

Body Posture Visualizer to Support Multimodal Learning Analytics

R. Munoz, *Member, IEEE*, T. S. Barcelos, R. Villarroel, R. Guíñez, E. Merino

Abstract—Learning analytics consists of gathering and analyzing data from students in order to understand complex aspects of the learning process and promote its improvement. Currently, to the best of our knowledge, there is a lack of tools aimed at displaying multimodal data in an integrated way for general purpose analysis. In this paper, we present a free software tool based on the Microsoft Kinect sensor for automatic capture, identification, and visualization of ten body postures for posterior analysis. It is also possible to incorporate the identification of new postures if necessary. Learning and recognition is based on the AdaBoost algorithm. Posture recognition reached accuracy rates as high as 80% for 8 of the 10 identified postures. Concerning the software usability, a heuristic evaluation with three specialists was performed, as well as a usability test with five volunteer students. Results indicated that the software interface, based on the metaphor of a video editor, may allow its effective use by end users, though some adjustments are still necessary, such as the terminology used in some commands and the help system.

Index Terms—Multimodal Learning Analytics, Kinect, Visualizer

I. INTRODUCCIÓN

EL desarrollo y evaluación de las competencias profesionales (también denominadas blandas), tales como el trabajo en equipo, la colaboración y comunicación de manera eficiente, son claves para un buen desempeño laboral [1]. El desarrollo de estas competencias, han sido puestas como prioridad en Agencias Acreditadoras de Educación en Ingeniería como lo son ABET (*Accreditation Board for Engineering and Technology*) o INQAAHE (*International Network for Quality Assurance Agencies in Higher Education*). Esto debido a que el entorno laboral requiere profesionales altamente competitivos, que posean las habilidades para enfrentar los nuevos desafíos. Sin embargo, el desarrollo y evaluación de estas competencias son un desafío para las instituciones de educación superior [2].

R. Munoz, Escuela de Ingeniería Informática, Facultad de Ingeniería, Universidad de Valparaíso, Chile. roberto.munoz@uv.cl

T. S. Barcelos, Laboratório de Tecnologias para Interação – Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Brasil. tsbarcelos@ifsp.edu.br R. Villarroel, Facultad de Ingeniería, Pontificia Universidad Católica de Valparaíso, Chile. rodolfo.villarroel@pucv.cl

R. Guíñez, Departamento de Desarrollo, Consultora, Tecnología e Innovación Fisfor Ltda., Chile. rodolfo.guinez@fisfor.cl

E. Merino, Facultad de Ingeniería, Universidad de Valparaíso, Chile. erick.merino@postgrado.uv.cl

La razón de esto se debe principalmente a dos aspectos. El primero, está asociado a que el docente, al menos los primeros años, generalmente poseen cursos numerosos y estudiantes heterogéneos en términos de habilidades blandas, y el segundo, porque la evaluación de estas competencias se debe realizar en los entornos de enseñanza aprendizaje donde se desarrollan estas competencias, y no a través de evaluaciones estandarizadas, como lo son los exámenes escritos o de selección múltiple [3].

Dados los desafíos mencionados anteriormente, es que surge la Analítica del Aprendizaje, la cual es la medición, recopilación, análisis y presentación de datos sobre los estudiantes, sus contextos y las interacciones que allí se generan, con el fin de comprender el proceso de aprendizaje que se está desarrollando y optimizar los entornos en los que se produce [4]. Una de las aristas de la analítica del aprendizaje se basa en el lenguaje no verbal, donde el aprendizaje se analiza mediante gestos, posturas corporales, sonidos y otros. A esta arista, se le conoce como Analítica Multimodal del Aprendizaje (*Multimodal Learning Analytics* – MLA), la cual pretende mantener la riqueza y la naturaleza altamente contextualizada del análisis cualitativo tradicional mediado por personas, pero con los beneficios añadidos de cuantificar esos datos en nuevas formas; y aprovechar sensores innovadores para capturar datos que no son fácilmente perceptibles a través de visión humana [5].

Por otra parte, los avances tecnológicos que ha provisto la informática en este último tiempo, ha sido de gran utilidad para la MLA, como por ejemplo la gestión de grandes cantidades de datos utilizando técnicas de Big Data y Minería de Datos [6]. A pesar de las grandes ventajas que entregan las tecnologías de la información, las investigaciones y desarrollos de la MLA siguen utilizando técnicas temporales, las cuales son útiles para aquellos que poseen experiencia en este contexto, no traspasándose al potencial educador.

Por lo antes expuesto, en este trabajo se presenta el diseño y desarrollo de una herramienta que permite la visualización y clasificación de posturas corporales utilizando la Microsoft Kinect. La herramienta propuesta (*Lelikëlen – ojos abiertos en Mapudungún*), permite detectar, almacenar y visualizar posturas corporales de los sujetos grabados, dando las opciones de agregar posturas personalizadas, exportar e importar escenas y un visualizador de las posturas detectadas en función del tiempo. Además, la herramienta propuesta permite exportar los datos generados para su utilización con herramientas de minería de datos.

El presente artículo se estructura de la siguiente forma. En

la sección II se presentan los trabajos relacionados. El diseño de la solución como su implementación son presentados en la sección III. En la sección IV es presentada la validación de la herramienta desarrollada, para finalizar, en la sección V son presentadas las conclusiones y trabajo futuro.

II. TRABAJOS RELACIONADOS

Con el objetivo de tener una visión general del estado actual de las investigaciones en el área de esta propuesta se realizó una búsqueda bibliográfica en las principales bases de datos asociadas a Ciencias de la Computación, IEEE (*IEEE Xplore*), ACM (*ACM Digital Library*) y Elsevier (*ScienceDirect*). Para ello se definió la cadena de búsqueda (“*learning analytic**” AND “*kinect*”). Se obtuvieron 14 resultados, los cuales se distribuyeron de la siguiente forma IEEE (3), ACM (5), ScienceDirect (6). Tras una lectura de los resultados provistos, ninguno de los trabajos obtenidos presentaba el desarrollo de una aplicación que buscara transferir la investigación a un público de mayor alcance. A continuación, se presentan aquellos que consideramos que contribuyen al área de aplicación.

A. Evaluación en la construcción de una torre

Worsley y Blikstein [7] presentan la evaluación del procedimiento que desarrollaban los estudiantes al construir una torre a mano. El trabajo se basa principalmente en el análisis de datos gestuales de 13 participantes por medio de la Microsoft Kinect. A los participantes se les facilitó materiales del día a día (palos de helado, cinta adhesiva, etc.) para que construyeran la torre más alta posible y que soportara una masa aproximada de 3 libras (1.3 kg). Las acciones realizadas por los participantes se clasificaron en *BUILD*, *PLAN*, *TEST*, *ADJUST* y *UNDO*. Tras el análisis se clasificó a los estudiantes en expertos y novatos en la manipulación de objetos.

B. Estimación de Habilidades de Presentación

En [8] se presenta un trabajo en la que se evalúa la presentación individual oral de 448 estudiantes. Para la captura de los datos se utilizó la Microsoft Kinect. Los criterios de análisis se basaron en el contacto visual y posturas corporales, en los cuales se logró una efectividad del 68% y 63% respectivamente. El objetivo de esta investigación fue identificar cuáles eran las características no verbales de los estudiantes y su influencia en una predicción positiva sobre el desarrollo de habilidades de presentación oral. Finalmente, los autores indican que los resultados pueden ser mejorados para luego desarrollar una herramienta donde el objetivo sea medir el desempeño de los estudiantes dándoles una retroalimentación rápida y concisa.

C. Análisis de Presentaciones Multimodales

Leong et al. [9] proveen una visión general sobre los trabajos relevantes sobre investigación multimodal en tareas y actividades educativas. Además, resumen su trabajo utilizando sensores de profundidad (utilizando Microsoft Kinect) en la generación de mediciones de habilidades de

comunicación, en particular presentaciones en público. Para finalizar los autores también introducen a un paquete *open-source* generado por ellos desarrollado en Python el cual permite facilitar la medición de lenguaje corporal.

D. Diseño de un sistema de reconocimiento de posturas utilizando el sensor Kinect

En [10] los autores utilizan el sensor 3D de la Microsoft Kinect con el objetivo de reconocer posturas corporales y así facilitar la interacción humano-robot. Para facilitar el reconocimiento los autores utilizan las posturas en base al alfabeto utilizado con banderas (*semaphore alphabet*). Para el desarrollo utilizan el SDK de Kinect con una red neuronal (*self-organizing map*) [11] entrenada con un algoritmo no supervisado. El sistema propuesto, después de reconocer la postura corporal y contrastarla con un crawl de información sobre esqueletos, envía un comando (potencialmente para un robot), sin embargo, los resultados obtenidos todavía son preliminares.

III. DISEÑO E IMPLEMENTACIÓN DE UN VISUALIZADOR DE POSTURAS CORPORALES

En esta sección se presenta la arquitectura e interfaces del sistema desarrollado, junto con las tecnologías utilizadas. Para ello, en la Figura 1, se presenta un esquema de alto nivel el cual permite facilitar el esquema de funcionamiento de *Lelikēlen*.

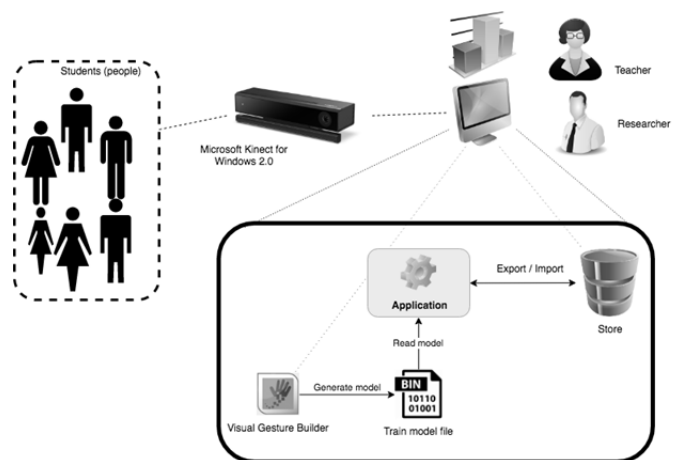


Fig. 1. Esquema de funcionamiento de alto nivel

Como se visualiza en la Figura 1, el sensor (Microsoft Kinect) es capaz de reconocer hasta seis personas simultáneamente. La Kinect envía los *frames* capturados mediante eventos a la aplicación, la cual con sus componentes internos es capaz de detectar las posturas realizadas por los sujetos en cada instante de tiempo, grabando y guardando la información obtenida en una capa de almacenamiento representada por el componente *Store*. Esta capa permite gestionar las rutas de los archivos “*Train model File*” y exportar e importar escenas. Internamente la aplicación contiene su propio “*Train model file*” base, el cual posee entrenamiento de 10 posturas predeterminadas. Para poder

agregar gestos personalizados a la aplicación, se debe utilizar el componente *Visual Gesture Builder*, generando el archivo “*Train model File*”, el cual debe ser registrado en la aplicación para su posterior uso. Esta flexibilidad permite que el software pueda ser utilizado por distintos tipos de usuarios, como investigadores, académicos, profesores, etc.

A continuación, en la sección A, se explica el funcionamiento en detalle de la aplicación desarrollada.

A. Detalle de la interacción de componentes del Visualizador

La aplicación desarrollada, utiliza el sensor de profundidad Microsoft Kinect para Windows 2.0, la cual captura un máximo de 6 sujetos en paralelo a una distancia máxima de 4.5 metros (14.7 pies). La aplicación utiliza el framework *KinectEx*, la cual a través del componente *KinectMedia* recibe cada *frame* generado por el sensor mediante un sistema de eventos. Con esta misma dependencia se identifican y grafican los esqueletos de cada sujeto. Una vez que el usuario selecciona grabar, el componente “*Recorder*” comienza a recibir los *frames*, y a almacenarlos en el archivo temporal “*Temp KDVR*”, bajo el formato establecido por *KinectEx*. Al mismo tiempo, “*Monitor*” utilizando al componente “*Gesture Detector*”, va detectando e identificando las posturas realizadas por los sujetos analizados, y almacenándolas en memoria principal. Una vez que termina el proceso de grabación, estos datos se almacenan en el archivo temporal “*Temp Scene Data*”, utilizando tecnología *SQLite*. Luego cuando se reproduce la escena previamente grabada, se invoca al componente *Player*, el cual lee ambos archivos temporales, y grafica en la línea de tiempo las posturas que contiene *Temp Scene Data*, y utilizando *KinectEx* reproduce la escena almacenada en *Temp KDVR*, graficando los esqueletos según el color asignado a cada persona. Las posturas en la línea de tiempo tienen el mismo color que el sujeto al que pertenecen. Para exportar la escena grabada, se invoca al componente *IO Helper*, el cual consolida ambos archivos temporales en uno solo, y lo almacena en una ruta definida por el usuario. Para importar escenas, se utiliza el mismo componente, el cual extrae de “*Scene File*” los archivos temporales, para luego ser utilizados por el componente “*Player*”.

Un esquema de la orquestación de los componentes que integran la aplicación es presentado en la Figura 2.

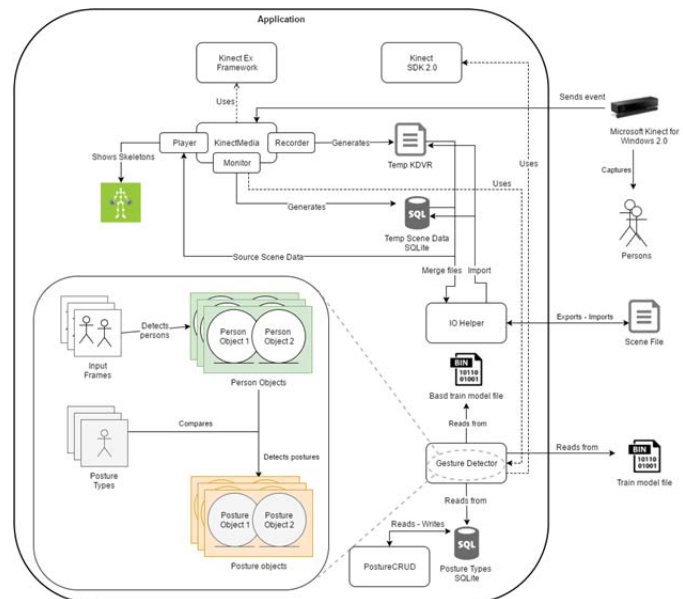


Fig. 2. Orquestación de componentes aplicación *Leikikelen*

Para poder visualizar las posturas corporales en la línea de tiempo, fue necesario implementar un sistema de intervalos de tiempo, los cuales se generan una vez terminada la grabación. Este proceso, que es llevado por *Gesture Detector*, es de suma importancia, ya que, al grabar, sólo se puede capturar la postura de un instante de tiempo específico, por lo tanto, para poder obtener una visualización clara de los eventos realizados por los sujetos, es necesario realizar una transformación de las posturas en todos los instantes de tiempo a intervalos definidos.

Para cada tipo de postura se realiza el algoritmo (Fig. 3), donde se recorre cada *MicroPosture* (postura detectada un instante), omitiendo las iteraciones donde el tipo de postura sea distinto al que se está buscando. El intervalo está conformado por un tiempo inicial, un tiempo final y un tipo de postura (*PostureType*). En la primera iteración se define el tiempo inicial y el final con el mismo instante de tiempo que posee la primera *MicroPosture*. Luego se continúa recorriendo las micro posturas hasta que se determine un instante de tiempo alejado del instante anterior con una distancia mayor a la del umbral permitido (definido en 2,0 segundos). Esto quiere decir que, si en 2,0 segundos no se detectó la postura que el algoritmo está buscando, entonces se cerrará el intervalo (asignando el tiempo final en el instante de tiempo anterior) y se procederá a graficar. Este proceso se repite, hasta recorrer todas las micro posturas asociadas cada sujeto. A continuación, en la Figura 3, se presenta el algoritmo asociado al recorrido de las micro posturas.

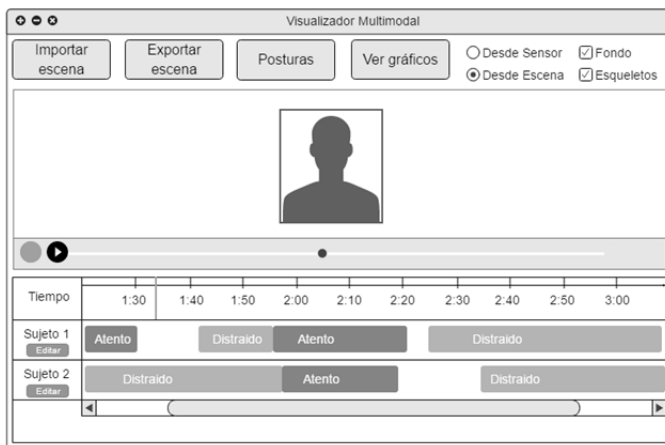
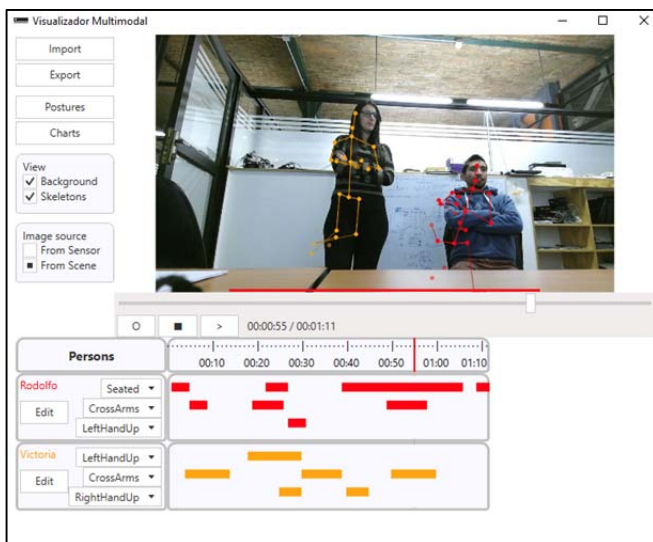
```
foreach (MicroPosture microPosture in person.MicroPostures)
{
    if (microPosture.PostureType.Name != currentPostureType.Name)
        continue;
    if (microPosture.GestureType == GestureType.Continuous)
        continue;
    if (lastMicroPosture == null)
    {
        initialMicroPosture = microPosture;
    }
    else if (microPosture.SceneLocationTime.Subtract(lastMicroPosture.SceneLocationTime) >= threshold)
    {
        postureIntervalGroup.addInterval(initialMicroPosture, lastMicroPosture);
        initialMicroPosture = microPosture;
    }
    lastMicroPosture = microPosture;
}
```

Fig. 3. Algoritmo reconocimiento micro posturas

B. Diseño de interfaces de usuario

El diseño de las interfaces de usuario es de gran relevancia para la aprobación y comprensión de la aplicación [12]. Esto debido a que afectan directamente en los resultados de la manipulación que tenga el usuario sobre ella, siendo un elemento determinante en la interacción con la aplicación [13].

Para el diseño de las interfaces se generaron *wireframes*, los cuales fueron utilizados por 5 profesionales del área de educación y así determinar quiebres comunicacionales en su utilización [14]. Una vez validadas las interfaces se procedió a su implementación final. Un ejemplo del proceso de pasar de *wireframe* a final es el presentado en las Figuras 4 y 5 respectivamente.

Fig. 4. Wireframe Analizador Multimodal (*Lelikëlen*)Fig. 5. Interfaz Principal Analizador Multimodal (*Lelikëlen*)

Es importante mencionar que la interfaz final varía en ligeros aspectos con respecto al prototipo presentado en la Figura 4, ya que se realizaron modificaciones menores posterior a la validación con los potenciales usuarios (entre ellos la posibilidad de exportar a SQL). Las funcionalidades del Analizador Multimodal desarrollado, se presenta en la sección C.

C. Funcionalidades Analizador Multimodal

Se desarrolló *Lelikëlen* con el objetivo de ampliar el número de potenciales usuarios, centrándose en el desarrollo de una interfaz usable, con funcionalidades claramente definidas.

Como se puede visualizar en la figura 5, la interfaz presenta 7 funcionalidades principales, las cuales se presentan a continuación.

1. *Import*: Permite importar un archivo en formato propio de extensión MVS (*Multimodal Visualizer Scene*) que ha sido grabado por el usuario para su posterior análisis.
2. *Export*: Permite exportar un archivo en formato MVS que ha sido grabado por el usuario para su análisis.
3. *SQL*: Permite exportar los datos generados y clasificados a un archivo formato SQL. Esta funcionalidad fue agregada para aquellos profesionales que deseen realizar otro tipo de análisis con los datos generados.
4. *Postures*: Permite ver qué posturas se encuentran entrenadas en la aplicación, además de permitir cargar nuevas posturas en archivos (extensión *gdb*) generados por el *Visual Gesture Builder* (VGB), la cual es una herramienta de la SDK de Microsoft XBOX.
5. *Charts*: Permite visualizar datos estadísticos asociado al video analizado. Entre estos se encuentran tiempo por persona en cada postura, gráficos por posturas, entre otras.
6. *View*: Permite seleccionar si el usuario quiere o no visualizar el fondo y esqueleto detectado por la Microsoft Kinect.
7. *Image Source*: Permite seleccionar cuál será la fuente de entrada para el análisis multimodal. Este puede ser un archivo MVS o del sensor Kinect.

D. Conjunto de posturas predefinidas

La herramienta desarrollada incluye un conjunto de 10 posturas predefinidas (clasificadores). Estas posturas se basan principalmente en las identificadas por [8], [15]. Las posturas generadas y entrenadas son: 1) Apuntando; 2) Brazos cruzados; 3) Mano arriba (ej: pregunta); 4) Mano en la cabeza; 5) Manos abajo; 6) Explicando con manos abiertas; 7) Manos en la cintura; 8) Mano en la cara (ej: mano en el mentón); 9) Explicando con una mano; y 10) Sentado. Las posturas sentado y mano arriba, fueron agregadas, ya que son de potencial utilidad para investigadores del área de educación, aunque no sean posturas comunes al momento de presentar.

Un ejemplo de cada postura se presenta en la Figura 6.

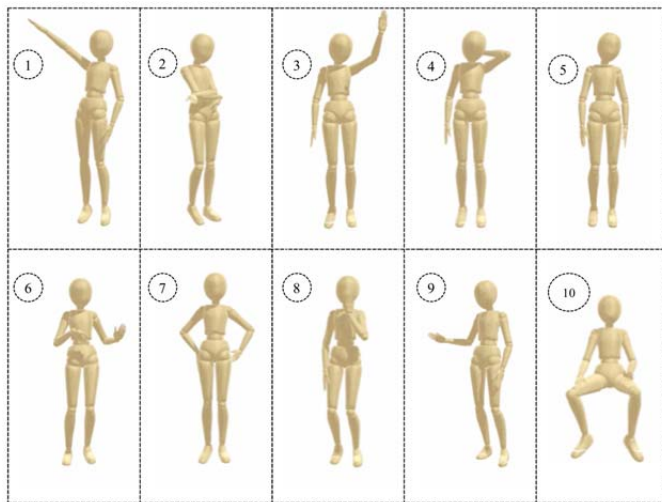


Fig. 6. Ejemplo posturas asociadas a clasificadores

Entrenamos y etiquetamos cada clasificador utilizando el VGB. Los sujetos que participaron en la generación de las escenas fueron de contexturas Ectomorfas, Endomorfas y Mesomorfas. Se utilizaron los 3 tipos de contexturas con el objetivo de generar clasificadores de amplia aplicación.

Por otra parte, VGB utiliza características tales como ángulos relativos formados por las articulaciones y movimiento de las articulaciones para aprender y generar los clasificadores utilizando el algoritmo AdaBoost [16]. La elección de AdaBoost frente a otros algoritmos de clasificación, se debió principalmente a que ha sido altamente utilizado (de manera positiva) para clasificación categórica de posturas simples [17].

Cada clasificador fue entrenado utilizando AdaBoost con 1.000 clasificadores débiles. Se ignoró en 9 de las posturas las articulaciones bajo la cintura para que la herramienta no discriminara si estaba sentado o parado, y sólo se utilizaron las articulaciones bajo la cintura al momento de entrenar el clasificador sentado. La Tabla I presenta las estadísticas calculadas, en términos de exactitud y precisión [18] de nuestros clasificadores. Para el cálculo, se utilizó la herramienta de análisis que provee el VGB. Es importante mencionar que se utiliza para el proceso un conjunto de frames de entrenamiento (entto) y uno de prueba para así evaluar el clasificador generado.

TABLA I
EXACTITUD Y PRECISIÓN CLASIFICADORES GENERADOS

#	Postura	#frames entto	#frames prueba	Exactitud (accuracy)	Precisión (precision)
P1	Apuntando	13.155	5.370	82,16%	73,57%
P2	Brazos cruzados	27.574	5.488	91,63%	86,89%
P3	Mano arriba	11.178	4.820	82,76%	82,57%
P4	Mano en la cabeza	17.507	8.191	84,01%	89,19%
P5	Manos abajo	26.468	5.724	80,29%	77,42%
P6	Exp. c/manos	20.586	4.623	90,77%	85,14%

abiertas					
P7	Manos en la cintura	16.150	6.396	82,96%	70,20%
P8	Mano en el mentón	13.847	5.289	72,40%	95,55%
P9	Exp. c/una mano	14.210	5.318	77,96%	68,04%
P10	Sentado	86.300	4.541	94,71%	90,24%

Un aspecto de importancia a evidenciar, es que la exactitud y precisión de los clasificadores binarios generados fue calculada con sujetos distintos a los de entrenamiento. Por ejemplo, con el clasificador Brazos cruzados se obtuvo una exactitud de 91,63% y una precisión del 86,89% al momento de clasificar con *frames* de prueba. Por otra parte, al visualizar en la Tabla I se puede evidenciar que el clasificador con menor exactitud y precisión fue el asociado la postura “explicando con una mano”, con un 77,96% y 68,04% de exactitud y precisión respectivamente. Sin embargo, esto sigue siendo positivo, puesto que la clasificación no requiere de una persona presente para su identificación, tal como sucedería con un análisis manual.

IV. VALIDACIÓN DE LA HERRAMIENTA

Para la validación de la herramienta se consideraron 2 tipos de pruebas. La primera fue asociada a la clasificación de posturas y la segunda a la facilidad de uso. A continuación, se presenta el diseño y resultado de ambas evaluaciones.

A. Evaluación Clasificación de Posturas

Para la evaluar la clasificación de posturas se diseñó una prueba la cual contempló una evaluación con 6 estudiantes en un ambiente real. Los sujetos fueron estudiantes (hombres) de Ingeniería en Informática de primer año de carrera, y que no pertenecían al conjunto de prueba ni entrenamiento.

La cámara fue ubicada frente a la tarima de presentación de 4.0 metros de ancho, a una distancia de 4.1 metros. En la Figura 7 se presenta un esquema e imagen del escenario utilizado. Un aspecto de alta importancia a evidenciar es que las grabaciones fueron realizadas en un ambiente no controlado, esto quiere decir con fondo sin contraste que facilite la detección e iluminación estándar (como se visualiza en la Figura 7). Esto fue realizado con el objetivo de evidenciar la potencialidad de uso de estas tecnologías en una sala de clases real.

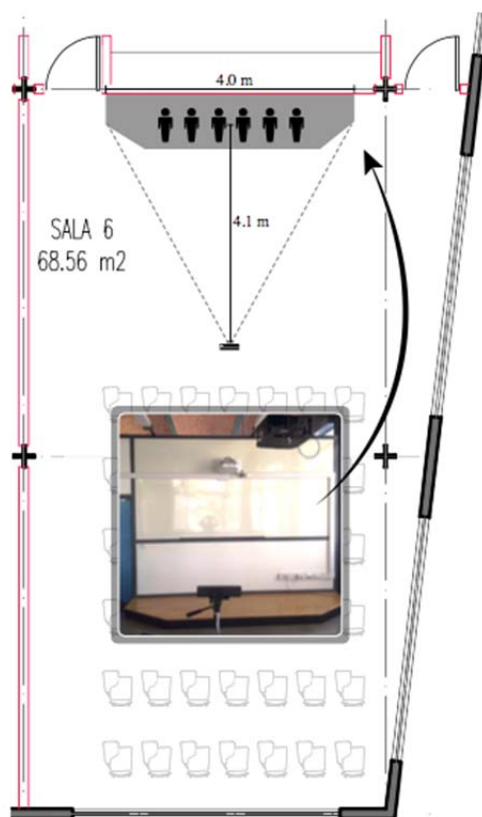


Fig. 7. Escenario de prueba (sala de clases real)

Cada estudiante debió ejecutar una secuencia de 10 posturas bases en 10 ocasiones (presentadas en la sección III punto D). Esto debió ser realizado en un intervalo máximo de 15 minutos. Es decir, cada estudiante debía ejecutar 100 posturas, clasificando de manera automatizada un total de 600 posturas (60 posturas por clasificador).

Un aspecto importante de mencionar es que los sujetos que realizaron las posturas no pertenecían al conjunto de entrenamiento. A continuación, en la Tabla II, se presenta la matriz de confusión al momento de clasificar de manera automatizada las posturas realizadas por los participantes.

TABLA II
MATRIZ DE CONFUSIÓN POSTURAS

#	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	50	0	1	0	0	0	0	2	1	0
P2	0	60	0	0	0	0	0	0	0	0
P3	3	0	47	1	0	0	0	0	5	0
P4	2	0	5	53	0	0	0	0	0	0
P5	0	0	0	0	59	0	0	0	0	0
P6	0	0	0	0	0	59	0	0	1	0
P7	0	0	0	0	4	0	55	0	1	0
P8	0	0	8	0	0	0	0	44	0	0
P9	5	0	0	0	0	0	0	1	49	0
P10	0	0	0	0	0	0	0	0	0	60

La matriz de confusión permite evidenciar que los clasificadores incorporados en *Lelikelen*, de acuerdo con la

evaluación realizada, clasifican de manera satisfactoria todas las posturas. Por una parte, en las posturas P8 y P3 y P5 presentan los menores índices de detección, sin embargo, éstos superan el 70%. Por ejemplo, la postura P8 (Mano en el mentón), un 73% fue clasificada de manera correcta, y un 13% fue clasificada de manera errónea. El caso en que fue clasificada de manera incorrecta se debió a que fue categorizada como la postura P3 (Mano arriba). Por otra parte, tan sólo en un 13% de las ocasiones esta postura no pudo ser asignada a una de las posturas predefinidas.

Para finalizar, al visualizar la matriz de confusión, es posible evidenciar que en ningún caso las posturas clasificadas de manera incorrecta por la herramienta fueron radicalmente diferentes, ya que todas seguían algún patrón base (ej: apuntando cuando que debiese haber sido clasificada como mano arriba).

B. Evaluación Usabilidad Herramienta

Para la evaluación de la herramienta se realizaron inspecciones y pruebas de usabilidad. Las primeras tuvieron como objetivo identificar potenciales problemas de usabilidad, por medio de evaluadores expertos en el área y la segunda tuvo como objetivo evidenciar problemas mientras los usuarios utilizaban la herramienta. A continuación, en los puntos 1 y 2 se presenta los resultados de ambas evaluaciones.

1) Evaluación heurística

Con el objetivo de complementar la validación de la herramienta, se realizó una evaluación heurística de la interfaz persona-computador de la aplicación. La evaluación heurística es una técnica de evaluación de usabilidad por medio de la inspección de la interfaz por especialidades que buscan identificar de forma rápida problemas que puedan potencialmente afectar la interacción del usuario [19]. La utilización de esta técnica es relevante considerando que se pretende que la aplicación sea utilizada en poco tiempo por investigadores y profesores interesados en técnicas de análisis multimodal.

Participaron de la evaluación tres especialistas, los cuales tienen en promedio más de 5 años de experiencia con la realización de evaluaciones heurísticas. Los tres especialistas trabajaron de manera individual, y para cada una de las diez heurísticas de usabilidad propuestas por Nielsen [19], dieron una calificación de 1 a 5, donde 1 indica una menor conformidad de la aplicación con la heurística y 5 indica total conformidad de la aplicación con la heurística. Además, fue solicitado que cada evaluador escribiese una breve justificación para la calificación indicada, inclusive indicado potenciales problemas de usabilidad vinculados a la heurística.

Para finalizar, en la Tabla IV, son presentados el promedio y la desviación estándar de las calificaciones asociadas a cada heurística.

TABLA IV
RESULTADOS EVALUACIÓN HEURÍSTICA

Heurística	Descripción	Promedio	Desv. Estándar
H1	Visibilidad del estado del sistema	4,33	0,58
H2	Utilizar el lenguaje de los usuarios	3,67	1,15
H3	Control y libertad para el usuario	4,67	0,58
H4	Consistencia y estándares	5,00	0,00
H5	Prevención de errores	3,33	1,15
H6	Minimizar la carga de la memoria del usuario	4,67	0,58
H7	Flexibilidad y eficiencia de uso	4,00	1,00
H8	Diseño estético y minimalista	4,33	1,15
H9	Ayudar a los usuarios a reconocer, diagnosticar y recuperarse de los errores	2,67	0,58
H10	Ayuda y documentación	2,00	1,00

En la Figura 8, se presenta un gráfico de radar resumiendo los promedios obtenidos en el cumplimiento de las 10 heurísticas.

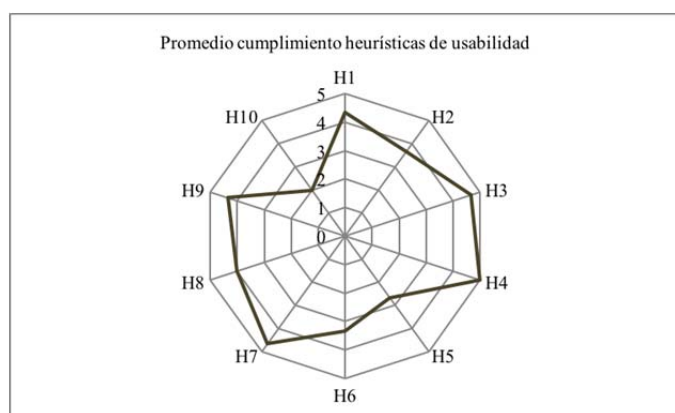


Fig. 8. Gráfico cumplimiento heurísticas de usabilidad

Los resultados de la evaluación heurística no deben ser tomados como un diagnóstico final de la calidad de la usabilidad del sistema, sino como un resultado que debe retroalimentar el proceso de desarrollo de su interfaz. Verificamos que las calificaciones más altas, fueron asociadas a las heurísticas “H4 - Consistencia y estándares” (5,00), “H6 - Minimizar la carga en memoria del usuario” (4,67) y “H3 - Control y libertad para el usuario” (4,67). Los principales comentarios de los evaluadores se refieren al hecho de que la aplicación sigue los estándares para interfaces en ambiente Windows y el uso de la metáfora de un editor de video con múltiples indicadores para su control, que probablemente será fácilmente identificado por los usuarios.

Por otro lado, las evaluaciones más bajas se refieren a las heurísticas “H10 - Ayuda y documentación” (2,00), “H9 - Ayudar a los usuarios a reconocer, diagnosticar y recuperarse de los errores” (2,67) y “H5 - Prevención de errores” (3,33). Esto se debe a la falta de un sistema de ayuda integrado a la interfaz, así como también el uso de terminologías poco usuales para los usuarios no habituados con el uso del sensor de la Microsoft Kinect y sus bibliotecas asociadas (por ejemplo “*from scene*” y “*from sensor*”). Tales ocurrencias, como los demás potenciales problemas encontrados en la evaluación, serán evaluados como tales en el uso de la herramienta mediante una prueba de usabilidad con usuarios.

2) Pruebas con usuarios

Se realizó una prueba de usabilidad con potenciales usuarios de la herramienta. Los participantes fueron 5 estudiantes universitarios que no conocían la herramienta ni su propósito. Su edad fluctuó entre 24 y 28 años (con un promedio de 25,6 años y desviación estándar de 1,52 años). Cuatro de los participantes eran de sexo masculino y una de sexo femenino.

A los participantes, de manera individual, se les realizó una presentación general de 5 minutos del propósito de la herramienta y luego, se les invitó a participar. Aquellos que participaron, debieron firmar un consentimiento informado de acuerdo a las recomendaciones de la Organización Mundial de la Salud.

Posteriormente, se les presentó una lista de 11 tareas de que debían realizar con la herramienta (disponible en la Tabla V). Esto fue realizado con el fin de evaluar los tiempos utilizados y tasa de errores por tarea [20]. Para el registro se utilizó el software Camtasia Studio [21], esto con el fin no interrumpir al usuario mientras realizaba la actividad.

TABLA V
LISTA DE TAREAS PRUEBA DE USABILIDAD

ID	Descripción
T1	Grabar escena con Microsoft Kinect (incluye detener la grabación)
T2	Agregar postura (nombre de 6 letras)
T3	Editar postura (nombre de 6 letras)
T4	Eliminar postura
T5	Visualizar gráficos (persona, grupo)
T6	Importar escena
T7	Visualizar escena (reproducir, detener, pausar, adelantar y retroceder)
T8	Modificar datos persona (nombre 6 letras, masculino, edad de 2 dígitos)
T9	Seleccionar posturas a visualizar en la línea de tiempo (2 posturas)
T10	Exportar escena (nombre de 8 letras)
T11	Exportar a SQL (nombre de 8 letras)

Se estimó el tiempo esperado de cada tarea por medio del modelo *Keystroke level* [22]. Este modelo permite predecir cuánto tiempo demoraría un experto en realizar una tarea sin errores utilizando un computador. Por ejemplo, para T5 *Visualizar gráficos* el usuario debería:

- Buscar el botón charts (*mental* = 1.2 s)
- Apuntar hacia el botón charts (*point* = 1.1 s)
- Presionar sobre el botón charts (*button* = 0.1 s)
- Soltar el mouse (*button* = 0.1 s)

Por tal razón, de acuerdo con este modelo el tiempo que tomaría al usuario sería:

$$mpbb = 2.5 s \quad (1)$$

Si bien los usuarios que ejecutaron las actividades no eran expertos en el uso de la herramienta, se utilizó esta métrica para contrastarlo con un usuario novato, puesto que de esta forma se puede evidenciar la facilidad de aprendizaje de la herramienta. Por otra parte, las métricas obtenidas después de ejecutada la prueba son presentadas en la Tabla VI.

TABLA VI

MÉTRICAS PRUEBA DE USABILIDAD

ID	Tiempo (segundos)			Errores (total)	
	Promedio	D Estándar	Esperado	Número	Ayuda
T1	15,5	5,2	5.0	4	4
T2	33,3	11,1	21.0	3	4
T3	15,6	6,8	16.4	0	0
T4	9,6	1,4	7.5	0	0
T5	11,0	3,4	2.5	3	0
T6	10,1	3,3	7.5	0	0
T7	16,6	2,8	12.3	0	0
T8	17,1	1,7	15.3	0	0
T9	12,9	1,7	5.0	0	0
T10	13,7	2,5	9.7	0	0
T11	14,2	2,5	9.7	0	0

Con respecto al resultado de la prueba de usabilidad, los resultados son alentadores. Esto debido a que el número de errores cometidos por los participantes es bajo, considerando que los usuarios no poseían experiencia con el uso de la herramienta y por otra parte son perfectibles con bajo esfuerzo. Por ejemplo, en la T1 los 4 participantes cometieron errores, al entrevistarlos la totalidad no comentó que se debía a que no sabían si estaba grabando o no la herramienta. Al explorar con mayor profundidad la razón, esto se debía a que los usuarios esperaban que al presionar “rec” esperaban que este cambiara de color o estado, tal como sucede en la mayoría de los sistemas de este estilo.

Por otra parte, en T2 al agregar postura, el concepto no era el claro, por tal los usuarios tuvieron que consultar al manual y pudieron completar la tarea con el mínimo esfuerzo. Por tal razón, si bien la opción agregar postura no será una tarea común para el usuario, dado a que la herramienta posee 10 posturas cargadas, al iniciar por primera vez la herramienta, se le ofrecerá al usuario una presentación rápida de la herramienta que permita comprender las funcionalidades principales de la misma.

En T5, si bien 3 de los 5 usuarios cometieron un error, esto se debió a que en la primera opción que buscaban era en el menú *view*. Sin embargo, los gráficos se encuentran en la opción *charts* del menú superior. Con respecto a esto, dado los resultados, moveremos la opción *charts* hacia el menú *view*.

Para finalizar, si bien la prueba fue realizada en un ambiente controlado, es posible visualizar que los tiempos no fueron excesivos al momento de analizar y realizar tareas de una complejidad media y en algunos casos alta.

V. CASO DE ESTUDIO: INTRODUCCIÓN A LA INGENIERÍA INFORMÁTICA

Con el objetivo de evaluar la herramienta en un entorno real solicitamos utilizar *Lelikëlen* en la asignatura de Introducción a la Ingeniería Informática, de la carrera de Ingeniería Informática de la Universidad de Valparaíso. La asignatura corresponde a una de las 6 contempladas en el primer semestre de la carrera. La selección de esta asignatura como caso de estudio se debió a básicamente 2 aspectos. El primero, en su plan contempla la realización y presentación oral de 3 proyectos grupales en tres áreas distintas (Unidad 1:

Desarrollo Web, Unidad 2: Microcontroladores, Unidad 3: Base de Datos). El segundo aspecto de la elección es que una de las competencias declaradas en su plan de estudio, corresponde a comunicar de manera oral de manera eficiente en distintos contextos profesionales.

Cada grupo debía estar conformado por un mínimo de 4 integrantes y un máximo de 5. Al final de cada unidad (Desarrollo Web, Microcontroladores, Base de Datos), los grupos debían presentar de manera oral (apoyados por una presentación) el trabajo realizado en un máximo de 5 minutos, generándose 20 grupos para la primera presentación, 12 para la segunda, y 13 para la tercera. Cabe mencionar que la diferencia entre grupos se debe a la baja retención de las carreras de Ingeniería en primer año.

Para la presentación, 2 estudiantes de cada grupo eran escogidos de manera aleatoria con una tómbola para así presentar. Por otra parte, las presentaciones fueron realizadas en el mismo ambiente que el presentado en la Figura 7. Un aspecto importante por mencionar es que la totalidad de los participantes, previo a la grabación, firmaron un consentimiento informado en la que se informaban los aspectos del estudio, así como aprobaban que sus datos fueran analizados de manera grupal.

Se analizaron las presentaciones realizadas (el año 2017), en la Unidad 1 (24 de abril), Unidad 2 (8 de junio) y Unidad 3 (11 de julio), utilizando el software generado, siendo los resultados presentados en las Figuras 9, 10 y 11 de manera visual a través de los gráficos de Árbol.



Fig. 9. Gráfico porcentaje del tiempo por postura Unidad 1



Fig. 10. Gráfico porcentaje del tiempo por postura Unidad 2

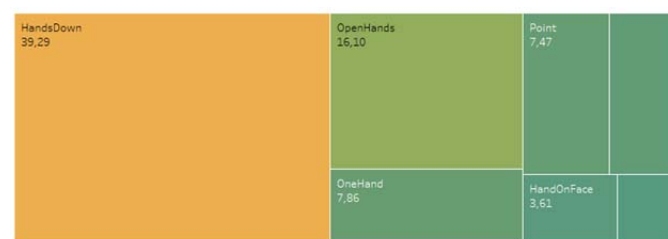


Fig. 11. Gráfico porcentaje del tiempo por postura Unidad 3

Al finalizar cada exposición, se les expuso en términos globales a los estudiantes los resultados de su presentación. Un aspecto importante para recalcar es que el foco de las exposiciones eran evidenciar el cómo habían resuelto un

problema que involucraba el uso de los contenidos de cada unidad temática del curso.

En términos generales los estudiantes en la primera presentación estuvieron cerca de un 50% del tiempo con las manos abajo. Esto puede deberse a que, para la mayoría, dado que son estudiantes de primer año, era su primera presentación oral en la Universidad. En la segunda y tercera presentación se puede visualizar una disminución de esta postura. Si bien, este cambio se puede deber a variados factores, uno de ellos, se puede atribuir a que después de entregarles los resultados de la primera presentación, también se dieron recomendaciones sobre características de una buena presentación oral.

Por otra parte, al momento de analizar las posturas asociadas generalmente a buenas posturas al momento de presentar, como lo son explicando con una y dos manos, además de la asociada apuntar, se mantuvieron relativamente estables en la primera y segunda presentación (superior al 30% del tiempo). Sin embargo, en la segunda presentación, esta se vio ligeramente disminuida. Entre la razón de este comportamiento, se pueden identificar diferentes factores. El primero, se puede deber a la complejidad de la Unidad y proyecto que tuvieron que presentar (integración de software microcontroladores), esto debido a que la integración de hardware con software no es común, al menos primer año de las carreras asociadas a la Ingeniería Informática. Por otra parte, se puede visualizar que en esta presentación, se clasificó la posición sentado. Si bien esta no es una postura que esté considerada al momento de presentar, lo que sucedió es que los estudiantes muchas veces, mientras presentaban, se inclinaban para así mostrar su maqueta (como la presentada en la Figura 12).



Fig. 12. Maqueta generada por los estudiantes

Por otra parte, un aspecto que también puede que haya influido en el resultado de su presentación, es que la docente de la Unidad no poseía experiencia en docencia universitaria. Por lo cual, puede que ella, haya intimidado de manera involuntaria a los grupos. Un análisis entre grupos es presentado en la siguiente sección. Sin embargo, si bien el foco de este artículo es presentar las potencialidades que posee incorporar una herramienta de *multimodal learning analytics* para poder otorgar retroalimentación a todas las partes del proceso.

Para finalizar, si bien el foco del curso no es desarrollar

competencias asociadas a presentaciones orales, sino que potenciar habilidades técnicas, las primeras están declaradas en el programa de asignatura. Por tal razón, el uso de una herramienta que permita retroalimentar al estudiante es de alto valor para ambas partes. Para los estudiantes dado que tienen una retroalimentación certera sobre su presentación oral, y para el docente porque puede tener una visión general del grupo curso.

VI. CONCLUSIONES Y TRABAJO FUTURO

Las disciplinas evolucionan cuando se pueden evaluar y medir. Bajo esta premisa, la incorporación de analítica del aprendizaje, en los contextos donde ocurren las transformaciones, es de vital importancia para la mejora constante de los procesos asociados a la educación. Si bien los resultados muestran que no basta sólo con declarar el desarrollo de competencias blandas, es importante que el foco de este trabajo fue demostrar que es posible incorporar este tipo de herramientas para así evidenciar de manera mucho más certera, el desarrollo o no de competencias.

Por tal razón, la analítica del aprendizaje es de vital importancia en el área de la enseñanza. Esto debido a que permite ser un apoyo al evaluar el desempeño de los estudiantes durante el proceso y no sólo al finalizar éste.

Por otra parte, las tecnologías de la información son de gran apoyo para la analítica del aprendizaje multimodal, ya que permiten ser un soporte para la medición de desempeños complejos, además de facilitar el almacenamiento de grandes volúmenes de datos y efectuar un análisis inteligente de ellos.

La ausencia de una herramienta que considere al educador como usuario final, motivó el desarrollo de esta aplicación. La aplicación desarrollada permite a un público masivo realizar un proceso de captura y visualización en un entorno de enseñanza-aprendizaje, presentando de forma comprensible las posturas corporales realizadas por los sujetos analizados, permitiendo también exportar los resultados para así aplicar, por ejemplo, minería de datos y asociar posturas y comportamientos y relacionarlos con desempeños.

Para finalizar, la arquitectura diseñada permite la integración de distintos sensores, los cuales han bajado considerablemente su costo, tales como GSR (*Galvanic Skin Response*), HRV (*Heart Rate Variability*), EEG (*Electroencephalogram*), entre otros. La integración de este tipo de sensores son el siguiente paso para la integración de la analítica multimodal en situaciones de enseñanza – aprendizaje y el trabajo futuro de este grupo de investigación.

REFERENCIAS

- [1] B. Lucas, J. Hanson, and G. Claxton, *Thinking like an engineer: implications for the education system*. Royal Academy of Engineering, 2014.
- [2] D. R. Laker and J. L. Powell, "The differences between hard and soft skills and their relative impact on training transfer," *Hum. Resour. Dev. Q.*, vol. 22, no. 1, pp. 111–122, Mar. 2011.
- [3] L. J. Shuman, M. Besterfield-Sacre, and J. McGourty, "The ABET Professional Skills? - Can They Be Taught? Can They Be Assessed?," *J. Eng. Educ.*, vol. 94, no. 1, pp. 41–55, Jan. 2005.
- [4] M. van Harmelen and D. Workman, "Analytics for learning and teaching," *CETIS Anal. Ser.*, vol. 1, no. 3, pp. 1–40, 2012.

- [5] M. Worsley, D. Abrahamson, P. Blikstein, S. Grover, B. Schneider, and M. Tissenbaum, "Situating multimodal learning analytics," Proc. Int. Conf. Learn. Sci. ICLS, vol. 2, pp. 1346–1349, 2016.
- [6] M. Worsley, "Multimodal learning analytics," in Proceedings of the 14th ACM international conference on Multimodal interaction - ICMI '12, 2012.
- [7] M. Worsley and P. Blikstein, "Towards the Development of Multimodal Action Based Assessment," in Proceedings of the Third International Conference on Learning Analytics and Knowledge, New York, NY, USA, 2013, pp. 94–101.
- [8] V. Echeverría, A. Avendaño, K. Chiluíza, A. Vásquez, and X. Ochoa, "Presentation Skills Estimation Based on Video and Kinect Data Analysis," in Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge, New York, NY, USA, 2014, pp. 53–60.
- [9] C. W. Leong, L. Chen, G. Feng, C. M. Lee, and M. Mulholland, "Utilizing Depth Sensors for Analyzing Multimodal Presentations: Hardware, Software and Toolkits," in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, New York, NY, USA, 2015, pp. 547–556.
- [10] C. W. Chang, M. D. Nian, Y. F. Chen, C. H. Chi, and C. W. Tao, "Design of a Kinect Sensor Based Posture Recognition System," 2014, pp. 856–859.
- [11] T. Kohonen, "The self-organizing map," Neurocomputing, vol. 21, no. 1–3, pp. 1–6, Nov. 1998.
- [12] C. Snyder, Paper prototyping : the fast and easy way to design and refine user interfaces. San Francisco, CA: Morgan Kaufmann Publishers, 2003.
- [13] J. J. Garrett, The Elements of User Experience: User-Centered Design for the Web. American Institute of Graphic Arts, 2003.
- [14] C. S. De Souza, The semiotic engineering of human-computer interaction. Cambridge, Mass.: MIT Press, 2005.
- [15] L. Chen, C. W. Leong, G. Feng, C. M. Lee, and S. Somasundaran, "Utilizing multimodal cues to automatically evaluate public speaking performance," 2015, pp. 394–400.
- [16] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," J. Comput. Syst. Sci., vol. 55, no. 1, pp. 119–139, Aug. 1997.
- [17] R. Lun and W. Zhao, "A Survey of Applications and Human Motion Recognition with Microsoft Kinect," Int. J. Pattern Recognit. Artif. Intell., vol. 29, no. 05, p. 1555008, Aug. 2015.
- [18] A. B. Badiru and L. Racz, Eds., Handbook of measurements: benchmarks for systems accuracy and precision. Boca Raton: CRC Press, 2016.
- [19] J. Nielsen, "Finding usability problems through heuristic evaluation," New York, NY, USA, 1992, pp. 373–380.
- [20] ISO 9241-210:2010 -- Part 210: Human-centred design for interactive systems." [Online]. Available: <https://www.iso.org/standard/52075.html>. [Accessed: 13-Apr-2013].
- [21] C. Cox, "From Cameras to Camtasia: Streaming Media Without the Stress," Internet Ref. Serv. Q., vol. 9, no. 3–4, pp. 193–200, May 2004.
- [22] S. K. Card, T. P. Moran, and A. Newell, "The keystroke-level model for user performance time with interactive systems," Commun. ACM, vol. 23, no. 7, pp. 396–410, Jul. 1980.



Roberto Muñoz es profesor adjunto en la Escuela de Ingeniería Civil Informática de la Universidad de Valparaíso, Chile e investigador asociado del Centro de Cognición y Lenguaje CIDCL, así como del Centro de Ingeniería en Salud, ambos de la Universidad de Valparaíso. El profesor Muñoz posee estudios doctorales en Ingeniería Informática, así como magister en Ingeniería Informática, Educación y Ciencias de la Computación. Actualmente ha publicado sobre 50 artículos científicos tanto en revistas de corriente principal, así como en prestigiosas conferencias de la disciplina. Sus áreas de investigación son: Informática en Educación, Interacción Humano-Computador e Análisis Multimodal del Aprendizaje.



de

Thiago Schumacher Barcelos es Licenciado en Ciencias de la Computación la Universidad de São Paulo (2002) y Magister en Ciencias de la Computación (2005) de la misma institución. Obtuvo un doctorado en Ciencia y Educación Matemática en la Universidade Cruzeiro do Sul (2014). En la actualidad es profesor e investigador en el Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Brasil. Sus áreas de investigación son: Informática en Educación, Interacción Humano-Computador e Ingeniería de Software.



Rodolfo Villarroel es profesor e investigador de la Escuela de Ingeniería en Informática de la Universidad Católica de Valparaíso, Chile. Obtuvo su doctorado en la Universidad de Castilla-La Mancha (2005), Ciudad Real, España, y su Magister en Ciencias de la Ingeniería Informática en la Universidad Federico Santa María (2000), Valparaíso, Chile. Sus áreas de investigación son: Ingeniería de Software, Calidad y Seguridad de la Información e Informática en Educación.



Rodolfo Guíñez es Ingeniero Civil en Informática de la Universidad de Valparaíso, Chile (2016). Actualmente es director General de la Consultora en Tecnología e Innovación Fisfor Ltda. - Chile. Sus áreas de investigación son: Analítica del aprendizaje y Big Data.



Erick Merino es Ingeniero en Informática de la Universidad de Valparaíso, Chile. Actualmente se encuentra realizando su magister en Ingeniería Biomédica. Su tesis de graduación fue asociada a detección de emociones en entornos de enseñanza - aprendizaje. Sus áreas de investigación son: Analítica del aprendizaje y Big Data.