, 2004]. También se han utilizado métodos de inspección tradicional como la caminata cognitiva, que se basa en un modelo psicológico de aprendizaje. Consiste en la construcción de escenarios de un prototipo y la simulación de interacción por parte de expertos como si estuvieran trabajando con las tareas en una interfaz construida [Wharton et al., 1994].

Aún cuando se han propuesto métodos que contribuyan a la evaluación de entornos virtuales, es poco el trabajo actual que permita asegurar que un entorno virtual es en efecto usable [Bello et al., 2014]. Esto se puede atribuir a que los entornos virtuales presentan retos únicos de usabilidad [Moreno-Ger et al., 2012]. Los entornos virtuales demandan pensamiento exploratorio y experimentación. Los errores y la frustración son parte del diseño de interacción. Un entorno virtual sin retos resulta aburrido y esto no está considerado en las métricas tradicionales que se enfocan en la productividad.

Uno de los objetivos de la evaluación de sistemas interactivos es minimizar la carga cognitiva del usuario para facilitar la realización de las tareas del sistema. En el caso de un entorno virtual, no necesariamente es apropiado minimizar la carga cognitiva; se trata de equilibrar la carga cognitiva para que la experiencia del usuario no resulte negativa, aunque en términos propios de usabilidad resulte contradictorio. La figura 3.1 muestra cómo influye el nivel de carga cognitiva en relación al desempeño de un usuario.

Este trabajo presenta una propuesta de método biométrico y automático de evaluación de usabilidad, extendiendo el concepto por la integración de las características del perfil del usuario y por otro lado las características generales del entorno virtual. Además de añadir el nivel de carga cognitiva presente en la interacción del usuario con un entorno, no en afán de disminuirla sino de equilibrarla para alentar al usuario a completar cada

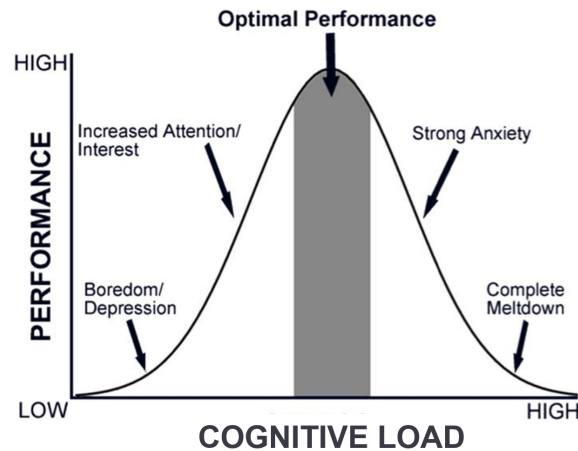


Figura 3.1: La relación entre el desempeño y la carga cognitiva no es lineal. Se observa que un bajo desempeño puede asociarse tanto a un bajo nivel como a un alto nivel de carga cognitiva y la interpretación es diferente. El desempeño óptimo al realizar una tarea requiere de un equilibrio en el nivel de carga cognitiva del usuario.

tarea.

3.2 Medición de actividad cognitiva por métricas oculares

El seguimiento del ojo introduce dos fuentes potenciales de información sobre el cerebro: los movimientos oculares y la dilatación de la pupila. Parámetros como el incremento de la dilatación de la pupila [Klingner et al., 2011] y decremento de la duración de fijaciones [Van Orden et al., 2001] son fuentes de información del esfuerzo mental [Holmqvist et al., 2011]. Partala y Surakka encontraron que la variación de la pupila puede ser un indicador del proceso afectivo y cognitivo [Partala and Surakka, 2003]. Las técnicas combinan información de tres fenómenos del ojo: el tamaño de la pupila, los movimientos y los parpadeos. Estos indicadores son utilizados por modelos matemáticos y estadísticos para estimar el estado cognitivo. Marshall [Marshall, 2002] genera un índice de actividad cognitiva (ICA) que es calculado como el número de incrementos inusuales en el tamaño de la pupila por segundo. De manera similar [Demberg, 2013], Demberg obtiene el índice

de actividad cognitiva en tareas duales a partir de calcular la frecuencia de las pequeñas dilataciones de la pupila.

Aunque es frecuente el análisis pupilométrico para determinar la actividad cognitiva, las métricas más utilizadas en la investigación son las fijaciones y los movimientos sacádicos [Sharma and Dubey, 2014]. Dempere-Marco [Dempere-Marco et al., 2002] analiza los movimientos sacádicos en la búsqueda visual utilizando modelos de Markov en una aplicación de soporte a la comprensión de imágenes. En [Rognin et al., 2004], Rognin correlaciona las fijaciones en combinación con los movimientos sacádicos y el número de parpadeos con la carga cognitiva en tareas de tráfico aéreo.

3.2.1 Seguimiento del ojo en IHC

El determinar la actividad cognitiva de una persona por medio de la respuesta ocular varía en dependencia de la tarea que se esté realizando [Charleton and O'Brien, 2002]. Hay una gran cantidad de trabajos enfocados a tareas complejas, específicamente a simulaciones de manejo de vehículos. En [Tsai et al., 2007], Tsai et. al, analizan diferentes métricas que incluyen la frecuencia de parpadeos, duración del parpadeo, frecuencia de fijaciones, duración de fijaciones y diámetro de la pupila, para evaluar el comportamiento cognitivo de un conductor, en afán de predecir su desempeño en tiempo real.

Aunque el seguimiento del ojo es frecuentemente usado en investigaciones cognitivas, no se ha explotado únicamente en la práctica para medir carga cognitiva [Wernaart, 2013]. El uso de herramientas de seguimiento ocular se ha visto en trabajos que analizan el comportamiento humano en relación a la atención visual [Granka et al., 2004], en el tratamiento de pacientes con desórdenes de ansiedad [Grillon et al., 2007], entre otros. Dentro del campo de la IHC, se ha explotado en el desarrollo de interfaces oculares [Brooks et al., 2013]. Un ejemplo es EyeDraw [Hornof et al., 2004], una aplicación que permite dibujar a niños con discapacidades motoras. Y además se ha usado como una métrica en la evaluación de usabilidad [Crowe and Narayanan, 2000] [Brooks et al., 2013].

También se han utilizado en conjunto con otras medidas para obtener otros aspectos cognitivos, por ejemplo, Van Gog et al. usan el seguimiento ocular en conjunto con reportes verbales [Van Gog et al., 2009].

A medida de los avances tecnológicos, se ha intentado aplicar el seguimiento ocular para responder cuestiones de usabilidad [Benel et al., 1991] desde el enfoque de los evaluadores y desde el enfoque científico. Schiessl [Schiessl et al., 2003] propone un sistema de seguimiento visual, utilizando dispositivos remotos para evaluar la usabilidad de sitios web y de aplicaciones multimedia. Ehmke y Wilson [Ehmke and Wilson, 2007] correlacionan patrones de comportamiento ocular y problemas de usabilidad en dos sitios web. Guan [Guan et al., 2006] correlaciona la información de las fijaciones con los resultados de la técnica de pensamiento en voz alta como métrica de usabilidad.

A pesar de la gran cantidad significativa sobre la relación entre la carga cognitiva y el comportamiento ocular, no parece haber un consenso entre la combinación de las métricas que se utilizan [Rosch and Vogel-Walcutt, 2013]. Es evidente la dependencia del dominio en el análisis de la actividad cognitiva por métricas oculares.

Investigaciones han demostrado que el movimiento ocular está influenciado por la visualización de la tarea [Henderson et al., 2013]. Por ejemplo, la duración promedio de fijación y la amplitud sacádica difieren en una tarea de lectura en comparación con la visualización de una imagen [Henderson and Hollingworth, 1998] [Rayner, 1998]. Del mismo modo difieren dependiendo de si una persona está observando una escena en particular o tratando de memorizar características de esa misma escena [Henderson and Hollingworth, 1999] [Castelhana et al., 2009].

Otro problema es la dependencia al dispositivo con el que se recogen los datos. La mayoría de "eye trackers" producen señales que representan la orientación del ojo dentro de la cabeza, o la posición del punto de visión sobre una pantalla a una distancia específica. En cualquier caso, se provee de coordenadas horizontales y verticales para cada muestra. Dependiendo de la frecuencia de muestreo y la duración de la sesión es el tipo de información que podemos obtener de los datos. De igual forma, el análisis de los

datos, en muchos casos, es en dependencia de las herramientas de software disponibles para cada dispositivo.

En esta investigación se analiza el nivel de carga cognitiva en dos tareas específicas para las cuales el comportamiento ocular es diferente entre sí. El análisis de las métricas oculares es en relación a la información de fijaciones, parpadeos y movimientos sacádicos por la naturaleza de las tareas. El dispositivo de hardware utilizado es una cámara web de alta resolución, por lo cuál la implementación del algoritmo de seguimiento y análisis de las métricas es libre del tipo del dispositivo.

Capítulo 4

Resultados

En este capítulo se exponen y discuten los resultados obtenidos en cada parte de la metodología seguida para esta investigación. Se muestra el modelo final obtenido con la técnica de sobremuestreo por ajuste de datos y la validación de la técnica por medio de un caso de estudio a un problema de la literatura con un modelo de red existente.

4.1 Implementación del algoritmo de seguimiento del ojo CECL+LSH

Como resultado del primer paso de la metodología señalada se obtuvo la implementación del algoritmo CECL+LSH que sirve de base para la obtención del nivel de carga cognitiva necesario en la extensión del concepto de usabilidad hacia UX.

El dispositivo de grabación utilizado es una cámara Logitech HD c920 con captura de video de 1080p a 30fps, campo visual de 78° y resolución de 15MP. El instrumento disponible para esta investigación tiene una frecuencia de 30 fps, que no es suficiente para obtener los cambios abruptos en la dilatación de la pupila y que son necesarios para el análisis de la carga cognitiva por este método.

El desempeño de CECL+LSH fue evaluado sobre la base de datos BioID, que contiene 1521 imágenes en escala de grises de 23 diferentes sujetos en diferentes lugares y diferentes condiciones de iluminación. Algunos incluso utilizando lentes o cubriendo los ojos con el cabello. El desempeño de la detección es definido como el radio de la distancia Euclidiana entre la distancia predicha y la distancia real de ambos ojos.

La evaluación se hizo en términos de precisión adoptando la medida estándar de error

Método	$e < 0.05$	$e < 0.1$	$e < 0.15$
Valenti and Gevers (2008)	84.10 %	90.90 %	93.80 %
Niu et al. (2006)	75.00 %	93.00 %	95.80 %
Christinacce et al. (2004)	57.00 %	96.00 %	96.50 %
Asteriadis et al. (2005)	44.00 %	81.70 %	92.60 %
Timm and Barth (2011)	82.50 %	93.40 %	95.20 %
Leo et al. (2014)	80.70 %	87.30 %	88.80 %
CECL	80.75 %	95.15 %	97.78 %
CECL+LSH	80.18 %	95.70 %	98.01 %

Tabla 4.1: Resultados de desempeño de CECL+LSH en comparación con la literatura.

normalizado (e):

$$e_{worst} = \frac{\max(d_l, d_r)}{\|C_l - C_r\|}$$

donde C_l y C_r son las posiciones euclidianas verdaderas del centro de cada ojo, y d_l y d_r son las distancias Euclidianas entre la predicción y la localización real del ojo izquierdo y derecho respectivamente. CECL tiene una precisión de 80.75 % para un margen de error de $e \leq 0.05$, 95.15 % para $e \leq 0.1$ y 97.78 % para $e \leq 0.15$. La precisión es constante después de $e \geq 0.15$.

CECL+LSH se comparó con 6 métodos de la literatura, obteniendo resultados competitivos. Los resultados se resumen en la tabla 4.1. CECL tiene un desempeño competitivo y una precisión aceptable dentro de los utilizados para la obtención de indicadores de movimiento ocular.

4.2 Carga cognitiva y entrenamiento virtual

Para evaluar el desempeño del entrenamiento virtual con el uso de las interfaces diseñadas en este trabajo, además de hacer una comparación directa entre los puntajes de la sesión inicial y final de cada participante, se analizó el nivel de carga cognitiva inicial y final. La evaluación del entrenamiento es independiente a la interfaz, por lo que la comparación se realizó únicamente entre las tareas (cognitiva y motora) en relación a los niveles de carga cognitiva. La tabla 4.2 muestra los resultados de la primera y la tercera sesión

Usuario	Sesión inicial (%)	Sesión final (%)
1	52.66	84
2	67	46.33
3	45.66	65.33
4	79.66	86.66
5	57.66	58.33
6	67.33	82.66
7	64.66	70.33
8	64.66	87
9	36.66	48
10	53.33	66
11	81.66	82.33
12	61.33	28
13	93	100
14	45.33	40.33
15	43.66	54

Tabla 4.2: Niveles de carga cognitiva de la primera y tercera sesión de entrenamiento cognitivo.

de los participantes en el entrenamiento cognitivo y la tabla 4.3 los resultados del entrenamiento motor.

La tabla 4.4 muestra el puntaje obtenido por cada usuario en cada sesión de entrenamiento para la tarea cognitiva, es decir, el desempeño del usuario al utilizar las interfaces diseñadas para esta tarea. La tabla 4.5 corresponde a los puntajes del entrenamiento motor.

Los resultados muestran diferencias notables en el desempeño de cada usuario en ambos tipos de entrenamiento. En las figuras 4.1 y 4.2 se hace una comparación entre el desempeño por puntaje y nivel de carga cognitiva de cada usuario. Es decir, el porcentaje de aumento entre los puntajes inicial y final y la relación entre la carga cognitiva medida en la primera y última sesión de entrenamiento.

Usuario	Sesión inicial (%)	Sesión final (%)
1	79.66	64.33
2	38.33	58
3	66.66	42.66
4	83	73.33
5	66.33	53
6	84	67.33
7	70.66	82.33
8	87	64.66
9	58.33	43.33
10	72	65.33
11	91.33	73.66
12	44	29.66
13	96.66	95.66
14	85	85
15	88.66	80

Tabla 4.3: Niveles de carga cognitiva de la primera y tercera sesión de entrenamiento motor.

Usuario	Sesión 1 (%)	Sesión 2 (%)	Sesión 3 (%)
1	25	9.3	9.5
2	0	17.1	17.7
3	16.6	29.1	10.9
4	64.2	13.5	30.3
5	0	81.8	75
6	0	0	18.1
7	15.3	13.2	15.6
8	7.6	15.7	0.5
9	0	37.5	27.7
10	0	15	8.7
11	25	7.5	8.8
12	14.2	72.5	77.1
13	30	30	31.5
14	0	0	0
15	57.1	0	80

Tabla 4.4: Puntajes del entrenamiento cognitivo.

Usuario	Sesión 1	Sesión 2	Sesión 3
1	410	567	654
2	916	1494	1530
3	450	625	666
4	1483	1507	1785
5	1164	1521	1458
6	1167	1518	1523
7	920	1172	1300
8	904	1198	1205
9	408	555	571
10	457	556	612
11	507	591	520
12	419	585	609
13	527	657	635
14	538	413	567
15	59	48	54

Tabla 4.5: Puntajes del entrenamiento motor.

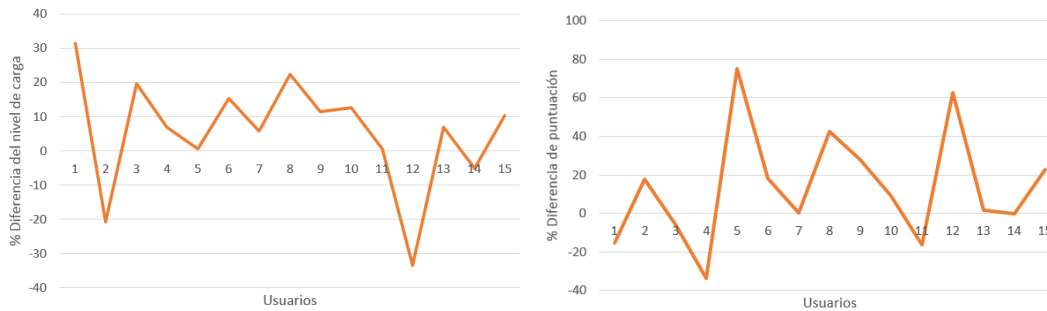


Figura 4.1: Comparación entre el desempeño por puntaje y por nivel de carga cognitiva del entrenamiento cognitivo

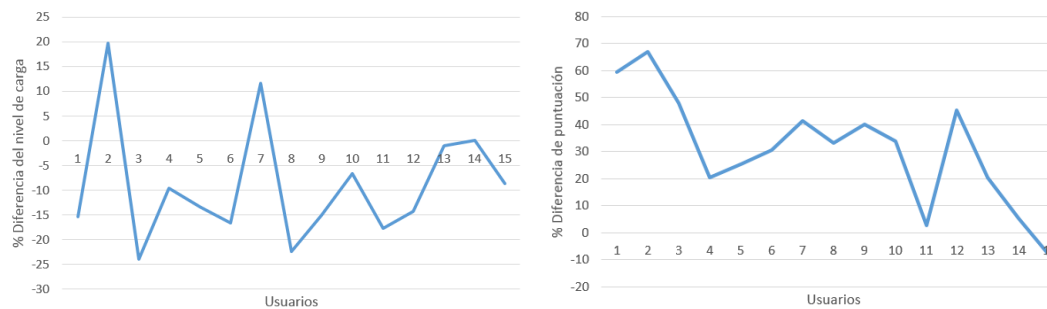


Figura 4.2: Comparación entre el desempeño por puntaje y por nivel de carga cognitiva del entrenamiento motor

4.3 Estudio de usabilidad de los entornos virtuales de entrenamiento

Se realizó un estudio de usabilidad a las interfaces desarrolladas para cada tarea. El objetivo fue la evaluación de los entornos virtuales de entrenamiento por un grupo de estudiantes de licenciatura en psicología. El estudio tiene un nivel de confianza del 85 % con un margen de error del 10 % para un grupo total de 27 estudiantes. A continuación se describen los puntos importantes del estudio:

- Los participantes en el estudio fueron 15 estudiantes de la facultad de psicología de la Universidad Popular Autónoma del Estado de Puebla (UPAEP).
- El entrenamiento para cada participante fue de 3 sesiones.
- Cada sesión tuvo una duración de 20 minutos.
- Se evaluaron 6 diferentes interfaces.
- El instrumento de evaluación de usabilidad fue el cuestionario *System Usability Scale* [Brooke, 1996].
- La asignación de interfaces se hizo de forma aleatoria.

4.3.1 Protocolo de la sesión de entrenamiento

- Los participantes firmaron una carta de consentimiento e información del procedimiento antes de la primera sesión de entrenamiento.
- Cada participante respondió el cuestionario de perfil de usuario que se muestra en la tabla 4.6.
- La primera parte de ponderación del cuestionario NASA-TLX fue aplicada al principio de la primera sesión de entrenamiento.

-
-
- Antes de comenzar el entrenamiento, cada participante recibió instrucciones claras sobre el objetivo y manipulación del entorno virtual.
 - Al comenzar la sesión se inicia la video grabación con la cámara web frente al participante.
 - El usuario interactúa con el sistema durante 20 minutos para finalizar la sesión.
 - Después del entrenamiento se aplicó la segunda parte del cuestionario NASA-TLX que corresponde a la evaluación.
 - Al finalizar las 3 sesiones de entrenamiento se aplica el cuestionario SUS a cada participante.

Cada participante fue usuario de dos interfaces asignadas aleatoriamente. La primera es una variante de la tarea cognitiva y la segunda corresponde a la tarea motora.

4.3.2 Cuestionario NASA-TLX

Esta técnica es la más utilizada actualmente para medir el nivel de carga cognitiva en la ejecución de tareas. La técnica ha sido probada en un gran número de tareas experimentales desde simuladores de vuelo hasta salas de control de generación de energía. Ha sido también estudiada en relación a tareas de laboratorio (por ejemplo, tareas de memoria de Sternberg, tiempo de reacción en tareas de elección, seguimiento, rotación mental, razonamiento gramatical, etc.).

Es un procedimiento de escalamiento que ofrece una medida general de carga mental basada en un promedio ponderado de los juicios que una persona da en seis factores y dimensiones:

- Demanda Mental
- Demanda Física

- Demanda Temporal
- Rendimiento
- Esfuerzo
- Nivel de Frustración

La aplicación de este instrumento requiere dos fases que se describen en las siguientes dos secciones.

4.3.2.1 Obtención de pesos o importancia iniciales

El objetivo de esta fase es la obtención de la importancia que, a priori, cada individuo asigna a cada una de las seis dimensiones como fuente potencial de carga mental. Esta fase es previa a la realización de la tarea y permite obtener los valores por los que se van a ponderar las estimaciones de carga, al calcular el índice global de carga mental de una determinada tarea o combinación de tareas.

1. En primer lugar, se establecen las 15 comparaciones binarias de las seis dimensiones, en las que el sujeto debe elegir, de cada par, la que percibe como mayor fuente de carga.
2. Para cada dimensión se obtiene un peso, que viene dado por el número de veces que ésta haya sido seleccionada en las comparaciones binarias. Este peso puede variar entre 0 (la dimensión no ha sido elegida en ninguna de las comparaciones) y 5 (la dimensión ha sido elegida en todas las comparaciones en las que aparecía).

4.3.2.2 Valoración

Una vez realizada la tarea o tareas de interés, el sujeto tiene que estimar, en una escala de 0 a 100, dividida en intervalos de 5 unidades, la carga mental de la tarea debida a cada una de las seis dimensiones.

Con los datos obtenidos en las dos fases, se puede calcular un índice global de carga mental de la tarea, aplicando la siguiente fórmula:

$$IC = \sum_{i=1}^6 \frac{p_i X_i}{15}$$

donde IC es el Índice de Carga, p_i es el peso obtenido para cada dimensión en la primera fase y X_i es la puntuación obtenida por la dimensión en la fase de valoración.

4.3.3 Cuestionario de usabilidad SUS

El cuestionario SUS [Brooke, 1996] fue desarrollado en 1986 como parte de la introducción de la ingeniería de usabilidad a los sistemas de oficina de Digital Equipment Co. Ltd. Su propósito era proporcionar una prueba fácil de completar (número mínimo de preguntas), fácil de evaluar y que permitiera establecer comparaciones cruzadas entre productos.

La escala SUS se utiliza generalmente después de que un usuario ha tenido la oportunidad de utilizar un sistema pero antes de que cualquier informe o discusión tenga lugar. Se solicita a los usuarios el registro inmediato de su respuesta a cada punto, en lugar de pensar largamente en los mismos.

La escala SUS es una escala de estilo *Likert* que genera un único número, representando una medida compuesta de la usabilidad del sistema global sometido a estudio. Hay que advertir que las puntuaciones independientes no son significativas por sí mismas.

Para calcular la puntuación del SUS, hay que sumar primero las contribuciones de cada punto. La contribución de cada punto valdrá entre 0 y 4. Para los puntos 1, 3, 5, 7 y 9, la contribución será la posición de la escala menos 1. Para los puntos 2, 4, 6, 8 y 10, la contribución será 5 menos la posición en la escala. Se multiplica la suma de los resultados por 2.5 para obtener el valor global del SUS. El resultado estará entre 0 y 100.

La escala SUS se construyó sobre un conjunto original de 50 puntos. Las ponderaciones se efectuaron para 20 usuarios sobre los 50 puntos para dos sistemas, uno diseñado para su uso por el usuario final y otro diseñado para el uso de los programadores de sistemas.

Pregunta	Opciones
Edad	(abierta)
Sexo	Hombre, Mujer
Experiencia computacional	Ninguna, Muy poca, Poca, Moderada, Mucha
Experiencia como usuario de videojuegos	Ninguna, Muy poca, Poca, Moderada, Mucha
Experiencia en el uso de instrumentos musicales	Ninguna, Muy poca, Poca, Moderada, Mucha
Experiencia en el uso del piano o teclado	Ninguna, Muy poca, Poca, Moderada, Mucha
Conocimientos y práctica mecanográfica	Ninguna, Muy poca, Poca, Moderada, Mucha

Tabla 4.6: Cuestionario de perfil general del Usuario

Interfaz	Resultado SUS (%)
Cognitiva Baja 1	67
Cognitiva Baja 2	57.5
Cognitiva Alta	50.8
Motora Baja 1	64.3
Motora Baja 2	75.6
Motora Alta	59.5

Tabla 4.7: Resumen de los resultados del cuestionario SUS para todas las interfaces

Estos dos sistemas fueron elegidos para representar los extremos de la usabilidad. Los 10 puntos seleccionados eran aquellos que aglutinaban las respuestas más consistentes y polarizadas. Los puntos seleccionados tienen todos una intercorrelación de entre 0.7 y 0.9.

4.3.4 Cuestionario de perfil de usuario

Con el objetivo de obtener información directa sobre las características del usuario, que permitan asociar el resultado de la evaluación de usabilidad y carga cognitiva presente en la interacción, se diseñó un cuestionario de 7 preguntas de opción múltiple como se muestra en la tabla 4.6.

Para cada interfaz desarrollada se obtuvo un resultado de usabilidad aplicando el cuestionario SUS a los usuarios de cada una. Los resultados de usabilidad del estudio se resumen en la tabla 4.7.

Como parte de los objetivos de este trabajo de investigación está el relacionar el nivel de carga cognitiva con el resultado de una evaluación de usabilidad tradicional. Nuestra

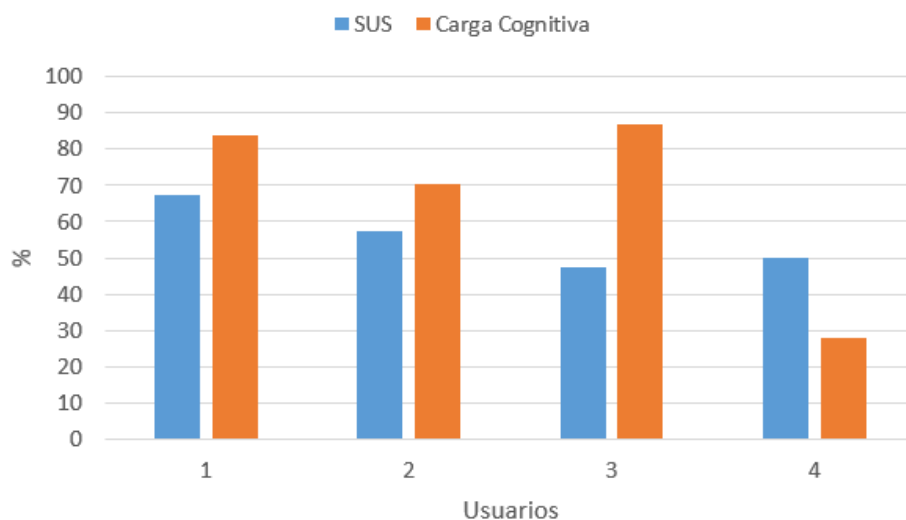


Figura 4.3: Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento cognitivo (Interfaz 1 del nivel más bajo de abstracción)

hipótesis es que dos personas pueden dar puntuaciones de usabilidad similares bajo condiciones cognitivas diferentes, por lo que ese resultado de evaluación del sistema necesita ser reinterpretado. En las figuras 4.3, 4.4 y 4.5 se compara el resultado de la evaluación de usabilidad de las 2 interfaces para la tarea cognitiva en el nivel de abstracción más bajo y en el nivel de abstracción más alto, respectivamente, con el nivel de carga cognitiva de cada usuario al finalizar el entrenamiento (tercera sesión). Así mismo, las figuras 4.6, 4.7 y 4.8 corresponden en el mismo orden a las interfaces de la tarea motora.

Un ejemplo de este fenómeno se presenta en los usuarios 3 y 4 de la interfaz 1 del nivel más bajo de abstracción para la tarea cognitiva. Aunque la percepción de utilidad del sistema es muy similar para ambos participantes, el usuario 3 presenta un nivel de carga cognitiva cercano al 90 %, mientras el usuario 4 está por debajo del 30 %. Esto podría interpretarse en atenciones diferentes a la mejora del diseño de interacción, debido al puntaje de usabilidad que ronda en 50 %. La sobre carga mental del usuario 3 podría ser interpretada como cansancio, mientras que un bajo nivel de carga cognitiva podría interpretarse como aburrimiento o desinterés. Sin embargo, si se consideran puntajes altos de usabilidad, como en el caso del usuario 2 y 5 de la interfaz para el nivel más

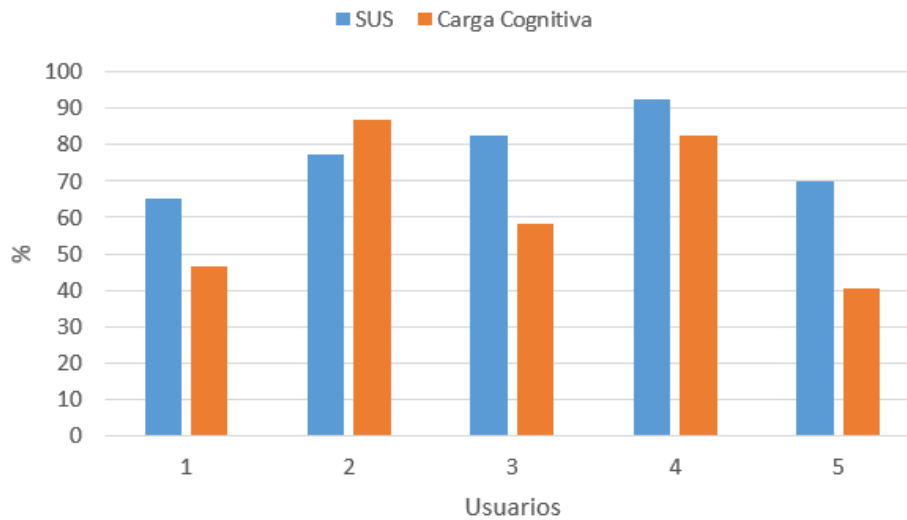


Figura 4.4: Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento cognitivo (Interfaz 2 del nivel más bajo de abstracción)

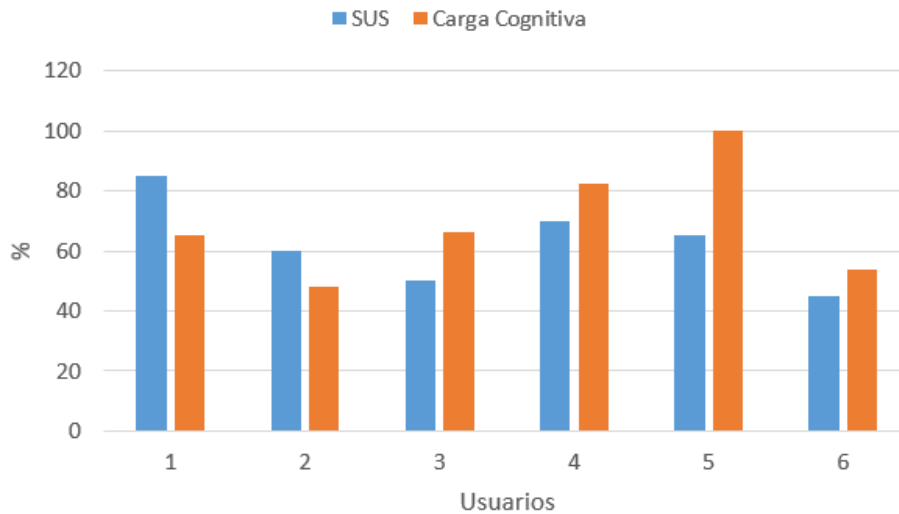


Figura 4.5: Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento cognitivo (Interfaz del nivel más alto de abstracción)

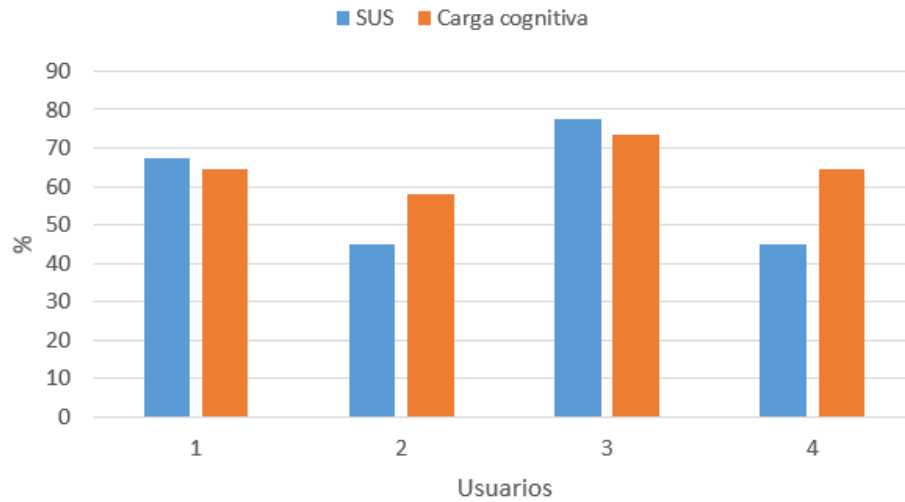


Figura 4.6: Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento motor (Interfaz 1 del nivel más bajo de abstracción)

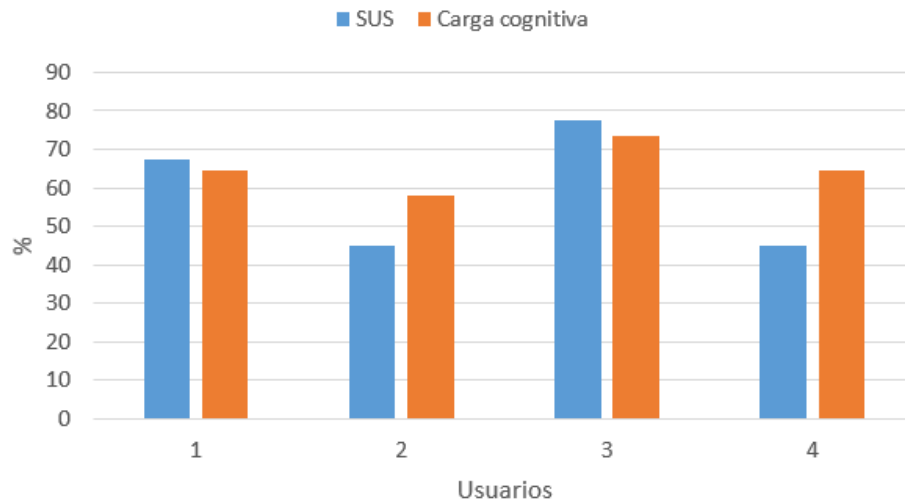


Figura 4.7: Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento motor (Interfaz 2 del nivel más bajo de abstracción)

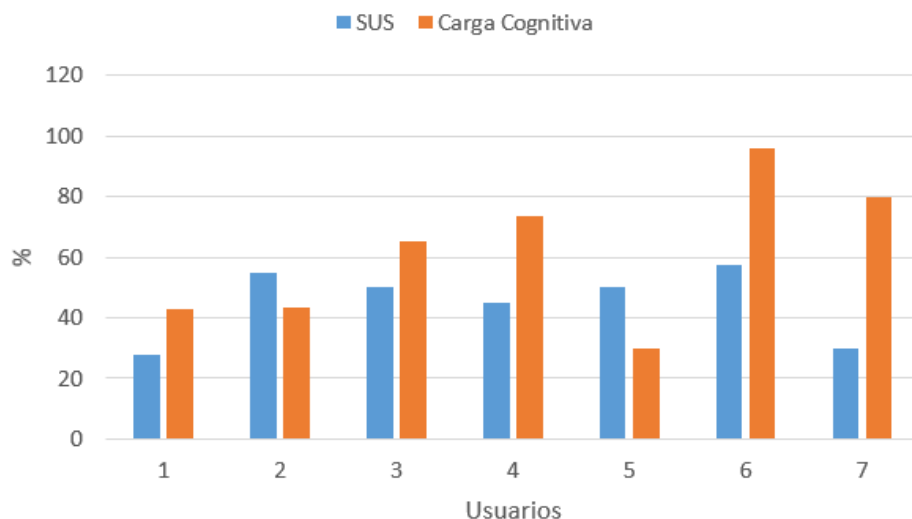


Figura 4.8: Comparación entre el resultado de usabilidad y la carga cognitiva al finalizar el entrenamiento motor (Interfaz del nivel más alto de abstracción)

alto de abstracción del mismo entrenamiento, los niveles de carga cognitiva deberían interpretarse como aspectos positivos de interacción.

4.4 Modelo de Red Bayesiana jerárquica del análisis de usabilidad

Se construyeron 2 modelos para representar las relaciones de dependencia probabilística entre las variables de los conjuntos de datos de las características del usuario y las características del sistema. Se utilizó el algoritmo PC sobre los conjuntos de datos originales y posteriormente sobre el conjunto de 5,000 muestras sintéticas generadas por el ajuste de datos a una distribución multivariada. En la figura 4.9 se observan las estructuras generadas para el conjunto de datos de características del usuario. La falta de información en el conjunto de datos original dificulta que el algoritmo PC encuentre relaciones de dependencia entre las variables. De igual forma la figura 4.10 muestra el caso de las características del sistema.

Cada una de estas redes corresponde una subred del modelo de datos final. La

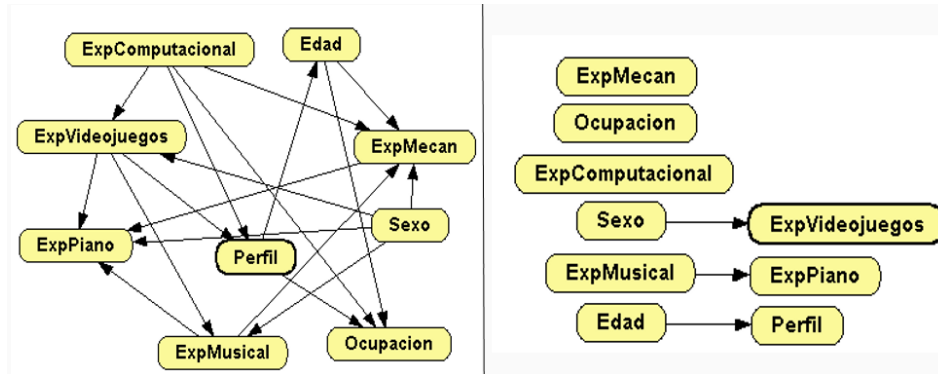


Figura 4.9: Estructura del perfil de usuario generada sobre el conjunto de datos con sobremuestreo (izquierda) y sobre el conjunto original (derecha)

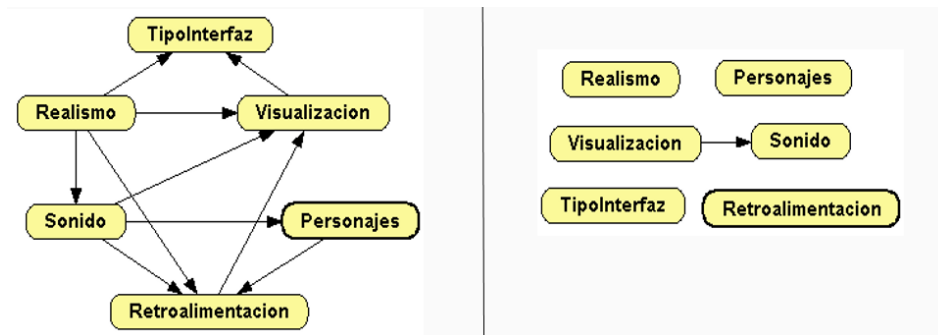


Figura 4.10: Estructura del perfil del sistema generada sobre el conjunto de datos con sobremuestreo (izquierda) y sobre el conjunto original (derecha)

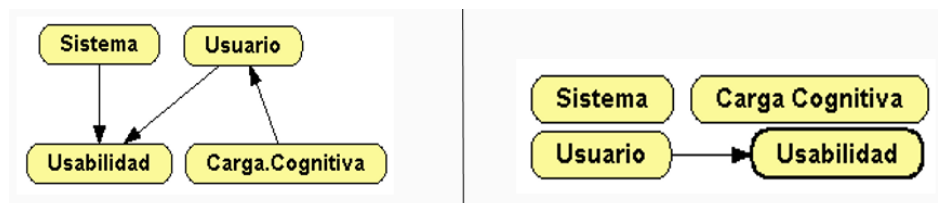


Figura 4.11: Estructura del modelo general sobre el conjunto de datos con sobremuestreo (izquierda) y sobre el conjunto original (derecha)

estructura de datos general de este modelo es independiente a los anteriores, ya que no existe relación entre las variables de cada conjunto de datos. La estructura final del modelo fue obtenida con la misma técnica de generación sobre muestras sintéticas. Los resultados entre el conjunto original y sintético se muestran en la figura 4.11. Se puede observar que las dependencias encontradas sobre el conjunto sintético son coherentes con la hipótesis inicial en comparación con el conjunto original.

4.5 Aprendizaje automático de la estructura de la red Bayesiana

Para validar la técnica de aprendizaje automático de redes bayesianas en conjuntos de datos pequeños por ajuste de datos a distribuciones multivariadas se aplicó y generó una red sobre un conjunto de datos de la literatura. El modelo corresponde al conjunto de Diagnóstico de cancer de mama de Wisconsin (WDBC) [Lichman, 2013], disponible en www.openmarkov.org.

El procedimiento fue el siguiente:

- Utilizando el conjunto de datos original se obtuvo un modelo base con el algoritmo PC bajo el software de OpenMarkov. El resultado se observa en la figura 4.12.
- Del conjunto de datos original de 569 registros se tomaron aleatoriamente 25 registros.
- Se utilizó el conjunto de datos reducido para aprender automáticamente la estruc-

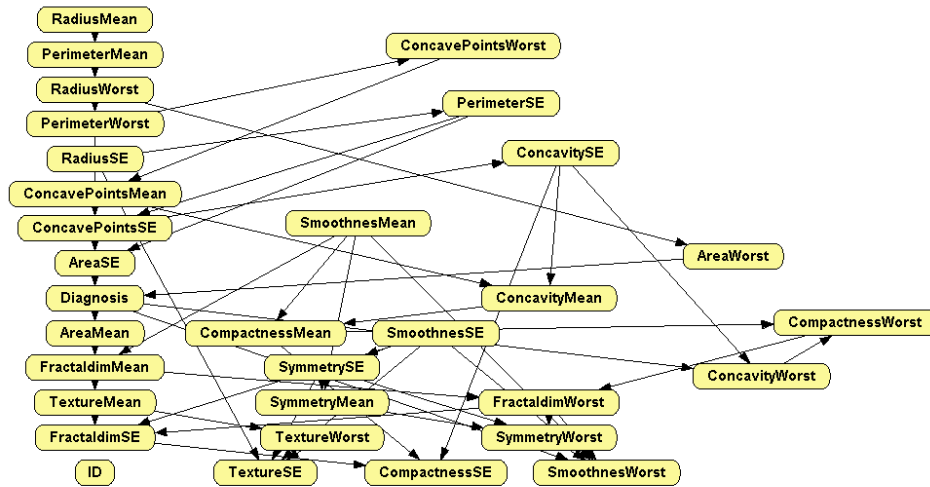


Figura 4.12: Estructura del conjunto original de datos WDBC por el algoritmo PC

tura de red bajo los mismos parámetros del algoritmo PC. El resultado se muestra en la figura 4.13.

- Sobre este conjunto se aplicó la técnica de ajuste de datos a una distribución normal multivariada y se obtuvieron 569 muestras sintéticas.
- Con el conjunto sintético se probó el algoritmo PC con los mismos parámetros que en los pasos anteriores y se obtuvo el modelo de red que se observa en la figura 4.14.

Se puede observar las similitudes en los enlaces entre variables del conjunto de datos para los casos generados con 569 muestras en contraste con el modelo obtenido con un conjunto de datos reducido. El modelo obtenido mediante la técnica de sobremuestreo resulta válido al compararse con el modelo base de este conjunto de datos.

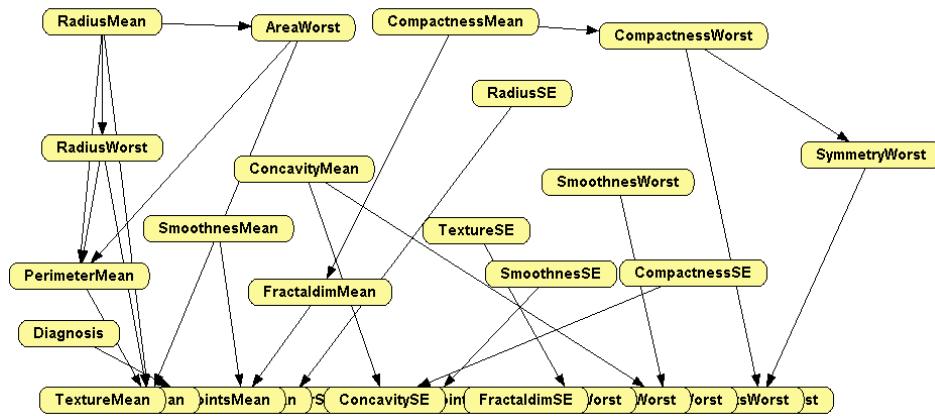


Figura 4.13: Estructura del conjunto reducido a 25 datos WDBC por el algoritmo PC

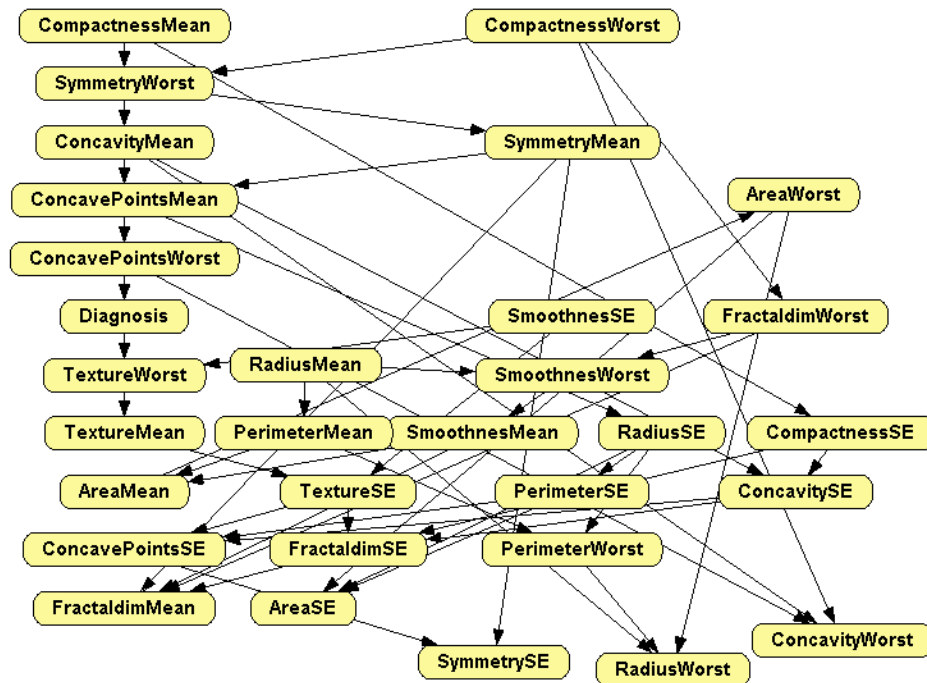


Figura 4.14: Estructura del conjunto sintético de datos WDBC por el algoritmo PC

Capítulo 5

Conclusiones

En este trabajo se propuso la integración del nivel de carga cognitiva que un usuario tiene al interactuar con un entorno virtual. El nivel fue obtenido por medio de un surrogado de actividad cognitiva, específicamente, por métricas oculares. La investigación presenta elementos para responder a las preguntas de investigación planteadas en el capítulo 1. Después de realizar los experimentos descritos en el capítulo 5 y del análisis de los datos obtenidos para cada parte del proceso de investigación se han llegado a distintas conclusiones correspondientes a las áreas de interés de este trabajo:

- **El proceso de evaluación de usabilidad de entornos virtuales puede ser modelado utilizando Modelos gráficos probabilistas.** La contribución final de esta investigación es una red bayesiana jerárquica que modela el proceso de evaluación de usabilidad de un sistema interactivo, integrando las características del perfil de usuario, las características del sistema y el nivel de carga cognitiva presente durante el momento de interacción. Este tipo de modelo pretende ser explicativo e incluso predictivo dada la naturaleza de los modelos gráficos probabilistas.
- **Integrar el nivel de carga cognitiva presente en la interacción del usuario con el sistema extiende el concepto de Usabilidad hacia un concepto más amplio de Experiencia de Usuario.** A partir de las observaciones a los resultados del estudio de usabilidad es posible concluir que los resultados de usabilidad ligados a diferentes situaciones cognitivas dan lugar a diferentes experiencias de usuario. Es decir, la integración de la carga cognitiva hace posible decidir si un resultado de usabilidad tiene un trasfondo positivo o negativo.

-
- **Es posible obtener métricas oculares para la medición de carga cognitiva utilizando dispositivos web de alta resolución.** En este trabajo de investigación el seguimiento del ojo se realizó utilizando una cámara web de alta resolución de uso doméstico. A pesar de la falta de un dispositivo apropiado para captar los cambios pupilométricos deseables en el análisis de carga cognitiva, se pudieron obtener métricas primarias (fijaciones y parpadeos) para la evaluación del nivel de carga cognitiva.
 - **Las métricas oculares adecuadas para la medición de actividad cognitiva dependen directamente de la tarea.** No existe una métrica óptima para generalizar la medición de la actividad cognitiva. Después del análisis de la literatura y de los datos obtenidos en el estudio de usabilidad realizado en este trabajo se puede concluir que el tipo de tarea a ejecutar determina el comportamiento ocular y consecuentemente las métricas ligadas a la actividad cerebral.
 - **Los algoritmos de aprendizaje automático de redes bayesianas tienen bajo desempeño con conjuntos de datos reducidos.** Los experimentos realizados en este trabajo con respecto a la generación automática de estructuras de red bayesiana con conjuntos de datos reducidos (menores a 25 entradas) resultaron en modelos carentes de dependencias entre variables o incongruentes con lo establecido por modelos válidos.
 - **Las técnicas de sobremuestreo por ajuste a distribuciones multivariadas mejoran el desempeño del aprendizaje automático de redes bayesianas.** Con los experimentos realizados en esta investigación se puede observar la mejora en el desempeño del algoritmo PC para todos los casos probados, comparando los modelos obtenidos con modelos validados en la literatura y en contraste con las pruebas realizadas sobre conjuntos reducidos.
 - **La selección del tipo de distribución multivariada para el sobremuestreo depende del tipo de variables de cada conjunto de datos.** En este trabajo

se utilizaron dos diferentes distribuciones multivariadas. Para el problema principal de esta investigación el ajuste de datos se realizó a una distribución multinomial, mientras que para el conjunto de datos WDBC la distribución que dio mejores resultados fue la distribución normal multivariada.

5.1 Trabajo Futuro

Tomando en cuenta los resultados de la presente investigación, como trabajo futuro se buscará extender la técnica de sobremuestreo aplicada a los algoritmos de aprendizaje automático de la estructura de la red bayesiana a otros dominios de aplicación.

Otro punto no explorado es la generación de una subred que seleccione automáticamente la mejor combinación de indicadores oculares con respecto a la tarea en la cual se investigará el medir la actividad cognitiva de una persona.

Es posible realizar un análisis al conjunto de datos obtenidos en el estudio de usabilidad para evaluar factores afectivos de la experiencia del usuario.

References

- [Agrafioti et al., 2012] Agrafioti, F., Hatzinakos, D., and Anderson, A. K. (2012). Ecg pattern analysis for emotion detection. *Affective Computing, IEEE Transactions on*, 3(1):102–115.
- [Alan et al., 2004] Alan, D., Janet, F., Gregory, A., and Russell, B. (2004). Human-computer interaction. *England: Pearson Education Limited*.
- [Albert and Tullis, 2013] Albert, W. and Tullis, T. (2013). *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes.
- [Alegria et al., 2015] Alegria, Y. A. M., Ordóñez, C. A. C., Saltiveri, A. G., and Huitr, A. S. (2015). Evaluating interactive systems from an emotional perspective. *Guillermo de Ockham*, 12(1):43–50.
- [Ardila, 1995] Ardila, A. (1995). Estructura de la actividad cognoscitiva: hacia una teoría neuropsicológica. *Neuropsichología Latina*, 1:21–32.
- [Bandler et al., 1982] Bandler, R., Grinder, J., and Andreas, S. (1982). Neuro-linguistic programmingTM and the transformation of meaning. *Real People, Moab*.
- [Bargas-Avila and Hornbæk, 2011] Bargas-Avila, J. A. and Hornbæk, K. (2011). Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2689–2698. ACM.
- [Bello et al., 2014] Bello, O. W., Ehira, D. N., Ayeni, J. K., Faruk, N., and Balogun, N.

-
- (2014). Towards the design of a synchronous virtual learning system. *Covenant Journal of Informatics and Communication Technology (CJICT) Vol*, 2:59–74.
- [Benel et al., 1991] Benel, D. C., Ottens, D., and Horst, R. (1991). Use of an eyetracking system in the usability laboratory. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 35, pages 461–465. SAGE Publications.
- [Bevan, 1995] Bevan, N. (1995). Usability is quality of use. *Advances in human factors ergonomics*, 20:349–349.
- [Bevan, 1998] Bevan, N. (1998). Iso 9241: Ergonomic requirements for office work with visual display terminals (vdts)-part 11: Guidance on usability. *TC*, 159.
- [Bowman et al., 2002] Bowman, D. A., Gabbard, J. L., and Hix, D. (2002). A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence: Teleoperators and Virtual Environments*, 11(4):404–424.
- [Branco et al., 2005] Branco, P., Firth, P., Encarnação, L. M., and Bonato, P. (2005). Faces of emotion in human-computer interaction. In *CHI'05 Extended Abstracts on Human factors in computing systems*, pages 1236–1239. ACM.
- [Brooke, 1996] Brooke, J. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.
- [Brooks et al., 2013] Brooks, M., Aragon, C. R., and Komogortsev, O. V. (2013). Perceptions of interfaces for eye movement biometrics. In *Biometrics (ICB), 2013 International Conference on*, pages 1–8. IEEE.
- [Carrillo et al., 2015] Carrillo, I., Meza-Kubo, V., Morán, A. L., Galindo, G., and García-Canseco, E. (2015). Emotions identification to measure user experience using brain biometric signals. In *Human Aspects of IT for the Aged Population. Design for Aging*, pages 15–25. Springer.
- [Castelhano et al., 2009] Castelhano, M. S., Mack, M. L., and Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3):6–6.

-
- [Chae et al., 2011] Chae, S., Kyung, G., and Nam, K. (2011). Measuring drivers' dynamic seating experience using pressure mats. In *Design, User Experience, and Usability. Theory, Methods, Tools and Practice*, pages 368–375. Springer.
- [Charleton and O'Brien, 2002] Charleton, S. and O'Brien, T. (2002). Measurement of cognitive states in testing and evaluation. *Handbook of human factors and evaluation*, pages 97–126.
- [Chorianopoulos and Spinellis, 2004] Chorianopoulos, K. and Spinellis, D. (2004). Affective usability evaluation for an interactive music television channel. *Computers in Entertainment (CIE)*, 2(3):14–14.
- [Crowe and Narayanan, 2000] Crowe, E. C. and Narayanan, N. H. (2000). Comparing interfaces based on what users watch and do. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 29–36. ACM.
- [Demberg, 2013] Demberg, V. (2013). Pupillometry: the index of cognitive activity in a dual-task study. In *Proceedings of the 35th annual meeting of the cognitive science society (cogsci-13)*. URL <http://www.coli.uni-saarland.de/vera/ICApaper.pdf>, volume 1.
- [Dempere-Marco et al., 2002] Dempere-Marco, L., Hu, X.-P., MacDonald, S. L., Ellis, S. M., Hansell, D. M., and Yang, G.-Z. (2002). The use of visual search for knowledge gathering in image decision support. *Medical Imaging, IEEE Transactions on*, 21(7):741–754.
- [D'Hertefelt, 2000] D'Hertefelt, S. (2000). Emerging and future usability challenges: designing user experiences and user communities. *InteractionArchitect.com*, 2.
- [Dix et al., 2004] Dix, A., Finaly, J., Abowd, G., and Beale, R. (2004). Human-computer interaction: Pearson education.
- [Ehmke and Wilson, 2007] Ehmke, C. and Wilson, S. (2007). Identifying web usability problems from eye-tracking data. In *Proceedings of the 21st British HCI Group*

-
- Annual Conference on People and Computers: HCI... but not as we know it-Volume 1*, pages 119–128. British Computer Society.
- [Finstad, 2010] Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, 22(5):323–327.
- [Fitzpatrick, 1999] Fitzpatrick, R. (1999). Strategies for evaluating software usability. *Articles*, page 1.
- [Folmer and Bosch, 2004] Folmer, E. and Bosch, J. (2004). Architecting for usability: a survey. *Journal of systems and software*, 70(1):61–78.
- [Fredericks et al., 2005] Fredericks, T. K., Choi, S. D., Hart, J., Butt, S. E., and Mital, A. (2005). An investigation of myocardial aerobic capacity as a measure of both physical and cognitive workloads. *International Journal of Industrial Ergonomics*, 35(12):1097–1107.
- [Gediga et al., 2002] Gediga, G., Hamborg, K.-C., and Düntsch, I. (2002). Evaluation of software systems. *Encyclopedia of computer science and technology*, 45(supplement 30):127–53.
- [Gil-Gómez et al., 2013] Gil-Gómez, J.-A., Gil-Gómez, H., Lozano-Quilis, J.-A., Manzano-Hernández, P., Albiol-Pérez, S., and Aula-Valero, C. (2013). Seq: suitability evaluation questionnaire for virtual rehabilitation systems. application in a virtual rehabilitation system for balance rehabilitation. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, pages 335–338. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [Granka et al., 2004] Granka, L. A., Joachims, T., and Gay, G. (2004). Eye-tracking analysis of user behavior in www search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 478–479. ACM.
- [Granollers i Saltiveri et al., 2004] Granollers i Saltiveri, T. et al. (2004). Mpiu+ a. una

metodología que integra la ingeniería del software, la interacción persona-ordenador y la accesibilidad en el contexto de equipos de desarrollo multidisciplinares.

- [Grillon et al., 2007] Grillon, H., Riquier, F., and Thalmann, D. (2007). Eye-tracking as diagnosis and assessment tool for social phobia. In *Virtual Rehabilitation, 2007*, pages 138–145. IEEE.
- [Group, 2000] Group, N. N. (2000). User experience - our definition. <https://www.nngroup.com/articles/definition-user-experience/>. Revisado: 2015-11-30.
- [Guan et al., 2006] Guan, Z., Lee, S., Cuddihy, E., and Ramey, J. (2006). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1253–1262. ACM.
- [Hart and Staveland, 1988] Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183.
- [Heckerman, 2008] Heckerman, D. (2008). A tutorial on learning with bayesian networks. In *Innovations in Bayesian networks*, pages 33–82. Springer.
- [Henderson and Hollingworth, 1998] Henderson, J. M. and Hollingworth, A. (1998). Eye movements during scene viewing: An overview. *Eye guidance in reading and scene perception*, 11:269–293.
- [Henderson and Hollingworth, 1999] Henderson, J. M. and Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 10(5):438–443.
- [Henderson et al., 2013] Henderson, J. M., Shinkareva, S. V., Wang, J., Luke, S. G., and Olejarczyk, J. (2013). Predicting cognitive state from eye movements. *PloS one*, 8(5):e64937.
- [Herskovits, 1991] Herskovits, E. (1991). *Computer-based probabilistic-network construction*. PhD thesis, Stanford University USA.

-
- [Hilera et al., 1999] Hilera, J. R., Otón, S., and Martínez, J. (1999). Aplicación de la realidad virtual en la enseñanza a través de internet. *Universidad de Alcalá, septiembre de*.
- [Holmqvist et al., 2011] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- [Holzinger, 2005] Holzinger, A. (2005). Usability engineering methods for software developers. *Commun. ACM*, 48(1):71–74.
- [Hornof et al., 2004] Hornof, A., Cavender, A., and Hoselton, R. (2004). Eyedraw: a system for drawing pictures with the eyes. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, pages 1251–1254. ACM.
- [ISO/IEC, 2001] ISO/IEC (2001). *ISO/IEC 9126. Software engineering – Product quality*. ISO/IEC.
- [Ivory and Hearst, 2001] Ivory, M. Y. and Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys (CSUR)*, 33(4):470–516.
- [John and Kieras, 1996] John, B. E. and Kieras, D. E. (1996). The goms family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3(4):320–351.
- [Kahneman, 1973] Kahneman, D. (1973). *Attention and effort*. Citeseer.
- [Kalawsky, 1999] Kalawsky, R. S. (1999). Vruse—a computerised diagnostic tool: for usability evaluation of virtual/synthetic environment systems. *Applied ergonomics*, 30(1):11–25.
- [Karat, 1997] Karat, J. (1997). User-centered software evaluation methodologies. *Handbook of human-computer interaction*, 2:689–704.
- [Kaye, 2007] Kaye, J. (2007). Evaluating experience-focused hci. In *CHI'07 extended abstracts on Human factors in computing systems*, pages 1661–1664. ACM.

-
- [Keinonen, 2003] Keinonen, T. (2003). Usabilidad de los productos interactivos.
- [Klingner et al., 2011] Klingner, J., Tversky, B., and Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology*, 48(3):323–332.
- [Kwon et al., 2003] Kwon, O.-W., Chan, K., Hao, J., and Lee, T.-W. (2003). Emotion recognition by speech signals. In *INTERSPEECH*. Citeseer.
- [Lallemand et al., 2015] Lallemand, C., Gronier, G., and Koenig, V. (2015). User experience: A concept without consensus? exploring practitioners’ perspectives through an international survey. *Computers in Human Behavior*, 43:35–48.
- [Lam and Bacchus, 1994] Lam, W. and Bacchus, F. (1994). Learning bayesian belief networks: An approach based on the mdl principle. *Computational intelligence*, 10(3):269–293.
- [Lichman, 2013] Lichman, M. (2013). UCI machine learning repository.
- [Maglogiannis et al., 2009] Maglogiannis, I., Vouyioukas, D., and Aggelopoulos, C. (2009). Face detection and recognition of natural human emotion using markov random fields. *Personal and Ubiquitous Computing*, 13(1):95–101.
- [Marshall, 2002] Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. In *Human factors and power plants, 2002. proceedings of the 2002 IEEE 7th conference on*, pages 7–5. IEEE.
- [Mayhew, 1999] Mayhew, D. J. (1999). The usability engineering lifecycle.
- [Mendoza and Novick, 2005] Mendoza, V. and Novick, D. G. (2005). Usability over time. In *Proceedings of the 23rd Annual International Conference on Design of Communication: Documenting & Designing for Pervasive Information*, SIGDOC ’05, pages 151–158, New York, NY, USA. ACM.
- [Möller et al., 2006] Möller, S., Englert, R., Engelbrecht, K.-P., Hafner, V. V., Jameson, A., Oulasvirta, A., Raake, A., and Reithinger, N. (2006). Memo: towards automatic

-
-
- usability evaluation of spoken dialogue services by user error simulations. In *INTERSPEECH*.
- [Montero et al., 2004] Montero, Y. H., Iazza, G., and Fernández, F. J. M. (2004). Diseño web centrado en el usuario: usabilidad y arquitectura de la información. *Hipertext.net*, (2):1.
- [Moreno-Ger et al., 2012] Moreno-Ger, P., Torrente, J., Hsieh, Y. G., and Lester, W. T. (2012). Usability testing for serious games: Making informed design decisions with user data. *Adv. in Hum.-Comp. Int.*, 2012:4:4–4:4.
- [Nielsen, 1994] Nielsen, J. (1994). *Usability engineering*. Elsevier.
- [Norman, 2004] Norman, D. A. (2004). Emotion design: Why we love (or hate) everyday things.
- [Parasuraman, 2003] Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science*, 4(1-2):5–20.
- [Partala and Surakka, 2003] Partala, T. and Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, 59(1):185–198.
- [Peter et al., 2007] Peter, C., Beale, R., Crane, E., and Axelrod, L. (2007). Emotion in hci. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 2*, pages 211–212. British Computer Society.
- [Petrantonakis and Hadjileontiadis, 2010] Petrantonakis, P. C. and Hadjileontiadis, L. J. (2010). Emotion recognition from eeg using higher order crossings. *Information Technology in Biomedicine, IEEE Transactions on*, 14(2):186–197.
- [Picard, 2000] Picard, R. W. (2000). Perceptual user interfaces: affective perception. *Communications of the ACM*, 43(3):50–51.
- [Pike et al., 2012] Pike, M., Wilson, M. L., Divoli, A., and Medelyan, A. (2012). Cues: Cognitive usability evaluation system. In *EuroHCIR*, pages 51–54.

-
- [Poels and Dewitte, 2006] Poels, K. and Dewitte, S. (2006). How to capture the heart? reviewing 20 years of emotion measurement in advertising. *Journal of Advertising Research*, 46(1):18–37.
- [Pons, 2012] Pons, I. P. (2012). *Terapia manual en el sistema oculomotor*. Elsevier España.
- [Rayner, 1998] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372.
- [Rizzo et al., 2007] Rizzo, M., Robinson, S., and Neale, V. (2007). Neuroergonomics: The brain at work.
- [Rognin et al., 2004] Rognin, L., Grimaud, I., Hoffman, E., and Zeghal, K. (2004). Assessing the impact of a new instruction on air traffic controller monitoring tasks. *Disponible à l'adresse http://www.eurocontrol.int/eec/gallery/content/public/documents/EEC_SSP_reports/2004/EEC_SSP_20*.
- [Rosch and Vogel-Walcutt, 2013] Rosch, J. L. and Vogel-Walcutt, J. J. (2013). A review of eye-tracking applications as tools for training. *Cognition, technology & work*, 15(3):313–327.
- [Ross et al., 2003] Ross, M. H., Kaye, G. I., and Pawlina, W. (2003). *Histology: a text and atlas: with cell and molecular biology*. Lippincott Williams & Wilkins.
- [Sánchez et al., 2008] Sánchez, J. G., Zea, N. P., Gutiérrez, F., and Cabrera, M. (2008). De la usabilidad a la jugabilidad: Diseño de videojuegos centrado en el jugador. *Proceedings of INTERACCION*, pages 99–109.
- [Sánchez and Vela, 2014] Sánchez, J. L. G. and Vela, F. L. G. (2014). Jugabilidad como medida de calidad en el desarrollo de videojuegos. In *CoSECivi*, pages 147–158.
- [Schiessl et al., 2003] Schiessl, M., Duda, S., Thölke, A., and Fischer, R. (2003). Eye tracking and its application in usability and media research. *MMI-interaktiv Journal*, 6:41–50.

-
- [Sharma and Dubey, 2014] Sharma, C. and Dubey, S. K. (2014). Analysis of eye tracking techniques in usability and hci perspective. In *Computing for Sustainable Global Development (INDIACom), 2014 International Conference on*, pages 607–612. IEEE.
- [Spirtes et al., 2000] Spirtes, P., Glymour, C. N., and Scheines, R. (2000). *Causation, prediction, and search*. MIT press.
- [Sucar, 2015] Sucar, L. E. (2015). *Probabilistic Graphical Models*. Springer.
- [Sutcliffe and Gault, 2004] Sutcliffe, A. and Gault, B. (2004). Heuristic evaluation of virtual reality applications. *Interacting with computers*, 16(4):831–849.
- [Tarta and Moldovan, 2006] Tarta, A. and Moldovan, G. (2006). Automatic usability evaluation using aop. In *Automation, Quality and Testing, Robotics, 2006 IEEE International Conference on*, volume 2, pages 84–89. IEEE.
- [Tonetto and Desmet, 2016] Tonetto, L. M. and Desmet, P. M. (2016). Why we love or hate our cars: A qualitative approach to the development of a quantitative user experience survey. *Applied Ergonomics*, 56:68–74.
- [Tsai et al., 2009] Tsai, T.-C., Chen, J.-J., and Lo, W.-C. (2009). Design and implementation of mobile personal emotion monitoring system. In *Mobile Data Management: Systems, Services and Middleware, 2009. MDM'09. Tenth International Conference on*, pages 430–435. IEEE.
- [Tsai et al., 2007] Tsai, Y.-F., Viirre, E., Strychacz, C., Chase, B., and Jung, T.-P. (2007). Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, space, and environmental medicine*, 78(Supplement 1):B176–B185.
- [Van Gog et al., 2009] Van Gog, T., Kester, L., Nieuwstein, F., Giesbers, B., and Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior*, 25(2):325–331.
- [Van Orden et al., 2001] Van Orden, K. F., Limbert, W., Makeig, S., and Jung, T.-P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Hu-*

-
- man Factors: The Journal of the Human Factors and Ergonomics Society*, 43(1):111–121.
- [Vermeeren et al., 2010] Vermeeren, A. P., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J., and Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pages 521–530. ACM.
- [Wernaart, 2013] Wernaart, G. (2013). Cognitive load measurement: Different instruments for different types of load?
- [Wharton et al., 1994] Wharton, C., Rieman, J., Lewis, C., and Polson, P. (1994). The cognitive walkthrough method: A practitioner’s guide. In *Usability inspection methods*, pages 105–140. John Wiley & Sons, Inc.
- [Woodward, 1998] Woodward, B. (1998). Evaluation methods in usability testing. *CS5326*. Disponible en <http://web.archive.org/web/20030213050921/www.swt.edu/~hd01/5326/projects/BWOODWARD.HTML>.
- [Zenk et al., 2008] Zenk, R., Franz, M., and Bubb, H. (2008). Emocard—an approach to bring more emotion in the comfort concept. *SAE International Journal of Passenger Cars-Mechanical Systems*, 1(2008-01-0890):775–782.
- [Zyda, 2005] Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer*, 38(9):25–32.