UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE
INGENIEROS DE TELECOMUNICACIÓN

Contribuciones a Técnicas de Interacción por Apuntamiento para Espacios Inteligentes

TESIS DOCTORAL

David Gómez Cordero
Ingeniero en Informática

2020

DEPARTAMENTO DE SEÑALES, SISTEMAS Y RADIOCOMUNICACIONES

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN

UNIVERSIDAD POLITÉCNICA DE MADRID

Contribuciones a Técnicas de Interacción por Apuntamiento para Espacios Inteligentes

TESIS DOCTORAL

Autor

David Gómez Cordero

Ingeniero en Informática

Directora

Ana María Bernardos Barbolla

Doctora Ingeniera de Telecomunicación

2020

DEPARTAMENTO: **SEÑALES, SISTEMAS Y RADIOCOMUNICACIONES**

Escuela Técnica Superior de Ingenieros de Telecomunicación

Universidad Politécnica de Madrid

TÍTULO DE LA TESIS: **CONTRIBUCIONES A TÉCNICAS DE INTERACCIÓN POR APUNTAMIENTO PARA ESPACIOS INTELIGENTES**

Autor: **David Gómez Cordero**

Ingeniero en Informática

Directora: **Ana María Bernardos Barbolla**

Doctora Ingeniera de Telecomunicación

Comite nombrado por el Rector de la Universidad Politécnica de Madrid,
el ____ de _____ de 20____

Presidente: _____

Miembro: _____

Miembro: _____

Miembro: _____

Secretario: _____

Después de la defensa de la Tesis el ____ de _____ de 20____,
en la Escuela Superior de Ingenieros de Telecomunicación, el comité está de
acuerdo en otorgar la siguiente calificación:

Presidente

Secretario

Miembros

A mis hijas, Ana y Julia

Ten en mente que tu propósito y tu plan para lograrlo puede ser modificado de vez en cuando...lo importante es que comprendas el significado de trabajar siempre con un objetivo en mente y con un plan bien estructurado.

NAPOLEON HILL

Si puedes soñarlo puedes hacerlo, recuerda que todo esto comenzó con un ratón. **WALT DISNEY**

Agradecimientos

Esta Tesis es fruto del trabajo, constancia y esfuerzo hasta llegar a convertirse en un reto personal. Mientras esta tesis ha ido tomando forma, he vivido bastantes cambios a nivel educativo, profesional y personal. Además, esta etapa ha estado marcada por un continuo aprendizaje.

Quiero empezar dando las gracias a mi directora de Tesis Ana María Bernardos por su ayuda y colaboración en la elaboración de este trabajo. Su desarrollo no ha sido fácil pero ha sabido orientar las investigaciones realizadas y los experimentos llevados a cabo hasta conseguir plasmarlo en el presente documento. No puedo olvidar al profesor Jose Ramón, que con su experiencia y capacidad consigue objetivos que en su planteamiento pueden parecer de *ciencia ficción*. Gracias a este tipo de personas hoy la Humanidad disfruta de avances tecnológicos que hace años eran impensables y que tan importantes son en nuestra actual vida cotidiana. Tener un recuerdo también al resto de profesores que forman el Grupo de Procesado de Datos y Simulación (GPDS), que en alguna ocasión les ha tocado participar en los experimentos de investigación estudiados en este trabajo.

Un apartado especial se merecen los que fueron mis compañeros del Grupo GPDS, y algunos a día de hoy, amigos. Guardo un bonito recuerdo de esta etapa y creo que gran parte de la culpa la tienen muchos de ellos. Desde desayunos compartidos, charlas de la vida, cafés de sobremesa hasta las cañas de los jueves. Aunque está lista es extensa, en ella se encuentran grandes personas como Edu, Iñigo, Rafa, Luca, Guille, David, Henar, Josue, Wang, Julie, Gonzalo, Ivan, Paula y Laura. No puedo olvidar que algunos de ellos tuvieron que *sufrir* mis *temidas* pruebas de usuario. Es el momento de recordar el dicho que dice: *es bueno sembrar una buena y sincera amistad y el tiempo te permitirá disfrutar de una gran cosecha*.

Tengo y quiero dar las gracias a mis padres, Julio y Paqui, que aunque suene a tópico ellos lo han dado todo por mí. Sin ellos, ni se podría leer este trabajo ni yo sería lo que soy actualmente. Simplemente gracias por todo.

Quien me conoce sabe que no podría faltar un apartado especial a mis abuelos. Entre lo que he aprendido de mi abuelo Manuel y lo que me quiere mi abuela Juliana hay material de sobra para escribir una segunda Tesis. Por desgracia, no tuve la suerte de poder disfrutar el tiempo necesario de mi abuelo Feliciano ni mi abuela Francisca, hubiese sido divertido, aunque saben que tienen todo mi cariño y recuerdo.

Cuando todo esto comenzó había dos *personitas* que no existían, mis niñas, mis hijas, Ana y Julia. Aunque no estuvieron en el inicio de este camino hoy son el principal motivo de alegría y de las preocupaciones de mi vida. Ellas con el poco tiempo que llevan aquí me han enseñado mucho y a la vez transformado. Gracias por aparecer y existir.

No puedo olvidarme de mi compañera en el camino, Iris. Como comenté en cierta ocasión, ella y yo nos complementamos perfectamente y formamos un buen equipo; a día de hoy, es un equipo de 4 miembros. Como dice la canción, *Ohana* significa familia y familia significa que estaremos juntos siempre.

Por último y no por ello menos importante, agradecer al resto de mi familia y amigos.

Resumen

El apuntamiento es un gesto que usan las personas para expresarse. El gesto de apuntar permite indicar que algo es relevante en un momento determinado para la persona que lo realiza. Este gesto unido a la tecnología apropiada puede transformar un simple espacio físico en un entorno inteligente. Se denomina entorno inteligente a un espacio en el que sensores y actuadores permiten a los usuarios formas de interacción diferentes a las habituales con los objetos de dicho espacio. Estas interacciones pueden consistir en encender y apagar una lámpara cuando una persona apunta hacia ella, cambiar el color de su luz mediante comandos de voz o la interacción con una imagen virtual mostrada en una pantalla mediante gestos.

Las formas más habituales de apuntar de una persona hacia un objetivo son dos, bien extendiendo el dedo índice de su mano acompañado habitualmente del movimiento de su brazo; o bien, con su mirada. Un sistema de interacción basado en apuntamiento debe ser capaz de reconocer el apuntamiento realizado, estimar el objetivo apuntado, para luego, si procede, permitir la interacción de la persona con el objetivo apuntando. Esta interacción debe basarse en una Interfaz Natural de Usuario (NUI, Natural User Interface), de modo que para el usuario el sistema sea totalmente transparente, es decir, que sea el gesto de apuntar el que desencadena la interacción. Tiene una especial relevancia, por tanto, el dispositivo o conjunto de dispositivos empleados para detectar el apuntamiento y habilitar la interacción.

Esta Tesis tiene como objetivos principales (I) el estudio de los sistemas de interacción basados en apuntamiento que actualmente existen, (II) establecer una conexión entre estos sistemas y los entornos inteligentes, y (III) el diseño de nuevos sistemas que permitan dicha conexión. Dada la naturaleza de la Tesis, cobra especial importancia el análisis realizado para la elección del dispositivo habilitador, pieza clave para una correcta integración de la persona con el entorno inteligente. Igualmente, caben destacar los resultados obtenidos del sistema final, así como su valoración en términos de rendimiento

y eficiencia. Este trabajo también incorpora una importante investigación en materia de la evaluación de la experiencia y aceptación del usuario respecto de la utilización de los sistemas de apuntamiento.

La investigación tiene su inicio en la comparación de las dos principales técnicas de detección del apuntamiento, la señalización manual y la simulación del seguimiento de la mirada. Para realizar dicha comparación se ha empleado como dispositivo habilitador Kinect debido a la idoneidad de sus características y capacidades. Para evaluar esto, se ha realizado un estudio desde el punto de vista del usuario con tareas de apuntamiento siguiendo la Norma ISO de Requisitos ergonómicos para trabajos de oficina con pantallas de visualización de datos (comprueba que este es el nombre correcto de la norma) (ISO 9241-9). Este estudio incluye pruebas de uso de un servicio para el manejo de objetos inteligentes. El diseño de la prueba y el análisis de los resultados obtenidos se ha realizado usando la Ley de Fitts. Esta Ley permite evaluar el rendimiento del usuario cuando utiliza sistemas basados en apuntamiento. La métrica propuesta se denomina índice de dificultad y relaciona la distancia al objetivo a seleccionar con su tamaño. También emplea la variable de rendimiento donde se relaciona el índice de dificultad citado con el tiempo de movimiento empleado para alcanzar el objetivo. La aplicación de esta Ley arroja importantes conclusiones: (I) los parámetros de rendimiento son similares para ambas técnicas de detección, (II) el tamaño del objeto a seleccionar afecta críticamente a la efectividad de los métodos, (III) la señalización manual es la alternativa mejor valorada por los usuarios, y (IV) la detección mediante seguimiento de la mirada destaca por la rapidez de respuesta del sistema.

El segundo aspecto fundamental objeto de la Tesis ha sido el diseño de un sistema de interacción basado en apuntamiento dentro de espacios inteligentes, con el fin de conocer las ventajas y desventajas del uso de estos sistemas en estos espacios. El sistema diseñado tiene las características de emplear señalización manual y usar el Microsoft Kinect como sensor habilitador de la interacción. El diseño comienza con el análisis del vector de detección del apuntamiento a emplear que indica la dirección del apuntamiento. Este queda definido por dos puntos del espacio, en este caso, estos puntos corresponden a posiciones de partes concretas de una persona por ejemplo la posición de una de sus manos o de su nariz. En total se estudian siete opciones distintas. El análisis refleja un error en la estimación del apuntamiento. Este error proviene de una estimación incorrecta de la altura, problemas en la calibración del sistema y errores derivados de las características personales del usuario. En esta Tesis, se propone, por tanto, un sistema mejorado de apuntamiento que disminuye el error de estimación general. Respecto a las propuestas

que se pueden encontrar en la literatura, la propuesta ofrecida en esta Tesis incorpora como novedad un vector de detección de apuntamiento múltiple. En concreto, es la combinación de dos de estos vectores, dependiendo de la visibilidad del dispositivo sensor en cada momento. Asimismo, este sistema incluye una mejor calibración del sistema (permite referenciar todas las posiciones o localizaciones que se empleen de acuerdo a un sistema de referencias global único) y la posibilidad de realizar una calibración personalizada por usuario (modifica el apuntamiento estimado de forma personalizada) respecto a la de otros sistemas. Se ha realizado un estudio completo del sistema propuesto para obtener tanto el rendimiento de este, como la experiencia de usuario a la hora de usarlo. Asimismo, se ha desarrollado un servicio final para manejar un videowall (mosaico de pantallas) a través de interacción mediante apuntamiento dentro de un espacio inteligente.

El trabajo realizado aplica las principales y más efectivas técnicas de estimación del apuntamiento en entornos inteligentes que se pueden encontrar en la literatura actual y solventa los errores que estas puedan incluir. Asimismo, incluye una exhaustiva evaluación de la experiencia del usuario con los sistemas de apuntamiento.

Las contribuciones de esta Tesis, en su ámbito teórico, pueden resumirse en:

- El estudio comparativo de alternativas de interacción natural por apuntamiento.
- La propuesta de técnicas para estimación del apuntamiento desplegable en espacios inteligentes.
- El análisis de las fuentes de error en las estimaciones de la señalización manual junto a un estudio de relevancia y búsqueda de soluciones para mejorar la precisión y exactitud del sistema de apuntamiento.
- La evaluación de la aceptación de la tecnología y la experiencia de usuario.

La contribución práctica de esta Tesis consta de:

- Una plataforma de gestión de blancos multimodal (preparada para realizar medidas de precisión y eficacia, en la que se pueden incorporar diversos sistemas de apuntamiento).
- La particularización de la plataforma anterior para gestión de videowall.

- Una metodología de prototipado de servicios de selección y activación.
- Diseño de estudios de validación con usuarios.

Abstract

Pointing is a representative gesture that people use to express themselves or to show interest in something. This gesture combined with a appropriate technology can transform a simple physical space into an smart space, which has sensors and actuators allowing users new interaction modes than usual with objects in that space. These interactions can be turning a lamp on and off when a person points at it, changing the colour of its light through voice commands, or interacting with a virtual image displayed on a screen through gestures.

The common ways as a person point towards an item are two, either by extending the index finger of his hand, usually accompanied by the movement of his arm; or, with his gaze. A pointing interaction system must be able to recognize the pointing, estimate the pointed target and allow the person-target interaction. This interaction must be based on a Natural User Interface (NUI), being transparent for the user, so the interaction is triggered with the gesture of pointing. Therefore, the device or devices set used to detect pointing and enable interaction has an important relevance.

The main objectives of this thesis are (I) the study of existing pointing-based interaction systems, (II) establish a connection between these systems and smart spaces and finally, (III) the design of new systems that allow this connection. Due to the nature of the Thesis, the analysis to choose the enabling device has special relevance, so it is a key element for a correct integration of the person within the smart space. Furthermore, the results obtained from the final system and its assessment in terms of performance and efficiency should be highlighted. This work also incorporates important research on the user experience evaluation and the user acceptance with the use of pointing-based interaction systems.

The research began in the comparison of the two main techniques of pointing detection, deictic pointing and gaze tracking. A Kinect device is used as the enabling device due to the suitability of its characteristics and capabili-

ties. Furthermore, a user study is proposed with pointing tasks as indicate ISO 9241-9 and a test a service to command smart objects. The analysis of the results is carried out by applying the Fitts Law that allows the evaluation of user performance when using pointing-based interaction systems. The difficulty index is the proposed metric, which relates the distance to the target with the target size. It also uses the performance, which relates the difficulty index with the movement time used to reach the objective. The application of this Law yields important conclusions: (I) the performance parameters are similar for both detection techniques, (II) the size of the object to be selected critically affects the effectiveness of the methods, (III) deictic pointing is the alternative best valued by users, and (IV) the detection based on gaze tracking stands out for the quick system response.

The second fundamental aspect that is the object of the thesis is the design of a pointing-based interaction systems within smart spaces, in order to know the advantages and disadvantages of using these systems in these spaces. The designed system uses deictic pointing and the Microsoft Kinect as an enabling sensor to allow the interaction. First an analysis of the possible vectors for detecting tacking, which indicates the pointing direction, is carried out This vector is defined by two points in space, in this case, these points correspond to positions of specific parts of a person, for example the position of one of her hands or her nose. In total, seven different options have been studied. The analysis reflects an error in the pointing estimation. This error comes from an incorrect estimation of the height, problems in the system calibration and errors derived from the personal characteristics of the user. In this Thesis, therefore, an improved targeting system is proposed that reduces the general estimation error. Regarding the proposals that can be found in the literature, the proposal offered in this Thesis incorporates as a novelty a multiple pointing detection vector. Specifically, it is the combination of two of these vectors, depending on the visibility of the sensor device in every instant. Likewise, this system includes a better system calibration (it allows to reference positions or locations that are used according to a single global reference system) and the possibility to perform a custom user calibration (modifies the estimated pointing in a personalized way) compared to other systems. A complete study of the proposed system has been carried out to obtain its performance and the user experience. In addition, a final service has been developed to manage a videowall (screens mosaic) through interaction by pointing in a smart space.

This work applies the main and most effective pointing estimation techniques in smart space available in the current literature and solves the errors that these may include. It also includes a thorough evaluation of the user

experience with the pointing systems.

The contributions of this Thesis, in its theoretical scope, can be summarized as:

- The comparative study of natural interaction alternatives by pointing.
- The proposal of techniques for pointing estimation within smart spaces.
- The analysis of the sources of error in deictic pointing estimation together with a study of relevance and search for solutions to improve the accuracy and precision of the pointing system.
- The evaluation of the acceptance of the technology and the user experience.

The practical contribution of this Thesis consists of:

- A multimodal target management platform (ready to perform precision and efficiency measurements, in which different systems pointing can be incorporated).
- The particularization of the previous platform for videowall management.
- A prototyping methodology for selection and activation services.
- Design of validation studies with users.

Índice general

1. Introducción	1
1.1. Contexto y motivación	1
1.2. Contribuciones	4
1.3. Metodología	7
1.4. Estructura del documento	9
2. Estado del arte	11
2.1. Objetivos del Capítulo	11
2.2. Métodos de detección de apuntamiento	11
2.2.1. Habilitadores tecnológicos (dispositivos y sensores) . .	11
2.2.2. Técnicas de detección	14
2.2.3. Técnicas de selección	20
2.2.4. Estudios comparativos de interacción mediante apun- tamiento basado en diferentes técnicas	21
2.3. Evaluación del rendimiento de sistemas de apuntamiento . . .	23
2.3.1. Procedimientos estandarizados	23
2.3.2. Otras metodologías	25
2.4. Aplicaciones y clasificación	26
2.5. Resumen	32
3. Comparación de técnicas de detección de objetivo apuntado	

basadas en señalización manual y posición-orientación de la cabeza	33
3.1. Objetivos del Capítulo	33
3.2. Conceptos previos del estudio	34
3.2.1. Ley de Fitts y ISO 9241-9	35
3.2.2. Escenario de interacción	39
3.3. Estimación del apuntamiento	40
3.3.1. Métodos de selección	40
3.3.2. Sistemas de coordenadas	43
3.3.2.1. Proceso de calibración de la cámara de Kinect	43
3.3.2.2. Transformaciones entre sistemas de coordenadas	44
3.4. Arquitectura del sistema de estimación del apuntamiento . . .	47
3.5. Diseño del estudio para la evaluación del rendimiento del sistema y la experiencia de usuario	51
3.5.1. Objetivos	51
3.5.2. Participantes	51
3.5.3. Dispositivos y Herramientas	52
3.5.4. Procedimiento y tareas	54
3.5.5. Diseño del experimento	55
3.6. Resultados del experimento	56
3.6.1. Análisis del rendimiento de la interacción	56
3.6.1.1. Tiempo de movimiento	56
3.6.1.2. Análisis basado en la Ley de Fitts	61
3.6.1.3. Dificultad percibida y aprendizaje	64
3.6.2. Confort, usabilidad y experiencia de usuario	66
3.6.2.1. Análisis del confort del usuario	66
3.6.2.2. Análisis de usabilidad y experiencia de usuario	68

3.6.2.3.	Influencia de la referencia visual en la experiencia de usuario y preferencia del mismo . . .	71
3.6.2.4.	Preferencias de usuario sobre la técnica de detección de apuntamiento	72
3.7.	Resumen	73
4.	Diseño de un sistema <i>mejorado</i> de apuntamiento basado en señalización manual	75
4.1.	Objetivos del Capítulo	75
4.1.1.	Revisión de los requisitos del sistema de apuntamiento	76
4.1.2.	Entorno de despliegue	78
4.1.3.	Experimentos para determinar las fuentes de error en la estimación del apuntamiento	78
4.2.	Selección del vector de apuntamiento	81
4.2.1.	Pruebas para la selección del vector de apuntamiento .	81
4.3.	Aspectos sobre la estimación de la coordenada en altura (Y) de los puntos que definen el vector de apuntamiento	87
4.3.1.	Estabilidad de la estimación de la altura de puntos de referencia (pupila)	87
4.3.2.	Efecto de pequeñas variaciones en la estimación de la altura (de la pupila) sobre la estimación de la coordenada en altura del apuntamiento	90
4.3.3.	Variaciones en la estimación de la altura (de la pupila) en función de la posición del usuario con respecto al sensor	92
4.3.4.	Variaciones en la estimación de la altura (de la pupila) en función de la inclinación del sensor Kinect	95
4.3.5.	Variación de la estimación de la altura para otros puntos de referencia: dedo y cabeza	99
4.3.6.	Variación de las coordenadas de apuntamiento en función del uso de gafas	102

4.4.	Variación del resultado del apuntamiento en función del diseño de la calibración	103
4.4.1.	Relación entre la posición del patrón de calibración y el resultado del apuntamiento	104
4.4.2.	Relación entre la inclinación del patrón de calibración y la estimación de las coordenadas en altura	105
4.4.3.	Relación entre la inclinación del sensor dispositivo Kinect y la inclinación de la calibración	108
4.4.4.	Propuesta de corrección de los efectos del diseño de la calibración. Análisis de la universalidad de la solución.	111
4.5.	Propuesta de un sistema de apuntamiento mejorado	118
4.6.	Resumen	120
5.	Estudio de evaluación del sistema de apuntamiento propuesto	123
5.1.	Objetivos del Capítulo	123
5.2.	Diseño del estudio de rendimiento, precisión del sistema y experiencia de usuario	124
5.2.1.	Objetivos	124
5.2.2.	Participantes	124
5.2.3.	Entorno e Infraestructura	125
5.2.4.	Procedimiento y tareas	127
5.2.4.1.	Pruebas de rendimiento y precisión del sistema de apuntamiento	129
5.2.4.2.	Pruebas de selección mediante apuntamiento	131
5.2.5.	Diseño del experimento	132
5.2.5.1.	Experimento dedicado a obtener el rendimiento y precisión del sistema de apuntamiento	133
5.2.5.2.	Experimento para comprobar el apuntamiento sobre tareas reales de selección	134
5.3.	Resultados	135

5.3.1.	Análisis de rendimiento y precisión	135
5.3.1.1.	Rendimiento : Tiempo de ejecución	135
5.3.1.2.	Precisión del sistema : Número de aciertos . . .	138
5.3.2.	Análisis de confort	140
5.3.3.	Análisis de usabilidad	141
5.3.4.	Análisis de la experiencia de usuario	142
5.4.	Resumen	144
6.	Conclusiones y trabajo futuro	145
6.1.	Conclusiones	146
6.1.1.	Resultados de la comparativa de técnicas de detección del apuntamiento basadas en posición-orientación de la cabeza y señalización manual	148
6.1.2.	Diseño de un sistema <i>mejorado</i> de apuntamiento basado en señalización manual	150
6.1.3.	Resultados del estudio de evaluación del sistema de apuntamiento <i>mejorado</i> propuesto	152
6.1.4.	Recomendaciones de diseño de sistemas de interacción mediante apuntamiento	153
6.2.	Trabajo futuro	155
A.	Cuestionario de evaluación de la experiencia de usuario	159
B.	Cuestionarios de evaluación de usuarios para las pruebas de videowall	173
C.	Modelo de autorización de los participantes de la prueba	179
	Glosario	183
	Bibliografía	187

Capítulo 1

Introducción

1.1. Contexto y motivación

El concepto de tecnología *tranquila* (*calm technology*) fue acuñado hace 20 años para referirse a la tecnología que monitoriza, pero no demanda nuestra atención [1]. Con el uso de esta tecnología no intrusiva se puede transformar un entorno habitual en un espacio inteligente. Además si se emplea esta Tecnología para detectar el apuntamiento de las personas se puede permitir la interacción del usuario dentro de este espacio con los objetos que lo componen.

La interacción con objetos dentro de los espacios inteligentes permanece vinculada a los paradigmas tradicionales, en la actualidad se han incorporado al mismo dispositivos tales como los teléfonos inteligentes [2]. Estos dispositivos permiten ser utilizados como controladores de todo tipo de artefactos conectados; como por ejemplo, lámparas y dispositivos multimedia. Al final se pretende obtener una interacción lo más natural posible del usuario con el entorno, por lo que, la tecnología para permitir la interacción debe ser invisible y estar integrada dentro del propio entorno. La interacción con el habla, el tacto, la mirada, la escritura o los gestos se considera *natural*. La integración de ésta, plantea nuevos problemas, nuevos retos y el potencial para los errores masivos y la confusión, que es necesario manejar para materializar sus posibilidades, como evidencia el estudio [3].

El apuntamiento es uno de los gestos humanos, conscientes o inconscientes, que con mayor naturalidad realiza una persona. El término *apuntar* según la Real Academia Española es la acción de señalar con el dedo o de cualquier

otra manera hacia sitio u objeto determinado, sinónimos de ello son *señalar* o *indicar*. Es un gesto universal reconocido que expresa el interés de una persona hacia algo que llama su atención [4]. La mayoría de las veces, el apuntamiento se usa como una forma de dirigir la atención de una persona o mostrar interés hacia algo de su entorno [5]. A priori, la interacción de un usuario mediante el apuntamiento empleando partes de su cuerpo parece la forma más natural de realizarlo. La forma más común de apuntar hacia un elemento es extendiendo el dedo índice hacia el objetivo a señalar (lo que en esta Tesis se referirá como *método de señalización manual*). Este gesto va acompañado de un movimiento más o menos pronunciado del brazo. Previo a dicho movimiento, la persona suele mirar hacia el objetivo [6].

Esta Tesis centra su estudio en como hacer viable un sistema de apuntamiento que transforme en interactivo un espacio haciendo uso del *apuntamiento*. La interacción mediante apuntamiento debe diferenciar dos aspectos fundamentales: la selección del objetivo y la activación del mismo.

En la Tesis también se abordará la evaluación del sistema. El término de usabilidad también está muy ligado a las interacciones de usuarios. No tiene ningún sentido desarrollar un tipo de interacción para un usuario que no sea operativa ni funcional. Este neologismo (del inglés *usability*, facilidad de uso) es definido por la Organización Internacional de Normalización (*International Organization for Standardization*, ISO) en su norma ISO 9241 como *la eficacia, eficiencia y satisfacción con la que un producto permite alcanzar objetivos específicos a usuarios específicos en un contexto de uso específico*. El grado o nivel de usabilidad es una medida empírica y relativa de la usabilidad basado en: pruebas empíricas, no se basa en opiniones subjetivas de un usuario, sino en pruebas de usabilidad llevadas a cabo; pruebas relativas, los resultados no son ni desechables ni aceptados, simplemente se valora si cumple unos objetivos o se compara con otros similares.

Por tanto, esta Tesis pretende ser un estudio original en el desarrollo de sistemas de apuntamiento. Además, se pretende que suponga a su vez una mejora con respecto los sistemas ya existentes. Se quiere marcar una diferencia y aportar con el uso de los sistemas de apuntamiento un valor añadido a los espacios para convertirlos en inteligentes. Esto conllevará analizar los problemas que acarrea el uso de estos sistemas. Otro aspecto importante en que hacer hincapié es la evaluación y aceptación de los usuarios en el uso de los sistemas de apuntamiento dentro de espacios inteligentes.

Asimismo se persigue que esta Tesis sirva como punto de partida a la hora de diseñar nuevos sistemas de apuntamiento y guíe a futuros investigadores a resolver los problemas derivados de su diseño o relacionados con el uso de

los mismos por parte de los usuarios finales.

Objetivos de la Tesis

- El objetivo principal de esta Tesis es estudiar la viabilidad de despliegue de los sistemas interactivos por apuntamiento en espacios inteligentes, asumiendo que éstos pueden habilitar nuevas formas de seleccionar y accionar objetos o recursos digitales más naturales y convenientes. Cuando una persona muestra interés hacia *algo* de su entorno, y por ejemplo quisiera interaccionar hacia ese objetivo, apuntará probablemente coordinando su brazo, su cabeza, sus ojos y/o cualquier otro gesto que sirva para mostrar su interés hacia ello.
- Desde el punto de vista tecnológico, el propósito principal en la implementación de un sistema de apuntamiento es conseguir la mejor solución en términos de precisión, universalidad y cobertura máxima. Estas técnicas se comparan también con métodos de selección de objetivos basados exclusivamente en seguimiento de mirada (simulada usando la posición y orientación de la cabeza). En la Tesis se observa la comparación realizada entre sistemas que emplean apuntamiento mediante señalización manual y seguimiento de la mirada arroja un rendimiento similar, aunque la preferencia de los usuarios que lo probaron apunta al empleo del apuntamiento mediante seguimiento de la mirada.
- Otro de los objetivos es construir una plataforma de pruebas de apuntamiento que siga los estándares aceptados en la medida de eficiencia, y en la que se puedan integrar algoritmos de apuntamiento. Esta plataforma debe servir para poder dar cabida a todas las pruebas que se necesiten realizar para obtener un análisis completo de los sistemas de apuntamiento. La Ley de Fitts evalúa el rendimiento del usuario durante el uso de sistemas de apuntamiento. El estándar definido en ISO 9241-9 (Requisitos ergonómicos para trabajos de oficina con pantallas de visualización de datos; Parte 9: Requisitos para dispositivos de entrada diferentes al teclado) recoge una serie de tareas que ayudará a obtener parámetros para valorar la calidad de la propia interacción.
- Sobre el conocimiento y utilizando las herramientas anteriores, el objetivo es diseñar un sistema robusto de apuntamiento, desplegable de manera sencilla en espacios inteligentes, e integrable con diferentes servicios.

- Finalmente es necesario que usuarios finales prueben y validen los aspectos relacionados con rendimiento, usabilidad y aceptación de la tecnología que ha sido empleada.

1.2. Contribuciones

Las contribuciones de esta Tesis se pueden clasificar en conceptuales-teóricas y experimentales-prácticas.

Contribuciones teóricas

Las contribuciones teóricas se refieren al estudio, diseño y propuesta de nuevas técnicas y algoritmos que mejoran la detección del apuntamiento en un espacio inteligente. Estas contribuciones son las siguientes:

1. Estudio comparativo de alternativas de interacción natural por apuntamiento.
2. Propuesta de técnicas para estimación de apuntamiento desplegable en espacios inteligentes.
3. Análisis de fuentes de error en las estimaciones de la señalización manual, estudio de relevancia y búsqueda de soluciones para mejorar la precisión y exactitud del sistema de apuntamiento.
4. Evaluación de la aceptación de la tecnología y la experiencia de usuario.

La primera contribución comprende el catálogo de las distintas alternativas de interacción natural existentes en la literatura, considerando los dispositivos que permiten desarrollarlas. Los dispositivos pueden ser portados por la propia persona o desplegarse en el entorno. El tipo de dispositivo condiciona la interacción que se realiza. En esta selección destaca el dispositivo-sensor Microsoft Kinect, que presenta una serie de ventajas frente a las presentadas por otros dispositivos.

En la segunda contribución se propone una técnica para estimar el apuntamiento dentro de un espacio inteligente, con el requisito de que se pueden instalar cámaras. La precisión es uno de los principales requisitos exigidos

para estos sistemas, ésta dependerá del servicio a ofrecer. Por otra parte, está la ubicuidad del sistema. Es decir, el apuntamiento en cualquier posición del usuario dentro del área de cobertura del sensor. El sistema debe tener un uso universal y servir para cualquier persona que lo quiera usar, funcionando en las mismas condiciones para todos ellos, además, su despliegue debe ser sencillo.

Por otra parte, durante la propuesta de la técnica de estimación del apuntamiento se detectan tres diferentes fuentes de error que distorsionan el resultado de apuntamiento estimado. Las fuentes de error detectadas son: distorsión en la estimación de la coordenada en altura, error inducido por el resultado de la calibración durante la estimación del vector, o características del usuario que generan error. Se estudian para obtener que las está provocando y con las conclusiones obtenidas, se propone una metodología para paliar los errores obtenidos durante la estimación del apuntamiento.

Durante las pruebas, se comprueba que la estimación de la coordenada en altura es errónea, aún sin estar el usuario en movimiento. Esto es, pequeñas variaciones de la estimación de la coordenada genera un error del apuntamiento que se agudiza cuando aumenta la distancia entre el usuario y el objetivo a seleccionar. El proceso de calibración que permite transformar las coordenadas del entorno en coordenadas del sistema también influye en la estimación del vector de apuntamiento.

Por último, es necesario que los usuarios prueben la tecnología y validen si la experiencia de usuario es adecuada y aceptable. Esto se realiza mediante una metodología estándar. Los tres aspectos a comprobar son confort, usabilidad y experiencia de usuario. El análisis de confort se realiza tomando como base el cuestionario del anexo C de ISO 9241-9. La evaluación de la usabilidad se lleva a cabo con la escala definida en el estudio [7]. Finalmente se evalúa la experiencia de usuario evaluando el *atractivo* (impresión general del sistema), *familiaridad* (facilidad para familiarizarse), *eficiencia* (idoneidad para resolver las tareas del usuario), *fiabilidad* (sensación de control), *estimulación* (excitación y motivación) y *novedad* (innovación) del sistema planteado. Todo ello basado en el cuestionario del estudio [8].

Contribuciones prácticas

Las contribuciones experimentales incluyen herramientas de apoyo, despliegues reales y estudios específicos que se llevaron a cabo para generar o evaluar algunas de las contribuciones anteriores, en el ámbito de los espacios

inteligentes, utilizando dispositivos COTS (*Commercial Off-The-Shelf*):

5. Plataforma de gestión de blancos multimodal, preparada para realizar medidas de precisión y eficacia, en la que se pueden incorporar diversos sistemas de apuntamiento.
6. Particularización de la plataforma anterior para gestión de videowall.
7. Metodología de prototipado de servicios de selección y activación.
8. Diseño de estudios de validación con usuarios.

El sistema de apuntamiento se sitúa dentro de un espacio inteligente, Figura 1.1. La imagen muestra una situación real donde el usuario dentro de este espacio puede seleccionar apuntando con su brazo una de las nueve pantallas que forman el videowall, la interacción consiste en mostrar sobre la pantalla seleccionada una imagen. Únicamente se va a visualizar la imagen sobre una de las pantallas. En esta figura se aprecia como un usuario apunta e interacciona con un objetivo dentro de este espacio. Los elementos a destacar son: i) el sensor Kinect situado en la parte superior para tener una visión amplia, ii) el equipo que controla este sensor, iii) los objetivos con los que usuario puede seleccionar o interaccionar, que en este caso son las diferentes pantallas que forma el videowall. La arquitectura aparece detallada en la Sección 3.4.

Este es un ejemplo del servicio final usando un sistema de apuntamiento implementado sobre el espacio. Este servicio está montado sobre la plataforma que gestiona objetivos *interaccionables* o *seleccionables*. Además, dependiendo del servicio que se implemente sobre ella sirve para realizar las medidas de precisión y eficacia con las que evaluar el sistema de apuntamiento.

Una parte importante de este tipo de prototipos es la selección de objetivos y si procede, la activación de los mismos. Para ello se ha planteado una metodología dentro del prototipo que permite realizar la selección de elementos del entorno. El paso siguiente es la propia interacción con el objetivo seleccionado y para ello se necesita activar el mismo. En este estudio se han implementado servicios domóticos como el control de una bombilla, permitiendo su encendido y apagado.

Una vez que se ha particularizado para implementar un servicio concreto, el propio servicio, así como el apuntamiento, se ha probado con diversas pruebas de usuarios. Como se comentaba con anterioridad se han personalizado estos estudios para cada caso.

Figura 1.1: Entorno del sistema de apuntamiento

1.3. Metodología

La Metodología define los pasos, detalla los productos, las técnicas y los métodos que hay que seguir para cumplir con los objetivos planteados. Por su parte, el ciclo de vida (o modelo de desarrollo, o modelo del proceso) define las etapas y actividades genéricas por las que es necesario pasar para cumplir los objetivos. En ambos casos, metodología y ciclo de vida, definen este conjunto de pasos o actividades. La principal diferencia es que la metodología especifica en detalle que pasos hay que dar y cómo deben ser estos dados; mientras que, el ciclo de vida especifica las fases por las que pasa de forma genérica. Ambos términos están relacionados. Una metodología puede seguir uno o más ciclos de vida. A su vez, un ciclo de vida no es exclusivo de una metodología y de esta forma varias metodologías pueden seguir el mismo ciclo de vida.

Este trabajo ha combinado principios propios del modelo basado en prototipos y del modelo basado en espiral. El modelo basado en prototipos lleva a la construcción de un prototipo, el cual tras ser analizado, se consigue un refinamiento de los requisitos, o bien, plantea unos nuevos. Por su parte, en el modelo basado en espiral desarrolla una serie de versiones incrementales. Durante las primeras iteraciones, la versión incremental podría ser un prototipo. Durante las últimas iteraciones, se producen versiones cada vez más completas del sistema diseñado. En esta Tesis, el primer modelo se sigue para aportar la construcción del prototipo, y el segundo modelo establece la forma

de construir el prototipo en diferentes fases.

La metodología escogida para el desarrollo y realización de esta Tesis, como se aprecia en la Figura 1.2, sigue las siguientes fases: revisión permanente del estado del arte, propuesta tecnológica, prototipado, validación y pruebas de usuarios. A continuación se explicará cada una de las fases en mayor detalle.

- **Planteamiento de requisitos y retos para sistemas de apuntamiento en espacios inteligentes.** En esta fase se plantean los requisitos que debe tener el sistema y los principales retos que se abordará con la construcción del mismo.
- **Revisión del estado del arte.** Una vez conocido lo que se intenta obtener, se debe realizar una revisión de lo ya existente. De esta forma poder ver si hay algo con lo que se pueda contribuir o emplear algo que ya exista.
- **Prototipado del entorno y sistema de apuntamiento.** Construcción del prototipo que cumplan los requisitos y aborde los retos planteados, teniendo en cuenta las decisiones que se hayan podido tomar con la revisión del estado del arte realizado.
- **Validación de prototipo.** Se valida que el prototipo ha sido construido correctamente.
- **Pruebas de usuario.** Diseño y realización de pruebas reales de usuario con el objetivo de la analizar principalmente la usabilidad del prototipo y la experiencia en el uso del mismo. La pruebas son llevadas a cabo por usuarios. El número de usuarios varía dependiendo de las necesidades concretas. En particular, esta Tesis usa un número diferente de usuarios de acuerdo a la prueba a realizar.
- **Resultados del estudio de usuario.** Se obtienen las conclusiones del estudio del prototipo y de las pruebas realizadas a usuarios reales. Comprobando que se han valorado todos los retos y abarcando todos los requisitos.

Figura 1.2: Metodología empleada

1.4. Estructura del documento

La estructura de este documento, dividido en 6 capítulos, se describe a continuación.

El Capítulo 2 hace una revisión del estado del arte. Este capítulo estudia los métodos de detección del apuntamiento considerando: los habilitadores tecnológicos, técnica de detección y selección. Se realiza, asimismo, una comparativa de diferentes interacciones. Además se recogen los diferentes métodos para evaluar el rendimiento en los sistemas de apuntamiento, bien a través de procedimientos estandarizados u otras metodologías. Por último, se repasan las diferentes aplicaciones que hacen uso de sistema de apuntamiento en la literatura y se les realiza una clasificación de las mismas.

El Capítulo 3 compara dos técnicas de detección, la primera empleando la posición y orientación de la cabeza; y la segunda, mediante señalización manual. Se entra en detalle en el dispositivo sensor que se emplea *Microsoft Kinect*. Además, se diseña una serie de servicios finales que son probados mediante estudios de usuarios para evaluar su rendimiento, confort, usabilidad y su experiencia.

El Capítulo 4 describe el diseño de un sistema de apuntamiento empleando señalización manual. A partir de los principales retos y requisitos a tener en cuenta en el diseño del mismo, se exploran las alternativas del vector de apuntamiento. Se estudia la precisión dada por estos vectores. Se realiza un estudio completo de los principales factores de error de estos sistemas. Finalmente se plantea el diseño de un sistema mejorado para implementar el apuntamiento.

El Capítulo 5 recoge el estudio realizado para evaluar el sistema de apuntamiento mejorado que se planteó en Capítulos 3 y 4 . El estudio se centra en analizar el rendimiento y precisión exactitud de este sistema, a través de pruebas llevadas a cabo por un grupo de usuarios. Además estudia el confort, usabilidad y la experiencia del usuario con el sistema.

El Capítulo 6 concluye la investigación desarrollada en esta Tesis. Este capítulo presenta un conjunto de conclusiones y aportaciones del trabajo presentado. Además define algunas posibles directrices de investigaciones futuras.

Al finalizar el documento se recogen algunos anexos que complementan al trabajo principal desarrollado a lo largo de la Tesis. Esta documentación complementaria comprende el modelo de los cuestionarios rellenados por los participantes para evaluar la experiencia de usuario del sistema propuesto en el Capítulo 3 (Anexo A); los cuestionarios de evaluación de usuarios para las pruebas de videowall del Capítulo 5 (Anexo B); por último, el modelo de las autorizaciones de los participantes para realizar las pruebas de usuario (Anexo C).

Capítulo 2

Estado del arte

2.1. Objetivos del Capítulo

El contenido de este Capítulo es una revisión de la literatura sobre la interacción mediante apuntamiento. El tema inicial del Capítulo son los métodos de detección: los dispositivos o sensores que se emplean, las técnicas que permiten desarrollar el apuntamiento, las técnicas de selección y una comparativa de dos interacciones diferentes. El otro gran bloque es el modo de evaluar el rendimiento de los sistemas de apuntamiento, bien sea mediante procedimientos estandarizados o no. También se pone el foco en las aplicaciones para las que puede emplearse la interacción mediante apuntamiento.

2.2. Métodos de detección de apuntamiento

2.2.1. Habilitadores tecnológicos (dispositivos y sensores)

Para llevar a cabo la estimación del apuntamiento es necesario emplear algún dispositivo o sensor que lo habilite. Hoy en día, existen en el mercado multitud de dispositivos comerciales dotados de sensores que permiten detectar apuntamiento con formas tan cotidianas como, brazaletes equipados con sensores inerciales, relojes o anillos. En esta sección trataremos los siguientes: dispositivo (mostrado en la Figura 2.1), *smartphone*, proyectores, gafas, *webcam* y relojes inteligentes. Aparte se detallan de forma simplificada los

dispositivos específicos para el seguimiento de la mirada.

Figura 2.1: Dispositivo kinect versión 2

Uno de los principales dispositivos es Microsoft Kinect. Éste tiene interés para este tipo de interacción debido a sus capacidades y a su bajo coste. Incorpora una cámara de color, un sensor de profundidad de infrarrojos y un micrófono. Permite el seguimiento simultáneo del movimiento de hasta seis usuarios y la estimación de la posición de 25 articulaciones por cada uno de ellos, como indica el estudio [9]. Han existido dos versiones en el mercado, en [10] se puede encontrar una comparativa entre ambas versiones de este sensor. Se comenzó a comercializar en el año 2010 y dejó de estar a la venta en 2016.

En la literatura, hay numerosas propuestas basadas en el sensor Kinect como dispositivo para habilitar el apuntamiento mediante señalización manual (a través del brazo, dedo, etc.). En general, el diseño de los sistemas de apuntamiento presenta problemas técnicos tales como encontrar el vector de apuntamiento más fiable o evitar las oclusiones. Entre los estudios que emplean Kinect están [11], [12], [13], [14], [4], [15], [16] y [17]. En trabajos previos, se ha probado las posibilidades que ofrece Kinect para el reconocimiento del apuntamiento en espacios inteligentes, con un despliegue de dos dispositivos, colocados adecuadamente en las esquinas del espacio [4]. Esta configuración permite implementar una interacción con objetos reales (en este caso, con una fila de lámparas) con una tasa de reconocimiento-interacción que está sobre el 90 % si el usuario se encuentra en el punto de visión óptimo de dos de los dispositivos. Por otra parte, en [18] se emplea el mismo dispositivo pero éste no se coloca de forma fija en un sitio de la sala, sino sobre un robot que se mueve. En [19] se une el uso de Kinect con un *smartphone* para

poder apuntar y actuar sobre pantallas virtuales. Otra de las ventajas que ofrece este dispositivo es que permite el reconocimiento de voz permitiendo la combinación de comandos [20], lo cual amplía las posibilidades de interacción sobre objetos apuntados.

El sistema diseñado en [21] permite a los usuarios seleccionar objetos dentro de un espacio 3D con la ayuda de un ratón 3D. Este utiliza las características del esqueleto captado por Kinect para calcular las posiciones de las articulaciones del usuario del codo y la muñeca que forman el vector direccional. Las mismas articulaciones son usadas para crear el vector de apuntamiento en [22] y así facilitar el manejo de una *smart TV*. El control de robots mediante gestos capturado por Kinect se describe en [23]; el usuario apunta hacia la posición del objetivo y el robot se mueve hacia dicha posición.

En [24], se vuelve a hacer uso de Kinect combinado con otros dispositivos: el usuario lleva unas gafas dotadas de un puntero láser que se emplea para apuntar hacia una imagen. En el entorno está instalada una webcam, colocada de forma que es capaz de reconocer la imagen que ha sido apuntada. Una vez reconocida la imagen, ésta puede ser manipulada a través de gestos de la mano haciendo uso del reconocimiento de gestos estáticos que ofrece el dispositivo Kinect, que reconoce gestos. Por tanto, para el diseño del prototipo intervienen 3 dispositivos diferentes: las gafas con el puntero láser para apuntar y así conocer la orientación de la cabeza, la webcam para ver donde está apuntando, ya que se proyecta el puntero sobre la imagen, y un sensor Kinect para reconocer los gestos que el usuario hace con la mano y permitir la manipulación de la imagen seleccionada. Los autores afirman que el movimiento de las manos puede emplearse para mejorar el apuntamiento que ellos implementan.

Las soluciones de interacción de seguimiento de la mirada se basan en dispositivos específicos que facilitan el seguimiento de los ojos (por ejemplo, Tobii X120, IBM eye tracker, ASLabs 3250 corneal reflection tracker, etc). En concreto, el artículo [25] desarrolla un análisis del apuntamiento usando como instrumento para controlar el seguimiento de la mirada el *Applied Science Lab. Model 3250R corneal reflection eye tracker*. En [26] para permitir el seguimiento de la mirada y el apuntamiento mediante el dedo se emplean el dispositivo *Tobii X120 Eye Tracker* y el dispositivo *Kinect*, respectivamente. Otra versión del anterior dispositivo *Tobii* es el usado en [27]. Los sistemas *EyeLink* o *Irisbond* necesitan tener visión de la pupila del usuario durante su funcionamiento. Proporcionan una alternativa de interacción a personas que tienen limitaciones motoras y del habla. El inconveniente es que necesitan continuamente tener visibilidad del ojo, lo que plantea una

serie de restricciones en el movimiento del propio usuario dentro del espacio. Generalmente, estos dispositivos requieren de unas condiciones específicas de visibilidad para detectar adecuadamente el apuntamiento (algunos de ellos trabajan en infrarrojos por lo que son muy sensibles a la iluminación) estando también restringida la movilidad del usuario. Los sensores usados para seguir la mirada no funcionan correctamente si el usuario no está situado frente al sensor, por lo que en muchas ocasiones estos sensores se montan sobre la cabeza del usuario. Esto conlleva que el usuario emplee un objeto específico para el apuntamiento restando naturalidad al gesto. Tratando de sacar el máximo provecho de los dispositivos comerciales, algunos trabajos tienen la intención de utilizar Kinect como sensor para obtener el seguimiento de la mirada. Por ejemplo, [28] usa su cámara RGB-D para captar el centro del iris del ojo del usuario y sus marcas faciales en 2 dimensiones sin emplear información de profundidad. Actualmente, la empresa *Apple* ha incorporado en su versión del sistema operativo macOS la posibilidad de controlar el ratón con movimientos de la cabeza del usuario.

Modificando la filosofía anterior encontramos el reloj *Watchpoint* presentado en [29]. Este dispositivo permite controlar los movimientos del cursor en una pantalla cercana a través de los movimientos del antebrazo en las cuatro direcciones: izquierda, derecha, arriba y abajo. Por otra parte, la interacción mediante el clic se realiza con gestos de la muñeca. El estudio [30] emplea el dispositivo *Watchpoint* anterior para compararlo con el nuevo concepto de apuntamiento *Twist, Point y Tap (TPT)* implementado a través aplicación Android portable.

2.2.2. Técnicas de detección

Esta sección estudia las técnicas usadas para detectar el apuntamiento. En concreto, se comienza las técnicas de señalización manual. Además se abarcarán los problemas que estas técnicas conllevan como, por ejemplo, las oclusiones. También, se revisan estudios que analizan y emplean el seguimiento de la mirada y el reconocimiento de gestos para estimar el apuntamiento. Por último, se muestra alguna técnica novedosa para conocer el apuntamiento, como el uso de detección de sombras.

Señalización manual

Las tecnologías para detectar el apuntamiento usando el apuntamiento con el brazo o mano (y los dedos de la misma) se han estudiado desde hace

bastantes años. Una muestra de ello, es que ya en el año 1980 se propone un sistema [31] que usa la orientación-posición del usuario para estimar el apuntamiento mediante señalización manual.

Los autores del estudio [12] proponen un nuevo método para detectar el extremo del dedo que se usa para apuntar. Los resultados experimentales demuestran que dicho método es robusto y eficiente para reconocer el gesto de señalar. Emplea el extremo del dedo y no la mano para definir el vector de apuntamiento con un algoritmo que se basa en el reconocimiento de la mano del usuario para poder extraer el *finger tip* (extremo del dedo). Según los resultados mostrados, la tasa de acierto medio es alta, 99.2 %, en la selección y manejo de luces virtuales.

El estudio [13] emplea el vector *codo-mano* al igual que [17]. La mano que no se emplea para apuntar, se usa para realizar un gesto de interacción con la pantalla. Más en detalle, el vector de apuntamiento empleado es el formado por el *codo-mano*, o el formado *cabeza-mano*. Para la fase de la realización de las pruebas, los autores usan ambos vectores para apuntar y además diferencian la posición de los objetos (en la pared, encima de la mesa o en el suelo) y el tamaño de los mismos (pequeño, mediano, grande). El estudio explora tres situaciones domésticas específicas: objetos sobre una mesa, objetos en el suelo, y objetos colocados en una pared, con dos técnicas de apuntamiento diferentes LHH (*head-hand*, cabeza-mano) y LEH (*elbow-hand*, codo-mano). El resultado de las pruebas muestra que la técnica LHH supera en precisión a la técnica de LEH. Esta afirmación es consistente con otros resultados de la literatura, como también se podrá apreciar en el Capítulo 3 de este documento. En cuanto a precisión, la interfaz denominada SEPO permite a los usuarios señalar y seleccionar objetos con una precisión de posición media de $9,6 \pm 1,6$ cm en las situaciones citadas anteriormente. Esto se traduce en poder seleccionar objetos similares a caja de cereales situada en una mesa, un cuaderno sobre un escritorio, un póster en la pared, una aspiradora en el suelo, etc. También se pregunta a los usuario por nuevos métodos de selección, como en las pruebas de usuario realizadas en los siguientes capítulos de esta Tesis (Capítulos 3 y 5). Sus conclusiones coinciden con las nuestras en añadir reconocimiento de voz a la selección. Con esto se consigue una precisión mayor de lo que se quiere apuntar, ya que se amplían las variables que se puede señalar, como puede ser por ejemplo, el color, la forma, etc.

El sistema denominado SELECON se diseña en [32]. Para la detección del apuntamiento usan unos sensores inerciales colocados en la mano y mediante transmisores *ultrawideband* permiten reconocer objetos en su área de cobertura (el objeto seleccionable está dotado también de *ultrawideband*). La Figura 2.2

muestra un esquema del funcionamiento del sistema. Por un lado, los autores reconocen el hecho de reconocer el gesto de apuntar y por otro lado, usan *ultrawideband* para obtener el objeto apuntado. Además, el sistema dispone de un alfabeto propio de gestos (al igual que la Tesis [33]), y permite al usuario realizar acciones sobre los objetos. Los resultados son óptimos, con tasas de éxito de un 84.5 % en cuanto a precisión se refiere en las tareas de selección, y un 97 % en tareas de reconocimiento de gestos.

Figura 2.2: Selección de dispositivos de IoT basada en gestos usando dispositivos portátiles [32].

Los autores de [18] presentan un método de control de un robot mediante gestos de apuntamiento. El usuario indica con el brazo donde debe moverse el robot, y éste se desplaza hacia la posición apuntada. Uno de los retos es hacer instintiva y natural la interacción con el robot. El apuntamiento manual detectado lo realiza el usuario con su brazo y el vector queda definido por *codo derecho-muñeca derecha*. La intersección de esta línea imaginaria con el suelo es lo que da la posición apuntada y es la posición objetivo a la que debe moverse el robot. Al igual que en la presente Tesis es necesario manejar varios sistemas de coordenadas y realizar las transformaciones posibles entre ellas. Antes de elegir ese vector también plantea construir el vector con otros puntos del cuerpo del usuarios, elegidos como puntos extremos del os vectores: *codo-muñeca*, *hombro-muñeca*, o *muñeca-palma de la mano*. Aunque los resultados de precisión de la desviación estándar dan un mejor resultado del vector *hombro-muñeca*, eligen el vector *codo-mano* porque tienen una desviación de la precisión menor.

También para control robótico, en [14] se presenta un método de apuntamiento basado en el reconocimiento del vector formado por el hombro y la

mano. El control de la velocidad a la que se debe mover el robot se indica con un gesto realizado con la mano que no apunta. La altura de esta mano respecto de la cabeza marca la velocidad a la que se mueve el robot. La Figura 2.3 muestra algunos puntos claves de este estudio como: los puntos del usuario que captura el dispositivo Kinect y los vectores de apuntamiento que emplea.

Figura 2.3: Puntos de un cuerpo humano capturado por Kinect y vector de apuntamiento en 2D [14].

En [4] para manejar y controlar recursos domóticos (en concreto luces e imágenes proyectadas en una pared) se colocan dos dispositivos Kinect en las esquinas superiores de la pared donde se proyectan las imágenes, con la correspondiente limitación de posición-espacio que esto conlleva. El apuntamiento debe realizarse de frente a los sensores para estos que capten correctamente el gesto del individuo. El vector de apuntamiento empleado es el formado por codo-muñeca del usuario, a su vez se calcula el vector que va del codo del usuario a los objetos disponibles (filtrando para no tener que calcular de todos los objetos). Con el ángulo que forma el vector codo-muñeca del usuario y el del codo-objeto, se elige cuál de los objetos será el seleccionado por el usuario. Las pruebas realizadas consistieron en apuntar a 9 luces colocadas en el techo y sobre 4 imágenes proyectadas. La tasa de satisfacción tiene unos valores medios de un 90 % de acierto.

Los autores de [34] muestran una interacción en 3 dimensiones con realidad virtual sobre una arquitectura de cueva proyectada. Los vectores para estimar la dirección de apuntamiento son los siguientes: *hombro-muñeca* y *codo-muñeca* y el vector *cadera-muñeca*. Por primera vez, aparece en la literatura el uso del vector *cadera-muñeca* que, a priori, no parece muy preciso. En general, los resultados muestran que el vector *codo-muñeca* es la opción más ruidosa (introduce muchos valores incorrectos). La opción *hombro-muñeca* es la más estable. Por su parte, la opción *cadera-muñeca* es ligeramente más

estable que el resto de opciones. En la práctica, la idoneidad de estas alternativas en tareas de selección en 3 dimensiones depende de una serie de factores. Cuando la precisión del apuntamiento no es crítica (por ejemplo, objetivos grandes no alcanzados), la condición *codo-muñeca* es una opción adecuada que requiere poco esfuerzo físico del usuario. En otros escenarios, las condiciones *hombro-muñeca* o *cadera-muñeca* son más adecuadas, pero los movimientos son menos naturales y requieren un mayor esfuerzo físico.

Por su parte, el artículo [16] describe un prototipo para hacer posible la interacción multimodal con contenidos multimedia que aparecen en una ventana virtual. El vector de apuntamiento empleado es el formado por los puntos que tienen como origen el hombro y la posición *hand-tip* de la mano (extremo de la mano).

La problemática de las oclusiones hay que tenerla en cuenta. Cuando se estima el vector de apuntamiento se usan dos posiciones, el problema es que una de ellas puede quedar ocluida por la otra, lo que introduce errores y estimaciones imprecisas. Los autores de [11] proponen un método de reconocimiento del apuntamiento por gestos sobre una gran pantalla basado en el modelo de esqueleto del dispositivo Kinect. Teniendo en cuenta el problema de la auto-oclusión, la posición detectada del hombro de una persona ayuda a estimar la posición de la mano, en el caso de que exista una oclusión de la mano.

En general, la literatura existente demuestra que las técnicas de interacción con la mano, actuando de forma libre, no son muy precisas. Los autores de [35] afirman que esto puede deberse a dos factores: las desviaciones causadas por el temblor natural de las manos y la limitada precisión de la puntería humana. Para abordar estas cuestiones, proponen una técnica de transformación adaptativa para mejorar el ajuste tradicional del mapeo de posición a posición. Se emplea un puntero láser para evaluar esta técnica, existiendo una reducción del tiempo de movimiento de un 19 % y una tasa de error de un 16 %.

Una forma novedosa de estimar el apuntamiento es mediante la detección de sombras dependiendo de la distancia a la que esté el usuario del objeto a seleccionar. El artículo [15] demuestra que es posible estimar el apuntamiento de un usuario sobre una pantalla virtual mediante el reconocimiento de sombras y la detección de la profundidad del usuario (a la distancia a la que se encuentra el usuario de la pantalla). Una muestra de este sistema puede verse en la Figura 2.4. Para ello emplean un dispositivo Kinect y le muestran al usuario sobre la pantalla con la que interactúan, la sombra generada estando la luz colocada en un sitio fijo.

Figura 2.4: Las sombras virtuales de los usuarios se superponen al contenido de la pantalla. Usando el apuntamiento detectado (círculo verde), los usuarios interactúan con los contenidos 2.4.

Seguimiento de la mirada

Una alternativa/complemento para la estimación del blanco apuntado es conocer hacia donde está mirando la persona (seguimiento de la mirada). Ésta, aparte de estimar el apuntamiento, permite conocer el interés que muestra una persona hacia cierto elemento del entorno, pudiendo ser empleada esta información para administrar las capacidades disponibles en los entornos inteligentes. En [36] puede encontrarse un completo estudio sobre técnicas de seguimiento de la mirada, junto a sus soluciones tecnológicas también se recogen sus limitaciones.

La estimación de la mirada habitualmente se calcula como la combinación de los datos procedentes de la posición de la cabeza de la persona y la posición de sus ojos [37]. También es posible seguir la pupila del usuario, con el objetivo de no tener que tener continuamente visibilidad del ojo. Esta técnica consiste en calcular la dirección de la mirada del usuario como una combinación de la posición de la cabeza y la orientación de la cara del mismo. En este sentido, estudios previos como [38] demuestran que la orientación de la cabeza contribuye en un promedio de un 68.9 % a la dirección de la mirada. Aparte, es más fácil de estimar que otros parámetros.

Con el objetivo de subsanar los problemas que conlleva realizar el exhaustivo seguimiento de la mirada, se intenta aproximar ese seguimiento de la mirada usando la posición-orientación de la cabeza. Actualmente, la posición-orientación de la cabeza se emplea para mejorar la estimación de la localización del ojo, como muestran [39], [40] y [41]. Por su parte, los autores de [42] explican que cuando los sujetos apuntan con el dedo a lugares predecibles o impredecibles marcados por la aparición de una luz, no hay diferencias entre la iniciación del movimiento de la cabeza o del ojo. En [43], los autores consideran suficiente el empleo de la posición-orientación de la cabeza para estimar el foco de atención de un conductor. El estudio obtiene

una exactitud en promedio del 88.7 % dentro de un escenario complejo como puede ser una reunión con varios participantes. Aunque la respuesta motora puede variar dependiendo de la situación y de la tarea [44], intentar realizar el seguimiento de los ojos empleando la posición-orientación de la cabeza puede ser una solución para algunas aplicaciones.

2.2.3. Técnicas de selección

Un aspecto importante en la interacción basada en apuntamiento es como seleccionar los objetivos apuntados. Esta tarea es particularmente delicada cuando la interacción es a través de interacción con seguimiento de mirada. Entre las técnicas encontradas en la literatura están: *primero en entrar* (*first-entry*) que consiste en que sea suficiente seleccionarlo una única vez y el *tiempo de permanencia* apuntando sobre un objeto (*dwell-time*); que requiere mantener la selección un periodo de tiempo sobre el objeto seleccionado. La técnica de permanencia mirando hacia un objetivo se analiza en [25] considerando el rendimiento del dispositivo tipo ratón como referencia. Los resultados se obtienen gracias a 16 usuarios que llevan a cabo tareas básicas de selección de círculos destacados sobre un panel con más círculos y con una selección mediante voz sobre un panel de letras. El tiempo de permanencia con la mirada para seleccionar objetos se fija en 150 ms por encima de los 116 ms estimados para el clic del ratón a pesar de la inmadurez comercial de la tecnología de seguimiento ocular. El *tiempo de permanencia* se emplea como método de selección en [27]. Este tipo de selección se usa para comparar con selección a través de la pulsación de una tecla. Los autores de [27] resumen los métodos de selección de 6 estudios diferentes. Los métodos que aparecen aquí son *tiempo de permanencia*, pulsación de tecla, sonreír o guiñar un ojo.

La combinación de gestos con la selección a través de la mirada también ha sido estudiado como solución de selección multimodal, por ejemplo, la combinación de la mirada con los movimientos o comportamiento de la cara. El estudio [45] emplea la sonrisa como método de selección. Las pruebas en este estudio son llevadas a cabo por cinco usuarios dotados de un prototipo montado sobre su cabeza para realizar el seguimiento del ojo, que tiene a su vez un electrodo que detecta los movimientos faciales como por ejemplo, la sonrisa. Durante este estudio no se observa degradación de la precisión del seguimiento de la mirada por los movimientos involuntarios de la cara durante la selección de objetos, aunque el estudio no proporciona una evaluación de la eficacia o la experiencia del usuario. En [46] también se utiliza la sonrisa como gesto de selección.

Por otra parte, la técnica combinada de señalización manual con la técnica del reconocimiento de voz se usa en [31]. En [47] se une el reconocimiento de voz con el apuntamiento para tareas de selección. Este último estudio utiliza el *iView XTM* de *SensoMotoric Instruments* para el seguimiento de los ojos; doce participantes realizaron una tarea de selección con la mirada, tanto para apuntar y seleccionar como para usar la voz. Sus autores afirman que el tiempo de permanencia apropiado para la selección del blanco fue de 1500 ms. El tiempo de respuesta es mejor con el uso de la técnica de seguimiento de la mirada frente a la solución que combina la anterior con el reconocimiento de voz. Sin embargo, el número de errores es mayor con el empleo únicamente de la técnica de seguimiento de la mirada.

2.2.4. Estudios comparativos de interacción mediante apuntamiento basado en diferentes técnicas

En la línea del trabajo presentado en el Capítulo 3 de este documento existen estudios previos comparando dos tipos de apuntamiento: posición-orientación de la cabeza y señalización manual. Los autores de [48] se centran en analizar métodos de apuntamientos usando la mirada y la mano para interactuar con objetos estereoscópicos dentro de entornos virtuales empleando el dispositivo Kinect. El vector de apuntamiento no se define a partir de la posición del brazo, pero toma la posición de la mano del participante proporcionada por Kinect y aplica una corrección para transformar el desplazamiento de mano en el desplazamiento del cursor. Para la detección del gesto de *clickear* se emplea un guante especial (*5DT Data Glove*). El resultado muestra que el tiempo de movimiento fue significativamente mayor para la interacción empleando el seguimiento de la mirada (5070 ms vs. 3280 ms). Este mismo comportamiento se repitió para la tasa de error, 37.6 % vs. 27.1 %. Los valores de *rendimiento del sistema* obtenidos fueron bastante bajos, tanto en el caso de apuntamiento con la mano (1.13 bits/s), como con la mirada (0.75 bits/s). El análisis de la usabilidad muestra mejor aceptación para el método de apuntamiento con la mano.

Por otra parte, los autores de [49] comparan la técnica de apuntamiento basado en la mirada con el apuntamiento con la mano como método de interacción en entornos de realidad virtual. Un dispositivo de seguimiento de la mirada binocular es montado dentro de una pantalla tipo *Virtual Research V8 Head Mounted Display* sobre la cabeza. Como herramienta de apuntamiento se emplea un dispositivo tipo ratón *6DoF*. El *tiempo de permanencia* apuntando sobre un objeto se emplea como método de selección para el apun-

tamiento con la mirada. Durante las pruebas 24 participantes interactuaron con 20 objetos geométricos distribuidos en un mundo virtual de 3 dimensiones. En ambos casos, se proporcionó *feedback* visual al usuario para conocer en todo momento la posición apuntada. Los autores encontraron que el seguimiento de la mirada era más lento que el realizado con la mano para objetos que se encontraban lejos.

En [25] se comparan las técnicas de apuntamiento anteriores (señalización manual vs mirada), y a pesar de las limitaciones tecnológicas, la conclusión del estudio plasma que a mayor distancia del objetivo, mayor ventaja en precisión y velocidad presenta el apuntamiento basado en seguimiento de mirada. Estos autores añaden que el apuntamiento mediante mirada es útil para apuntar hacia regiones lejanas, y se intuye poco práctico para las tareas de apuntamiento muy preciso.

El apuntamiento mediante seguimiento de la mirada y señalización manual es empleado en [26] para manejar la interfaz que aparece en un monitor. La Figura 2.5 explica de forma esquemática el funcionamiento del sistema. La selección de objetivos se realiza cuando el usuario muestra la mano contraria a la empleada para apuntar. La mirada combinada con el apuntamiento con el dedo parece ofrecer una tasa de error menor que si solo se emplea la primera técnica. Las diferencias de tiempo de movimiento no son estadísticamente significativas. Es importante señalar que el apuntamiento no se calcula desde la posición del brazo, sino desde la posición de la mano en sí (el sistema está diseñado para interacción con ordenador a distancia muy corta).

Figura 2.5: La mirada controla un área de interacción (que se muestra como un rectángulo en el monitor). El seguimiento con la punta de los dedos se usa para controlar el puntero dentro de ese área de interacción

El artículo [27] compara también el seguimiento de la mirada con la detección de la pupila del ojo y el apuntamiento usando la posición de la cabeza sobre pantallas de tamaño reducido como puede ser la de un *smartphone*. Los

resultados demuestran que el apuntamiento usando la posición de la cabeza no se ve afectado por apuntar sobre pantallas de tamaño pequeño.

2.3. Evaluación del rendimiento de sistemas de apuntamiento

2.3.1. Procedimientos estandarizados

Un modo de evaluar los sistemas de apuntamientos es utilizar la Ley de Fitts y la Norma *ISO 9241-9:2000* publicada por la Organización Internacional de Estandarización. Aunque se explican estos dos aspectos en detalle durante el desarrollo de la Sección 3.2.1, se realizará una pequeña introducción en esta Sección.

La Ley de Fitts permite modelar el acto de apuntar. Desde que se formuló por primera vez en el año 1954, se han publicado versiones diferentes de la misma, siendo la mas popular la formulación de Shannon (detalle en 3.2.1, ecuación Ecuación 4). Los principales parámetros contemplados en esta Ley son: índice de dificultad (ID, *Index of difficulty*), rendimiento (TP, *Troughput*) y el tiempo de movimiento (MT, *Movement time*). También en la sección 3.2.1 se encuentran las ecuaciones que definen estos parámetros respectivamente: Ecuación 1, Ecuación 2 y Ecuación 3)Por otra parte, el estándar *ISO 9241-9:2000* recoge los requisitos ergonómicos para el trabajo de oficina con terminales de pantalla visual para dispositivos de entrada que no son teclado.

Por ejemplo, los estudios: [50] [25] [51] [26] [52], [53], [54] [48] y [30] usan la Ley de Fitts para evaluar las tareas de apuntamiento de las pruebas que desarrollan, usando como principios subyacentes la norma ISO 9241-9:2000. Un factor común en estos estudios es el calculo del ID usando la formulación propuesta por Shannon. Teniendo en cuenta los estudios que utilizan los principios de la norma ISO 9241-1, el TP mínimo obtenido es 0.75, obtenido por el dispositivo *Wii mote* en tareas de 3 dimensiones [51] o con en apuntamiento calculado con posición-orientación de la cabeza empleando Kinect [48]. Excluyendo el ratón, el TP máximo se obtiene empleando el dispositivo *MS Xbox* en [54]. Con respecto a la bondad del ajuste del modelo de regresión para el apuntamiento mediante señalización manual (con el dedo), r^2 apenas alcanza 0.27 para las tareas en 2D con Kinect 2D [26] y sube hasta 0.51 en [52]. La comparación con *Wii mote* realizada en [51] obtiene un valor de

r^2 de 0.73 para tareas también de 2 dimensiones. En [25] se muestra el escaso ajuste de la técnica de seguimiento de la mirada a la ley de Fitts ($r^2 = 0.02$). Tanto *Wiimote* [51] como Kinect V1 [52] parecen funcionar mejor que el ratón en para tareas de 3 dimensiones, aunque el ratón obtiene un ajuste menor que otros estudios para tareas de 2 dimensiones.

Con respecto a la evaluación del rendimiento, en el estudio [53] se evalúa el apuntamiento realizado por un usuario con una de sus manos y la otra mano se usa para seleccionar haciendo un imaginario *click* en el aire, de acuerdo a la estrategia seguida en [26] y [55]. Los autores comparan el rendimiento de los dos métodos de apuntamiento (codo-muñeca frente a hombro-muñeca) empleando la Ley de Fitts. La conclusión del estudio es que el vector de apuntamiento basado en hombro-muñeca obtiene mejores resultados (valor medio de 24.3 %) que el basado en codo-muñeca. Los objetivos de este estudio tienen un diámetro que va de 39 cm a 4.6 cm, ordenados según la distancias de la circunferencia variando de los 28 cm a los 79.5 cm, y se seleccionan a una distancia de 3.1 m. En dicho estudio, no se incluye ni la evaluación del ajuste del modelo ni el análisis del TP, ni el análisis de la usabilidad o de la experiencia del usuario.

En [30], también realiza una comparativa de métodos de apuntamiento. Para el primero de los métodos se emplea un dispositivo reloj dotado de sensores, *Watchpoint* y del concepto de apuntamiento *Twist, Point y Tap (TPT)* habilitado gracias a una aplicación Android portable. Según los resultados mostrados el concepto TPT es menos propenso a errores que *Watchpoint* con igual rendimiento. El estudio aplica las tareas de apuntamiento para 3 tamaños objetivos con 3 distancias diferentes y empleando cada uno de las interacciones a comparar. Obtiene un parámetro de correlación para el TPT de 0.86 frente al valor para *watchpoint*. La tasa de error medio en la combinación de distancia y tamaño del objetivo es de 6.3 % en TPT y de 13.8 % para el *watchpoint*.

El rendimiento para tareas de apuntamiento en 2D y 3D del dispositivo Kinect V1 se analiza en [52] combinado con un dispositivo tipo *ratón*. La norma ISO 9241-9 se aplica para definir la tarea de evaluación multidireccional en 2D, mientras que el 3D se evalúa a través de una interfaz cúbica con objetivos en las esquinas. Para las tareas 2D con Kinect, el TP es un 39 % menor que usando el dispositivo tipo ratón; mientras que para las tareas en 3D, el ratón proporciona un TP con un valor de un 9 % menor que el conseguido para el dispositivo Kinect. Estos mismos autores estudian también el comportamiento del dispositivo Wii, obteniendo unos resultados similares [51]. Por su parte, [54] recoge un análisis comparativo de diferentes dispositivos de

entrada modernos para juegos y que permiten realizar un disparo en primera persona dentro de espacios de juegos en 3D. Los dispositivos estudiados son: un ratón, videoconsola *Microsoft XBox 360*, videoconsola *Sony PlayStation move* (un sistema de control de videojuegos por movimiento) con cámara de precisión para la detección del tracking del ojo, y por último, el dispositivo Kinect. Las pruebas de este estudio fueron realizadas por 12 participantes y los objetivos se colocaron en ubicaciones predeterminadas pero impredecibles en un entorno 3D a 3 distancias diferentes (5, 10 y 15 metros). Los resultados indican que el ratón era el mejor dispositivo de acuerdo al tiempo de movimiento, con el controlador del juego en un cercano segundo puesto. La velocidad de error fue mayor para el controlador que para el ratón. El rendimiento de los dispositivos de entrada 3D (Move y Kinect) fue peor, en particular Kinect funcionó peor en términos de tiempo de movimiento para objetivos más cercanos.

En la literatura revisada, se han encontrado pocos estudios que analicen estos aspectos (comodidad, usabilidad y experiencia del usuario) para técnicas de apuntamiento, en especial sobre el apuntamiento mediante señalización manual y seguimiento de la mirada empleando la posición y orientación de la cabeza. Por ejemplo [55] realiza un análisis del confortl y [48] emplea una escala de *Borg* de 10 valores para medir el esfuerzo del usuario al usar ese tipo de sistemas de apuntamiento.

2.3.2. Otras metodologías

En esta sección se recogen otras pruebas dirigidas al uso y valoración del sistema por parte de los usuarios.

En las pruebas desarrolladas en [24] a los usuarios se les muestra un objetivo circular sobre la pantalla en posiciones aleatorias. Se les pide señalar cada objetivo a través de un cursor y con las técnicas desarrolladas en el artículo. Cada participante repitió esta simple tarea varias veces. Las distancias entre dos objetivos temporalmente adyacentes eran idénticas para mantener el mismo esfuerzo físico de mover el cursor para cada tarea. Los resultados muestran que, con la ayuda del seguimiento facial, el 82,6 % de las tareas se realizaron con un solo intento de movimiento del cursor, mientras que el 74,4 % de las tareas solo con la mano requirieron dos o tres intentos. Esto indica que el uso del seguimiento facial redujo significativamente la distancia de movimiento del cursor eventual y, por lo tanto, el tiempo total de la tarea.

Las pruebas del sistema planteado en [14] se estructuran en dos tipos de

experimentos. El primer experimento pretende medir el ángulo de error entre la dirección asignada y el resultado en la situación real. El experimento es llevado a cabo por siete personas, tomando como referencia 7 puntos ideales que corresponden con los puntos cardinales durante 10 intentos. El segundo experimento controla la navegación del robot haciendo un recorrido de una sala real.

Las pruebas de [19] consisten en seleccionar una imagen y desplazarla dentro de un rectángulo. Se comparan diferentes tamaños de imagen y distintas distancias, para manejar y obtener los resultados, comparando el tiempo de selección de las 4 técnicas, con dos tamaños de imagen y el tiempo de movimiento basado a su vez en 2 distancias diferentes. En general, los resultados muestran que el manejo de imágenes pequeñas conlleva tiempos de selección más altos para todas las técnicas empleadas. En cuanto al tiempo de movimiento para las 2 distancias, el uso del *smartphone* junto al dispositivo Kinect, muestra un tiempo de movimiento mayor para distancias más pequeñas frente a distancias mayores. En general esta técnica combinada es buena para la interacción con objetivos de tamaño grande, pero no es muy adecuada para tareas de interacción con objetivos pequeños. También existe una latencia en la implementación de Kinect que puede justificar esta última afirmación.

El experimento de [11] que realizan los usuarios del prototipo consiste en mover el cursor de un ratón a través de 5 puntos definidos en la pantalla. Este estudio se centra en mejorar el apuntamiento con gestos y el reconocimiento de gestos realizado con los dedos. Los resultados obtenidos son bastante satisfactorios, obteniendo una tasa de error de 13 píxeles en pantallas de un total de 1920x1080 y una tasa de reconocimiento de gestos próxima al 100 %, en concreto del 98.3 %.

En [27] se divide a los participantes de las pruebas en dos grupos de acuerdo a la experiencia previa con la tecnología que se quiere probar: expertos y no expertos. La división se ve reflejada en los resultados obtenidos también.

2.4. Aplicaciones y clasificación

Las tecnologías de apuntamiento se aplican con diferentes usos: controlar objetos inteligentes [4]; también, dirigir robots [23], [18] y [14]; aumentar información sobre vídeos [16]; controlar dispositivos electrónicos como la televisión [22]; navegar dentro de espacios virtuales [56]; el control de objetos

domóticos [12]; y el manejo del control del vuelo de drones como se muestra en [20] .

Por otra parte, dentro de estos espacios inteligentes, otro de los usos comunes es la interacción con grandes pantallas. Ejemplos de estas aplicaciones se encuentran en [11], donde es posible manejar pantallas mediante apuntamiento. En la misma línea, los autores de [24] diseñan un prototipo para el manejo de imágenes mostradas en grandes pantallas. Orientado al manejo de imágenes sin contacto físico con pantallas, y en concreto, dentro del área médica está el estudio [57]. Por su parte, [19] también explora técnicas de interacción natural para el manejo de varias pantallas y uniéndolas en una única. Otro ejemplo es [58]. En este artículo, sus autores desarrollan un prototipo para el manejo de forma natural de imágenes virtuales dentro de un entorno virtual gracias al reconocimiento de gestos de la mano implementando realidad aumentada. La aplicación del estudio es el manejo de imágenes virtuales colocadas en una pantalla virtual también en el entorno del usuario.

La tabla 2.1 recoge algunos de los estudios que han ido apareciendo a lo largo de este capítulo destacando sus aspectos más relevantes. Los aspectos considerados relevantes han sido :

- Tipo de entorno.
- Tipo de apuntamiento (detección del apuntamiento) y el método de selección implementado.
- Dispositivo o dispositivos empleados durante el estudio.
- Gestión de oclusiones.
- Tasa de error, importante para saber si es adecuado como técnica de apuntamiento.
- Apuntamiento sobre imágenes proyectados u objetos reales.

Esta pequeña muestra de 15 estudios destaca la importancia del desarrollo de servicios para espacios reales usando técnicas de apuntamiento, dejando a un lado los espacios virtuales. En este caso, 12 de los 15 estudios totales se centran en este tema. Los tres restantes siguen estudiando los espacios virtuales.

Con respecto al método de apuntamiento utilizado predomina el apuntamiento con el brazo, estando muy presente el reconocimiento de gestos. El dispositivo empleado más usado es Kinect, 10 de estos 15 estudios lo emplean.

Otros se centran en otras soluciones comerciales como *Leap Motion* o prototipos diseñados *ad hoc* como es el caso de [24] con sus gafas. La prevención de las oclusiones solo se abarca en uno de los 15 estudios, en los otros no se trata, bien porque se ignora o porque la naturaleza del sistema no aplica. El último de los aspectos es conocer si el apuntamiento se realiza sobre objetos proyectados o objetos reales, estando muy ligado al tipo del entorno donde se desarrolla el servicio. En líneas generales, los estudios de la tabla están más centrados en interaccionar con objetos reales.

Tabla 2.1: Tabla resumen de estudios sobre apuntamiento

Título	Referencia	Tipo de entorno	Tipo de apuntamiento (vector) y método de selección	Dispositivo empleado	Evita las oclusiones	Tasa error	Apuntamiento sobre imágenes proyectados o objetos reales
Method for user interface of large displays using arm pointing and finger counting gesture recognition	[11]	Espacio inteligente real	Apuntamiento con el brazo (mano derecha-hombro) y reconocimiento de los gestos del dedo	Kinect	Sí	En apuntamiento: 13 px en 1920x1080 y en reconocimiento de gestos: 98.3 % aciertos	Si
Human computer interaction using pinning gesture based on an adaptative virtual touch screen	[12]	Espacio inteligente real	Apuntamiento con la punta del dedo (finger tip)	Kinect	No	99.2 % media de tasa de acierto de selección sobre pantalla virtual.	Si
SEPO: Selecting by Pointing as an Intuitive Human-Robot Command Interface	[13]	Espacio inteligente real	Apuntamiento empleando 2 configuración posibles: el vector cabeza-mano dominante o vector codo-mano dominante. Además existe un gesto con la otra mano para interactuar.	Kinect	No	precisión promedio de posición 3D de 9.6 cm. Exactitud de posición promedio de 9.6 ± 1.6 cm en situaciones del hogar, eso significa que podemos seleccionar con éxito objetos similares a: caja de cereal en la mesa de la cocina, una libreta en el escritorio etc.	No, sobre objetos reales (del espacio 3D)
Inmersive 3D environment for data centre monitoring base don gesture based interaction	[59]	Espacio virtual	No apuntamiento – Reconocimiento de gestos	Leap motion y gafas de realidad virtual	No aplica	No definido	No aplica
SeleCon: Scalable IOT device selection and control using hand gestures	[32]	Espacio inteligente real	Reconocimiento de apuntamiento con la mano y reconocimiento de alfabeto de gestos	Sensores inerciales y receptores ultrawide-band.	No aplica	84.5 % de precisión para la selección de dispositivos y 97 % del reconocimiento de gestos	No sobre objetos reales

Hand gesture controls for image categorization in immersive virtual environments	[58]	Espacio virtual	Reconocimiento de gestos con la mano (incluido apuntamiento con la mano)	Dispositivo captura de movimiento portable en la mano y dispositivo montado sobre la cabeza (gafas) para realidad virtual	No aplica	No definido	Manejo de imágenes virtuales sobre pantalla virtual (dentro del entorno virtual)
Facial Tracking-Assisted Hand Pointing Technique for Wall-Sized Displays	[24]	Espacio real inteligente (manejo de pantalla)	Apuntamiento y reconocimiento de gestos	Gafas normales con puntero láser(para apuntar), webcam para ver donde apunta el puntero láser y Kinect para el reconocimiento de gestos de la mano	No	con la ayuda del seguimiento facial, el 82,6 % de las tareas se realizaron con un solo intento de movimiento del cursor, mientras que el 74,4 % de las tareas solo con la mano requirieron dos o tres intentos.	Manejo de imágenes sobre grandes displays-pantallas de forma real.
Foundations of Visual Linear Human-Robot Interaction via Pointing Gesture Navigation	[18]	Espacio real (manejo de robot)	Apuntamiento mediante reconocimiento vector codo-muñeca	Kinect sobre robot	No	Muestra resultado de precisión en distancia no tasas de errores o aciertos.	Manejo de robots mediante apuntamiento.
3D pointing gesture recognition for human-robot interaction	[14]	Espacio real (manejo de robot)	Apuntamiento mediante reconocimiento de vector hombro-mano,	Kinect	no	El resultado no está indicado con un porcentaje de error, depende del tipo de experimento (explicado en el detalle de la explicación)	Apuntamiento mediante reconocimiento de vector hombro-mano, y control de la velocidad de movimiento del robot mediante reconocimiento de gestos con la otra mano que no apunta.
Depth based shadow pointing interface for public displays	[15]	Entorno real de interfaz en pantalla virtual)	Apuntamiento mediante reconocimiento de sombras y profundidad	Kinect	No	No proporciona resultados	Manejo de interfaces virtuales gracias a apuntamiento a través del reconocimiento de sombras.

A Kinect-based system to enable interaction by pointing in Smart spaces	[4]	Entorno real (interacción con objetos reales del entorno)	Apuntamiento mediante reconocimiento del vector codo-muñeca	Kinect	No	La tasa de satisfacción y error durante el apuntamiento a diferentes objetos del espacio inteligente (con valores medios de un 90 % de acierto y un 10 % de fallo)	Interacción con recursos de domótica del entorno, se emplea el vector codo-muñeca para apuntar y se usa también el vector codo-objeto. El ángulo que forman estos dos sirve para saber cuál es el objeto apuntado dentro de los existentes.
A system for multimodal interaction with enabled virtual Windows	[16]	Interacción en espacio real	Apuntamiento mediante vector hombro-extremo de la mano	Kinect	No	No aparece	Interacción en espacio real sobre imagen virtual en pantalla.
A multimodal interaction system for big displays	[20]	Interacción en espacio real	Apuntamiento mediante vector hombro-extremo derecho de la mano.	Kinect	No	No aparece	Manejo de grandes pantallas mediante apuntamiento y voz.
HydraScope: Interaction and Application of Wall Sized Display	[19]	Interacción en espacio real	Interacción remota apuntamiento mediante reconocimiento de gesto con el brazo y la mano(en el uso de Kinect) y varias técnicas con el smartphone.	Smartphone y Kinect	No con Kinect, si con smartphone	No aparece tasa, usa Ley de fits	Manejo de varias pantallas que forman una grande mediante interacción remota usando el Smartphone y Kinect.
A mutli-projector CAVE sytem with commodity hardware and gesture-based interaction	[34]	Entorno virtual proyectado.	Apuntamiento y selección mediante reconocimiento del vector hombro-muñeca, codo-muñeca (ambos lado dominante) y cadera-mano.	Kinect	No	No aparece	Interacción con objetos proyectados en la cueva que es proyectada.
Smartwatch-based pointing interaction	[30]	Espacio real	Apuntamiento con la muñeca y reconocimiento de los gestos del dedo	Smartwatch, aplicación portable Android	No	Tasa de error de 6.3 %para la aplicación portable Android y 13.8 % del watchpoint	Si para control de un cursor sobre una pantalla
Recognition and localisation of pointing gestures using a RGB-D camera	[17]	Espacio inteligente real	Apuntamiento con el brazo (codo-mano)	Kinect	No	Tasa de acierto de 89.59 % y 79.92 % para brazo izquierdo y derecho respectivamente sobre matrices 2x3 y de 73.57 % y 68.99 % en matrices 3x4	Sobre tablero o matriz real de 1290mm x 1910mm

2.5. Resumen

Este Capítulo ha revisado el estado del arte en aspectos relacionados con la interacción mediante apuntamiento. El Capítulo comienza esta revisión considerando los métodos de detección del apuntamiento, a partir de las diferentes alternativas de dispositivos o sensores que se pueden usar para desarrollar este tipo de interacción. Estos dispositivos van desde dispositivos experimentales hasta productos comerciales como relojes, anillos o gafas inteligentes. En esta revisión se ha puesto especial atención en el dispositivo comercial Kinect, un dispositivo dotado de una cámara RGB y un sensor de profundidad junto a un micrófono. En función del dispositivo escogido, el entorno y del servicio a implementar hay que escoger la técnica de detección del apuntamiento. Las alternativas consideradas están basadas en señalización manual y en emplear la mirada del usuario. En este punto, el sistema será capaz de calcular cual es el objetivo apuntado por el usuario. Ahora debe completarse la interacción pudiendo seleccionar el objetivo apuntado. Las diferentes alternativas existentes para poder realizar esta selección también se abordan en la Sección 2.2.3. Finalmente se realiza una comparativa de la interacción mediante apuntamiento implementado con diferentes técnicas de detección y selección.

Por otra parte, es necesario elegir cómo se evalúa el rendimiento de los sistemas que implementan interacción mediante apuntamiento. Para ello existen procedimientos estandarizados y no estandarizados. Los parámetros empleados para llevar a cabo la valoración de los resultados miden el tiempo, la precisión, los aciertos del usuario, etc. Los procedimientos y tareas que se desarrollan para conseguir los valores de estos parámetros.

La última Sección está orientada a conocer las aplicaciones de la interacción por apuntamiento. Además, se recoge en una tabla un resumen de algunos de los estudios que han aparecido a lo largo del Capítulo para destacar de forma más clara los siguientes aspectos (considerados importantes para este tipo de interacción): tipo de entorno, detección del apuntamiento y método de selección empleado, si realiza gestión de oclusiones, tasa de error, y por último, si el apuntamiento se realiza sobre imágenes reales o proyectadas.

Capítulo 3

Comparación de técnicas de detección de objetivo apuntado basadas en señalización manual y posición-orientación de la cabeza

3.1. Objetivos del Capítulo

En este Capítulo se exploran y comparan las técnicas de detección del objetivo apuntado basado en señalización manual (dirección obtenida de la relación física del cuerpo del usuario frente al entorno) y seguimiento de la mirada estimada de la posición y orientación de la cabeza. Se describe la solución tecnológica sobre la que se han implementado las diferentes estrategias de detección haciendo uso del dispositivo Microsoft Kinect en su versión 2.0. Con esta base tecnológica, se ha realizado un estudio de usuario para evaluar y comparar el rendimiento y confort del sistema de apuntamiento (usando la Norma ISO 9241-9 [60]) y analizar su usabilidad (medido en los términos de la escala de usabilidad del sistema definida en [7]) y la experiencia de usuario (evaluada a través del cuestionario de experiencia de usuario de [8]). Este estudio proporciona un análisis integrado del rendimiento del sistema, junto con la usabilidad y la experiencia del usuario para las dos técnicas mencionadas anteriormente.

Cada una de las dos técnicas de detección del apuntamiento emplean un vector diferente para estimar el objetivo apuntado. La Figura 3.1 recoge una descripción gráfica de cada uno de los vectores que se usan en cada estas

dos técnicas. La imagen de la izquierda (Figura 3.1.a) refleja el vector que se emplea para la técnica de señalización manual. Este vector tiene como origen el hombro y como destino el dedo de la mano del mismo lado. La imagen de la derecha (Figura 3.1.b) describe gráficamente el vector que se emplea en la técnica de simulación del seguimiento de la mirada basada en la estimación de la posición y orientación de la cabeza. Esta técnica tiene origen la posición del centro geométrico de la cabeza y la dirección del mismo viene determinada por la orientación de la cara.

Figura 3.1: Descripción gráfica de los vectores empleados en la técnica de detección del objetivo apuntado: a) señalización manual; y, b) estimación del seguimiento de la mirada con la posición y orientación de la cabeza.

Con la comparación de estas dos técnicas, y por consiguiente los vectores empleados se persigue encontrar una técnica fiable para implementar un sistema de apuntamiento que sirva para estimar los posibles objetivos apuntados por los usuarios.

Parte de los resultados de este Capítulo se recogen en [61]. Inicialmente se realiza una introducción a la Ley de Fitts y a la Norma ISO 9241-9 como marcos de referencia para la comparación de sistemas de apuntamiento; además, se define el entorno donde se coloca el sistema de apuntamiento, su arquitectura y se detalla el proceso de estimación del apuntamiento dentro del sistema. Posteriormente se presenta el estudio para evaluar el rendimiento del sistema diseñado y se presentan los resultados obtenidos.

3.2. Conceptos previos del estudio

Como el objetivo del Capítulo es estudiar el rendimiento de varias técnicas de apuntamiento, es necesario definir el marco de comparación (y, por ende,

el sistema tecnológico) que se utilizará para llevarla a cabo. Esta sección describe la metodología formal en la que se basa el estudio y la infraestructura con la que se desarrollan las pruebas. La primera parte está centrada en la metodología que se ha seguido, principalmente la Ley de Fitts y la norma ISO 9241-9. La segunda parte describe en detalle el entorno diseñado donde se ha desplegado el servicio de apuntamiento y se han llevado a cabo las pruebas.

3.2.1. Ley de Fitts y ISO 9241-9

La aplicación de la Ley de Fitts permite evaluar el rendimiento del usuario cuando utiliza sistemas de apuntamiento. Los datos recogidos para aplicar dicha Ley provienen de la realización de las tareas propuestas en el Estándar ISO 9241-9 [60].

En 1954, Paul Fitts propuso una forma de medir la dificultad a la hora de realizar una tarea de apuntamiento, con el objetivo de evaluar el sistema motor de una persona en sus funciones sensoriales, perceptibles y perceptivo-motoras [62]. La métrica propuesta por Fitts recibió el nombre de *índice de dificultad* (ID , con bits como unidad, ecuación Ecuación 1). Este ID está inspirado en la teoría de procesamiento de información, y pone en relación la distancia al objetivo a seleccionar (D , o la señal) con su tamaño o ancho (W , el ruido). Además, Fitts propuso una métrica complementaria, el *índice de rendimiento* (posteriormente denominado rendimiento únicamente, TP de su denominación en inglés *throughput*) para medir el rendimiento humano durante las tareas de apuntamiento. Esta última métrica combina el índice de dificultad de las tareas con el *tiempo de movimiento* (MT , en inglés *movement time*) para alcanzar el objetivo (ecuación Ecuación 2).

$$ID = \log_2\left(\frac{2D}{W}\right) \quad (\text{Ecuación 1})$$

$$TP = \frac{ID}{MT} \quad (\text{Ecuación 2})$$

En el área del conocimiento de la Interacción Persona-máquina *Human Computer Interaction* se aceptó la Ley de Fitts como un modelo válido de métrica y se convirtió en una forma habitual de ajustar modelos predictivos

para estimar el rendimiento de un sistema. La Ley de Fitts es un modelo predictivo de movimiento humano que predice que el tiempo requerido para desplazarse hacia un objetivo es una función lineal de la relación entre la distancia al blanco y su ancho.

Un movimiento durante una tarea modelada según Fitts puede descomponerse en dos fases:

- Movimiento inicial, rápido pero impreciso hacia el blanco.
- Movimiento final, más lento, pero más preciso, para llegar al blanco.

La ecuación Ecuación 3 relaciona el *tiempo de movimiento* con el *índice de dificultad*, siendo las constantes empíricas a y b *intercepción* y *pendiente*, respectivamente. La variable a describe la aceleración y b es la intersección con eje Y y define el retardo. Debido a la variabilidad del rendimiento humano, la intercepción no será cero, pero deberá ser un valor pequeño próximo a cero (siendo, <400 ms si es positivo o -200 ms si es negativo). En este sentido, los valores altos de este parámetro (intercepción) sugieren que hay un problema con la metodología seguida [63].

$$MT = a + b \cdot ID \quad (\text{Ecuación 3})$$

La bondad del ajuste realizado es asociado con la correlación (r) o el cuadrado de la misma (coeficiente de determinación, r^2). El coeficiente de determinación estima la proporción de la varianza de una variable que es obtenida de otra. Por regla general, un valor de r^2 por encima de 0.8 es interpretado como un buen ajuste, mientras que, si el valor está por debajo de 0.5 se considera un ajuste débil. La formulación más extendida y usada en el campo de la Interacción Personas-máquina del *índice de dificultad* fue propuesto por Mackenzie en el año 1992 [64]. Esta nueva formulación recibió el nombre de la formulación de Shannon, y está inspirada en el Teorema de Shannon-Hartley que relaciona el ID a la información transmitida al realizar la tarea (ecuación Ecuación 4, en *bits*).

$$ID = \log_2\left(\frac{D}{W} + 1\right) \quad (\text{Ecuación 4})$$

La Ley de Fitts tiene dos aplicaciones directas: la primera es usar un modelo ajustado para predecir el tiempo que el usuario emplea en seleccionar los

objetivos (esto es útil en el diseño de interfaces), mientras la segunda consiste en comparar y evaluar los dispositivos de apuntamiento. Actualmente, el primer estudio que aplica la Ley de Fitts para comparar diferentes dispositivos de apuntamiento, concretamente el ratón, *joystick* y las teclas de dirección del teclado, es el presentado en [65] en el año 1978. Este estudio usa una nueva formulación del rendimiento (o *throughput*) .

La Organización Internacional de Estandarización (*ISO, International Organization for Standardization*) lanzó en el año 2000 el estándar *ISO 9241-9:2000* [60]. Dicho estándar recoge los requisitos ergonómicos para el trabajo de oficina con terminales de pantalla visual; concretamente, la parte 9 hace referencia a dispositivos de entrada que no son teclado. Con el objetivo de homogeneizar la evaluación de los dispositivos de apuntamiento se plantean una serie de tareas adaptadas a nuestras pruebas. Concretamente, el Anexo B del estándar *ISO 9241-9* se basa en la Ley de Fitts (usando la formulación de Shannon que ya ha sido citada con anterioridad, para el *ID* y los ajustes de la precisión propuestos por Crossman) y mide la eficiencia y efectividad de la interacción. Adicionalmente, el Anexo C de dicho estándar, sugiere un cuestionario para medir el confort del usuario durante la interacción. La primera tarea que se realiza en el estudio de usuario al que se refiere este capítulo, es una tarea con movimientos multidireccionales en 2 dimensiones como la descrita en el Anexo B de la Norma ISO. En este caso, un usuario tiene el control del puntero moviéndolo entre los diferentes objetivos dispuestos de forma circular. La tarea comienza cuando el usuario selecciona el objetivo situado en la posición de más altura y termina cuando todos los objetivos han sido seleccionados secuencialmente (dada una secuencia predefinida previamente) al menos una vez. La Figura 3.2 es la recogida en el Anexo B de la ISO 9241-9 y la Figura 3.3 es la que ejemplifica la prueba realizada en el estudio de usuario que se está tratando en este capítulo. Habitualmente los objetivos son de forma circular, aunque esto no está restringido y pueden tener otra forma. La configuración de la tarea permite probar diferentes índices de dificultad modificando el radio de la plantilla donde se colocan los objetivos y el propio radio de éstos. La norma ISO sugiere calcular el *TP* basándose en la propuesta de Fitts. Además, en [63] existen recomendaciones para realizar el estudio posterior a la recogida de los datos, todo ello basado en la Ley de Fitts. Existen numerosos estudios para la tarea de apuntar y seleccionar usando diferentes dispositivos como el ratón, *trackballs*, *stylus*, *joysticks*, pantallas táctiles, *touchpads* [63], incluso usando la inclinación de los dispositivos móviles [66].

Figura 3.2: Tarea multidimensional del Anexo B de ISO 9241-9

Figura 3.3: Tarea multidimensional desarrollada en este estudio de usuario

3.2.2. Escenario de interacción

En esta Tesis se busca analizar la viabilidad de disponer de un sistema de apuntamiento que permita a un usuario interactuar en un espacio. Para realizar la comparación de técnicas de apuntamiento, se simplificará este escenario inicial y se analizará el rendimiento 3D de los sistemas sobre una proyección 2D en la pared de una sala. Como ya se ha introducido anteriormente, el dispositivo habilitador es *Kinect* versión 2.0. Este dispositivo está compuesto por una serie de sensores, siendo capaz de proporcionar un seguimiento estable de las articulaciones de los usuarios.

El esquema del despliegue que permite realizar las pruebas corresponde al mostrado en la Figura 3.4. El espacio real de pruebas corresponde a una sala de 20 m^2 con una pared frontal de 9.6 m^2 usada para proyectar. El dispositivo Kinect ha sido situado sobre una estructura en el medio del eje X (del sistema de coordenadas de la sala) de la pared frontal, a una altura de 1.6 m desde el suelo. La localización del dispositivo tiene el objetivo de facilitar la visibilidad de la cara del usuario en todo el espacio independientemente de la altura del usuario, así como, el seguimiento de la orientación de la cara para conseguir unos resultados precisos de dicha estimación. Las proyecciones en la pared se realizan con un proyector¹; en algunos casos, la proyección se complementa con objetos físicos para dotar de mayor realismo a la tarea.

El área de movimiento en la que el sistema funciona viene limitada por el campo de visión del dispositivo Kinect 2.0, que se encuentra en el rango de 0.5 m a 4.5 m tomando el dispositivo como referencia. En este sentido, la versión empleada del dispositivo respecto de su versión anterior, permite a los usuarios estar más cerca del dispositivo. Durante el experimento, el usuario realizará sus movimientos a una distancia aproximada de 2 m desde el dispositivo dentro del ángulo donde Kinect garantice los resultados óptimos para estimar los dos tipos de apuntamientos considerados.

Los autores de [4] demuestran que es factible con esta tecnología estimar un apuntamiento 3D basado en señalización manual dentro de entornos inteligentes. El objetivo ahora es evaluar el rendimiento de la interacción y la usabilidad del sistema de señalización, antes de configurar la solución tecnológica definitiva para la interacción en 3D.

¹Proyector Sony modelo VPL-HW65ES <http://www.sony.es/pro/product/projectors-homecinema/vpl-hw65es/features/>

Figura 3.4: Despliegue para pruebas. Se aprecian las áreas de cobertura óptima del dispositivo Kinect junto a la aplicación de la norma ISO 9241-9 proyectada sobre la pared de la sala.

3.3. Estimación del apuntamiento

Esta sección describe la forma en que el sistema de apuntamiento, partiendo de los datos recogidos por el dispositivo Kinect, puede estimar el apuntamiento empleando señalización manual y simulación del seguimiento de la mirada.

3.3.1. Métodos de selección

La estimación de apuntamiento por señalización manual se obtiene tomando las posiciones de las articulaciones calculadas por Kinect, proporcionadas por el kit de desarrollo software que proporciona el fabricante. En este caso, el vector de apuntamiento conecta el hombro derecho con el extremo del dedo índice de la mano derecha del usuario (Figura 3.1) .

Respecto a la estimación simulando el seguimiento de la mirada, que

combina la posición de la cabeza y la orientación de la cara del usuario, Kinect es capaz de obtener la posición de la cabeza como otra articulación más. La Figura 3.1.b muestra el vector buscado. Esta idea ya fue estudiada por Mora y Odobez en [67], investigando la estimación del seguimiento de la mirada usando los movimientos de la cabeza del usuario con el dispositivo Kinect en su primera versión. El API de *face tracking* (*detección de la cara*) disponible en el SDK de Kinect proporciona la dirección-orientación de la cara del usuario (ver Figura 3.5). El objetivo es estimar la dirección de la mirada de un usuario cuando se le pide que seleccione un objeto con la mirada. Esto se calculará como una combinación de la posición de la cabeza y la orientación de la cara². La dirección de la mirada habitualmente coincide con la orientación de la cara, evidentemente, el movimiento de los ojos cambia la dirección de la mirada sin ser necesario cambiar la orientación de la cara ni mover la cabeza. Esto resulta especialmente relevante cuando la persona mira hacia distancias lejanas, y es menos frecuente cuando las distancias hacia las que se miran están próximas a la persona. Como ya se comentó, los sistemas de seguimiento de mirada precisos requieren de una visibilidad permanente de los ojos. Este requisito suele ser muy complicado mantenerlo en el tiempo y llevarlo a cabo en los escenarios con servicios inteligentes que se está planteando para este trabajo.

Figura 3.5: Esquema de las coordenadas 3D respecto del sistema de coordenadas de la cámara. Fuente: [61].

El seguimiento de la mirada se calcula con la posición de la cabeza de las 3 coordenadas del espacio tridimensional y los 87+13 puntos de 2 dimensiones

²El cuaternión que proporciona la orientación de la cara está disponible a través de la función `Microsoft.Kinect.Face.FaceFrameResult.FaceRotationQuaternion` del SDK de Kinect, <https://msdn.microsoft.com/en-us/library/jj130970.aspx>

de la malla de la cara (contorno de las cejas, ojos, nariz y boca, apuntando el centro del ojo, las esquinas de la boca, el centro de la nariz y el contorno de la cara), ver detalles en Figura 3.6. La orientación de la cara es dada por los tres ángulos (*pitch*, *roll* y *yaw*, ver Figura 3.7) que indican la orientación de la cabeza en los ejes de referencia de las 3 dimensiones con un rango comprendido entre los valores -180° y 180° . El cuaternión que representa la orientación viene definido por la ecuación Ecuación 5. Al normalizar el anterior cuaternión, y convertirlo en una matriz de rotación el resultado corresponde a la matriz que aparece en la ecuación Ecuación 6. Usando el sistema de coordenadas de la cámara (espacio de cámara), el vector que indica la orientación de la cara estará en la dirección $\vec{v} = [0, 0, -1]$. El vector de orientación de la cara ($\vec{v}_{ori}$) además es $\vec{v}_{ori} = R * \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$ referenciado al espacio de la cámara (Figura 3.5).

$$q = w, x, y, z = w + xi + yj + zk \quad (\text{Ecuación 5})$$

$$q = \begin{bmatrix} 1 - 2y^2 - 2z^2 & 2xy - 2wz & 2xz + 2wy \\ 2xy + 2wz & 1 - 2x^2 - 2z^2 & 2yz - 2wx \\ 2xz - 2wy & 2yz + 2wx & 1 - 2x^2 - 2y^2 \end{bmatrix} \quad (\text{Ecuación 6})$$

Figura 3.6: Puntos detectables de la cara a través del API de Kinect. Fuente: <https://blogs.msdn.microsoft.com/kinectforwindows/2014/01/31/mysteries-of-kinect-for-windows-face-tracking-output-explained/>

Figura 3.7: Ángulos de la orientación de la cabeza. Fuente: <https://web.asu.edu/sites/default/files/cnsg/files/c38.pdf>

3.3.2. Sistemas de coordenadas

El sistema de detección de apuntamiento va a trabajar con cuatro diferentes sistemas de coordenadas de referencia (ver Figura 3.9). Estos sistemas son:

1. *Sistema de coordenadas de la cámara .*
2. *Sistema de coordenadas de Kinect .*
3. *Sistema de coordenadas basado en patrón de calibración.*
4. *Sistema de coordenadas global.*

Estos cuatro sistemas permiten referenciar las coordenadas necesarias desde la imagen que toma el dispositivo en coordenadas del mundo real. El proceso de transformación se explica en el siguiente apartado de esta Sección. Sin embargo, antes de ello es necesario realizar el proceso de calibración de la cámara, en este caso del dispositivo Kinect. Este proceso permite estimar las distancias requeridas en el mundo real desde las imágenes capturadas a través de la obtención de los parámetros intrínsecos y extrínsecos de la cámara.

3.3.2.1. Proceso de calibración de la cámara de Kinect

El proceso de calibración de la cámara se ha realizado usando la herramienta de calibración disponible para Matlab³. Este proceso es el mismo

³Calibration toolbox for Matlab: http://www.vision.caltech.edu/bouguetj/calib_doc/

que se describe en [16] y en [4] y detallado en la web: http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/example.html. El proceso permite calibrar la cámara y obtener sus parámetros característicos. Estos parámetros se dividen en dos grupos de parámetros: intrínsecos y extrínsecos. Los parámetros extrínsecos estiman la ubicación y la orientación de la cámara dentro del espacio real y se obtienen a partir de los parámetros intrínsecos de la cámara. Los parámetros intrínsecos (el eje principal, el centro óptico y la distancia focal) están relacionados con su geometría interna y las características ópticas de la propia cámara. Estos parámetros mantienen su valor constante si las características y las posiciones relativas de la óptica y el sensor de imagen no varían. Los parámetros intrínsecos señalan las coordenadas en el sistema de referencia de la cámara.

El patrón de calibración empleado es un tablero con un ajedrez de cuadrados (5x5) con 60 mm de lado. Con la cámara de Kinect se han tomado 21 imágenes diferentes de dicho patrón colocado en diferentes posiciones y orientaciones, como se puede ver en la Figura 3.8. Después de procesar las imágenes en Matlab, se extraen las esquinas del patrón seleccionando manualmente las cuatro esquinas exteriores del patrón. Una vez hecho esto, se obtiene los parámetros intrínsecos del sensor.

Figura 3.8: Ejemplos de imágenes tomadas para extraer los parámetros intrínsecos durante el proceso de calibración

3.3.2.2. Transformaciones entre sistemas de coordenadas

Para calcular las coordenadas apuntadas es imprescindible tratar las consecutivas transformaciones que son necesarias realizar entre los diferentes sistemas de coordenadas que se manejan. Las transformaciones a realizar son (ver Figura 3.9):

1. Transformación entre el *sistema de coordenadas de la cámara* y el *sistema de coordenadas de Kinect*.

2. Transformación entre el *sistema de coordenadas de Kinect* y el *sistema de coordenadas basado en patrón de calibración*.
3. Transformación entre el *sistema de coordenadas basado en patrón de calibración* y el *sistema de coordenadas global*.

Figura 3.9: Transformaciones de coordenadas entre los diferentes sistemas de coordenadas existentes

En resumen, el objetivo es obtener las coordenadas del objetivo en el sistema de coordenada global apuntado en un formato comprensible para diseñar una aplicación final que haga uso de ello.

3.3.2.2.1. Transformación 1 - Del *sistema de coordenadas de la cámara* al *sistema de coordenadas de Kinect* En primer lugar, los puntos esenciales (posición del hombro y de la cabeza) y los vectores de dirección (tanto del brazo como el de la mirada) son calculados respecto al *sistema de coordenadas de la cámara* (o *espacio de la cámara*; Figura 3.9.a). Estas coordenadas hay que referenciarlas respecto del *sistema de coordenadas de Kinect* (también llamado *espacio del esqueleto*), que corresponde al sistema de coordenadas en 3 dimensiones definido en el SDK de Kinect. Como se aprecia en la Figura 3.9.b) el origen del sistema de coordenadas se sitúa en el centro del sensor de infrarrojos del dispositivo-sensor Kinect. Aunque los ejes de ambos sistemas de coordenadas satisfacen la regla de la mano derecha, el primero presenta una rotación de 180° del eje z sobre el segundo. En resumen, un punto definido en el *espacio de la cámara*, $p_c = [x_c, y_c, z_c]$, se transforma al punto definido en el *espacio del esqueleto*, $p_k = [x_k, y_k, z_k]$, con la siguiente equivalencia: $p_k = [-x_c, -y_c, z_c]$. Con lo que respecta a un vector definido en el *espacio de la cámara*, $\vec{v}_c = [\vec{x}_c, \vec{y}_c, \vec{z}_c]$ se transforma en un vector referenciado al *espacio de Kinect*, $\vec{v}_k = [\vec{x}_k, \vec{y}_k, \vec{z}_k]$, como se indica a continuación: $\vec{v}_k = [-\vec{x}_c, -\vec{y}_c, \vec{z}_c]$

Este sistema de coordenadas no resulta suficiente ya que la aplicación final se desarrolla para un espacio real, por tanto, se necesita seguir haciendo

transformaciones entre sistemas para conseguir referenciar las coordenadas respecto del espacio real.

3.3.2.2.2. Transformación 2 - Del *sistema de coordenadas de Kinect* al *sistema de coordenadas basado en patrón de calibración*

En este punto del proceso, se introduce un nuevo *sistema de coordenadas basado en patrón de calibración* (ver Figura 3.9.c), el cuál tendrá su origen en la esquina derecha inferior del patrón de calibración empleado. La transformación entre las coordenadas referenciadas al *sistema de coordenadas Kinect* y el *sistema de coordenadas basado en el patrón de calibración* utiliza una matriz de rotación ($R_{c_{ext}}$) y un vector de traslación ($T_{c_{ext}}$) obtenido durante un proceso de calibración previo ya explicado. De esta forma, el punto anterior p_k , referenciado en el *espacio del esqueleto*, $p_k = [x_k, y_k, z_k]$, se transforma en un punto del *sistema de coordenadas basado en patrón de calibración*, de la siguiente forma: $p_p = R_{c_{ext}}^{-1}(p_k - T_{c_{ext}})$. En el caso de tratar vectores, la traslación no se aplica y el resultado de la transformación queda de la siguiente forma: $\vec{v}_p = [\vec{x}_p, \vec{y}_p, \vec{z}_p] = R_{c_{ext}}^{-1}\vec{v}_k$.

3.3.2.2.3. Transformación 3 - Del *sistema de coordenadas basado en patrón de calibración* al *sistema de coordenadas global*

Para continuar con el proceso de transformación y poder referenciar las coordenadas respecto a un sistema relativo al espacio, se introduce un cuarto sistema de coordenadas, denominado *sistema de coordenadas global* (ver Figura 3.9.d). Este último sistema se usa tanto para referenciar los puntos calculados del espacio real en 3 dimensiones, como para los objetos distribuidos en dicho espacio. La transformación a este *sistema de coordenadas global* se calcula aplicando una traslación, $T_{pg} = [x_{pg}, y_{pg}, z_{pg}]$, que corresponde a la distancia entre los orígenes de ambos sistemas de coordenadas medido en cada uno de los ejes de forma empírica dentro del escenario real (el *sistema de coordenadas basado en patrón de calibración* y el *sistema de coordenadas global*) y una rotación del eje y , R_{pg} , que corresponde a una combinación de una rotación de 90° sobre el eje z y una rotación de 180° sobre el eje x como muestra la ecuación Ecuación 7. Por lo tanto, un punto referenciado en el *sistema de coordenadas basado en patrón de calibración* se transforma en un punto en el sistema el *sistema de coordenadas global* siguiendo la ecuación Ecuación 8. Como en casos anteriores, si en vez de un punto, lo que se quiere transformar

es un vector, la equivalencia es: $\vec{v}_g = [\vec{x}_g, \vec{y}_g, \vec{z}_g] = R_{pg} \vec{v}_p$

$$R_{pg} = R_{x180} R_{z90} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (\text{Ecuación 7})$$

$$P_g = [x_g, y_g, z_g] = R_{pg} P_p + T_{pg} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} + \begin{bmatrix} x_{pg} \\ y_{pg} \\ z_{pg} \end{bmatrix} \quad (\text{Ecuación 8})$$

El último paso a llevar a cabo es la proyección del vector sobre el plano objetivo, todo ello referido al *sistema de coordenadas global*. Este plano se define previamente en el *sistema de coordenadas global* en el punto donde el eje z es nulo. El resultado de la proyección devolverá las coordenadas del objetivo apuntado sobre este plano. Una vez disponemos del origen del vector referenciado en el *sistema de coordenadas global*, $O_g = [x_g, y_g, z_g]$ y el vector de apuntamiento, $\vec{v}_g = [\vec{x}_g, \vec{y}_g, \vec{z}_g]$, la proyección a realizar aparece definida en la ecuación Ecuación 9. Como el plano donde se proyecta está situado en las coordenadas con el valor del eje z a cero, se puede obtener el valor de la pendiente, m , como se indica a continuación: $z = z_g + m \vec{z}$.

$$\text{Punto proyectado} = O_g + m \vec{v}_g = \begin{bmatrix} x_g \\ y_g \\ z_g \end{bmatrix} + m \begin{bmatrix} \vec{x}_g \\ \vec{y}_g \\ \vec{z}_g \end{bmatrix}, \text{siendo } m \text{ el valor de la pendiente} \quad (\text{Ecuación 9})$$

3.4. Arquitectura del sistema de estimación del apuntamiento

Esta sección se centra en la arquitectura del sistema para estimar el apuntamiento diseñado. El sistema permite realizar el seguimiento de un usuario que realiza acciones de señalización dentro del entorno (en la actual configuración, el sistema es capaz de rastrear un único usuario, aunque la opción de multiusuario es factible). El sistema de estimación realiza los cálculos explicados en las secciones anteriores y proporciona a las aplicaciones finales

los puntos estimados de apuntamiento. La Figura 3.11 recoge los principales componentes del sistema, a saber:

- **Componente de adquisición.** Este componente usa las librerías del SDK de kinect para obtener los datos del esqueleto de los individuos dentro del área de visión del dispositivo (6 personas como máximo). Todos los datos son obtenidos referidos al *espacio de la cámara*.
- **Componente de selección de usuario.** Este componente realiza la selección del usuario. El usuario rastreado corresponderá con el que tenga mejor calidad de seguimiento, habitualmente el más cercano al dispositivo Kinect. La salida proporcionada serán los puntos del esqueleto del usuario empleados para definir el apuntamiento.
- **Componente de estimación de apuntamiento.** Este componente agrupa la lógica para calcular los vectores de dirección del apuntamiento en los tres módulos siguientes:
 - *Módulo de estimación de apuntamiento basado en señalización manual.* El vector de apuntamiento viene definido por la posición del hombro (punto origen del vector) y la posición del extremo del dedo índice de la mano derecha (punto final del vector). El componente proporciona como salida tanto el punto origen como el vector de apuntamiento referido al sistema de coordenadas del *espacio de la cámara*.
 - *Módulo de estimación de apuntamiento basado en simulación de la mirada.* La salida de este módulo corresponde al origen, posición de la cabeza, y al vector de dirección de la cara. Al igual que en caso anterior referido al sistema de coordenadas del *espacio de la cámara*.
 - *Módulo transformación de coordenadas entre sistemas de coordenadas.* Este modulo transforma las coordenadas desde el *sistema de coordenadas de la cámara* al *sistema de coordenadas global*, como se explico en la Sección 3.3.2.
- **Componente de presentación.** Este componente proyecta el vector del apuntamiento sobre un plano en el *sistema de coordenadas global* con el objetivo de determinar el objetivo apuntado (en esta implementación, sobre la pared). Los resultados se ofrecen gracias a un interfaz de programación de aplicaciones (*API, Application Programming Interface*). Además, en esta misma *API* se comparte el sistema de coordenadas global.

- **Aplicaciones.** Los clientes software consumen la información de salida del componente de presentación. La comunicación entre la parte servidora y estas aplicaciones cliente se realiza mediante la tecnología basada en sockets⁴. En este caso se ha empleado el lenguaje de programación C# para implementar el sistema de estimación del apuntamiento por la compatibilidad que ofrece con las herramientas de Microsoft, y en concreto, con el propio dispositivo Kinect. El servidor se ejecuta sobre un sistema operativo Windows 8.1.

El diagrama de secuencia de la Figura 3.10 muestra los principales pasos entre los elementos de la arquitectura que se han explicado previamente.

Figura 3.10: Diagrama de secuencia del proceso de estimación de la proyección.

⁴La tecnología usando *sockets* permite que dos programas (posiblemente, ejecutados en equipos distintos) puedan intercambiar un flujo de datos, de manera fiable y ordenada. Un *socket* queda definido por un par de direcciones, IP local y remota, un protocolo de transporte y un par de números de puerto, local y remoto.

Figura 3.11: Arquitectura con los datos de salida para los diferentes bloques que la forman

3.5. Diseño del estudio para la evaluación del rendimiento del sistema y la experiencia de usuario

3.5.1. Objetivos

El objetivo del estudio es comprobar el rendimiento y la experiencia de usuario con los métodos de interacción mencionados, señalización manual y simulación del seguimiento de la mirada. Esta sección está orientada a describir los principales aspectos del estudio de usuario realizado, mostrando la metodología seguida, que incluye tanto variables dependientes como independientes, tareas y mediciones hechas. El estudio de usuario persigue principalmente:

1. Comparar el rendimiento, usabilidad y experiencia de usuario de los dos métodos de detección definidos.
2. Identificar los obstáculos que presentan ambos métodos en un servicio desarrollado en un entorno inteligente.
3. Resaltar los aspectos a mejorar en este tipo de sistemas de interacción.
4. Analizar los métodos de interacción multimodal preferidos por los usuarios.

3.5.2. Participantes

El estudio de usuario fue llevado a cabo en el *Laboratorio de Experiencias de Espacios del futuro*, que está localizado en el Centro de Apoyo a la Innovación del Campus de Montegancedo perteneciente a la Universidad Politécnica de Madrid. En dicho estudio participaron 20 personas en total, 17 hombres y 3 mujeres, con edades comprendidas entre los 22 y 51 años ($\mu = 30,3; \sigma = 7,07$) escogidas del propio personal del Centro de innovación. Entre los participantes, 9 de ellos habían usado el dispositivo Kinect con anterioridad, y además, 5 personas de estos lo hacían de forma habitual (varias veces a la semana). Esta composición de la muestra a la hora de evaluar el rendimiento podría permitir hacer la distinción entre usuarios expertos y no expertos.

3.5.3. Dispositivos y Herramientas

El entorno de pruebas ha sido descrito de forma detallada en la Sección 3.2.2. Se describe a continuación la infraestructura necesaria para hacer posible la realización de las tareas a realizar por el usuario que requieren también del manejo y control de algunos objetos inteligentes que están conectados a una plataforma que controla las funciones domóticas del espacio. Por otra parte, también es necesario usar una infraestructura que emplea tecnología DLNA (*Digital Living Network Alliance*⁵) disponible en el Laboratorio para manejar los contenidos multimedia. Además, existe un segundo equipo para ejecutar las aplicaciones cliente. Las aplicaciones cliente (desarrolladas en Java) proporcionan el soporte para poder completar las tareas del estudio de usuario, concretamente, la tarea de selección de posiciones multidireccionales definidas en la ISO 9241-9 y la tarea de manejar objetos inteligentes reales.

En la primera tarea, la relacionada con la norma ISO, el usuario tiene que seleccionar un total de 12 objetivos de igual tamaño que están colocados sobre un círculo y que son proyectados en la pared (ver Figura 3.12). Es importante señalar que la tarea multidireccional contemplada en la ISO 9241-9 es una tarea estándar, que tiene como objetivo medir y comparar el rendimiento del apuntamiento entre las diferentes tecnologías empleadas. La aplicación está preparada para manejar un número variado de objetivos con diferentes tamaños, tanto para el objetivo en sí, como para el círculo donde se colocan estos.

La aplicación se configura para proporcionar un *feedback* de la señalización al usuario, se muestra una bola de menor tamaño que los objetivos y de color rojo señalando el punto donde el usuario está apuntando. El método de selección escogido para esta primera tarea es el método llamado primero en entrar (*first-entry*). Esto significa que nada más seleccionar el objetivo, éste cambia de color y se resalta el siguiente objetivo a seleccionar de color azul (el objetivo a seleccionar siempre estará relleno de color azul y solo existirá uno).

Para el posterior procesamiento de los datos se almacena el movimiento del cursor en un fichero de *log* en formato XML; se almacenan los tiempos del movimiento de cada objetivo y las posiciones del cursor, para después calcular la velocidad, el rendimiento (*throughput*) y otros parámetros que resulten de interés. Este procesamiento también se realiza de forma automática.

⁵**Digital Living Network Alliance** es un grupo de compañías eléctricas que trabajan juntas para promover y desarrollar guías de interoperabilidad que permitan compartir contenido digital entre dispositivos.

La segunda tarea implica manejar objetos inteligentes; consiste en que el usuario interactúe un número de veces fijo (10 en este caso) con dos objetos inteligentes diferentes, en particular, una bombilla Philips Hue y un reproductor multimedia de música con tecnología Bluetooth. Ambos objetos están colocados en la pared, en diferentes ángulos del área de proyección (aproximadamente a 30° y 240° , respectivamente) para simular una interacción real en 3D dentro de un espacio inteligente. La aplicación, desarrollada usando tecnología Java, recibe el punto estimado de apuntamiento del sistema y comprueba si está dentro del área de influencia (círculo que cubre el volumen del objeto; para el reproductor multimedia, 295×454 píxeles - 56×86 cm; para la bombilla, 164×122 píxeles - 31×23 cm de alguno de los dos objetos anteriores).

El método de selección empleado en esta segunda tarea es el llamado *dwell-time* que consiste en apuntar sobre un objetivo y mantenerlo apuntando durante un periodo de tiempo predefinido. El periodo de tiempo que el usuario debe mantenerse apuntando hacia el objetivo se ha configurado en 1200 ms, en línea con estudios previos como por ejemplo [47], dirigido a la estimación de la posición de la cabeza. Si el usuario se mantiene durante ese tiempo apuntando sobre el objeto, la acción será lanzada sobre el objeto real. Esto es también controlado de forma interna por la aplicación, la cuál al ver que el objeto ha sido seleccionado envía un comando usando arquitectura REST (REST, *Representational State Transfer* es un estilo de arquitectura software para sistemas hipermedia distribuidos) proporcionada por la infraestructura de este Laboratorio para el control de los objetos. Un ejemplo de esta segunda tarea puede verse en la Figura 3.13.

Figura 3.12: Aplicación ISO 9241-9. Algunas de las vistas son: a) implementación de la interfaz para la tarea ISO 9241-9 con los 12 objetivos a seleccionar, el próximo objetivo aparece resaltado en color azul. b) vista de las rutas seguidas por el movimiento del usuario una vez que la tarea ha sido completada (Este movimiento no es visible durante la realización de la prueba). c) Configuración del experimento, con la vista de la interfaz de la tarea ISO proyectada sobre la pared.

Figura 3.13: Interacción con objetos inteligentes a través de señalización manual. En esta imagen el usuario interactúa con la bombilla *Philips Hue*.

3.5.4. Procedimiento y tareas

El estudio de usuario se ha diseñado para que todos los participantes completen las mismas tareas usando ambas técnicas de detección del objetivo apuntado, la señalización manual y la de simulación del seguimiento de la mirada. Después de firmar un formulario de consentimiento (Anexo C), se expone a los usuarios el objetivo de las pruebas, remarcando que la destreza del usuario no es de interés. También se les indica que se pueden mover libremente en el espacio siempre manteniendo un mínimo de 1.5 m de distancia respecto del dispositivo Kinect (el área efectiva de movimiento es de $7.5 m^2$, $2.5 \times 3 m$). Además, a los usuarios se les pidió que expresasen de forma verbal (*Thinking aloud*) sus pensamientos en relación al uso del sistema. La prueba se graba para cada usuario, por si era necesario extraer alguna información adicional o aclarar algún resultado. Cada usuario invirtió una media de tiempo de alrededor de 1 hora para finalizar la prueba de forma completa.

Como se muestra en la Figura 3.14, los participantes, tras la introducción que se les realizó, rellenaron un cuestionario dirigido a conocer su experiencia previa con el dispositivo. No se incluyó una fase previa de entrenamiento. El sistema actualmente es capaz de trabajar en entornos inteligentes sin entrenar.

Como se recoge en la Figura 3.14, la prueba sigue una secuencia predefinida, primero se completa la tarea que corresponde a la ISO 9241-9 y después

la correspondiente a manejar objetos inteligentes (cada una de ellas, con ambos métodos de interacción propuestos). El tipo de interacción a realizar primero se eligió de forma aleatoria para evitar los efectos relacionados con el orden de la elección. Asimismo, cada tarea es repetida con feedback visual y sin feedback visual, visualizando y no visualizando el punto de referencia del movimiento (a través de un punto rojo proyectado en la pared). Después de cada iteración, el usuario evalúa la dificultad de la tarea realizada. También el usuario valora el confort a la hora de usar el sistema rellenando el cuestionario del ANEXO C (ver Anexo A) tras completar la tarea ISO con referencia visual. Al finalizar la tarea de forma completa (dos técnicas de apuntamiento disponiendo y no disponiendo de referencia visual) el usuario contesta unas preguntas para comparar el rendimiento de ambas técnicas y expresar sus preferencias al respecto.

Posteriormente, tras finalizar la primera tarea, correspondiente a la ISO 9241-9, el usuario realiza la tarea de controlar los objetos inteligentes, apunta e interacciona con ellos. Esta tarea se realiza únicamente proporcionando una referencia visual en todo momento del sitio apuntado para las dos técnicas de apuntamiento. Tras realizar la prueba los usuarios evalúan la usabilidad del sistema y la experiencia de usuario rellenando un cuestionario como el que aparece en el anexo A basado en [7] y [68]. Aunque la usabilidad y la experiencia de usuario son conceptos que están muy ligados, es importante realizar esta separación ya que el segundo concepto busca ir más allá de la medición del rendimiento humano, y explora más a conciencia la satisfacción de uso [8].

3.5.5. Diseño del experimento

Inicialmente, la tarea ISO 9241-9 relaciona la técnica de detección del apuntamiento (señalización manual y simulación de la mirada) y el tamaño del objetivo a apuntar (tres diámetros: 60, 130 y 200 píxeles, equivalente a 114, 247 y 380 milímetros cuando se proyectan en la pared). Además, se consideró disponer de un feedback visual en cada momento, marcándolo en cada momento.

Para la tarea de la norma definida por la ISO con referencia visual el usuario realizó 6 iteraciones en total (3x2). Una vez que se escogió la técnica de apuntamiento se repitió 5 veces la prueba para cada uno de los tamaños del objetivo. En cuanto a la tarea de manejar objetos inteligentes, se pidió a los usuarios que interactuarán 10 veces con los dos objetos disponibles, a su voluntad, con la única restricción de hacerlo al menos una vez con cada uno.

El número total de pruebas para la tarea de la ISO fue de 20 usuarios x 2 técnicas de apuntamiento de detección x (3 tamaños de objetivos x 5 repeticiones *con feedback* visual) + (3 tamaños de objetivos x 1 repetición *sin feedback* visual) ; finalmente un total de 720. Por su parte, la tarea de manejar objetos inteligentes fue realizada por 20 usuarios x 2 técnicas de detección de apuntamiento x (10 repeticiones con *feedback* visual + 3 repeticiones sin *feedback* visual) ; finalmente con un total de 520 iteraciones.

3.6. Resultados del experimento

Los resultados se presentan en dos partes. La primera parte está orientada a analizar el rendimiento para las técnicas de detección del apuntamiento utilizando la Ley de Fitts, junto con aspectos como el aprendizaje o la dependencia de la posición del participante. La segunda parte se refiere al análisis del confort, usabilidad y experiencia del usuario durante la realización de las pruebas.

3.6.1. Análisis del rendimiento de la interacción

Los datos empleados para recoger los resultados de esta sección fueron recogidos durante la realización de la tarea de la ISO 9241-9. Todos los usuarios completaron de forma satisfactoria las 30 iteraciones.

3.6.1.1. Tiempo de movimiento

El tiempo de movimiento (*MT*, del inglés *Movement Time*) es el tiempo que el usuario tarda en alcanzar el objetivo desde la posición anterior. Se ha realizado un estudio con dos factores (tipo de interacción, 2 en total, y el tamaño del objetivo, 3 en total) y una variable dependiente (tiempo de movimiento) para determinar si existen diferencias significativas en el tiempo de movimiento. En el estudio se han empleado pruebas del análisis de la varianza para medidas repetidas⁶, en adelante ANOVA, para los dos factores anteriores (un total de 2 x 3 condiciones) y así conseguir una verificación

⁶el *análisis de la varianza* (ANOVA, Analysis Of Variance, según denominación inglesa) es una colección de modelos estadísticos y sus procedimientos asociados, en el cual la varianza está dividida en ciertos componentes debidos a diferentes variables explicativas.

Figura 3.14: Flujo del estudio de usuario mostrado de forma temporal: Presentación y pruebas del perfil del usuario, tarea ISO, tarea de manejar objetos inteligentes.

formal de los resultados. Previamente a la hora de realizar estas pruebas se comprobó la *hipótesis de esfericidad* asumida en una prueba de ANOVA con medidas repetidas. La prueba de Mauchly muestra que la hipótesis de la esfericidad tiene que ser rechazada para los datos existentes (no se puede decir que las varianzas de las diferencias entre todos los pares posibles de grupos

son iguales), por lo que se ha introducido la corrección de Greenhouse-Geisser.

Al comparar los valores promedio de la duración del *tiempo de movimiento* para la técnica de apuntamiento de señalización manual y la de simulación de mirada se comprueba que son bastante similares (ver tabla 3.1). La prueba ANOVA de medidas repetidas confirma que la hipótesis nula de igualdad de medias para el tiempo de movimiento no puede ser descartada ($F(1,19) = 0,425, p = 0,522$). Este resultado mantiene la coherencia con el análisis del tiempo en finalizar las tareas, la cual difiere en 1062 ms comparando la técnica de señalización manual (16311 ms) y la de simulación de mirada (17373 ms).

Tabla 3.1: Tiempo de movimiento (MT) por tipo de interacción

Interaction	Mean	Std. Error	95 % Confidence Interval	
			Lower Bound	Upper Bound
Deictic	1351,458	78,355	1187,458	1515,457
Gaze	1399,049	49,891	1294,627	1503,472

Los tiempos de movimiento para los diferentes tamaños de los objetivos usando ambas técnicas de apuntamiento aparecen en la Figura 3.15. El tiempo de movimiento se incrementa a medida que el tamaño del objetivo es más pequeño para ambas técnicas de apuntamiento, siendo más evidente en el caso del apuntamiento con *simulación de la mirada*. La prueba realizada determinó que el promedio de los tiempos de movimiento varía significativamente, entre los tres tamaños de objetivos habilitados ($F(1,306, 24,808) = 342,623, p \ll$). Empleando la corrección de Bonferroni, se determina que el tiempo de movimiento es significativamente diferente entre los tres tamaños de objetivo ($p \ll$ para cada combinación), siendo esto consistente con los intervalos de confianza al 95 % de la Figura 3.15. La conclusión es que el tamaño del objetivo influye en la efectividad de la interacción; teniendo en cuenta el tamaño de nuestros objetivos proyectados, las superficies interactivas con un radio de más de 190 mm pueden ser fácilmente accesibles en ambos tipos de apuntamiento cuando los usuarios están a mas de 2 metros del objetivo. Los tiempos de movimiento obtenidos son más pequeños que los que se obtienen en el estudio [48].

La Figura 3.16 recoge la variación del tiempo de movimiento respecto de la posición del objetivo calculado como el ángulo θ (ángulo θ medido en el sentido de las agujas del reloj desde la línea vertical de la Figura 3.12). La Figura 3.16 recoge la variación del tiempo de movimiento respecto de la posición del objetivo calculado como el ángulo θ . Cuando los objetivos son

Figura 3.15: Relación entre tamaño del objetivo y tiempo de movimiento para ambas técnicas de interacción. Barras de error mostradas en $\pm 1SD$.

de gran tamaño, el tiempo de movimiento se mantiene estable con pequeños incrementos en los ángulos cercanos a 0° ($30^\circ - 330^\circ$) y 180° ($150^\circ - 210^\circ$) para ambos tipos de apuntamientos. Estos pequeños incrementos se ven acentuados cuando el tamaño del objetivo decrece. La desviación estándar en los 180° es significativamente mayor para el apuntamiento basado en simulación de la mirada cuando se apunta a objetivos pequeños. Una prueba ANOVA de medidas repetidas realizada para los tres factores (tipo de interacción, tamaño del objetivo y ángulo al mismo, un total de $2 \times 3 \times 11$ condiciones) confirma que es posible rechazar la hipótesis nula con respecto a la posición del objetivo (ángulo) ($F(10,190) = 18,125, p \ll 0,000004$). Por tanto, la diferencia de medias al considerar el tamaño objetivo y el ángulo de interacción es también estadísticamente relevante ($F(2,191, 41,623) = 13,944, p = 0,000014$).

Por su parte, la Figura 3.17 muestra el tiempo de movimiento en promedio para las dos técnicas de apuntamiento usadas. De acuerdo a esta Figura, el apuntamiento mediante señalización manual parece ser más independiente de la localización del objetivo, mientras que, el apuntamiento por simulación de la mirada tiene un comportamiento simétrico con respecto a los diferentes ángulos, siendo más relevantes las diferencias entre ángulos consecutivos. También se aprecia que el apuntamiento hacia objetivos situados en las posiciones con un ángulo 180° son difíciles de alcanzar con la mirada, ya que el dispositivo Kinect presenta una inestable detección de la dirección de la cara cuando el usuario mira hacia abajo. En el apuntamiento mediante señalización manual, el objetivo más complicado de alcanzar se encuentra localizado en el ángulo de 330° . Las posteriores pruebas realizadas sobre estos datos

Figura 3.16: Relación entre el tiempo de movimiento, eje vertical de las gráficas, y el ángulo donde está colocado el objetivo, eje horizontal de las gráficas; teniendo en cuenta el tamaño del objetivo para ambas técnicas de interacción. Barras de error mostradas en $\pm 1SD$.

usando la corrección de Bonferroni muestran que el tiempo de movimiento hacia algunos ángulos (180° , 210° , 90° y 270°) varían significativamente (siendo este tiempo mucho mayor o menor) del tiempo para alcanzar otros ángulos (esto se produce al menos en 6 de ellos).

Figura 3.17: Gráfico de perfil de los ángulos para ambas técnicas de interacción. Los ángulos aparecen ordenados teniendo en cuenta la secuencia de interacción realizada durante la ejecución de la tarea.

3.6.1.2. Análisis basado en la Ley de Fitts

Los datos han sido ajustados al modelo de Fitts de acuerdo a la ecuación Ecuación 1, usando la definición de Shannon de la ecuación Ecuación 4 (para mayor detalle ver apartado 3.2.1 relativo a la Ley de Fitts y la ISO 9241-9). La Figura 3.18 presenta la curva ajustada a través de la regresión lineal por mínimos cuadrados para ambos métodos de interacción, considerando todos los movimientos en el objetivo número 1 (ver Figura 3.12 del diseño de prueba) para todas las condiciones y los participantes de las pruebas. En estos resultados se han descartado los *outliers* como recomienda el estudio [63]; representando estos menos de un 5 % de las muestras totales en los datos recogidos en ambas técnicas de apuntamiento. Resaltar que alguno de estos puntos se debe al cansancio o confusión momentánea del usuario durante la realización de las pruebas, como se ha podido comprobar tras la visualización de los vídeos.

Figura 3.18: Relación entre el índice de dificultad (ID), eje horizontal de las gráficas, y el tiempo de movimiento (MT), eje vertical de las gráficas, al objetivo en milisegundos, tanto para apuntamiento por señalización manual (a) como por simulación de la mirada (b). El ajuste por regresión permite ajustar a la curva desde la nube de puntos representados.

La tabla 3.2 resume los parámetros del ajuste realizado. El valor de la intercepción está sobre los 400 ms para ambas técnicas de apuntamiento, siendo consisten con los límites que se consideran aceptables en [63]. Esta intercepción probablemente habría sido menor si se hubiese realizado un proceso previo de entrenamiento. El valor del parámetro r^2 de la regresión lineal entre el tiempo de movimiento (MT) y el índice de dificultad (ID) es bajo para ambos tipos de apuntamiento aunque el valor para la señalización manual es ligeramente superior que para el apuntamiento por simulación de la mirada. Este resultado está en consonancia con los resultados del estudio [26], en

el cual, el valor de r^2 en la interacción mediante señalización manual es de 0.27 y combinándolo con mirada es de 0.21.

Tabla 3.2: Resumen del ajuste de regresión

	% outliers	ID min	ID max	a (intercept)	b (slope)	r^2	Throughput (bps)
Deictic	4.82	1.77	4.01	523.94	-162.26	0.31	2.45 (+0.70)
Gaze	4.67	1.77	4.03	485.29	10.55	0.26	2.3 (+0.44)

Se prueba con un modelo modificado del ajuste, ya que se ha comprobado que el ajuste conseguido es deficiente. Esta modificación se realiza de acuerdo a la proposición de [69] que propone una ID modificada para incluir las variaciones del ángulo θ como se refleja en la ecuación Ecuación 10.

Aunque los autores del estudio [69] afirman una mejora significativa en su experimento, este enfoque no mejoró el modelo de ajuste para los datos de las pruebas de este estudio en ninguna de las técnicas de apuntamiento empleadas. El mejor valor de r^2 obtenido para el modelo lineal fue 0.15 para apuntamiento por simulación de la mirada y 0.18 para señalización manual, con un valor de c de 0.1.

El parámetro de rendimiento de la Ley de Fitts (*throughput*) reúne varias mediciones del tiempo de movimiento en una sola estadística combinando la velocidad y la exactitud, siendo bastante útil para comparar varios sistemas diferentes. Este valor aparece también reflejado en la tabla 3.2, el cual ha sido calculado para cada uno de los participantes (como el rendimiento conseguido por el participante para todas las condiciones propuestas durante las pruebas) y luego promediado, obteniendo un rendimiento global (matemáticamente reflejado en la ecuación Ecuación 11, donde n es el número de usuarios y m el número de condiciones durante los movimientos).

$$ID = \log_2 \left(\frac{D}{S} + 1 \right) + C \sin \theta \quad (\text{Ecuación 10})$$

$$TP = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{m} \sum_{j=1}^m \frac{ID_{ij}}{MT_{ij}} \right) \quad (\text{Ecuación 11})$$

Se observa que los valores del parámetro del rendimiento para ambas técnicas de detección del apuntamiento son similares y se pueden comparar con

los conseguidos para la interacción a través de dispositivos como *joysticks* (rango de valor de 1.8 a 2.3) o *touchpads* (rango de valor de 0.99 a 2.9) considerando estudios que también emplean la ISO 9241-9 [63]). La diferencia entre el rendimiento obtenido del sistema de interacción mediante señalización manual no difiere lo suficiente del obtenido en el sistema de interacción que simula el seguimiento de la mirada. Esto se confirma tras realizar el *Test-t* de muestras pareadas ($t(19) = 1,0676, n.s.$).

La Figura 3.19 muestra las diferencias en los valores obtenidos de rendimiento para los diferentes usuarios, habiendo diferencias entre ellos. Haciendo la división entre usuarios expertos (aquellos que han utilizado Kinect antes, 9 usuarios) y no expertos (un total de 11 usuarios) a partir de las respuestas obtenidas de preguntas iniciales realizadas a los participantes, se concluye, tras realizar un *Test-t* que no es posible establecer una diferencia significativa en el rendimiento medio respecto a la experiencia con el dispositivo, ni para la señalización manual ($t(18) = -1,88, p = 0,076$) ni para el apuntamiento por simulación de la mirada ($t(18) = -0,452, p = 0,6566$).

Figura 3.19: Parámetro de rendimiento (*throughput*, TP), eje vertical de las gráficas, para cada usuario en el eje horizontal usando ambas técnicas de apuntamiento. Las barras de error mostradas corresponden a $\pm 1SD$.

En relación con esto y con el fin de evaluar el efecto de la posición del usuario dentro de la zona de cobertura del dispositivo se ha diseñado un experimento adicional en el que un único usuario experto ha completado la tarea ISO para ambos métodos de apuntamiento, y todos los tamaños del objetivo (5 iteraciones) en siete posiciones predefinidas sobre el suelo (ver Figura 3.20.a). Según se observa con las barras de error de la Figura 3.20.b, no se aprecian diferencias relevantes para ambos métodos de apuntamiento; valores para el apuntamiento por señalización manual ($\mu = 870.94$ ms; $\sigma = 144.19$ ms) y valores para el apuntamiento mediante simulación de la mirada ($\mu = 1403.34$ ms; $\sigma = 162.13$ ms). El valor medio del tiempo de movimiento (MT) calculado para ambos tipos de interacción es ligeramente superior cuando el usuario se encuentra en las posiciones más alejadas al dispositivo,

mientras que los valores mínimos de este parámetro se obtienen en posiciones centrales. Lo contrario sucede con el rendimiento (TP, *throughput*), ver Figura 3.20.c, en seguimiento simulado de la mirada ($\mu = 2.5$ bps, $\sigma = 0.26$) y mediante señalización manual ($\mu = 3.74$ bps y $\sigma = 0.44$).

Figura 3.20: a) Configuración del espacio para el experimento por posiciones(b) Tiempo de movimiento para ambas técnicas de apuntamiento en las 7 posiciones configuradas para el experimento. (c) Parámetro de rendimiento (*throughput*, TP) para ambas técnicas de apuntamiento. Las barras de error mostradas corresponden a $\pm 1SD$

3.6.1.3. Dificultad percibida y aprendizaje

Los participantes de las pruebas evaluaron la dificultad que percibieron haciendo la tarea de la ISO. El facilitador de las pruebas les preguntaba por la dificultad percibida después de completar cada una de las iteraciones, la respuesta estaba en el rango entre 1 (muy fácil) y 5 (muy difícil). La tabla 3.3 refleja valores estadísticos relativos a este aspecto dependiendo del tipo de interacción empleada. A la vista de esta tabla, se aprecia que apuntar usando

la señalización manual es ligeramente más fácil que con la simulación de la mirada. Además, la dificultad se ve incrementada a medida que el objetivo se hace más pequeño, este comportamiento se reproduce para ambas técnicas de apuntamiento. Esta diferencia de dificultad tampoco es estadísticamente significativa; Se ha aplicado un test de rangos con signos de Wilcoxon (para datos emparejados) a los datos recogidos en la quinta iteración y todos los tamaños de objetivo. El resultado muestra que no es posible rechazar la hipótesis nula ya que no hay diferencia entre las medianas con un 5 % de nivel de significación. Con este mismo nivel de significación, la prueba de suma de rangos de Wilcoxon (para datos no emparejados) muestra que no hay diferencias entre lo que se denomina en este estudio usuarios y no expertos para cualquiera de los tamaños disponibles de los objetivos.

Tabla 3.3: Dificultad percibida (1 muy baja, 5 muy alta)

Circle size	Gaze			Pointing		
	Mean	Median	Mode	Mean	Median	Mode
Small	2.49	3	3	2.2	2	2
Medium	1.85	2	1	1.5	1	1
Big	1.55	1	1	1.34	1	1
Total	1.96	2	1	1.68	1	1

Otros de los aspectos estudiados es la evolución del tiempo para completar las tareas y la valoración de la dificultad distinguiendo el tamaño del objetivo, como se muestra en la Figura 3.21. De esta Figura se deduce que el tiempo para completar la tarea cuando el objetivo tiene un tamaño grande es bastante constante entre las diferentes iteraciones, mientras que para tamaños más pequeños los valores de este tiempo son más variables entre las iteraciones. Este efecto es más destacado en el caso de usar apuntamiento mediante señalización manual. Adicionalmente se preguntó a los participantes por su percepción de aprendizaje durante el uso de este sistema. Muchos de ellos (18 de 20 participantes) coincidieron en afirmar que habían mejorado su apuntamiento con el número de iteraciones de la prueba. Por el contrario, 2 de ellos únicamente aseguraron haber conseguido un leve aprendizaje durante las primeras iteraciones. Estos participantes sostuvieron que este hecho se debía a que una vez conocida la dinámica de la prueba era más fácil apuntar.

Figura 3.21: a) Tiempo promedio en completar la tarea por iteración.
b) Promedio de la dificultad percibida por iteración.

3.6.2. Confort, usabilidad y experiencia de usuario

Esta segunda parte del análisis está orientada a conocer de las impresiones y preferencias de los usuarios respecto a las técnicas de apuntamiento probadas.

3.6.2.1. Análisis del confort del usuario

Los usuarios respondieron a un cuestionario adaptado (ver Anexo A) del anexo C de la ISO 9241-9⁷ sobre la comodidad a la hora de emplear estas técnicas de apuntamiento. El cuestionario planteado tras realizar la primera tarea relativa a la ISO, emplea la escala de Likert para evaluar este aspecto, variando los valores de 1 (poco confortable) a 7 (muy confortable). Los datos recogidos en la tabla 3.4 muestran como el apuntamiento mediante señalización manual parece ser más preciso que el apuntamiento por simulación del seguimiento de la mirada, aunque requiere aplicar una fuerza mayor y provoca un mayor esfuerzo. Por el contrario, el apuntamiento mediante señalización manual se aprecia ligeramente más suave que el otro tipo de apuntamiento. En cuanto a la velocidad de respuesta, ambas técnicas obtienen una evaluación similar, sobre los 5 puntos. Respecto de la fatiga, también evaluada en este cuestionario, el apuntamiento a través de señalización manual se valo-

⁷ISO 9249-1: http://www.iso.org/iso/catalogue_detail.htm?csnumber=30030

ra fatigoso para los dedos, la muñeca y el cuello; mientras el realizado con simulación de seguimiento de la mirada concentra la fatiga en los ojos, obteniendo una valoración similar que la fatiga de los dedos para la otra técnica de apuntamiento.

La prueba de los rangos con signos de Wilcoxon aplicada a estos datos demuestra que es posible rechazar la hipótesis nula donde no hay diferencia entre las medianas con un nivel de significación al 5 % para los aspectos comunes evaluados en ambas técnicas de interacción. Los resultados de esta prueba para cada aspecto común son: fuerza de actuación requerida ($W = 11$, $p = 0.0083$), precisión ($W = 35.5$, $p = 0.0458$), rendimiento general ($W = 13.5$, $p = 0.0014$) y fatiga del cuello ($W = 14.5$, $p = 0.015$). Esta hipótesis nula no es posible rechazarla para el resto de aspectos.

Tabla 3.4: Resultados del cuestionario de evaluación del confort

	Gaze		Deictic	
	Mean	Standard deviation	Mean	Standard deviation
Required actuation force	2.35	1.14	3.25	1.55
Interaction smoothness	4.2	1.67	3.7	1.66
Interaction effort	3.5	1.36	3.85	1.57
Accuracy	4.05	1.61	4.7	1.08
Speed of use	5.1	1.41	5.1	1.33
General comfort	4.55	1.39	4.2	1.06
Overall performance of the input device	3.3	1.34	5	1.17
Neck fatigue	5.45	1.28	6.4	1.14
Eye fatigue OR fingers fatigue	5.95	1.23	5.8	1.85
Wrist fatigue	NA	NA	5.75	1.55
Arm fatigue	NA	NA	3.25	1.68
Shoulder fatigue	NA	NA	4.3	1.84

El estudio [55] emplea el anexo C de la ISO 9241-9 [60], como en este caso, para evaluar la satisfacción del usuario y la fatiga al hacer uso de

un sistema de interacción que emplea las dos manos del usuario, una para apuntar y la otra para desencadenar acciones a través de un vocabulario predefinido. Los resultados obtenidos por sus autores son comparables en términos de precisión o velocidad a los obtenidos en este estudio, sin embargo, los resultados de la fatiga obtenidos en estas pruebas son peores. Este hecho puede ser debido al elevado número de repeticiones que se requieren para completar la tarea de la ISO.

3.6.2.2. Análisis de usabilidad y experiencia de usuario

La usabilidad y la experiencia de usuario ha sido evaluada con la tarea de manejar objetos inteligentes del entorno. Esta tarea obliga al usuario a interactuar 10 veces con dos objetos, al menos una vez con cada uno de ellos. Los dos objetos elegidos son una bombilla inteligente posicionada a los 30°, tomando como referencia los objetivos de la tarea ISO, y un radio con bluetooth colocada en el ángulo de 240°. Un 75 % de los participantes que realizaron las pruebas consiguieron completar esta tarea apuntando mediante señalización manual. En el caso de apuntamiento por simulación de la mirada, este porcentaje se amplía hasta un 99 %.

Por lo que respecta a la evaluación de la usabilidad, se ha empleado la *Escala de usabilidad del sistema* definida en [8]. Se solicitó a los participantes una calificación para diez cuestiones referentes a la complejidad del sistema, frecuencia de uso o aprendizaje, con una valoración de cada una de ellas que iba del 1 (totalmente en desacuerdo) al 5 (totalmente de acuerdo) en relación con la escala de *Likert*. Con las respuestas recogidas se calcula un valor de esta *Escala de usabilidad del sistema* que va desde un valor de 0 hasta 100. El valor obtenido para el apuntamiento por señalización manual ha sido de 71.66 puntos frente a 54.58 puntos para el apuntamiento por seguimiento de la mirada, lo que conduce a afirmar que la primera técnica de apuntamiento (señalización manual) consigue una calificación razonablemente buena en contraposición a la otra técnica (seguimiento de la mirada), con una calificación menor. La tabla 3.5 recoge un resumen de los datos obtenidos.

De nuevo, se ha aplicado un test de los rangos con signo de Wilcoxon para detectar las diferencias entre estos datos en ambos tipos de apuntamiento, asumiendo una distribución simétrica de las diferencias sobre la mediana, con un nivel de significación del 5 %. La prueba muestra que existe una diferencia significativa en la declaraciones sobre la aceptación del sistema para su uso frecuente ($W = 4,5$, $p = 0,0011$), complejidad ($W = 11$, $p = 0,0273$) y facilidad de uso ($W = 14,5$, $p = 0,0029$), la necesidad de soporte técnico

para utilizar el sistema ($W = 28$, $p = 0.0156$), la percepción de inconsistencias del sistema ($W = 116$, $p = 0.0119$) y la confianza de los usuarios a la hora de usar el sistema ($W = 16.5$, $p = 0.0447$). El apuntamiento mediante señalización manual consigue una mejor valoración de estos aspectos que la técnica por simulación de seguimiento de la mirada.

Tabla 3.5: Cuestionario estándar de puntuación de usabilidad. Cada ítem es valorado entre un 1 (bastante en desacuerdo) y un 5 (muy de acuerdo)

		Mean		STD	
		Gaze	Deictic	Gaze	Deictic
1	I think that I would like to use this system frequently	2.2	3.3	1.1	0.9
2	I found the system unnecessarily complex	3.1	2.5	1.4	1.2
3	I thought the system was easy to use	2.6	4.0	1.2	1.1
4	I think that I would need the support of a technical person to be able to use this system	1.9	1.5	0.8	0.6
5	I found the various functions in this system were well integrated	3.4	3.9	1.1	0.9
6	I thought there was too much inconsistency in this system	3.0	2.0	1.2	0.9
7	I would imagine that most people would learn to use this system very quickly	3.4	3.9	1.2	0.8
8	I found the system very cumbersome to use	2.9	2.7	1.3	1.1
9	I felt very confident using the system	2.7	3.7	1.3	1.3
10	I needed to learn a lot of things before I could get going with this system	1.7	1.5	1.1	1.1

En cuanto a la evaluación de la experiencia de usuario (UX, *user experience*) se ha empleado el cuestionario propuesto en [7]. El cuestionario (ver Anexo A) maneja 6 escalas diferentes evaluando el aspecto atractivo del sistema (impresión general del sistema), claridad (facilidad para familiarizarse), eficiencia (idoneidad para resolver las tareas del usuario), fiabilidad (sensación de control), estimulación provocada (excitación y motivación) y lo novedoso del sistema (innovación) a través de 26 preguntas. La tabla 3.6 recoge los valores promedios de las valoraciones de los 20 participantes de la prueba para ambos tipos de interacción. A simple vista, el apuntamiento mediante señalización manual consigue mejores valoraciones que la otra técnica de apuntamiento por mirada. Nuevamente se vuelve a emplear un test de los rangos con signo de Wilcoxon (con un 5 % de nivel de significación) para mostrar las diferencias de forma formal de 10 aspectos en 5 de las escalas anteriores (la escala relativa a la claridad es la única que no presenta diferencias). En particular, tres cuestiones relacionadas con la escala de eficiencia muestran diferencias: velocidad ($W = 16$, $p = 0.0065$) y eficiencia ($W = 136.5$, $p = 0.0231$). El resto de las escalas difieren en dos cuestiones, con la excepción de la estimulación que sólo presenta diferencias en una cuestión. Al igual que ocurría en anteriores situaciones la puntuación para el apuntamiento por señalización manual supera a la basada en simulación de

seguimiento de la mirada.

Tabla 3.6: Clasificación para la escala del cuestionario de la experiencia del usuario (rango de -3 a 3). resultados de referencia y *Cronbach's Alpha*

	Scale	Mean	STD	Comparison to benchmark	Interpretation of benchmark rating	Cronbach's Alpha
Deictic	Attractiveness	1.46	0.86	Above average	25 % of results better, 50 % of results worse	0.90
	Perspicuity	2.03	0.64	Excellent	In the range of the 10 % best results	0.7
	Efficiency	1.125	1.00	Above Average	25 % of results better, 50 % of results worse	0.77
	Dependability	0.912	1.00	Below Average	50 % of results better, 25 % of results worse	0.65
	Stimulation	1.32	0.84	Above Average	10 % of results better, 75 % of results worse	0.75
	Novelty	1.38	0.77	Excellent	In the range of the 10 % best results	0.44
Gaze	Attractiveness	0.42	1.46	Bad	In the range of the 25 % worst results	0.93
	Perspicuity	1.35	1.10	Above Average	25 % of results better, 50 % of results worse	0.75
	Efficiency	0.15	1.46	Bad	In the range of the 25 % worst results	0.89
	Dependability	0.03	1.23	Bad	In the range of the 25 % worst results	0.73
	Stimulation	0.7	1.14	Below Average	50 % of results better, 25 % of results worse	0.83
	Novelty	1.36	0.99	Excellent	In the range of the 10 % best results	0.79

Además de esto, se ha medido la consistencia interna de los datos (como de estrecha es la relación entre un conjunto de datos) usando el coeficiente *Alpha* propuesto por Cronbach en 1951 [70]. El cálculo de este coeficiente muestra que los datos referentes al apuntamiento con seguimiento de la mirada son suficientemente consistentes (habitualmente, un valor que está sobre 0.7 es considerado suficiente). Por lo que respecta al apuntamiento mediante señalización manual, este coeficiente tiene un valor significativamente por debajo de 0.7, lo que puede ser una indicación de que algunos factores de la escala a evaluar no fueron correctamente entendidos por algunos de los participantes.

Adicionalmente, el uso de las herramientas del sitio web cuestionarios de experiencia de usuarios⁸ permiten comparar el rendimiento del sistema, gracias a que disponen de los resultados de 163 estudios que previamente se realizaron para analizar la experiencia del usuario sobre software de negocios, páginas web, comercio *online* y redes sociales. Como recoge la tabla 3.6, el apuntamiento por señalización manual sólo está por debajo del promedio total en las cuestiones relacionadas con la fiabilidad. El apuntamiento por simulación del seguimiento de la mirada continúa proporcionando resultados peores que los tomados como referencia.

⁸Herramientas para cuestionarios sobre experiencias de usuarios: <http://www.ueq-online.org>

3.6.2.3. Influencia de la referencia visual en la experiencia de usuario y preferencia del mismo

Aparte de las pruebas anteriores realizadas con referencia visual, en la tarea ISO se repite la misma prueba una única vez, para cada tamaño objetivo, sin la referencia visual del punto rojo en movimiento. Esta repetición de la prueba, bajo esta condición, se debe a que en un despliegue real es complicado proporcionar una referencia visual del apuntamiento del usuario. Otra motivación para repetir esta prueba es el resultado mostrado en [71], donde se afirma que el número de errores se ve incrementado cuando no se proporciona una referencia del movimiento al usuario; en realidad, muchos usuarios no fueron capaces de completar la tarea. El mejor rendimiento obtenido bajo esta condición se obtuvo con el objetivo tamaño mayor, siendo el porcentaje de participantes que finalizaron la prueba un 70 % y un 55 % usando apuntamiento a través de señalización manual y a través de simulación del seguimiento de la mirada, respectivamente .

El tiempo para completar la tarea se ve incrementado al no existir referencia visual. La tabla 3.7 recoge los tiempos en terminar la tarea para ambas condiciones. En el caso del apuntamiento por señalización manual, los tiempos se ven multiplicados por un factor con un valor mínimo de 3.7 a 4.1 según el tamaño; mientras que para la técnica de simulación de seguimiento de mirada este factor varía de 3.9 a 2.6. La dificultad que percibe el usuario bajo esta condición, sin referencia visual, también se ve incrementada, siendo aproximadamente el doble (con un factor de multiplicación entre 1.8 y 2.2) en ambos tipos de apuntamiento. Aun así, el empleo del apuntamiento por señalización manual se aprecia más difícil.

Por otra parte, la mayoría de los usuarios prefieren disponer de una referencia visual, tanto para apuntar mediante señalización manual (18 de 20) como para cuando se usa la simulación del seguimiento de la mirada (19 de 20). El resultado de preferencia es 5.45 usando señalización manual, y 5.65, al emplear simulación del seguimiento de la mirada cuando se realiza la tarea de la ISO. Estos dos valores están en una escala del 1 al 7, siendo el 7 *muy preferido*. Este valor sube hasta 6.3, en el caso de apuntamiento por simulación del seguimiento de la mirada, cuando se realiza la tarea de manejar objetos inteligentes.

Tabla 3.7: Dificultad percibida (1:bajo-5:alto) y tiempo para completar la tarea ISO con y sin referencia visual.

		Gaze pointing		Deictic pointing	
		Perceived difficulty	Task completion time(ms)	Perceived difficulty	Task completion time(ms)
With visual cue	Small	2.49	24402	2.2	25539
	Medium	1.85	15540	1.5	13745
	Big	1.55	12177	1.34	11650
	Average	1.96	17373	1.68	16311
Without visual cue	Small	4.45	95416	4.3	101428
	Medium	3.35	50404	3.6	74342
	Big	2.55	31677	3.05	43662
	Average	3.45	59166	3.65	73144

3.6.2.4. Preferencias de usuario sobre la técnica de detección de apuntamiento

Aparte de la preferencias de los usuarios que ya se han estudiado, se les plantea una serie de cuestiones abiertas para conocer las preferencias del tipo de apuntamiento que prefieren (ver Anexo A).

Respecto al apuntamiento mediante seguimiento de simulación de la mirada, los 20 participantes coincidieron en calificar esta técnica como agradable, útil, fácil de usar, simple, atractiva, rápida y precisa. Uno de ellos señaló que la forma de apuntar usando esta técnica no era todo lo natural que se puede esperar (se dio cuenta de que el sistema funcionaba mejor cuando se apuntaba usando como punto de referencia la nariz que los ojos). Otro de ellos informó de que el sistema era bastante impreciso en lo que al apuntamiento se refiere, cuando se apunta hacia posiciones en la mitad inferior. Este método de apuntamiento sorprendió a algunos de los participantes con respecto a la velocidad de respuesta y precisión. Por otra parte, todos los usuarios coinciden en afirmar que el apuntamiento mediante señalización manual es fácil de ejecutar y agradable, aunque un uso prolongado del sistema resulta cansado. Al comparar ambas técnicas, este último apuntamiento resulta más estable (menor movimiento de la referencia visual, punto rojo que marca el apuntamiento del usuario).

El apuntamiento por simulación del seguimiento de la mirada es el método preferido de 13 de los participantes del estudio, 5 de ellos prefieren apuntar usando la señalización manual, y los 2 últimos restantes optan por elegir el método de apuntamiento de acuerdo al propósito de la interacción. Un número pequeño de participantes (5 de 20) no incorporarían el apuntamiento en su vida diaria, por ejemplo para activar elementos de su hogar y realizar acciones. De todos los participantes, 4 de ellos usarían apuntamiento para interacciones cortas, como por ejemplo tener un control remoto sin necesidad

de disponer de un dispositivo adicional.

También se preguntó a los usuarios sobre la posibilidad de combinar el apuntamiento con el reconocimiento de voz (por ejemplo, apuntar a una bombilla y decir encender) para completar ordenes sobre objetos inteligentes. Un grupo reducido de los participantes (4 sobre los 20 totales) encontraron esta iniciativa redundante, mientras otro grupo afirmó que esto añadiría a la interacción realizada mayor utilidad, seguridad, naturalidad, agilidad, versatilidad y mayor fiabilidad. A ninguno de ellos les agradó la idea de usar solamente el reconocimiento de voz, excluyendo el apuntamiento. El reducido grupo restante (2 de 20) no proporcionó su opinión debido a que nunca antes habían probado esta combinación. Por otra parte, uno de los usuarios de las pruebas sugirió invertir el orden de la interacción, primero decir la orden y en segundo lugar, apuntar. Entre los problemas del reconocimiento de voz para la interacción que trasladó uno de los participantes se encuentra el problema que se pudiesen ocasionar en entornos ruidosos y la dificultad en el caso de entornos multiusuario que pueda realizar el sistema.

3.7. Resumen

Este capítulo estudia las técnicas de detección del apuntamiento basadas en señalización manual y simulación del seguimiento de la mirada. Además, realiza una comparativa de estas técnicas de detección.

Una vez planteados los objetivos del capítulo, se realiza la introducción a una serie de metodologías antes de comenzar con el estudio, empezando por la Ley de Fitts. Dicha ley nos permite evaluar el rendimiento de un usuario al hacer uso sistemas en los que la interacción este basada en apuntamiento, calculando una serie de parámetros como son el índice de dificultad, el índice de rendimiento y el tiempo de movimiento. La segunda herramienta es el Estándar ISO 9241-9, el cuál plantea una serie de tareas para llevar a cabo la evaluación del sistema de apuntamiento. En concreto el Anexo B de este Estándar usa la Ley de Fitts. Seguidamente, se introduce el escenario donde se va a desarrollar el sistema de interacción y se van a llevar a cabo las pruebas del mismo por parte de los usuarios.

La estimación del apuntamiento se explica en la Sección 3.2.2. El dispositivo empleado, Kinect, captura los puntos escogidos del usuario en un sistema de coordenadas interno y propio de la cámara del sensor; hasta que el sistema devuelve el punto apuntado por el usuario en el sistema de coordenadas del

entorno real; para ello es necesario hacer una serie de transformaciones.

Posteriormente, la Sección 3.4 detalla la arquitectura del sistema de apuntamiento diseñado aplicaciones que habilitan las pruebas de usuario.

Una vez planteado el sistema y explicada su arquitectura, despliegue y las operaciones que realiza se descubre el diseño del estudio para realizar la evaluación del sistema y la valoración de la experiencia de usuario. Este diseño se lleva a cabo en la Sección 3.5. El estudio detalla los participantes implicados, los dispositivos y las herramientas usadas, los procedimientos y tareas realizadas y el desarrollo completo del experimento.

En la última Sección se recogen los resultados del experimento planteado en la Sección previa. Los resultados se han dividido en dos grandes bloques: 1) análisis del rendimiento de la interacción; y 2) el análisis del confort, usabilidad y la experiencia de usuario. El primer bloque usa los conceptos recogidos en la Ley de Fitss. El segundo bloque está centrado en el Estándar ISO 9241-9.

En el próximo Capítulo se diseña un sistema de apuntamiento *mejorado* usando detección del apuntamiento mediante señalización manual que es el método que, a la luz de los resultados es mejor opción en términos de comodidad, facilidad de uso y respuesta a la experiencia de usuario ya que de acuerdo al índice de dificultad y rendimiento ambos tienen resultados similares. Este próximo Capítulo se alimentará de los resultados obtenidos en este Capítulo 3 para dar soluciones y mejoras a las características estudiadas.

Capítulo 4

Diseño de un sistema *mejorado* de apuntamiento basado en señalización manual

4.1. Objetivos del Capítulo

El Capítulo está dedicado a describir el diseño de un sistema de apuntamiento *mejorado* mediante *señalización* manual. Para el desarrollo del mismo es importante tener en cuenta los requisitos que estos sistemas deben cumplir para ello se hace inicialmente una revisión de los mismos.

La estimación del apuntamiento incluye un error que tiene su origen en diversos factores. En este Capítulo se estudian y se plantean soluciones para mitigarlo. En primer lugar, se analizará cuál es el vector de apuntamiento manual más exacto entre las diferentes alternativas que se plantean. Partiendo de los resultados obtenidas para el vector ojo dominante-mano, se analizarán factores que pueden hacer que la estimación del apuntamiento sea inexacta e imprecisa.

El ojo dominante de una persona es aquel que controla la visión de profundidad, mientras el otro controla la visión periférica y espacial. Con la combinación de ambos el cerebro recibe una imagen tridimensional. El usuario empleará este ojo, por ejemplo, para mirar por un microscopio ya que es con el que enfocará mejor. Un proceso habitual para determinar el ojo dominante consiste en colocar las manos extendidas formando un agujero en forma de triángulo y fijar la vista en una imagen un objeto a través del

agujero. Luego, se guiña uno de los ojos cada vez, aquel ojo con el que la imagen sigue centrada dentro del agujero es el ojo dominante ¹.

Finalmente se propone un sistema de apuntamiento *mejorado* que mitiga los errores del proceso de estimación del apuntamiento. Una de estas mejoras es realizar una calibración personalizada al usuario que mejore el apuntamiento realizado, denominada **calibración de usuario**, que no debe confundirse con la calibración del sistema. Por su parte, la **calibración del sistema** transforma el sistema de coordenadas que maneja el sensor al sistema de coordenadas definido para el entorno real.

4.1.1. Revisión de los requisitos del sistema de apuntamiento

Los requisitos para considerar un sistema de apuntamiento adecuado para su uso, son los siguientes:

- **Exactitud.** Uno de los principales requisitos exigidos a un sistema de apuntamiento es que sea exacto. La exactitud se define como la distancia entre el centro de masas del objeto apuntado y el punto estimado por el sistema de apuntamiento. En definitiva, la exactitud hace referencia a la proximidad entre el valor real y el estimado.

Relacionado a la exactitud del sistema se encuentra la definición de precisión para el sistema de apuntamiento. Este concepto está orientada a conocer la dispersión de los valores estimados en varias estimaciones; obteniendo mayor precisión del sistema a menor desviación de los datos estimados.

Aunque en la vida cotidiana estos dos términos se usen indistintamente hay que remarcar que tienen significados distintos. La diferencia de estos dos conceptos puede entenderse mejor de forma gráfica como puede verse en la Figura 4.1. La precisión muestra las veces que un mismo valor es estimado frente a la exactitud que indica lo cerca que está el valor estimado del que debería ser.

Por su parte la Norma ISO 5725-1 (*Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions*) usa los términos de veracidad y precisión como métodos de medida. La *veracidad* puede asemejarse al término de exactitud;

¹www.vista-laser.com/como-saber-cual-ojo-dominante/

Figura 4.1: Diferencia de los términos exactitud y precisión de modo gráfico. Fuente de la imagen : https://es.wikipedia.org/wiki/Precisi3n_y_exactitud

que aparece definida en dicha Norma como la proximidad al valor de referencia del valor medio obtenido de un conjunto de mediciones realizadas. Sin embargo, la precisión se refiere a la proximidad entre los valores obtenidos de un conjunto de mediciones hechas.

- **Ubicuidad.** Con este requisito, el sistema debe ser capaz de reconocer el apuntamiento independientemente de la posición en que se coloque el usuario dentro del área de acción del sistema de apuntamiento, por lo que debe cubrir la mayor parte de espacio posible y todas las posturas de una persona. En definitiva, el sistema debe permitir al usuario libertad de movimiento dentro del espacio inteligente.
- **Universalidad.** El sistema debe ser capaz de servir para cualquier persona. Es decir, permitir reconocer el apuntamiento de cualquier individuo, independientemente de sus características personales y su forma de apuntar. En este sentido se encuentran ciertos factores a los que el sistema debería ser independiente y que pudiesen interferir en la estimación del apuntamiento. Estos podrían ser la forma de apuntar que es particular del individuo, su ojo dominante, o bien, el uso de gafas.
- **Facilidad de despliegue.** El último requisito a tener en cuenta es que el sistema de apuntamiento se despliegue de modo fácil. Aunque el proceso de calibración del sistema será un proceso necesario e imprescindible, existen otros procedimientos de los que se podría prescindir. Estos procedimientos de los que se pudiese prescindir serían las posibles mejoras que se pudiesen hacer al sistema para conseguir mejores resultados, como por ejemplo, una calibración concreta por usuario. Esto implicaría mayor tiempo a la hora de ponerse un usuario a utilizar el sistema, aunque se conseguirían mejores resultados.

4.1.2. Entorno de despliegue

El entorno donde se realiza el despliegue del sistema de apuntamiento aparece recogido en el esquema de la Figura 4.2. Este espacio corresponde al área de cobertura del dispositivo Kinect. En el suelo se han definido unas posiciones para abarcar la cobertura en distancia y en ángulo-orientación del sensor. Cada una de las posiciones están identificadas con una letra (cada letra identifica el ángulo-orientación tomando como referencia el sensor, también llamadas *líneas de ángulo*, y son 3: A,B,C) y un número (cada número distingue la distancia al sensor).

4.1.3. Experimentos para determinar las fuentes de error en la estimación del apuntamiento

Se van a realizar algunos experimentos para estudiar cuales son los posibles factores que pueden estar influyendo en el error. Estos se estructuran como:

1. **Aspectos sobre la estimación de la coordenada en altura (Y) de los puntos que definen el vector de apuntamiento.** Durante la estimación de la coordenada en altura sucede algún fenómeno que provoca un sesgo en el error. Este fenómeno se estudia en la Sección 4.3. Los aspectos estudiados en ella son:
 - (a) Estabilidad de la estimación de la altura de puntos de referencia (pupila) (Sección 4.3.1).
 - (b) Efecto de pequeñas variaciones en la estimación de la altura (de la pupila) sobre la estimación de la coordenada en altura del apuntamiento (Sección 4.3.2).
 - (c) Variaciones en la estimación de la altura (de la pupila) en función de la posición del usuario con respecto al sensor (Sección 4.3.3).
 - (d) Variaciones en la estimación de la altura (de la pupila) en función de la inclinación del sensor Kinect (Sección 4.3.4).
 - (e) Variaciones en la estimación de la altura de otros puntos de referencia (cabeza, dedo) en función de la localización del usuario (Sección 4.3.5).
 - (f) Variación de las coordenadas de apuntamiento en función del uso de gafas (Sección 4.3.6).

2. **Variación del resultado del apuntamiento en función del diseño de la calibración.** Además se estudiará la influencia de la calibración empleada con los problemas que conlleva y la posible solución en la Sección 4.4. Los diferentes aspectos estudiados para conocer la influencia son:

- (a) Relación entre la posición del patrón de calibración y el resultado del apuntamiento (Sección 4.4.1).

Figura 4.2: Esquema del entorno de despliegue. A la izquierda aparece la sala, con el sensor en posición superior-centro y el esquema con las posiciones predefinidas aparece a la derecha. Destacar que en las posiciones A y E el sensor no capta la cara, por lo que no se han recogido datos. El ángulo está centrado en la posición del sensor y proyectada en el suelo. En el lado izquierdo de la imagen aparecen los ejes X,Y,Z en el punto que se ha tomado como origen de los mismos. El eje Z corresponde a la profundidad en horizontal y el eje Y corresponde a la vertical (altura).

- (b) Relación entre la inclinación del patrón de calibración y la estimación de las coordenadas en altura (Sección 4.4.2).
- (c) Relación entre la inclinación del sensor dispositivo Kinect y la inclinación de la calibración (Sección 4.4.3).
- (d) Propuesta de corrección de los efectos del diseño de la calibración y el análisis de la universalidad de la solución (Sección 4.4.4).

Los experimentos realizados se han planteado cada uno de ellos con la siguiente estructura:

1. Objetivo.
2. Preparación y tareas.
3. Resultados obtenidos.
4. Conclusiones y conexión con otros experimentos.

Estos experimentos no son experimentos exhaustivos, sino de exploración para determinar y descubrir diferentes comportamientos de los sistemas de apuntamiento. Su objetivo es tener un pre-diseño que permita confirmar intuiciones e información previa existente.

Premisas de las características de los usuarios participantes de las pruebas

Todos los participantes de las pruebas son personas cuyo ojo dominante es el derecho. Además son personas diestras, es decir, tienen como tendencia natural usar el lado derecho de su cuerpo. En este punto, se puede hacer las siguientes generalizaciones para el tratamiento de estos conceptos en esta Tesis:

- El ojo derecho es el ojo dominante.
- La mano derecha es la mano hábil de los participantes.
- Se comprueba durante las pruebas que el hombro derecho es el hombro más hábil de los participantes.

4.2. Selección del vector de apuntamiento

El vector escogido para estimar el apuntamiento es un aspecto esencial en estos sistemas. Este vector aproxima la dirección objetivo del usuario, a partir de las coordenadas físicas captadas de los sensores de medida.

Como se adelantó en el Capítulo del Estado del Arte en su Sección 2.2.2 hay diferentes alternativas, cada una con resultados diferentes en lo que se refiere a exactitud. Los vectores utilizables dependen también de los sensores de medidas. En el despliegue que se plantea en este Capítulo, es posible comparar el funcionamiento del sistema, utilizando, por ejemplo, los siguientes vectores:

	Punto origen	Punto destino
a)	Ojo izquierdo	Mano derecha (dedo índice)
b)	Ojo derecho	Mano derecha (dedo índice)
c)	Hombro derecho	Mano derecha (dedo índice)
d)	Punto extremo de la nariz (<i>nose top</i>)	Mano derecha (dedo índice)
e)	Centro esférico de la cabeza	Centro de la cara
f)	Centro esférico de la cabeza	Mano derecha (dedo índice)

4.2.1. Pruebas para la selección del vector de apuntamiento

Con el objetivo de evaluar qué vector tiene mejor resultado en términos de exactitud, dos usuarios realizaron una prueba de comprobación, ya que la revisión de la Literatura y de las pruebas del Capítulo 3 , se espera que el vector ojo-mano sea el más eficiente.

Las pruebas realizadas determinan la exactitud del apuntamiento empleando estos 6 vectores. La tarea a realizar en esta prueba consistió en que el participante apuntó hacia un único objetivo fijo situado en la pared de proyección durante un tiempo. El usuario tenía la libertad de moverse por el entorno dentro del área de cobertura del dispositivo Kinect, siempre que todas las localizaciones estuviesen en el área de cobertura, y sin dejar de apuntar al objetivo en ningún momento. En cada instante, se tomaron las posiciones de las diferentes partes del cuerpo del usuario para poder estimar

cada uno de los vectores de la tabla 4.1. Con esta información, se realizó la proyección de cada uno de los vectores sobre la pared. El error calculado es la distancia al centro del objetivo en cada una de sus coordenadas.

Las pruebas son realizadas por dos usuarios distintos. Los resultados del error de exactitud (diferencia en milímetros) al centro del objetivo en cada una de las dos coordenadas (X e Y) sobre la pared donde se encuentra situado se muestran en la tabla 4.2. Para la obtención de los valores de las tablas, las muestras se promedian sin diferenciar ángulo ni posiciones relativas entre usuario y blanco. El objetivo es apreciar el funcionamiento global del sistema en todo el espacio de cobertura del sensor.

	Usuario	Diferencia en coordenada X		Diferencia en coordenada Y	
		Promedio (μ)	Desviación estándar (σ)	Promedio (μ)	Desviación estándar (σ)
Vector ojo izquierdo-mano	Usuario 1	349	30	-230	61
	Usuario 2	325	15	-85	19
Vector ojo derecho-mano	Usuario 1	-7	29	-186	60
	Usuario 2	87	16	-73	18
Vector hombro-mano	Usuario 1	-397	30	822	60
	Usuario 2	-277	16	543	21
Vector <i>nose top</i> -mano	Usuario 1	171	31	-252	62
	Usuario 2	212	16	-108	19
Vector cabeza-cara	Usuario 1	-53	49	1.532	46
	Usuario 2	-91	37	793	37
Vector cabeza-mano	Usuario 1	137	25	21	52
	Usuario 2	174	15	6	17

Tabla 4.2: Error en cada una de las coordenadas de 2 dimensiones para cada vector en milímetros. El signo del resultado indica la dirección de desvío del error (derecha o izquierda, + o -, para coordenada X; arriba o abajo, + o -, para coordenada Y)

A la vista de los resultados extraídos para el Usuario 1 (tabla 4.2) el mejor resultado promedio del error para la coordenada X se obtiene usando el vector *ojo derecho-mano* con un valor medio de -7 milímetros y una desviación de los datos de 29 milímetros. Es un valor promedio muy aceptable ya que es un error que no llega a 1 centímetros. Sin embargo, con este mismo vector, el error en coordenada Y aumenta notablemente teniendo un valor promedio de -186 milímetros con una desviación alrededor de los 60 milímetros. Aunque no es un error excesivo, no es lo suficiente para poder emplear un sistema de apuntamiento exacto como el que se está intentando diseñar. El vector *cabeza-cara* también obtiene un buen valor del error promedio en X situándose en -53 milímetros y desviación estándar de 49 milímetros. En el caso del Usuario 2, los mejores valores promedios del error en la coordenada X también se obtienen para estos dos vectores. Sin embargo, el error es superior al anterior usuario. En el caso del vector *ojo derecho-mano* es de 87 milímetros (desviación de los datos de 16 milímetros) y para el vector *cabeza-cara* es de -91 milímetros (desviación típica de 37.6 milímetros). Destacan

los valores desorbitados de error para la coordenada Y que se da para estos 2 usuarios y los mismos vectores, en concreto, el vector *hombro-mano* y el vector *cabeza-cara*. En el caso del Usuario 1 está sobre los 80 centímetros y 150 centímetros; para el otro Usuario, sobre 50 centímetros y 80 centímetros. Este es uno de los puntos de partida para entender que algún fenómeno está afectando a la estimación de la coordenada Y, por lo que es necesario analizar y estudiar para entender su comportamiento, como se realizará en la Sección 4.3.

Se representa de forma gráfica lo expuesto anteriormente para un único usuario escogiendo los 3 vectores con mejores resultados: *ojo derecho-mano* (Figura 4.3), *cabeza-cara* (Figura 4.4) y *cabeza-mano* (Figura 4.5). A la vista de estas gráficas se que para el error en coordenada X hay un importante sesgo al evaluar el vector *cabeza-mano*. En el caso de la coordenada Y, se produce un sesgo destacable en la proyección del vector *cabeza-cara*. En los vectores en los que interviene la cabeza se produce un error importante en alguna de las dos coordenadas. La proyección del vector *ojo derecho-mano* tiene un comportamiento más correcto aunque existe un error y se introduce un sesgo pero mucho menor que en el anterior caso.

A la vista de los resultados y por los motivos anteriormente expuestos se comenzará a trabajar con el vector *ojo derecho-mano*, comprobando su exactitud, si se adecua a las necesidades del sistema e intentando proporcionar una solución a problemas que se puedan encontrar durante este estudio.

Figura 4.3: Vector Ojo derecho-mano. Error de proyección en cada una de las coordenadas X e Y para uno de los usuarios y diferentes vectores. El error se calcula como la diferencia entre la coordenada objetivo y el punto apuntado estimado.

Figura 4.4: Vector Cabeza-cara. Error de proyección en cada una de las coordenadas X e Y para uno de los usuarios y diferentes vectores. El error se calcula como la diferencia entre la coordenada objetivo y el punto apuntado estimado.

Figura 4.5: Vector Cabeza-mano. Error de proyección en cada una de las coordenadas X e Y para uno de los usuarios y diferentes vectores. El error se calcula como la diferencia entre la coordenada objetivo y el punto apuntado estimado.

4.3. Aspectos sobre la estimación de la coordenada en altura (Y) de los puntos que definen el vector de apuntamiento

En esta sección se va a estudiar el error de la coordenada en altura (coordenada Y). Para ello se realizarán una serie de experimentos de como afecta el cálculo de la altura de los puntos origen y destino del vector de apuntamiento empleado:

1. Estabilidad de la estimación de la altura de puntos de referencia (pupila). Estudiado en el apartado 4.3.1.
2. Efecto de pequeñas variaciones en la estimación de la altura (de la pupila) sobre la estimación de la coordenada en altura del apuntamiento. Estudiado en el apartado 4.3.2.
3. Variaciones en la estimación de la altura (de la pupila) en función de la posición del usuario con respecto al sensor. Estudiado en la Sección 4.3.3.
4. Variaciones en la estimación de la altura (de la pupila) en función de la inclinación del sensor Kinect. Estudiado en la Sección 4.3.4.
5. Variaciones en la estimación de la altura de otros puntos de referencia (cabeza, dedo) en función de la localización del usuario. Estudiado en la Sección 4.3.5.

Además se ha estudiado la posible influencia de elementos personales del usuario que pueda estar afectando a la estimación de la altura. Se analizará si el uso de ese vector y la estimación del apuntamiento se ve afectado por el uso de gafas en la Sección 4.3.6. Esto se aprecia a partir de las diferencias de los resultados obtenidos entre usuarios.

4.3.1. Estabilidad de la estimación de la altura de puntos de referencia (pupila)

a) Objetivo

El objetivo de este experimento es comprobar si en una misma posición con el usuario inmóvil, la estimación de la coordenada Y de la pupila del ojo

derecho sufre cambios. En principio, si el usuario no se mueve (no introduce error producido por el movimiento) la coordenada Y captada no debe variar, y esto debe repetirse para todas las posiciones dentro del espacio.

b) Preparación y Tareas

La prueba consiste en calcular la coordenada Y de la pupila de un usuario que permanece en cada una de las posiciones de referencia descritas en la Sección 4.1.2.

c) Resultados

El resultado de la prueba demuestra que este valor varía con el tiempo cuando esto no debería suceder. Las gráficas de la Figuras 4.6, 4.7 y 4.8 son los histogramas con el valor de la coordenada Y en la posición 11 en cada una de las líneas de las 3 líneas de ángulos respecto a la visión del sensor (ver el esquema del entorno de la Figura 4.2). A la vista de las gráficas, los rangos de valores de la variable captada varía alrededor de unos 7 centímetros. Si fuese algo esporádico podría considerarse un *outlier*, pero se aprecia que no corresponde a una única muestra de un valor extremo. El comportamiento es similar en el resto de posiciones.

d) Conclusiones

Los resultados ponen en evidencia que para un mismo punto la estimación de Y varía en el tiempo de una toma de muestras a otra, cuando esto no debería ocurrir.

Figura 4.6: Histogramas para la coordenada Y de la pupila del ojo derecho de un único usuario. Frecuencia de valor de esta coordenada Y en las posiciones 11 en las línea de ángulo de visión A (ver el esquema del entorno de la Figura 4.2). En cada eje horizontal están los diferentes valores que se recogen en milímetros y en el eje vertical su frecuencia durante la toma de datos. Todas corresponden a la posición 11 (A11).

Figura 4.7: Histogramas para la coordenada Y de la pupila del ojo derecho de un único usuario. Frecuencia de valor de esta coordenada Y en las posiciones 11 en las línea de ángulo de visión B (ver el esquema del entorno de la Figura 4.2). En cada eje horizontal están los diferentes valores que se recogen en milímetros y en el eje vertical su frecuencia durante la toma de datos. Todas corresponden a la posición 11.

Figura 4.8: Histogramas para la coordenada Y de la pupila del ojo derecho de un único usuario. Frecuencia de valor de esta coordenada Y en las posiciones 11 en las línea de ángulo de visión C (ver el esquema del entorno de la Figura 4.2). En cada eje horizontal están los diferentes valores que se recogen en milímetros y en el eje vertical su frecuencia durante la toma de datos. Todas corresponden a la posición 11.

4.3.2. Efecto de pequeñas variaciones en la estimación de la altura (de la pupila) sobre la estimación de la coordenada en altura del apuntamiento

a) Objetivo

Tras el experimento anterior se comprueba que verdaderamente el sensor capta diferentes valores de la coordenada Y (altura) para un usuario inmóvil en una posición. El objetivo de esta prueba es comprobar el impacto de esta inestabilidad sobre la estimación del vector *ojo derecho-mano* y en el cálculo de la proyección del objeto apuntado. Se van a introducir pequeñas variaciones sobre la medida de la pupila. Enlazándolo con los términos de precisión y exactitud, este experimento estudia la precisión del sistema.

b) Preparación y Tareas

La prueba consiste en : **1)** obtener el valor estimado para el origen del vector *ojo derecho-mano*, la pupila del ojo derecho; **2)** estimar de forma manual el punto destino del vector (mano del usuario); **3)** estimar el apuntamiento usando el vector que tiene como origen y destino los puntos extraídos en 1 y 2; **4)** tomando como referencia el punto calculado en 3, estimar el apuntamiento si existe una ligera alteración de la coordenada Y del punto origen.

El usuario se sitúa en una posición prefijada enfrente del dispositivo Kinect (línea B de la Figura 4.2). Se asume igual comportamiento en todos los puntos de referencia. En este caso se opta por la posición B11, la más alejada del dispositivo (4 metros). Con el usuario inmóvil, se estima las coordenadas de la pupila de su ojo. Como se ha comprobado anteriormente, no devuelve un valor único por lo que se calcula el valor promedio. Este valor promedio estimado para la pupila del ojo derecho en la posición B11 tiene las siguientes coordenadas: $(x, y, z) = (2109, 1611, 4390)mm$.

La estimación manual del punto final del vector, mano del usuario, se realizará de la siguiente forma: la misma coordenada X que la de la pupila del ojo derecho; una coordenada Y menor que la de la pupila del ojo derecho porque apunta hacia un punto que está situado por debajo de su cabeza; y una coordenada Z, 70 centímetros menos que la coordenada Z de la pupila (la coordenada Z es la profundidad y la mano está delante de forma alineada con el cuerpo sin ningún tipo de ángulo). Con estas premisas, el valor de la posición del dedo derecho en la posición B11 del entorno y que es el punto destino del vector corresponde a las coordenadas: $(x, y, z) = (2109, 1200, 3690)mm$.

Con estos dos puntos, la proyección del apuntamiento sobre el plano situado en $(x, y, z) = (0, 0, 0)mm$ que corresponde a la pared, es el punto $(x, y) = (2109, 0, -966, 6)mm$. Tomando los datos anteriores como datos de partida, se variará la coordenada Y de modo manual para comprobar su impacto en el punto estimado.

c) Resultados

Los resultados se recogen en la tabla 4.3. En esta tabla, se incrementa la coordenada Y en 1,3,5,10,15,20 y 30 mm. Se aprecia que con 30 mm de variación hay una diferencia de error de 150 mm en la coordenada Y.

d) Conclusiones

Se comprueba que pequeñas variaciones de la coordenada de origen provoca grandes diferencias en la estimación del apuntamiento. Es decir a pequeñas variaciones del apuntamiento el sistema se comporta de un modo impreciso, su precisión disminuye.

Tabla 4.3: Resultado de la proyección variando la coordenada Y del punto origen del vector ojo derecho-mano. Todas las cantidades están en milímetros y el signo indica la dirección.

Valor incrementado al inicial(milímetros)	Origen			Punto proyectado en la pared		Valor absoluto del error con respecto al punto inicial
	x	y	z	x	y	
0 (inicial)	2109	1611	4390	2109	-966.55	0
1	2109	1612	4390	2109	-971.82	5.27
3	2109	1614	4390	2109	-982.37	15.81
5	2109	1616	4390	2109	-992.91	26.35
10	2109	1621	4390	2109	-1.019.27	52.71
15	2109	1626	4390	2109	-1.045.62	79.07
20	2109	1631	4390	2109	-1.071.98	105.42
30	2109	1641	4390	2109	-1.124.70	158.14

4.3.3. Variaciones en la estimación de la altura (de la pupila) en función de la posición del usuario con respecto al sensor

a) Objetivo

En esta sección se estudia la relación existente entre la altura del origen del vector *ojo derecho-mano*, la distancia al dispositivo sensor (coordenadas **Y**, altura, y **Z**, distancia, del punto origen del vector) y el ángulo que forma el dispositivo respecto de la posición del usuario. Además, durante estas pruebas se introducirá la posibilidad de que sea la calibración del sistema un factor que afecte a la estimación, por lo que las pruebas se repetirán para dos calibraciones distintas.

b) Preparación y Tareas

La prueba va a consistir en que un único usuario va a recorrer todos los puntos predefinidos del entorno (ver Figura 4.2), apuntando hacia un único objetivo. En cada uno de los puntos se va recoger la coordenada Y y Z de la pupila del ojo derecho, exclusivamente cuando el usuario esté realizando el gesto de apuntar. Se excluyen las muestras cuando el usuario se encuentre en otras situaciones: moviéndose de una posición a otra (estas muestras no se graban), preparando el gesto de apuntar que es subiendo el brazo o bajando su brazo (esto es posible porque el usuario durante las pruebas solo sube su mano para apuntar y usando un umbral para la coordenada Y de la mano, se sabe que muestras son sólo específicas de apuntamiento). Además se estiman estas dos coordenadas empleando dos calibraciones diferentes: calibración 1 donde el patrón se colocó a una distancia del sensor de 2743 mm (coordenada Z) y la calibración 2 donde el patrón se colocó a la distancia de 4500 mm al

sensor (coordenada Z). En la Sección 4.4.1 se explicará el motivo por el que se emplean estas dos calibraciones.

Para cada calibración, la relación entre las coordenadas Y (altura) y Z (distancia) se analiza ajustando las nubes de puntos en las diferentes posiciones a una recta mediante una regresión lineal. Se opta por la aplicación de una regresión lineal, ya que a primera vista, la nube de puntos analizada parece ajustarse a una recta. La bondad del ajuste se mide evaluando los estadísticos del coeficiente de determinación (r^2) y el error cuadrático medio ($rmse$). El coeficiente de determinación mide la proporción de variabilidad total de la variable dependiente (Y) respecto a su media que es explicada por el modelo de regresión².

c) Resultados

Las gráficas de la Figura 4.9 representan la relación entre la coordenada Y y Z para 2 calibraciones diferentes (de acuerdo a la distancia a la que se colocó el patrón), en cualquiera de las 33 posiciones definidas en el espacio (11 posiciones en cada uno de los 3 ángulos que forma la posición del usuario con el dispositivo sensor, ver plano de la Figura 4.2). La gráfica muestra una curva con el ajuste mediante regresión lineal. Cada nube de puntos corresponde a las muestras tomadas en cada una de las posiciones. La definición de las rectas de ajustes tienen el formato: *Linear Model Poly1: $f(x) = p1 \cdot x + p2$, Coefficients (with 95 % confidence bounds)*. Lo que corresponde a: *altura(coord. Y) = $p1 \cdot distancia(coord. Z) + p2$* . Los valores para la Calibración 1 son $p1 = -0.04035(-0.04079, -0.03992)$ y $p2 = 1795(1794, 1797)$. Mientras que para la Calibración 2 son : $p1 = 0.03602(0.03575, 0.03629)$ y $p2 = 1539(1538, 1540)$.

La tabla 4.4 recoge los dos estimadores empleados para conocer la bondad del ajuste: el error cuadrático medio ($rmse$) y el coeficiente de determinación (r^2). El mejor ajuste se consigue para la línea de posiciones, marcadas como B, cuyo ángulo entre el sensor y el usuario es nulo. El r^2 es de 0.988 siendo un valor muy próximo a 1. En concreto un 98.8 % de la variabilidad de altura a su promedio es explicado por el modelo de regresión obtenido en ese caso. Además el otro estadístico en ese caso $rmse$ tiene un valor muy bajo. En general, el valor de r^2 para casi todos ronda el valor del 0.9, puede decirse existe una relación lineal entre las variables estudiadas.

²Web con explicación del ajuste coeficiente de determinación: <https://la.mathworks.com/help/curvefit/evaluating-goodness-of-fit.html>

Figura 4.9: Gráficas con la relación entre Coordenada Y y Z en la fase de apuntamiento. La primera imagen son los resultados para la Calibración 1 y la segunda imagen con la Calibración 2.

d) Conclusiones

Los resultados gráficos y los datos estadísticos obtenidos muestran una relación lineal entre la coordenada Y (altura) y la coordenada Z (distancia).

Otro factor destacable es que las gráficas dependen de la calibración empleada. A priori, el comportamiento debe ser similar e independiente de la

Tabla 4.4: Ajuste realizado con regresión lineal. Siendo **rsquare** (r-squared, coefficient of determination) y **rmse** (Root mean squared error)

	Calibración 1		Calibración 2	
	rsquare	rmse	rsquare	rmse
línea A	0.859	10.036	0.879	13.004
línea B	0.968	6.786	0.988	3.820
línea C	0.984	5.183	0.967	5.601
Todas las muestras	0.918	10.688	0.935	8.985

calibración empleada pero esto no sucede así. Como es de esperar, la calibración empleada influye en los resultados del apuntamiento. Por ello se dedicará la Sección 4.4 a estudiar esta influencia.

4.3.4. Variaciones en la estimación de la altura (de la pupila) en función de la inclinación del sensor Kinect

a) Objetivo

El objetivo es comprobar si el ángulo de inclinación del dispositivo sensor Kinect influye a la hora de captar las coordenadas Y, altura, de un punto. Esto viene motivado porque el sensor puede situarse en cualquier punto del entorno, por ejemplo, si se coloca en una posición superior de la sala el sensor deberá estar inclinado hacia abajo para poder captar al usuario. Lo contrario sería colocarlo en el suelo y que tuviese que tener una inclinación superior para poder captar al usuario en todo momento. Lo ideal es situarlo en un punto intermedio de altura pero esto puede llegar a ser poco adecuado para ciertos servicios. Esta prueba pretende estudiar este fenómeno y comprobar su influencia en la estimación de Y.

b) Preparación y Tareas

Las pruebas se van a desarrollar colocando el sensor con tres inclinaciones distintas, aquí esta los esquemas para estas tres situaciones:

- **Sin inclinación** : Figura 4.10.
- **Con inclinación superior**: Figura 4.11.
- **Con inclinación inferior**: Figura 4.12.

Figura 4.10: Esquema de Kinect sin inclinación

Figura 4.11: Esquema de Kinect con inclinación superior

Figura 4.12: Esquema de Kinect con inclinación inferior

El proceso real de preparación del sensor puede verse en la Figura 4.13. Con una herramienta niveladora, se ha dado al sensor la inclinación deseada respecto de la superficie donde estaba situado. Para intentar evitar posibles errores ajenos a lo que se está estudiando, el dispositivo se sitúa a la altura de la pupila del ojo del usuario que realizará las pruebas, con ello los datos de altura en crudo obtenidos directamente del sensor deben ser iguales o muy próximos a cero. Hay que tener en cuenta que este proceso es manual y que puede existir una mínima inclinación del dispositivo aunque no es deseable.

Figura 4.13: Preparación del entorno para nivelar kinect.

El usuario se moverá a través de línea en el espacio donde no exista ángulo entre la situación del dispositivo y la posición del usuario, la que queda enfrente del dispositivo, de la posición más cercana a la más alejada del dispositivo, siempre dentro del área de cobertura. Procederá situándose durante 3 segundos en una posición de esa línea y dando un paso hacia atrás, así hasta la posición más alejada cubierta por el dispositivo. En todo momento se recogerán los datos en crudo, sin procesar, obtenidos por el dispositivo sensor durante la prueba.

c) Resultados

Los datos sin tratar son recogidos directamente de dispositivo para cada una de las 3 inclinaciones de Kinect planteadas, con un ajuste mediante regresión lineal realizado, se recogen en las gráficas de las Figuras 4.14, 4.15 y 4.16.

Figura 4.14: Resultado de datos recogidos directamente desde Kinect sin inclinación.

Figura 4.15: Resultado de datos recogidos directamente desde Kinect con inclinación superior.

Figura 4.16: Resultado de datos recogidos directamente desde Kinect con inclinación inferior.

d) Conclusiones

Los resultados muestran :

1. La altura (coordenada Y) estimada varía con la distancia a la que se encuentre el usuario. Varía en todos los casos, sin inclinación y con cualquier inclinación del dispositivo. La variación cuando el dispositivo no tiene inclinación es de unos 2 centímetros en 4 metros de distancia.
2. La inclinación de la recta de la coordenada Y con respecto de Z depende de la calibración empleada. Se profundiza en este aspecto en la Sección 4.4.3.

4.3.5. Variación de la estimación de la altura para otros puntos de referencia: dedo y cabeza

a) Objetivo

Esta prueba persigue comprobar si la estimación de la coordenada de altura del dedo tiene un comportamiento similar que la estimación de la coordenada de altura de la pupila del ojo. Si se reproduce su comportamiento, podría aplicarse la misma corrección en ambos casos.

b) Preparación y Tareas

Las pruebas son desarrolladas por 3 usuarios distintos. La tarea a realizar por cada usuario es similar a la explicada en la Sección 4.3.4. Cada uno de los participantes se moverá a través de línea enfrente del sensor, de la posición más cercana a la más alejada del dispositivo, siempre dentro del área de cobertura del sensor. El usuario se situará durante 3 segundos en una posición de esa línea e irá dando un paso hacia atrás, así hasta la posición más alejada cubierta por el dispositivo. En todo momento se recogerán las muestras obtenidas por el dispositivo referente a la coordenada Y del ojo, dedo o mano y cabeza.

c) Resultados

El comportamiento es similar para los 3 usuarios por lo que reflejaremos en este documento las gráficas de uno de ellos (ver Figuras 4.17 4.18). Las gráficas muestran los resultados para la altura, coordenada Y, en cada uno de los puntos del espacio en los que se colocó el usuario, tanto para los datos en crudo como cuando las coordenadas son transformadas al sistema de coordenadas global. En estas gráficas no se discrimina sólo por estos puntos donde le usuario se detiene, sino que también se aprecian los cambios de posición del participante.

d) Conclusiones

A la vista de los gráficos de las Figuras 4.17 y 4.18, la coordenada Y de la mano tiene el mismo comportamiento que la coordenada Y de la pupila del ojo derecho. Además, el comportamiento se repite para los 2 sistemas de coordenadas considerados. Por tanto, se puede podrá aplicar la misma corrección a ambos puntos.

Figura 4.17: Comportamiento de coordenada y de la mano y pupila ojo derecho para un único usuario. Corresponde a las coordenadas de los datos en crudo recogidos por Kinect.

Figura 4.18: Comportamiento de coordenada y de la mano y pupila ojo derecho para un único usuario. Corresponde a las coordenadas ya transformadas al sistema global de referencia.

4.3.6. Variación de las coordenadas de apuntamiento en función del uso de gafas

a) Objetivo

Esta subsección está centrada en analizar, en el caso de que exista, el impacto en la estimación del apuntamiento de elementos propios del usuario y ajenos al sistema de apuntamiento. En concreto, se va a estudiar a un usuario portador de gafas. El objetivo es descubrir si las gafas distorsionan la estimación del ojo.

b) Preparación y Tareas

La prueba consiste en poner a un usuario durante un tiempo de 10 segundos a apuntar hacia un objetivo fijo en una posición del espacio donde la visibilidad sea buena. Esta prueba la realizará 2 veces, una sin gafas y otra con ellas. La posición escogida es la posición central identificada como B-5 del entorno (ver Figura 4.2). El motivo de no repetir esta prueba en distintas posiciones es porque las pruebas están centradas en conocer las diferencias entre llevar y no llevar gafas, y no en la propia capacidad del sensor para captar las distintas posiciones.

c) Resultados

Con los valores de apuntamiento estimados durante 10 segundos, los datos recogidos no van a reducirse a un único valor y van a existir varios dentro de un mismo rango, por lo que, se calcula el promedio de estos valores. El promedio de la altura calculada se recoge en la tabla 4.5. En lo que respecta al valor medio de la coordenada X parece que no hay diferencia destacable, la diferencia del promedio es de 1.7 milímetros; sin embargo, si existe disparidad en los datos. En cuanto a la coordenada Y, hay una diferencia de 6.9 milímetros entre usar gafas y no usarlas. Por el contrario, no hay tanta diferencia en la desviación estándar de sus datos.

Tabla 4.5: Valores promedios estimados para el apuntamiento con el vector *ojo derecho-mano* en la posición central del espacio (B-5), con el usuario llevando y sin llevar gafas. Todas las medidas son en milímetros

El usuario sí lleva gafas				El usuario no lleva gafas			
Promedio μ de coord. X proyectada	Desviación estándar σ para coord. X	Promedio μ de coord. Y proyectada	Desviación estándar σ para coord. Y	Promedio μ de coord. X proyectada	Desviación estándar σ para coord. X	Promedio μ de coord. Y proyectada	Desviación estándar σ para coord. Y
-24.9	22.9	-42.7	30.9	-26.6	11.9	-35.8	25.2

d) Conclusiones

Si que parece que influye, aunque los errores que se revelan parecen poco relevantes con respecto a los introducidos por otras fuentes.

4.4. Variación del resultado del apuntamiento en función del diseño de la calibración

Por los experimentos y observaciones realizadas parece existir un error introducido por el propio proceso de calibración. El proceso de calibración es imprescindible para el funcionamiento del sistema ya que realiza la transformación entre los diferentes sistemas de coordenadas (Sección 3.3.2.1). Durante esta sección se persigue comprobar si verdaderamente se introduce el error.

Esta sección está orientada a conocer la influencia de la distancia e inclinación del patrón empleada durante el proceso de calibración del sistema sobre el uso del sistema de apuntamiento. Si existe, será necesario proporcionar herramientas para corrección de los posibles errores introducidos. Para ello se han llevado a cabo los siguientes experimentos:

1. Relación entre la posición del patrón de calibración y el resultado del apuntamiento. Estudiado en la Sección 4.4.1
2. Relación entre la inclinación del patrón de calibración y la estimación de las coordenadas en altura. Estudiado en la Sección 4.4.2
3. Relación entre la inclinación del sensor dispositivo Kinect y la inclinación de la calibración . Estudiado en la Sección 4.4.3

Finalmente, con las conclusiones obtenida de los experimentos anteriores se propone una corrección para obtener un proceso de calibración sin errores en la Sección 4.4.4.

4.4.1. Relación entre la posición del patrón de calibración y el resultado del apuntamiento

a) Objetivo

Con este experimento se persigue conocer si existe influencia de la posición donde se coloca el patrón durante el proceso de calibración del sistema en el resultado final del sistema de apuntamiento. El proceso de calibración completo fue explicado en la Sección 3.3.2.1. En la Sección 4.3.3 ya se introdujo el uso de dos calibraciones diferentes diferenciadas por la posición donde se colocaba el patrón.

b) Preparación y Tareas

La prueba realizada consiste en que un usuario recorre todos los puntos definidos del entorno (ver esquema de despliegue en la Figura 4.2). En cada uno de ellos el usuario debe apuntar hacia un único objetivo proyectado en el centro de la pared de proyección. Con los datos recogidos se estima el apuntamiento con el vector *ojo derecho-mano* que este usuario ha realizado. Además, en cada uno de los puntos del espacio se van a estimar dos apuntamientos basados en el uso de dos calibraciones diferentes. Estas calibraciones se diferencian por el punto en el que se colocó el patrón durante el proceso de calibración. Durante ambas calibraciones las coordenadas X y Z donde se situó el patrón fueron las mismas (centrado en el espacio y sobre el suelo, la coordenada del punto origen en milímetros tienen el valor $(X,Z) = (2100,90)$); sin embargo, la coordenada Z (distancia) elegida para la calibración 1 fue Z de 2743 milímetros y con un valor de 4500 milímetros para la calibración 2.

c) Resultados

Con el apuntamiento realizado, se calcula la distancia euclídea entre las coordenadas del punto de apuntamiento estimado y las coordenadas del centro del objetivo en cada de las posiciones donde se colocó el usuario. Los resultados no muestran que exista influencia de la posición del patrón en el resultado del apuntamiento. La Figura 4.19 muestra la gráfica con la distancia euclídea obtenida en la posición B5 (ver plano en Figura 4.2) que corresponde a la posición más céntrica del espacio.

d) Conclusiones

Esto es consistente con los resultados del experimento anterior. No se aprecia diferencia en los resultados obtenidos para ambas calibraciones.

Figura 4.19: Error medido como la distancia euclídea entre las coordenadas del punto estimado y las coordenadas del objetivo en la posición central de la sala. La fase de apuntamiento es únicamente la parte central de la gráfica, el resto corresponde al movimiento del usuario entre posiciones. Imagen superior usando la calibración 1 e imagen inferior usando la calibración 2.

4.4.2. Relación entre la inclinación del patrón de calibración y la estimación de las coordenadas en altura

a) Objetivo

Como se recogía en la sección anterior la distancia del patrón al sensor no parece influir en el resultado del apuntamiento pero si en el ajuste frente a distancia. Esta sección analiza si la inclinación del patrón a la hora de realizar el proceso de calibración del sistema (proceso detallado en la Sección 3.3.2.1) tiene alguna influencia en el resultado final.

b) Preparación y Tareas

Se va a disponer de tres calibraciones distintas de acuerdo a la inclinación con la que se coloca el patrón durante el proceso de calibración. La inclinación del patrón es el ángulo que forma el patrón con respecto del suelo. La posición en la que se sitúa el patrón es la misma para las 3 calibraciones (posición de origen del patrón $(x,y,z)=(2100,90,4500)$ milímetros). En la Figura 4.20

aparece un esquema que explica las diferentes inclinaciones que identifican las 3 calibraciones escogidas:

1. **Calibración 1- C1**, el patrón recto sobre el suelo, 90 grados respecto del suelo.
2. **Calibración 2-C2**, el patrón inclinado hacia delante, respecto de la visión del dispositivo sensor.
3. **Calibración 3-C3**, el patrón inclinado hacia atrás, respecto de la visión del dispositivo sensor.

Figura 4.20: Croquis con la inclinación del patrón en 3 calibraciones diferentes

Durante la prueba, el usuario se desplaza a través de línea imaginaria en el espacio donde no exista ángulo entre la situación del dispositivo y la posición del usuario, enfrente del dispositivo, desde la posición más cercana a la más alejada, y estando siempre dentro del área de cobertura (ver Figura 4.2). El participante de la prueba se situará y apuntará hacia un único punto central de la pared durante 3 segundos y dará un paso atrás, así hasta la posición más alejada cubierta por el dispositivo. En cada una de estas posiciones se recoge la posición del ojo derecho del usuario y se transforma a coordenadas globales aplicando los 3 procesos de calibración identificados anteriormente.

c) Resultados

Con los datos obtenidos se crean las gráficas que relacionan la coordenada en altura con la distancia al dispositivo; es decir, coordenada Y respecto de Z del punto del ojo derecho del usuario. Aquí, como ya sucedió en la Sección 4.3.4 se repite el fenómeno de crecimiento y decrecimiento de la recta resultado del ajuste. Con la Calibración 1 (patrón recto sobre el suelo) y la Calibración 2 (patrón inclinado hacia delante) la coordenada Y decrece al aumentar la distancia. Sin embargo con la Calibración 3 (patrón inclinado hacia atrás) el comportamiento de Y es el contrario, crece con la distancia. La Figura 4.21 representa un esquema del comportamiento comentado.

Figura 4.21: Esquema del comportamiento de crecimiento y decrecimiento con 3 diferentes calibraciones.

El error obtenido es mayor cuando se ha puesto el patrón inclinado hacia atrás (Calibración 3). Inclinando el patrón hacia atrás (Calibración 3) hay casi 70 centímetros en los 3 metros del espacio que se ha movido el usuario, e inclinandolo hacia delante (Calibración 2) hay una diferencia de 20 centímetros solamente. Al situar el patrón sin ningún tipo de inclinación con respecto del suelo el error es únicamente de 7 centímetros, un valor del error inapreciable comprado con el resto.

d) Conclusiones

Este experimento indica que un error en la estimación del apuntamiento asociado directamente a la inclinación del patrón, relacionado con el proceso

de calibración, modulable de manera lineal.

4.4.3. Relación entre la inclinación del sensor dispositivo Kinect y la inclinación de la calibración

a) Objetivo

Las pruebas realizadas en la Sección 4.3.4 están dirigidas a averiguar si el ángulo de inclinación del dispositivo sensor Kinect, influye a la hora de captar la coordenada Y, altura. Sus resultados demostraron que para los mismos datos de entrada y dos procesos de calibración del sistema distintos, la relación entre la coordenada Y y la coordenada Z es contraria. Para una de las calibraciones la coordenada Y, altura, aumenta con la coordenada Z, distancia del usuario al dispositivo. Para la otra calibración la coordenada Y, altura, disminuye con la coordenada Z, distancia del usuario al dispositivo. La inclinación del dispositivo no influye en la relación entre estas dos variables, sino que esto varía dependiendo de la calibración empleada. En resumen, durante las pruebas de esta Sección 4.3.4, se detectó que la calibración influye.

Las pruebas de la Sección 4.4.2 están orientadas a comprobar si la inclinación del patrón empleado en el proceso de calibración influye a la hora de estimar el apuntamiento. En los resultados se vuelve a evidenciar que dependiendo de la calibración empleada, la relación entre la coordenada Y, altura, y coordenada Z, distancia, es totalmente opuesta. Con una de las calibraciones, la coordenada Y aumenta con Z. Para la otra calibración sucede lo contrario, disminuye la coordenada Y al aumentar la coordenada Z. Esto conduce a pensar en que es la calibración escogida la que afecta a este hecho. Además, parece que la inclinación del patrón pueda tener algo que ver. Durante las pruebas comentadas en el anterior párrafo se desconoce cuál es la inclinación real del patrón.

b) Preparación y Tareas

Con las dos premisas anteriores, se va a plantear una prueba para ver cuál es la implicación del dispositivo sensor sobre este fenómeno. Se realiza un conjunto de pruebas con 3 usuarios, donde cada uno de ellos se moverá a través de línea en el espacio donde el ángulo sea cero entre la situación del dispositivo y la posición del usuario, de la posición más cercana a la más alejada del dispositivo, siempre dentro del área de cobertura del dispositivo. Procederá situándose durante 3 segundos en una posición de esa línea y dando un paso hacia atrás, así hasta la posición más alejada cubierta por el

dispositivo.

En este experimento el dispositivo sensor se coloca sin inclinación, como se explica en la Sección 4.3.4. Los 3 usuarios son escogidos de acuerdo a su altura. La pupila del ojo del primer usuario se sitúa a la misma altura que el dispositivo (7 milímetros por encima del punto del sensor), la pupila del segundo usuario se sitúa por encima del dispositivo; por último, la pupila del tercer usuario se sitúa por debajo.

c) Resultados

Se obtienen los datos en crudo que recoge directamente el sensor y se plasman de forma gráfica para apreciar si se repite el comportamiento en las figuras 4.22, 4.23 y 4.24. En las figuras anteriores no se elimina el movimiento del usuario, las partes más oscuras corresponden al momento que el usuario está inmóvil. Sobre estos datos se realiza un ajuste mediante regresión lineal de los puntos. Como se puede ver, cuando el punto a captar está por debajo el ajuste lineal. Existe una dependencia negativa entre la variables Y (altura) y Z (distancia); y en los otros casos, es positiva. A la vista de esta prueba ya el propio sensor introduce este fenómeno sin intervenir ningún tipo de calibración.

Figura 4.22: Resultados modificando la inclinación de Kinect mostrando datos en crudo recogidos del sensor. Altura del punto a captar por debajo de Kinect(colocado totalmente horizontal).

Figura 4.23: Resultados modificando la inclinación de Kinect mostrando datos en crudo recogidos del sensor. Altura del punto a captar similar a Kinect (7 mm por encima, colocado totalmente horizontal).

Figura 4.24: Resultados modificando la inclinación de Kinect mostrando datos en crudo recogidos del sensor. Altura del punto a captar por encima de Kinect, totalmente horizontal.

d) Conclusiones

Por tanto, con los estudios realizados se plantea que este comportamiento se debe a la combinación de las siguientes variables: la orientación que considera la calibración por la inclinación del patrón y la verdadera orientación del sensor. El planteamiento es el de la Figura 4.25. Cuando la orientación considerada por la calibración es superior a la real del sensor, la coordenada Y (altura) decrece al aumentar Z (distancia). Sin embargo, cuando la orientación considerada por la calibración es inferior a la real del sensor, la coordenada Y (altura) crece al aumentar Z (distancia).

Figura 4.25: Explicación gráfica del fenómeno del crecimiento y decrecimiento

4.4.4. Propuesta de corrección de los efectos del diseño de la calibración. Análisis de la universalidad de la solución.

La inclinación del patrón durante el proceso de calibración introduce un error en la estimación del apuntamiento. Es necesario conseguir que el patrón esté perpendicular al suelo en el momento de la calibración del sistema. Como esto a veces de forma manual es difícil conseguirlo y no siempre es posible, puede ser que sea necesario manipular los datos obtenidos del proceso de calibración para conseguir estimar una coordenada Y de altura correcta. Se va a considerar una calibración del sistema aceptable a la que proporcione un valor de altura igual a una distancia de 2 metros desde la pared (o dispositivo) y a una distancia de 4 metros de la pared (o dispositivo). Siempre va a tener una pequeña variación, pero esta debe ser mínima, por ejemplo unos 2 centímetros de diferencia en la altura se puede considerar un valor aceptable, dado que anteriormente se manejaban hasta 10 centímetros de diferencia.

Elección de la mejor calibración

La primera opción a valorar es disponer de varias calibraciones y comprobar si alguna proporciona unos resultados de estimación aceptables y suficientes.

Para este experimento se van a generar 5 calibraciones distintas intentando colocar el patrón de forma manual lo más recto posible del suelo, formando con este 90° . De antemano se conoce que esto es difícil y que seguramente nos aproximemos al objetivo de los 90° sin conseguirlo.

El usuario realizará la prueba que viene siendo habitual en secciones previas. El usuario se moverá a través de línea imaginaria en el espacio donde no exista ángulo entre la situación del dispositivo y la posición del usuario, de la posición más cercana a la más alejada, y estando siempre dentro del área de cobertura. Durante la misma no es necesario que el usuario realice el gesto de apuntamiento, simplemente se parará durante 3 segundos y mirará hacia la pared de proyección. A continuación irá alejándose paso a paso, realizando paradas de 3 segundos cada vez, hasta que se salga fuera del área de cobertura del sensor.

Los datos recogidos son la coordenada Y del ojo derecho del usuario, en crudo; es decir, sin realizar ningún tratamiento. Para cada una de las calibraciones, 5 en total, obtendremos una gráfica de la coordenada Y respecto de la coordenada Z que es la distancia al dispositivo. La Figura 4.26 muestra la gráfica con la calibración que proporciona el mejor resultado, con un error mínimo estimado en altura, entre el punto más cercano y el más lejano, de unos 7 centímetros.

Aún realizando varias intentos de calibración, no se consiguen los resultados de apuntamiento planteados inicialmente como aceptables. Por lo que se considera difícil obtener una buena calibración empleando un método manual.

Figura 4.26: Resultado de transformar los datos en crudo al sistema de coordenadas global. En la gráfica aparece la coordenada Y con respecto Z en un recorrido del usuario de lo más cercano a sensor hacia lo más lejos. En este caso es la calibración que muestra mejores resultados de las 5 usadas.

Corregir automáticamente los errores del proceso de calibración

La colocación del patrón sin ningún tipo de inclinación de forma manual parece un método impreciso. No es posible asegurar que el patrón durante el proceso de calibración del sistema se sitúe a 90° respecto al suelo. Como solución se propone manipular la matriz de rotación empleada en el proceso de calibración del sistema (ver Sección 3.3.2). El algoritmo planteado es el siguiente:

1. Partimos del resultado de una calibración cualquiera.
2. Se repite la prueba del apartado anterior (Sección 4.4.4).
3. Con los datos anteriores, se comprueba la relación entre la altura (coordenada Y) y la distancia al dispositivo (coordenada Z) aplicando el proceso de transformación a coordenadas global donde participa la calibración del punto 1.

4. Se calcula la diferencia en la coordenada Y, la coordenada Z, y el ángulo

formado $theta$.

5. Se corrige la calibración para variar la rotación de forma mínima. Sólo es necesario realizar una pequeña rotación sobre uno de los ejes, en este caso el eje X.

$$theta = \arcoTangente(diferenciaencoord.Y, diferenciaencoord.Z) \quad (\text{Ecuación 1})$$

$$R_{comp} = \begin{Bmatrix} 1 & 0 & 0 \\ 0 & \cos(theta) & \sin(theta) \\ 0 & -\sin(theta) & \cos(theta) \end{Bmatrix} \quad (\text{Ecuación 2})$$

$$Rc_{ext} = R_{comp} * Rc_{ext_0} \quad (\text{Ecuación 3})$$

Se aplicará a continuación la corrección sobre la calibración anterior que proporcionaba mejores resultados. En este caso el ángulo $theta$ se calcula como: $theta = \arcoTangente(67, 3122)$. La Figura 4.27 muestra los datos de la coordenada Y con respecto de Z con la corrección planteada (la Figura 4.26 contiene la gráfica sin incluir mejora). Si la corrección en una distancia de unos 4 metros hay una diferencia de unos 7 centímetros en la estimación de la altura. Sin embargo, aplicando la corrección el error de estimación disminuye a 2 centímetros. Con este resultado ya puede considerarse una calibración aceptable.

Figura 4.27: Datos de la coordenada Y corrección de la calibración.

Extrapolar solución particular de un usuario a cualquier usuario

Veamos si este ángulo calculado previamente con los datos de un usuario puede servir para cualquier calibración y así cumplir con el principio de universalidad. Es decir si un único ángulo calculado una única vez y aplicado durante la configuración inicial del sistema puede extenderse a cualquier usuario.

La prueba del apartado anterior va a ser realizada por dos nuevos usuarios. En este punto, tenemos 3 usuarios en total, el usuario 1 es el que realizó la prueba inicialmente y con el que se obtuvo los datos el ángulo corrector. Los usuarios 2 y 3 son los que realizarán estas pruebas. Todos los usuarios tienen alturas diferentes. A los datos de estos dos últimos usuarios se les aplicará la corrección del usuario 1. Los resultados gráficos se recogen en la Figura 4.28 para el usuario 2 y en la Figura 4.29 para el usuario 3. Los puntos más oscuros de las gráficas corresponden a los puntos donde se paró el usuario, el resto corresponde a cuando estaba en movimiento.

A la vista de las gráficas la corrección parece funcionar y se podría extrapolar de un usuario concreto a un usuario cualquiera. En ambos casos las diferencias de estimaciones de altura en distancia de 4 metros se reducen a

unos 2 centímetros.

Figura 4.28: Relación de coordenada altura con respecto distancia con la corrección genérica comprobada para usuario 2 (diferente usuario del que se usó para calcular el ángulo corrector). Gráfica superior sin corrección y gráfica inferior es el resultado de aplicar la corrección a la calibración usada.

Figura 4.29: Relación de coordenada altura con respecto distancia con la corrección genérica comprobada para usuario 3 (diferente usuario del que se usó para calcular el ángulo corrector). Gráfica superior sin corrección y gráfica inferior es el resultado de aplicar la corrección a la calibración usada.

4.5. Propuesta de un sistema de apuntamiento mejorado

Una vez estudiadas las diferentes alternativas de vectores para implementar un sistema de apuntamiento, los problemas encontrados durante su implementación, se propone en esta Sección un sistema de apuntamiento de señalización manual mejorado.

Esta propuesta incluye el uso de dos vectores para estimar el apuntamiento mediante señalización manual. Por una parte, y en el caso de que haya visibilidad de la cara del usuario se empleará el vector *ojo dominante-mano* para apuntar. En el caso de no disponer de esos datos por falta de visibilidad de la cara, se usará el vector *cabeza-mano*. La combinación de estos dos vectores amplía la posibilidad y la cobertura del sistema. Si alguno de estos puntos no es visible simplemente el sistema no podrá detectar el apuntamiento.

La Figura 4.30 corresponde al esquema del sistema propuesto. Las entradas al sistema de estimación de apuntamiento son: 1) Los datos adquiridos desde el sensor en crudo; 2) El proceso de calibración o **calibración del sistema**, al que se aplica la corrección del sistema de la Sección 4.4.4 para ajustar los valores de la coordenada Y, altura. Estas dos entradas se dirigen al módulo de estimación del apuntamiento, donde se realiza la transformación de coordenadas. Se estiman los dos vectores : vector *ojo dominante-mano* o vector *cabeza-mano* y, posteriormente, dependiendo de la visibilidad del sensor se opta por emplear una de estas dos opciones. La salida del sistema corresponde a las coordenadas estimadas para el punto apuntado por el usuario. El punto estimado puede ser modificado si se le aplica la calibración de usuario.

La **calibración de usuario** ayuda a corregir el error que introduce el sistema durante el apuntamiento del usuario. Esta corrección es única para cada usuario. La fase de calibración consiste en que el usuario apunta sobre un objetivo proyectado en la pared (que tiene unas coordenadas x e y concretas). Cuando el usuario cree que está apuntando a este objetivo proyectado, se guardan las verdaderas coordenadas (x,y) sobre la pared a las que está apuntando. Con esto tenemos dos puntos de la pared, el *punto usuario* al que la persona ha apuntado sobre la pared ($P_{usuario} = x_{usuario}, y_{usuario}$) y el punto objetivo ($P_{objetivo} = x_{objetivo}, y_{objetivo}$) donde se sitúa el objetivo proyectado en la pared. Esto se repite con 4 objetivos diferentes. Usando cada uno de los objetivos obtenemos una diferencia de distancias en las dos coordenadas, tanto en x como en y ($Diferencia_x = x_{usuario} - x_{objetivo}$; $Diferencia_y =$

$y_{usuario} - y_{objetivo}$);). Por tanto al final del proceso tenemos 4 diferencias en X y cuatro diferencias en Y. Para cada coordenada se calcula la media de las diferencias y se obtiene un valor promedio. Con ello se tiene un promedio de diferencias en X y otro en Y. Estos dos valores de promedio son los que se emplean para corregir el apuntamiento. Es decir, durante las pruebas reales y fuera de esta fase de calibración, sobre las coordenadas del punto obtenido del apuntamiento realizado por el usuario, se suma o se resta, según proceda, en una de las dos coordenadas (X o Y) para obtener las nuevas coordenadas corregidas.

Figura 4.30: Propuesta del proceso del sistema de apuntamiento

4.6. Resumen

El Capítulo está enfocado en proponer un sistema de apuntamiento usando señalización manual *mejorado* respecto de los resultados el anterior Capítulo 3. La propuesta del sistema comienza esbozando los requisitos que se consideran importantes para obtener un sistema de apuntamiento adecuado. Los requisitos son: exactitud, ubicuidad, universalidad y facilidad de despliegue. Además, se detalla como es el esquema del entorno donde se va a desplegar el sistema de apuntamiento.

Hay una sección dedicada a elegir el vector de apuntamiento que se empleará en el sistema de apuntamiento. Para ello, se escogen seis vectores diferentes, candidatos a ser usados en el sistema de apuntamiento. Tras la realización de unas pruebas para ver su idoneidad, se elige el vector que tiene como origen el ojo del usuario y como destino la mano.

Al inicio del estudio se intuye que la estimación del vector apuntamiento se ve afectada por una serie de factores. Para ello, a lo largo del capítulo se realiza una serie de experimentos que nos ayuda a confirmar estas intuiciones y nos sirve de partida para el diseño.

El primer bloque de experimentos se agrupan en los aspectos referentes a tener una repercusión sobre la estimación de la coordenada en altura de los puntos que definen el vector de apuntamiento. Para ello se realizan una serie de experimento para comprobar: i) la estabilidad de la estimación de la altura de la pupila; ii) el efecto de las pequeñas variaciones en la estimación de la altura de la pupila sobre la estimación de la coordenada en altura del apuntamiento estimado; iii) las variaciones en la estimación de la altura de la pupila en función de la posición del usuario con respecto al sensor; iv) las variaciones en la estimación de la altura de la pupila en función de la inclinación del sensor Kinect; v) las variaciones de la estimación de la altura de los otros puntos de referencia del vector de apuntamiento en función de la localización del usuario; por último, vi) la variación de las coordenadas de apuntamiento en función del uso de gafas por parte del usuario.

El segundo bloque está orientado a descubrir las variaciones del apuntamiento estimado en función del diseño de la calibración realizado. Para ello, se analizan a través de experimentos las siguientes tres relaciones: i) la relación entre la posición del patrón de calibración y el resultado del apuntamiento; ii) la relación entre la inclinación del patrón de calibración y la estimación de las coordenadas en altura; y iii) la relación entre la inclinación del sensor Kinect y la inclinación de la calibración escogida. Con todo lo es-

tudiado se propone una corrección en el diseño de la calibración y se estudia la universalidad de dicha solución.

La última sección propone un sistema de apuntamiento *mejorado* que corrige los errores detectados y su influencia sobre la estimación del apuntamiento. El Capítulo siguiente pretende realizar un estudio para evaluar el sistema de apuntamiento *mejorado* que se propone.

Capítulo 5

Estudio de evaluación del sistema de apuntamiento propuesto

5.1. Objetivos del Capítulo

El objetivo del Capítulo es evaluar el sistema con las mejoras planteadas en el Capítulo 4 a través de una serie de pruebas. La finalidad de las pruebas es medir el rendimiento y precisión del apuntamiento implementado en el sistema en un entorno realista, para realizar tareas sobre contenidos proyectados (apuntamiento 2D) e interactuar con objetos (apuntamiento 3D) y comprobar la viabilidad de usarlo en servicios reales. El apuntamiento es estimado usando la posición del ojo dominante como punto origen y el vector formado por esa posición y la posición de la mano (dedo índice). En el caso de no poder captar este punto origen (porque en ese momento la cara no sea visible por el dispositivo) se toma como origen la posición del centro de la cabeza, y el vector empleado tendrá como origen la posición de la cabeza y como destino, el dedo índice de la mano derecha. Las tareas a realizar en las pruebas se explican en la Sección 5.2.4 siguiendo ideas similares a las desarrolladas en la Sección 3.2.1 dedicada a la Ley de Fitts y Norma ISO 9241-9.

Para conseguir un mejor rendimiento y precisión del sistema, es posible realizar una fase previa de personalización de la calibración (para cada usuario) para mejorar la estimación del apuntamiento realizado por el sistema.

5.2. Diseño del estudio de rendimiento, precisión del sistema y experiencia de usuario

5.2.1. Objetivos

El estudio de usuario está orientado a:

1. Obtener evidencia del rendimiento en cuanto al tiempo de ejecución.
2. Obtener evidencia de la precisión referido al número de aciertos.
3. Conocer los efectos de usar una calibración personalizada por usuario.
4. Estudiar el confort del usuario usando el sistema.
5. Usabilidad del sistema.
6. Analizar la experiencia de usuario.

5.2.2. Participantes

El estudio de usuario se desarrolló en el *Laboratorio de Experiencias de Espacios del Futuro*, localizado en el Centro de Apoyo a la Innovación del Campus de Montegancedo perteneciente a la Universidad Politécnica de Madrid. En dicho estudio participaron 8 personas, 7 hombres y 1 mujer, con edades comprendidas entre los 21 y 45 años voluntarias entre el propio personal del Centro; en la elección de los participantes no se ha tenido en cuenta su experiencia anterior con estos sistemas ni su ocupación. Tras la entrevista previa realizada a estos participantes antes de las pruebas, se extrae que todos ellos habían usado el dispositivo Kinect, y varias de ellos se identifican como usuarios experimentados con el dispositivo.

La decisión de emplear 8 usuarios busca encontrar el equilibrio de obtener resultados fiables con la eficiencia en el desarrollo de las pruebas (ya que éstas tiene una duración no breve). De acuerdo al estudio planteado por Nielsen [72], los principales problemas del sistema en cuanto a usabilidad se consiguen con un número no superior a 5 usuarios y la ejecución de tantas pruebas pequeñas como sea posible. Con el primer usuario se aprende casi un tercio de todo lo que hay que saber sobre la usabilidad del diseño. Al realizar las pruebas el segundo usuario, se descubre que este usuario repite los mismos errores que el primer usuario, por lo que se superpone o que

aprende, pero habrá algo nuevo. Así que el segundo usuario agrega cierta cantidad de información nueva, pero no tanto como lo hizo el primer usuario. El tercer usuario repetirá errores pero generará una pequeña cantidad de datos nuevos, aunque no tanto como el primero y el segundo usuario. A medida que se agregan usuarios, aprende cada vez menos porque seguirá viendo las mismas interacciones una y otra vez. Después del quinto usuario, se pierde recursos en repetir lo que ya se ha observado pero sin aprender nuevas cosas. Aun así, en esta Tesis se han empleado 8 personas ya que sus características personales permiten cubrir diversa variabilidad inter-usuarios, como por ejemplo, diferentes alturas, empleo de gafas, etc.

5.2.3. Entorno e Infraestructura

Entorno

Las pruebas se desarrollan en dos espacios diferentes: 1) espacio de las pruebas para comprobar el rendimiento y precisión del sistema de apuntamiento; y, 2) el espacio de las pruebas de selección mediante apuntamiento.

El primero de ellos, espacio de las pruebas para comprobar el rendimiento y precisión del sistema, dispone de unas posiciones predefinidas, donde debe situarse el usuario (Figura 5.1) dentro del entorno real que es el que aparece en la imagen de la Figura 5.2.a.

Figura 5.1: Plano-situación del lugar donde se llevaron a cabo las pruebas

El segundo espacio, donde se han desarrollado las pruebas de selección mediante apuntamiento, es un entorno con 9 pantallas unidas formando un videowall. El espacio de movimiento del usuario en este entorno corresponde al área de cobertura del dispositivo Kinect, sin tener posiciones predefinidas. Una imagen de este lugar es el recogido en la Figura 5.2.b.

Figura 5.2: a) Entorno real de las pruebas de rendimiento y precisión del sistema. b) Entorno real de las pruebas de selección mediante apuntamiento

Infraestructura

La Figura 5.3 representa la arquitectura cliente-servidor del sistema empleando un diagrama UML. El componente servidor maneja el sensor Kinect y contiene la lógica que estima el apuntamiento. Por otra parte, el componente cliente contiene la aplicación que utiliza el apuntamiento para mostrar la interacción en el espacio proyectado o interactuar con los objetos del entorno. Además la aplicación va a ser la encargada de relacionarse con los dispositivos del entorno, por ejemplo, si la aplicación consiste en proyectar, este cliente se relacionará con el dispositivo proyector físico de contenidos.

Figura 5.3: Diagrama de la infraestructura del sistema planteado.

5.2.4. Procedimiento y tareas

Los participantes del estudio de usuario realizaron las mismas tareas empleando el apuntamiento mediante señalización manual. La realización de las pruebas y la toma de opiniones fueron autorizadas por los usuarios. En el Anexo C se encuentra el modelo de autorización que ha sido firmado por cada uno de los participantes de las pruebas. El autor de esta Tesis explicó a los usuarios los objetivos de las pruebas que iban a realizar; siempre recalcando que éste no era analizar la habilidad del usuario sino el rendimiento, confort del sistema y su experiencia con el mismo. Además, se les indicó a los usuarios que mostrasen sus sensaciones durante las pruebas para que pudiesen ser grabadas con el fin de usarlo si era necesario. Cada uno de los participantes invirtió aproximadamente 2 horas en completar las pruebas.

Figura 5.4: Diagrama del flujo de trabajo del usuario. Sombreado en amarillo, las pruebas de rendimiento y precisión del sistema de apuntamiento en el entorno de proyección de la tarea ISO 9241-9; y en azul, las pruebas de selección mediante apuntamiento en el entorno del videowall.

El flujo de trabajo del usuario se muestra en la Figura 5.4. Como muestra este flujo, tras firmar el consentimiento y recibir las instrucciones; se obtiene cual es el *ojo dominante* del participante. El ojo dominante de una persona es aquel que tiene mayor agudeza visual. En la Sección 4.1 se explicó este término y el proceso para detectarlo. Conociendo el ojo dominante se configura el sistema para saber qué posición (ojo izquierdo u ojo derecho) debe ser tomada como origen del vector.

La siguiente acción es el *entrenamiento* del usuario en el uso del sistema. Para ello, se realizan diferentes pruebas para que el usuario se acostumbre a apuntar sobre objetos virtuales proyectados. Los círculos virtuales aparecen de forma aleatoria en unas posiciones predefinidas en la pared similar a la tarea de la Norma ISO 9241-9. A continuación se realiza la *calibración del usuario* personalizada y propia del mismo. Los resultados aplicando la calibración del usuario, explicado en la Sección 4.5, se almacenarán aunque estos datos podrán ser aplicados o no en la estimación final del apuntamiento. Para poder comprobar si esto tiene algún efecto, se realizará la comparativa aplicando la calibración personalizada por usuario o sin aplicarla.

En este momento las pruebas se dividen en las siguientes dos fases:

1. Las pruebas de rendimiento y precisión del sistema de apuntamiento (Sección 5.2.4.1).
2. Las pruebas de selección mediante apuntamiento (Sección 5.2.4.2).

Estas dos fases serán realizadas por todos los usuarios en este orden.

5.2.4.1. Pruebas de rendimiento y precisión del sistema de apuntamiento

Cada usuario va a realizar las tareas de seleccionar círculos virtuales proyectados sobre una pared siguiendo las directrices que marca la Norma ISO 9241-9 (Sección 3.2.1. El esquema de pruebas es el que se detalla en el flujo de trabajo 5.4. Aplicando la calibración personalizada por usuario, el usuario se situará en cada una de las 5 posiciones definidas en 5.1 y en cada una de ellas debe realizar la tarea de seleccionar 25 círculos proyectados sobre la pared de proyección. En cada ronda, los círculos a seleccionar tendrán 3 diferentes tamaños de radio (80,55 y 30 píxeles que equivalen a 15, 10 y 5 centímetros cuando están proyectados sobre la pared). El espacio de entre

círculos; y por tanto, la dimensión de los mismos aparece en detalle en la tabla 5.1. Se repetirá estas pruebas sin aplicar la calibración personalizada por usuario.

Tabla 5.1: Espacio entre bolas

Radio(píxeles)	Radio(centímetros)	Diámetro (2*radio)(centímetros)
80	15	30
55	10	20
30	5	10

Hasta ahora se ha realizado todas las tareas de selección sin mostrar en la proyección todo el *grid* de posibles posiciones donde se pueden proyectar los círculos como se aprecia en la Figura 5.5.a. Ahora, se vuelve a repetir la prueba pero únicamente en la posición central (Figura 5.1 posición *B-central*), con el sistema calibrado de forma personalizada y sin personalizar, mostrando el grid con las posiciones de proyección posibles (Figura 5.5.b) para los tres tamaños comentados y 25 círculos proyectados por cada tamaño.

Figura 5.5: Imagen selección de objetivos sin mostrar todo el grid (a) y mostrando todo el grid (b)

Durante la tarea de apuntamiento en cada posición los objetos virtuales irán apareciendo de forma aleatoria en diferentes posiciones posibles de proyección, que están predefinidas, de la pared. El usuario debe seleccionar los objetos apuntando con su brazo hacia ellos. Para seleccionar dicho círculo se dispondrá de un tiempo máximo (configurado el tiempo de activación de las bolas en 3 segundos). Si durante ese tiempo no ha conseguido seleccionar el objeto, este desaparecerá de esa posición y aparecerá en una nueva posición hasta el número máximo de objetivos, configurado en 25.

Con la realización de estas pruebas se registran los valores de tiempo y el número de aciertos, usados posteriormente para valorar el rendimiento y precisión del sistema de apuntamiento.

5.2.4.2. Pruebas de selección mediante apuntamiento

Las pruebas de selección se realizan sobre un videowall. Un videowall no es más que un conjunto de pantallas distribuidas en una superficie, colocada una junto a la otra. En este caso está formado por 9 pantallas colocadas en forma de matriz 3x3, unidas una a la otra, de forma que las 9 pantallas trabajan como si fuese una única. Un esquema de esta distribución aparece superpuesto sobre el entorno real del videowall en la Figura 5.6). El apuntamiento se utilizará para activar cada una de las pantallas de la matriz 3x3 de esta Figura. Este videowall combinado con el apuntamiento puede servir, por ejemplo, para determinar la interacción con contenidos o recursos específicos en un centro de control.

Figura 5.6: Distribución de pantallas en el videowall

En este caso, se ha mantenido la distribución de las pantallas y se trabaja con cada una de ellas de forma individual. De esta forma, se ha considerado el tamaño de pantalla como el mínimo de granularidad para la división del videowall. El usuario puede apuntar sobre cada una de ellas para interactuar con la misma. La selección de la pantalla se realiza después de un tiempo mínimo de apuntamiento sobre ella. La primera selección de alguna de estas pantallas implica que, sobre la pantalla, aparecerá una imagen predefinida. Las siguientes interacciones permiten mover dicha imagen al resto de pantallas. Es decir, el usuario selecciona la pantalla a la que quiere trasladar la imagen.

En este caso, el usuario tiene libertad para moverse por el espacio de cobertura del dispositivo Kinect (que es el que captura las posiciones para estimar el apuntamiento), y debe seleccionar al menos una vez cada una de las pantallas.

El orden de los métodos de selección que deberá usar es el siguiente:

1. Gesto con la mano izquierda (gesto apertura de mano).
2. Orden mediante voz.
3. Método dwell-time (mantener un periodo de tiempo el apuntamiento sobre el objetivo).

Posteriormente, el usuario evaluará el confort, la experiencia de usuario y la usabilidad del sistema. La Figura 5.7 recoge imágenes captadas durante la realización de estas pruebas.

Figura 5.7: Momentos durante las pruebas de los métodos de selección sobre las pantallas

5.2.5. Diseño del experimento

El experimento diseñado consta de una serie de tareas que cada participante realizó y que han sido explicadas en la Sección anterior 5.2.4.

El experimento, al igual que las tareas planteadas, ha sido diseñado en dos partes:

1. Experimento dedicado a obtener el rendimiento y precisión del sistema de apuntamiento (Sección 5.2.5.1).

2. Experimento para comprobar el apuntamiento sobre tareas reales de selección (Sección 5.2.5.2).

5.2.5.1. Experimento dedicado a obtener el rendimiento y precisión del sistema de apuntamiento

Durante las *pruebas de rendimiento y precisión del sistema de apuntamiento* los participantes del experimento deben apuntar sobre diferentes círculos virtuales que aparecen proyectadas en la pared. Se producen un total de 30 iteraciones (5x3x2) diferentes que es de la combinación de, a) 5 posiciones prefijadas, ver Figura 5.1 con el plano de las posiciones predefinidas; b) 3 tamaños de círculo proyectados diferentes; c) 2 posibilidades de calibración, aplicando la personalización de la misma o sin aplicarla. Estas 30 iteraciones primeras se realizan sin que el usuario visualice el grid completo con las diferentes posibilidades donde puede aparecer el círculo (ver Figura 5.5.a).

Cada usuario repite la tarea un total de 6 iteraciones (1x3x2) de nuevo ya que, a) el usuario se coloca una única vez en la posición central del plano de las posiciones predefinidas (posición central-B de la Figura 5.1); b) 3 tamaños de círculo diferentes; c) 2 posibilidades de calibración, aplicando la personalización de la misma o sin aplicarla. Estas 6 iteraciones se realizan con el usuario visualizando el grid completo con las diferentes posibilidades donde puede aparecer el círculo sobre la pared de proyección(ver Figura 5.5.b).

En total, se producen 36 iteraciones con las condiciones ya explicadas anteriormente. Al finalizar cada prueba se cuentan los aciertos, que son las veces que el usuario ha seleccionado dicho círculo virtual. En caso contrario, se cuenta un fallo.

Las variables que se estudiarán durante el experimento se clasifican como:

- Variables independientes
 - Tamaño de objetivo.
 - Posición del usuario.
 - Calibración personalizada por usuario.
 - Experiencia previa del usuario.

- Variables dependientes
 - Tiempo de ejecución.
 - Número de aciertos.

Estas pruebas se realizarán empleando los datos de corrección obtenidos en la fase de calibración. Para comprobar que la calibración tiene efecto se realiza la iteración anterior una vez por cada tamaño sin emplear los datos de la fase de calibración.

Podría incorporarse un *feedback* visual para que el usuario supiese dónde está apuntando realmente. Se entiende que esto condicionaría al usuario y dificultaría por ese motivo no se ha empleado. Además llevaría al usuario a mover la bola a través de la pared, y no se conseguiría la acción natural de apuntar que es el objetivo. También se puede filtrar el apuntamiento para suavizar y evitar bruscos movimientos pero esto incorpora un retardo que para este objetivo empeora las pruebas.

5.2.5.2. Experimento para comprobar el apuntamiento sobre tareas reales de selección

Posteriormente se prueba el apuntamiento con una aplicación final que es consiste en la interacción del usuario con el videowall. Además, se probarán 3 diferentes métodos de selección. Esto se evalúa tras rellenar el cuestionario del Anexo B.

El orden de los métodos de selección a probar es el siguiente:

1. Gesto con la mano izquierda (gesto apertura de mano).
2. Orden mediante voz.
3. Método dwell-time (mantener un periodo de tiempo el apuntamiento sobre el objetivo).

Después de realizar las pruebas, el usuario rellenará un cuestionario para evaluar el confort, la experiencia de usuario y la usabilidad del sistema similar a los empleados en el Capítulo 3. Además responderá a unas preguntas de forma libre para conocer su opinión acerca de sus preferencias entre los métodos de selección empleados y sobre el apuntamiento realizado.

Las variables que se estudiarán durante el experimento se clasifican como:

- Variables independientes
 - Posición del usuario.
 - Calibración personalizada por usuario.
 - Experiencia previa de usuario
- Variables dependientes
 - Métodos de selección
 - Confort del sistema de apuntamiento en entornos reales de ejecución
 - Usabilidad.
 - Experiencia de usuario

5.3. Resultados

En esta Sección se van a analizar el rendimiento del sistema y su precisión; usando para ello, el tiempo de ejecución y el número de aciertos respectivamente. Estas variables serán comparadas en diferentes circunstancias para conocer si existe o no una relación entre ellas. Posteriormente se mide el confort percibido por el usuario al emplear el sistema. Finalmente, se estudia la usabilidad y experiencia de usuario con este sistema.

5.3.1. Análisis de rendimiento y precisión

En esta sección se va a analizar el rendimiento en cuanto al tiempo de ejecución de las pruebas y la precisión referida al número de aciertos. El tiempo de ejecución corresponde al tiempo que usuario invierte en realizar la prueba completa. Por otra parte, la calibración del sistema para cada usuario se detalló en la Sección 4.5, y es una de las variables dependientes incluidas en este estudio.

5.3.1.1. Rendimiento : Tiempo de ejecución

La primera variable a analizar es el tiempo de ejecución que es considerado el tiempo medio en finalizar la prueba de cada uno de los usuarios.

El promedio de tiempos de todos los usuarios para finalizar las pruebas con calibración ha sido de 35433,11 milisegundos (con un promedio de desviación típica, σ , de 8360,76) y para el sistema sin calibrar de 36944,62 milisegundos (con un promedio de desviación típica, σ , de 10057,09). La diferencia entre estos dos promedios es mínima (es de 1 segundo aproximadamente en global). Esta diferencia de tiempos entre usar el sistema sin calibrar y calibrado es pequeño aunque son mejores los resultados para el sistema calibrado. A partir de ahora podemos fijarnos en los valores obtenidos al usar el sistema con corrección de la calibración.

Ahora se plantearán una serie de hipótesis y mediante métodos estadísticos para obtener conclusiones al respecto. De acuerdo al *boxplot* de la Figura 5.8, se aprecia que las distribuciones son bastante sesgadas; por lo tanto, no se puede asumir normalidad y al no haber muchos datos se opta por test de hipótesis no paramétricos.

Figura 5.8: *Boxplot* con el tiempo para completar las pruebas para cada uno de los usuarios.

5.3.1.1.1. Relación entre el tiempo de ejecución y la calibración personalizada del sistema

Este planteamiento tiene las siguientes hipótesis:

- Hipótesis nula H_0 : No existe diferencia entre el tiempo medio para terminar las pruebas, empleando y sin emplear calibración personalizada.
- Hipótesis alternativa H_1 : Existe diferencia entre el tiempo medio para terminar las pruebas, con el sistema calibrado y sin calibrar.

Se aplica un test de signos (*sign test*) sólo asintótica a todos los datos de todas las pruebas realizadas. El valor de probabilidad (p-valor) obtenido es de 0.312 siendo superior al valor de significación del 0.05, por lo tanto no se puede rechazar la hipótesis nula.

Se repite el test, únicamente para los resultados cuando el objetivo a seleccionar era el de 30 píxeles (el más pequeño). Al igual que ocurre con el caso anterior no se puede rechazar la hipótesis nula, ya que el p-valor es 0.08 siendo superior al valor de significación del 0.05 y a un nivel de confianza del 95 %.

5.3.1.1.2. Relación entre el tiempo de ejecución y la experiencia previa del usuario con Kinect

Este caso plantea si la experiencia previa del usuario con Kinect del usuario influye en el tiempo para terminar las pruebas, por lo que se plantean las siguientes hipótesis:

- Hipótesis nula H_0 : No influye la experiencia previa del usuario con Kinect en el tiempo medio para terminar las pruebas.
- Hipótesis alternativa H_1 : Sí influye la experiencia previa del usuario con Kinect en el tiempo medio para terminar las pruebas.

Se emplea un test de Kruskal Wallis para ver si influye la experiencia previa del usuario con Kinect para realizar la prueba. En este caso, el p-valor obtenido es muy pequeño y por lo tanto menor que el valor de significación α . Esto lleva a rechazar la hipótesis nula.

5.3.1.1.3. Relación entre el tiempo de ejecución y el tamaño del objetivo a seleccionar

Este análisis comprueba si el tamaño del blanco influye en el tiempo empleado por un usuario para finalizar la prueba. Las hipótesis planteadas son:

- Hipótesis nula H_0 : El tiempo necesario para completar la prueba es independiente del tamaño del blanco a seleccionar.
- Hipótesis alternativa H_1 : El tiempo necesario para completar la prueba depende del tamaño del blanco a seleccionar.

Para hacer este análisis se emplea un test de Friedman. En este caso, al igual que en el anterior, el p-valor es muy pequeño y por lo tanto menor que el valor de significación, α , de 0.05. Esto lleva a rechazar la hipótesis nula, la cuál decía que el tiempo para completar la prueba es independiente del tamaño del blanco a seleccionar.

En el caso de los tiempos de terminar las pruebas para el sistema calibrado, los resultados demuestran que es más fácil seleccionar objetivos de tamaño mayor como era de esperar. El patrón de menor tiempo de finalización de las pruebas a medida que aumenta el tamaño del objetivo se repite en 7 de los 8 participantes de las pruebas. Éste es el comportamiento esperado antes de realizar las pruebas. Solamente con uno de los participantes este patrón no se cumple. Para este participante los tiempos obtenidos son: 37.30 segundos para objetivos de 30 píxeles, 28.70 segundos para objetivos de 55 píxeles, 31.02 segundos para objetivos de 80 píxeles. Este fenómeno puede ser debido a un hecho puntual sucedido durante las pruebas ya que tampoco existe una gran diferencia. Otro dato a destacar es la desviación típica, σ , de los tiempos de finalización de las pruebas. Para los datos tomados en tamaños pequeños (30 píxeles), el valor es muy grande comparado a los obtenidos para los objetivos de tamaño mayor (80 píxeles). Esto puede estar debido a la dificultad a la hora de seleccionar un objetivo y que al cambiar entre círculos 2D objetivos, la separación entre ellas es mayor para tamaños pequeños.

5.3.1.2. Precisión del sistema : Número de aciertos

Hasta ahora se ha usado como referencia el tiempo de ejecución de las pruebas para comprobar el rendimiento del sistema, a partir de ahora también se va a poner el foco en la precisión del mismo. La precisión del sistema se mide como el número de aciertos promedios en cada una de las pruebas. Se considera un acierto cuando el usuario consigue seleccionar el blanco-objetivo (haciendo que la intersección del vector de apuntamiento con el plano de proyección caigan en la superficie acotada del mismo).

En el caso de sistemas calibrados, el número medio de aciertos es de 22.47 de un máximo de 25 con una desviación típica, σ de 3.45. Cuando se usa el sistema sin calibrar es de 21.7 con una σ de 4.03. Aunque la diferencia es mínima, al igual que ocurre con el tiempo es mejor emplear el sistema calibrado que el no calibrado.

Al igual que se ha realizado con la variable del tiempo en el punto anterior, en este apartado nos centramos en la variable número de aciertos, *hits*,

llevados a cabo por los usuarios.

5.3.1.2.1. Relación entre el número de aciertos y la calibración personalizada por usuario del sistema

Este planteamiento tiene las siguientes hipótesis:

- Hipótesis nula H_0 : La calibración personalizada del sistema no introduce mejoras en la precisión de sistema basada en el número de aciertos conseguidos por los usuarios.
- Hipótesis alternativa H_1 : Sí existe diferencia entre el número de aciertos conseguidos por los usuarios, con el sistema calibrado de modo personalizado y sin calibrar de modo personalizado.

Para contrastar la igualdad de medias de muestras relacionadas se aplica un test de signos a todos los datos de todas las pruebas realizadas. El p-valor obtenido es de 0.30 siendo superior al valor de significación del 0.05, por lo tanto, no se puede rechazar la hipótesis nula.

Se repite el test, únicamente para los resultados cuando el objetivo a seleccionar era el de 30 píxeles (el más pequeño). En este caso sí se rechaza la hipótesis nula, ya que el p-valor obtenido es de 0.049, siendo menor que el valor de significación que es de 0.05.

5.3.1.2.2. Relación entre el número de aciertos y la experiencia previa del usuario

En este caso se plantea que la experiencia previa del usuario con Kinect influye en número de aciertos que consigue en la realización de las pruebas, las hipótesis planteadas son las siguientes:

- Hipótesis nula H_0 : No influye la experiencia previa del usuario con Kinect en el número de aciertos conseguidos por los mismos (precisión).
- Hipótesis alternativa H_1 : Si influye la experiencia previa del usuario con Kinect en el número de aciertos (precisión).

Se emplea un test de Kruskal Wallis para ver si influye la experiencia previa del usuario a la hora de conseguir seleccionar los blancos. El valor obtenido de p con la ejecución del test es de 0.076, siendo superior al valor de significación de 0.05. Por lo tanto, no se puede rechazar la hipótesis nula.

5.3.1.2.3. Relación entre el número de aciertos y tamaño del objetivo a seleccionar

Por último, se comprueba si el tamaño del blanco a seleccionar tiene relación con el número de aciertos que consiguen los usuarios. Las hipótesis planteadas son:

- Hipótesis nula H_0 : El número de aciertos es independiente del tamaño del blanco a seleccionar.
- Hipótesis alternativa H_1 : El número de aciertos depende del tamaño del blanco a seleccionar.

Para hacer este análisis se usa un test de Friedman. En este caso, el valor obtenido de significación estadística p es muy pequeño y por lo tanto menor que el valor de confianza fijado a 0.05. Dado este resultado se rechaza la hipótesis nula y se puede decir que el número de aciertos conseguido parece influenciado por el tamaño del blanco a seleccionar.

Respecto al número de aciertos, el patrón citado cuando se evaluaban los tiempos de finalización, se mantiene. Es decir, a medida que aumenta el tamaño del objetivo, la tasa de aciertos también aumenta. Como en el caso de los tiempos, para uno de los usuarios esto no sucede y la tasa de aciertos en el tamaño mayor es más pequeña que en el tamaño medio (22.67 frente a 24.5 de un máximo posible de 25 aciertos). Aquí sin embargo la desviación típica, σ , de la tasa de aciertos no sigue un comportamiento concreto respecto al tamaño del objetivo.

5.3.2. Análisis de confort

Una vez realizadas las pruebas, el usuario responde a un cuestionario sobre la comodidad a la hora de usar el sistema. Estas cuestiones se toman del Anexo C de ISO 9241-9, tabla C1 (escala de evaluación independiente) filtrando alguna de las 12 cuestiones que no aplican en nuestro caso. Se emplea la escala de Likert, entre 1 y 7, en cada uno de las cuestiones estos valores extremos tiene un significado diferente (ver tabla 5.2). Los resultados se recogen en la tabla 5.2. Entre las 8 habilidades a evaluar, las tres que obtienen una valoración más próxima a 7 son la precisión del sistema (valoración de 5 con una desviación típica, σ , 1.19), la velocidad de empleo (valoración de 5.28 con una desviación de los datos de 1.60) y la fatiga provocada en el cuello durante el uso del sistema (valoración de 6.62 con una desviación de

los datos de 0.74, es decir, que se produce poca fatiga sobre el mismo y es bastante confortable en ese sentido). Las dos habilidades menos valoradas, que indican un buen comportamiento, son: cansancio en el brazo al usar el sistema (valoración de 2.12 con una desviación de los datos de 0.64), la suavidad durante la interacción (valoración de 2.62 con una desviación de los datos de 1.30) . Por último, las siguientes 3 habilidades han conseguido valores promedios rondando el centro de la escala: fuerza requerida durante la interacción (valoración de 3.75 con una desviación de los datos de 1.75), esfuerzo (valoración de 4.12 con una desviación de los datos de 1.35), comodidad general (valoración de 4.62 con una desviación de los datos de 1.06).

Tabla 5.2: Tabla análisis de confort

	Interpretación valor mínimo de escala(1)	Interpretación valor máximo de escala(7)	Valor me- dio	Desviación estándar
Fuerza de actua- ción requerida	Muy incómo- do	Muy cómodo	3,75	1,75
Suavidad de empleo	Muy duro	Muy suave	2,62	1,30
Esfuerzos requeri- dos para la interac- ción	Muy alto	Muy bajo	4,12	1,35
Exactitud	Muy inexacto	Muy exacto	5	1,19
Velocidad de em- pleo	Nada acepta- ble	Muy Acepta- ble	5,28	1,60
Comodidad general	Muy incómo- do	Muy cómodo	4,62	1,06
Fatiga del cuello	Muy alta	Ninguna	6,62	0,74
Fatiga del brazo	Muy alta	Ninguna	2,12	0,64

5.3.3. Análisis de usabilidad

Al igual que en el caso del confort, la usabilidad también se analiza tras la realización de las pruebas. Para llevar a cabo esta evaluación se ha empleado la *escala de usabilidad del sistema* que aparece explicado en [7]. Los participantes calificaron 10 aserciones con una escala de 1 (completamente en desacuerdo) a 5 (muy de acuerdo). Los valores medios calculados con las respuestas de los diferentes usuarios obtenidos para cada ítem se muestran en la tabla 5.3. Siguiendo las indicaciones del estudio plasmadas en [7] se calcula la puntuación de la escala de usabilidad (Standard Usability Scale,

SUS), donde se obtendrán valores entre 0 y 100. El valor obtenido de todos los usuarios participantes para la usabilidad es de 81.88 ($\sigma = 9,04$ puntos), siendo este valor muy alto próximo al valor máximo.

Tabla 5.3: SUS valores medios

		Valor medio μ	Desviación estándar σ
1	I think that I would like to use this system frequently	3,88	0,64
2	I found the system unnecessarily complex	1,13	0,35
3	I thought the system was easy to use	4,38	0,52
4	I think that I would need the support of a technical person to be able to use this system	1,38	0,52
5	I found the various functions in this system were well integrated	4,00	0,76
6	I thought there was too much inconsistency in this system	1,75	0,89
7	I would imagine that most people would learn to use this system very quickly	4,13	0,35
8	I found the system very cumbersome to use	1,88	0,64
9	I felt very confident using the system	3,75	0,89
10	I needed to learn a lot of things before I could get going with this system	1,25	0,46

5.3.4. Análisis de la experiencia de usuario

La experiencia de usuario es medida empleando el cuestionario propuesto en [8]. El estudio para la evaluación mantiene 6 escalas diferentes que evalúan el *atractivo* (impresión general del sistema), *familiaridad* (facilidad para familiarizarse), *eficiencia* (idoneidad para resolver las tareas del usuario), *fiabilidad* (sensación de control), *estimulación* (excitación y motivación) y *novedad* (innovación) del prototipo diseñado, planteando 26 preguntas. La Tabla 5.4 muestra los valores medios obtenidos para cada uno de estos puntos. De forma, gráfica estos mismos resultados se recogen en la Figura 5.9.

Empleando las herramientas que ofrece el sitio web de *UEQ*¹, los valores medios para cada escala se establecen en relación con los valores existentes

¹UEQ website: <http://www.ueq-online.org/>

de un conjunto de datos de referencia. Este conjunto de datos contiene de 4818 personas recogidos de 163 estudios sobre diferentes productos (software empresarial, páginas web, tiendas en la red, redes sociales). La comparación de los resultados del producto evaluado con los datos del punto de referencia permite obtener conclusiones sobre la calidad relativa del producto evaluado en comparación con otros productos. Esto también aparece reflejado en las dos últimas columnas de la Tabla 5.4 y Figura 5.9. Este análisis con esos datos de referencia hay que tomarlos con cuidado ya que los análisis no están evaluando el mismo producto pero sirve de guión o apoyo a la interpretación de resultados. A la vista de la comparativa planteada, los resultados de este estudio en las 6 escalas estudiadas se encuentran por encima de la media.

En este punto lo ideal es compararlo con los sistemas estudiados en el Capítulo 3. En la Sección 3.6.2.2 se encuentra la tabla con la experiencia de usuario analizada para los dos sistemas anteriores. La valoración del sistema mejorado propuesto es superior para todas las escalas estudiadas.

Tabla 5.4: Valores del cuestionario para cada escala en el análisis de la experiencia de usuario

Scale	Mean	Std. Dev.	Comparación al benchmark	Interpretacion del benchmark
Attractiveness	1,792	0,954	Excellent	In the range of the 10 % best results
Perspicuity	2,219	0,891	Excellent	In the range of the 10 % best results
Efficiency	1,781	0,986	Excellent	In the range of the 10 % best results
Dependability	1,125	0,655	Above Average	25 % of results better, 50 % of results worse
Stimulation	1,625	1,069	Good	In the range of the 10 % best results
Novelty	1,781	1,081	Excellent	In the range of the 10 % best results

Figura 5.9: Gráfico para cada escala con el análisis de la experiencia de usuario

5.4. Resumen

Tras la propuesta de un sistema de apuntamiento mejorado realizada en el Capítulo 4, en este Capítulo 5 se desarrolla un estudio para evaluar el anterior sistema. El diseño del estudio persigue obtener el rendimiento, precisión y la experiencia de usuario del sistema de apuntamiento con su uso en un entorno real, y valorar el uso de métodos de selección por parte de los participantes de estudio. La estructura del diseño es similar a la planteada en el Capítulo 3. Se comienza con una descripción de los participantes que van a realizar el experimento. A continuación se explica la infraestructura usada y el entorno sobre el que se desarrolla este experimento. Una sección engloba los procedimientos y tareas que deberán ser realizados por los participantes de las pruebas. Por último, se detalla como ha sido diseñado el experimento. El experimento conlleva el uso del sistema enmarcado en un servicio final real como es el manejo de nueve pantallas formando un *videowall*.

La sección de resultados analiza el rendimiento y precisión del sistema, el confort, la usabilidad y la experiencia de usuario con el mismo. Las dos variables empleadas durante el análisis de la precisión y el rendimiento han sido el tiempo de ejecución para completar las pruebas y el número de aciertos obtenidos en ellas. Estas variables se han combinado a la hora de obtener resultados con situaciones como la calibración del usuario personalizada, la experiencia previa del participante con el dispositivo sensor y el tamaño de los objetivos.

Capítulo 6

Conclusiones y trabajo futuro

En la Tesis se analizan y proponen sistemas que permiten la interacción de los usuarios mediante apuntamiento dentro de espacios inteligentes. La interacción se detecta y modela a partir de una infraestructura sensorizada que permite al usuario no llevar ningún sensor. El dispositivo escogido en esta Tesis es Microsoft Kinect en su versión 2.0 que integra cámara RGB, sensor de profundidad (consiste en un proyector láser infrarrojo combinado con un sensor monocromático que captura videos 3D) y micrófono. Este dispositivo tiene la ventaja de tener un bajo coste y gracias a él, se puede realizar un despliegue dotando a un espacio de interactividad. Actualmente esta versión de Kinect ya no se encuentra a la venta, sin embargo, existe una nueva evolución de este dispositivo de la misma empresa, Microsoft, e integrado con su plataforma Azure. Este nuevo dispositivo recibe el nombre de *Azure Kinect unit* con su kit de desarrollo *Azure Kinect DK* ¹. Este combina un sensor de profundidad avanzado con una matriz de micrófonos espaciales, una cámara de vídeo, un sensor de orientación y SDK.

El sistema de apuntamiento cobra sentido cuando se integra con una aplicación, como por ejemplo de control remoto de dispositivos (luces, altavoces, etc). Se muestran prototipos para el manejo de imágenes, control de robots, control de objetos inteligentes y juegos. En esta Tesis se han propuesto varias aplicaciones, que ha sido evaluadas por usuarios, para obtener una evidencia de experiencia y usabilidad. Además, las pruebas con usuarios han permitido valorar la eficiencia, el rendimiento, la precisión y exactitud.

Este Capítulo final recoge las conclusiones obtenidas y revisa algunos

¹Documentación de Azure kinect KD: <https://docs.microsoft.com/es-es/azure/Kinect-dk/>

problemas de investigación para avances futuros.

6.1. Conclusiones

El apuntamiento, como elemento de interacción natural o de gestión de recursos digitales (por ejemplo selección y activación de contenidos), es un factor diferencial a la hora de desarrollar servicios dentro de espacios inteligentes (Capítulo 2). En la literatura se han explorado las tecnologías y sistemas que habilitan las técnicas de apuntamiento; así cómo, las metodologías de evaluación a la hora de desarrollar sistemas basados en apuntamiento. Destaca el apuntamiento mediante señalización manual como la forma más intuitiva de interacción con objetos y con una mejor precisión y exactitud.

El catálogo de dispositivos comerciales o experimentales que integran cámaras, sensores de profundidad y otros sensores necesarios para detectar posición y orientación de individuos, para que estos no tengan que estar instrumentados, es amplio como se muestra en la revisión realizada en el Estado del Arte del Capítulo 2. El dispositivo Kinect es una alternativa habitual en los estudios que consideran el despliegue de servicios inteligentes. Por ejemplo, en la clasificación recogida en la tabla 2.1 del Capítulo 2, un 67 % de los estudios emplean este dispositivo. Kinect incorpora una cámara de color, un sensor de profundidad de infrarrojos y un micrófono. Entre las ventajas que presenta está su bajo coste, su capacidad para captar las posiciones y los movimientos de uno o varios usuarios (hasta un máximo de 6) y la poca intrusión que supone para el usuario. El dispositivo se puede situar en algún punto del entorno sin interferir con el flujo de personas. Uno de los principales inconvenientes como con cualquier cámara es que hay que calibrar el dispositivo dentro del entorno, éste es un proceso largo, que necesita rehacerse si la posición del sensor varía. Una vez calibrado, el dispositivo queda configurado para ese entorno concreto. En este sentido el objetivo de *fácil despliegue* no llega a cubrirse en su totalidad. Sin embargo, aún con estas limitaciones, Kinect es un perfecto dispositivo para poder aumentar de forma inteligente un espacio y permitir la interacción de las personas con el entorno que les rodea.

Otro aspecto relevante dentro del apuntamiento es el método de selección de los objetivos. Las dos técnicas de selección más utilizadas de los estudios revisados en el Capítulo 2 son: *primero en entrar* (*first-entry*) y el *tiempo de permanencia* apuntando sobre un objeto (*dwell-time*).

La evaluación de los sistemas de apuntamiento se realiza con diferentes métodos. En este ámbito, destaca el uso de la Norma ISO 9241:2000 (*UNE-EN ISO 9241-9: Requisitos ergonómicos para trabajos de oficina con pantallas de visualización de datos (PDV). Parte 9: Requisitos para dispositivos de entrada diferentes al teclado.*) que describe las tareas a realizar por los usuarios durante los experimentos. Por su parte la Ley de Fitts permite evaluar y cuantificar los resultados obtenidos durante los experimentos; así como, los estudios de usabilidad empleados, escala de valoración de estos estudios, etc. Por tanto, la aplicación de la Ley de Fitts nos permite evaluar el rendimiento del usuario al usar sistemas de apuntamiento mientras los datos son recogidos de la realización de las tareas propuestas en el Estándar ISO 9241-9.

El catálogo de los vectores de estimación del blanco apuntado presentes en la literatura incluye diversas alternativas como: la unión del centro de la cabeza con la mano dominante, la unión del centro de la cabeza con el de la cara, el vector formado por uno de los ojos y el dedo índice de una de las manos, el formado por el hombro y el dedo índice de la mano o el que nace de la unión de la nariz con la mano usada para apuntar.

Una de las técnicas a destacar es la que asimila el objetivo al que mira el usuario con el objetivo apuntado es el seguimiento de la mirada. Una forma de simular el seguimiento de la mirada, sin necesidad de *trackear* continuamente la pupila del ojo es emplear la posición y orientación de la cabeza. Desde los años 90, existen dispositivos muy precisos que permiten realizar el seguimiento de los ojos, sin embargo, estos son muy intrusivos para los usuarios y sensibles a la posición donde se sitúan. Con el planteamiento previo no se consiguen resultados óptimos pero si se llega a una exactitud suficiente para implementar servicios interactivos con objetivos. Algunos estudios añaden detección de gestos para mejorar la expresividad y ganar conseguir una ganancia en exactitud.

A continuación se recogen las conclusiones comparativas de las técnicas de detección del apuntamiento basadas en posición-orientación de la cabeza y señalización manual. A partir de los resultados obtenidos, en esta Tesis se ha propuesto un sistema que mejore los problemas detectados y se ha realizado un estudio sobre el mismo para analizar los resultados.

6.1.1. Resultados de la comparativa de técnicas de detección del apuntamiento basadas en posición-orientación de la cabeza y señalización manual

Los resultados del estudio presentado en el Capítulo 3 señalan que los dos tipos de apuntamiento propuestos (señalización manual y simulación del seguimiento de la mirada) consiguen un rendimiento similar, a pesar de que en primera instancia el apuntamiento mediante señalización manual se consideró más eficiente. Aún así, los participantes apuntaron que este último sistema es una alternativa con mayor robustez. Sin embargo, la alternativa preferida por ellos es el apuntamiento a través de simulación del seguimiento de la mirada. Durante la realización de este estudio también se han planteado otras cuestiones como la dependencia del rendimiento respecto del sistema con el tamaño del objetivo a apuntar, la capacidad de aprendizaje, la dificultad percibida o la influencia de disponer de una referencia visual.

En términos de *efectividad*, ambas técnicas de apuntamiento han obtenido un resultado similar, considerando la efectividad como el tiempo de movimiento invertido en completar la tarea ISO mutidireccional de apuntamiento. La aplicación de la Ley de Fitts confirma este hecho, ya que ambas técnicas obtienen un valor de rendimiento (TP, *throughput*) similar, con un valor de 2.45 para la señalización manual y 2.3 para la simulación del seguimiento de la mirada. El empleo de la técnica de *primero en entrar* (*first-entry*) para seleccionar los objetivos permite comparar el tiempo de movimiento entre ellos de forma estricta (sin introducir ningún tiempo adicional, como por ejemplo, el que se introduciría empleando un movimiento de selección específico). Los valores de rendimiento obtenidos permiten compararlos con los alcanzados usando otros dispositivos de interacción, por ejemplo, *touchpads* o *joysticks*.

Otro de los factores estudiados es la localización del usuario en el espacio, principalmente la posición respecto del dispositivo, en este caso Kinect versión 2.0. El resultado pone en evidencia la dificultad de detección a distancias lejanas del dispositivo (mayores de 4 metros), aunque el sistema trabaja de forma estable para ambas técnicas de apuntamiento empleadas.

Además se ha estudiado la influencia del *tamaño del objetivo* en la eficiencia de la interacción, como era de esperar ésta se reduce cuando los objetivos son de tamaño pequeño.

Otro de los aspectos que influyen en la eficiencia es el *ángulo* en el que se localizan los objetivos a señalar ya que condiciona la visibilidad que tiene el dispositivo Kinect del usuario; por ejemplo, es difícil estimar la dirección-

orientación de la cara cuando el usuario mira hacia posiciones inferiores. Por lo tanto, esta evaluación en 3 dimensiones (3D) es útil a la hora de diseñar un espacio inteligente de apuntamiento con objetos inteligentes, entendiendo estos como volúmenes en 3D que se pueden seleccionar.

El ajuste obtenido con la aplicación de la Ley de Fitts muestra un ajuste bajo, el cuál concuerda con previos estudios similares de señalización manual. Por lo tanto, se ve que el modelo usado puede ser mejorado para ajustar de forma más adecuada el comportamiento de los sistemas de interacción que emplean el dispositivo Kinect. En esta línea, podría tenerse en cuenta el ángulo que se forma entre el usuario y el objetivo para la formulación del índice de dificultad.

Por otra parte, los usuarios apreciaron una mejora su habilidad a la hora de apuntar con el transcurso de las pruebas. Este hecho fue más evidente cuando los objetivos eran de tamaño pequeño, suponiendo una reducción del tiempo invertido para finalizar una tarea. Cuando el tamaño es mayor, el tiempo se mantiene estable; mientras que, para tamaños de objetivos menores, el tiempo disminuye. En la práctica, esto se traslada a que si el tamaño objetivo es suficientemente grande, el usuario utilizará el método de interacción con un máximo nivel de eficiencia incluso desde su primer uso. Obviamente, objetivos muy grandes pueden ser incompatibles con determinadas aplicaciones

Respecto a la comodidad al usar el sistema, la preferencia de acuerdo a las respuestas de los usuarios es la que está basada en señalización manual. Aún siendo la preferida, la opinión general es que requiere de una mayor fuerza para actuar con el sistema que emplear el seguimiento de simulación de la mirada.

El método preferido para apuntar a objetos en espacios inteligentes, de forma habitual, fue el apuntamiento mediante señalización manual por ser más fácil de usar y parecer más fiable que el de simulación de seguimiento de la mirada. Sin embargo, el apuntamiento usando esta última técnica responde con mayor rapidez; por lo que tiene mayor efectividad. Al analizar las preferencias de los participantes, el seguimiento de la mirada se perfila como el preferido para un 65 % de los usuarios, mientras que sólo el 25 % elige el apuntamiento mediante señalización manual. La explicación para esto podría ser, una vez más, el diseño del estudio: tras un tiempo prologando realizando apuntamiento con la mano, los usuarios pasa a probar un método de mínimo esfuerzo. Esto puede conducir a buscar alternativas para la interacción y reducir el esfuerzo necesario, por ejemplo, combinando mas de una tecnología (la combinación de apuntar y voz) aunque hay que tener en cuenta que la

fatiga sólo es perceptible si el sistema se utiliza intensamente

El método de activación del blanco (por ejemplo el modo de que se dispare una acción sobre el objeto apuntado) fue *dwell time*. Éste consiste en apuntar durante un periodo de tiempo sobre el objeto a seleccionar. Este periodo de tiempo no debe ser muy grande, ya que podría resultar cansado y lento para el usuario que espera inmediatez, por lo que, debe ajustarse tanto como sea posible, de acuerdo al servicio que se está desarrollando. Una alternativa puede ser utilizar la voz para realizar las ordenes sobre los objetos planteados de tal manera que no se exija tiempo de permanencia. Esto fue propuesto a nuestros participantes y inicialmente aprobado por el 65 % de ellos, aunque sería necesario probarlo de forma real para conseguir una validación correcta.

La referencia visual que se le proporciona a los usuarios durante la interacción en el espacio se perfila como un aspecto relevante para el diseño de este tipo de sistemas. Por este motivo, es necesario evaluar este aspecto y ver las implicaciones de la referencia visual que se ha usado durante las pruebas (en forma de punto rojo). Según el análisis realizado, el uso de esta referencia visual facilita la interacción sustancialmente, en términos de errores, tiempo en completar las tareas y dificultad percibida. La diferencia de estas medidas es menor cuando el objetivo a apuntar es de tamaño grande, por lo tanto, la influencia del objeto volumen-tamaño del objetivo es un aspecto que también se debe tener en cuenta a la hora de diseñar el espacio de interacción. Sin embargo, esta referencia visual no es viable para muchos servicios.

6.1.2. Diseño de un sistema *mejorado* de apuntamiento basado en señalización manual

Una vez realizado el análisis de los sistemas de apuntamiento que emplean técnicas de detección del apuntamiento usando la posición-orientación de la cabeza del usuario y señalización manual, se ha propuesto un sistema de apuntamiento para solventar o mitigar los problemas encontrados durante el análisis previo.

Los requisitos considerados en el diseño de un sistema de apuntamiento son: exactitud, ubicuidad, universalidad y facilidad de despliegue. Los resultados y conclusiones obtenidos para cada uno de estos requisitos con este el sistema *mejorado* propuesto se recogen en el cuadro 6.1.

La detección del apuntamiento se realiza usando un vector de apuntamiento. En esta Tesis se han considerado siete alternativas en total, que se

Tabla 6.1: Tabla resumen con el cumplimiento de los requisitos.

Requisito	Verificación
Exactitud	Se consigue unos resultados de exactitud para la coordenada X de unos 5 cm de error y para la coordenada Y (altura) de 6 cm
Ubicuidad	Se comprueba un correcto funcionamiento del sistema en todo el área de despliegue. Aunque la mejor respuesta del sistema se obtiene cuando el usuario se encuentra en el centro del área de cobertura.
Universalidad	El sistema ha funcionado para los 8 usuarios de las pruebas. Además la corrección que se plantea en la Sección 4.4.4 permite que una única solución sirva para cualquier usuario.
Facilidad de despliegue	El despliegue del sistema es fácil, necesario buscar una posición para colocar el dispositivo Kinect y que se consiga la mejor visión posible. El proceso más tedioso está en realizar el proceso de calibración del sistema. Durante este proceso, habrá que colocar el patrón sin ninguna inclinación. Como se ha comprobado, a priori, esto es muy complicado y se proporcionan las herramientas para poderlo realizar de forma automática.

han estudiado en la Sección 4.2 del Capítulo 4. La conclusión a la que se ha llegado para obtener un rendimiento aceptable y un sistema preciso es emplear el vector que tiene como origen la pupila del ojo dominante y como destino la mano.

Una vez elegido este vector de apuntamiento se estudian qué factores pueden afectar a la estimación del apuntamiento derivando en un mal funcionamiento del sistema y, por tanto, no proporcionando los resultados esperados. Para ello, en el Capítulo 4 se han estudiado las posibles fuentes de error que originan errores en la estimación final del apuntamiento.

El primer análisis realizado comprueba que existe una errónea estimación de la coordenada en altura (coordenada Y). Relacionado con este hecho se ha buscado la relación entre la altura y la distancia al dispositivo para poder modelar el error. Tanto de un modo gráfico, como con los indicadores estadísticos adecuados, se ha modelado la relación de estas dos variables con una recta de regresión. Entre las fuentes de error que provocan una mala estimación de la coordenada Y se encuentra la forma de realizar el proceso de calibración. En este sentido, la inclinación del dispositivo diferencia el

resultado de la calibración y por tanto también la estimación de la coordenada Y. Además, dependiendo de la inclinación y la calibración obtenida, los resultados también se ven afectados. El mismo comportamiento para la coordenada Y del ojo sucede para la posición del dedo.

La relación entre la altura del punto a detectar y la distancia de este al dispositivo se puede modelar mediante una regresión lineal de estas dos variables.

El proceso de calibración del sistema es otra de las fuentes de error detectadas y estudiadas. Tanto la posición donde se coloca el patrón, como la inclinación del mismo son factores determinantes para obtener una buena precisión del apuntamiento. Comprobado que el proceso acarrea estos errores, se propone una metodología para mitigarlos. El empleo de esta metodología reduce el error a 2 centímetros en la altura con 4 metros de distancia, cuando antes el error era de 10 centímetros para la misma distancia. Este error, por lo comprobado en los experimentos realizados, tiene influencia sobre el apuntamiento estimado por el sistema.

El diseño propuesto incorpora la corrección de la coordenada Y de altura. Además, se añade la característica de emplear el vector *cabeza-mano* como alternativa al vector *ojo dominante-mano* cuando el sensor no disponga de la visibilidad de la cara. Inicialmente es posible configurar el ojo que se debe usar de acuerdo a las características del usuario.

6.1.3. Resultados del estudio de evaluación del sistema de apuntamiento *mejorado* propuesto

El estudio de usuario planteado en el Capítulo 5 para probar el sistema *mejorado* tiene una serie de limitaciones. En la primera parte de las pruebas donde se estudia el rendimiento del sistema, el participante no ha tenido la libertad de movimiento para colocarse en todo el espacio, sino que se han escogido una sola por orientación con un total de 5 posibles. Las pruebas se han desarrollado en el mismo orden para todos los usuarios, no se ha variado su ordenación de un usuario a otro.

Respecto al tiempo empleado por los usuarios para finalizar las pruebas, tiempo ejecución, los resultados obtenidos no concluyen que exista una diferencia entre el sistema calibrado de forma personalizada al usuario como sin esta calibración. Sin embargo, en lo que respecta a la relación entre el tiempo empleado y la experiencia previa del usuario con este dispositivo, si

se aprecia una influencia entre ambas variables. A mayor experiencia previa del usuario, menor tiempo para completar las pruebas. Otra de las variables que afecta al tiempo empleado por el usuario en finalizar las pruebas es el tamaño del objetivo que debe seleccionar en cada momento. En este sentido, el tiempo crece al decrecer el tamaño del objetivo.

Las mismas pruebas analizando el tiempo se han llevado a cabo para medir la precisión empleando el número de aciertos conseguido por los usuarios. En este caso, el resultado del experimento no muestra una relación entre el número de aciertos y el uso de una calibración personalizada por usuario. Tampoco se puede argumentar nada relativo a la relación del número de aciertos y la experiencia previa del usuario. Sin embargo, si se aprecia una relación directa entre el número de aciertos y el tamaño de los blancos a seleccionar durante el desarrollo de las pruebas.

En cuanto al análisis del confort, las pruebas usan el cuestionario planteado en el Anexo C del estándar ISO 9241-9. De las 9 habilidades analizadas, las tres que se posicionan en términos medios son: fuerza requerida, el esfuerzo realizado, y la comodidad en general. Las que obtienen una mejor valoración (tabla 5.2 del Capítulo 5) son: la precisión del sistema y la velocidad de empleo, como aspecto negativo destaca la fatiga que se produce en el cuello durante la interacción. Menor importancia tiene el cansancio provocado en el brazo, la suavidad durante la interacción y el rendimiento percibido durante la interacción.

De acuerdo a la usabilidad del sistema de apuntamiento, los usuarios valoran de forma muy positiva la usabilidad del mismo. El valor obtenido es de 82/100 puntos.

La experiencia de usuario se evalúa usando seis escalas diferentes. De estas escalas, cuatro de ellas (*atractivo, familiaridad, eficiencia y novedad*) consiguen un resultado excelente. La valoración de la escala *estimulación* tiene un resultado bueno. Por último, la escala con peor valoración, *fiabilidad*, se mantiene por encima de la media del banco de datos empleado de base.

6.1.4. Recomendaciones de diseño de sistemas de interacción mediante apuntamiento

El desarrollo de esta Tesis permite plantear una serie de recomendaciones a la hora de diseñar sistemas de interacción mediante apuntamiento:

1. Es necesario disponer de un dispositivo sensor que habilite la interac-

ción detectando apuntamiento. De este dispositivo se espera que no sea intrusivo para el usuario; por tanto, es preferible que se sitúe en el entorno y no tenga que llevarlo consigo el propio usuario. Como es lógico deberá disponer de cámaras, sensores de profundidad u otros que sean capaces de detectar la posición de los individuos y su orientación. Además la posición donde se sitúe debe ser adecuada para tener la mejor visibilidad del entorno.

2. Es preferible usar el apuntamiento mediante señalización manual frente al apuntamiento mediante simulación del seguimiento de la mirada; siendo este más intuitivo y permitiendo una mayor precisión y exactitud con la interacción con los objetos. Del estudio realizado se extrae que el primer modo de apuntamiento citado (señalización manual) es el método más fácil de usar y fiable. Sin embargo, el segundo modo de apuntamiento (simulación mediante seguimiento de la mirada) tiene una mayor rapidez de respuesta. Además destacar que la primera opción tiene una gran desventaja y es que presenta una mayor fatiga en el uso frente a la segunda opción.
3. En relación con el ítem anterior se encuentra el vector de estimación del blanco a emplear relacionado con el apuntamiento escogido. En el caso del apuntamiento mediante seguimiento de la mirada se podrían emplear aquellos formados por los siguientes puntos extremos: la unión del centro de la cabeza con la mano dominante, la unión del centro de la cabeza con el de la cara, el vector formado por uno de los ojos y el dedo índice de una de las manos. En el caso de la señalización manual se puede los definidos por el hombro y el dedo índice de la mano o el que nace de la unión de la nariz con la mano usada para apuntar.

De acuerdo a esta Tesis en el caso de señalización manual se recomienda usar el vector que tiene como origen la pupila del ojo dominante y como destino la mano, para obtener un rendimiento aceptable y conseguir un sistema preciso. No obstante, el empleo de este vector y según el dispositivo escogido puede traer consigo errores como son una mala estimación de la coordenada en altura.

Además puede haber momento en que no se disponga de la información de la cara del usuario y es conveniente tener una alternativa a este vector. Una buena opción es emplear el vector ojo dominante-mano.

Para poder llevar esto a cabo es necesario descubrir el ojo dominante de la persona y permitir personalizar el sistema con este aspecto.

4. Si el servicio final lo permite es conveniente proporcionar al usuario

una referencia visual del blanco apuntado. En este sentido, habilitar esta referencia facilita la interacción evitando errores, disminuyendo el tiempo de interacción y la dificultad a la hora de emplear el sistema.

5. Escoger adecuados blancos por su volumen-tamaño para el espacio de interacción, ya que estos dos factores tienen influencia en el rendimiento y respuesta de los sistemas de apuntamiento.
6. Escoger una técnica de selección que se adapte al servicio final a implementar. Las dos técnicas de selección más usadas son primero en entrar (*first-entry*) y el tiempo de permanencia apuntando sobre un objeto (*dwell-time*).
7. Un aspecto importante es evaluar el sistema y servicio final con una metodología de evaluación adecuada. Esta tesis considera adecuado para ello de la Norma ISO 9241:2000 (UNE-EN ISO 9241-9: Requisitos ergonómicos para trabajos de oficina con pantallas de visualización de datos (PDV). Parte 9: Requisitos para dispositivos de entrada diferentes al teclado.) que describe las tareas a realizar por los participantes de los experimentos de evaluación. Esta Norma combinada con la Ley de Fitts nos permite medir el rendimiento del usuario al emplear el sistema de apuntamiento.
8. La experiencia de usuario es un aspecto esencial que debe ser estudiado una vez implantado el prototipo de sistema. Esta debe evaluar al menos los siguientes seis aspectos : su atractivo, su familiaridad, la eficiencia, su innovación, estimulación y su fiabilidad.
9. En ningún momento durante el diseño hay que perder de vista los requisitos considerados en esta Tesis para un diseño correcto del sistema de apuntamiento. Estos son: exactitud, ubicuidad, universalidad y facilidad de despliegue.
10. Hay que buscar potenciar al máximo la utilidad de la interacción mediante el apuntamiento con el servicio final implementado.

6.2. Trabajo futuro

Los resultados de experimentos de esta Tesis han surgido diversos aspectos que deben continuar estudiando para facilitar estos métodos de interacción.

Un tema específico son los métodos de selección. Se pueden desarrollar nuevos métodos de selección novedosos y menos intrusivos como es el reconocimiento de la voz. Otra de las alternativas podría ser emplear métodos basados en *brain* con algún dispositivo que interprete la señales del cerebro y actuar de forma automática.

Otro aspecto estudiado y sobre el que se podría avanzar en el estudio es sobre la influencia del *tamaño del objetivo* en la eficiencia de la interacción. Esto conduce al planteamiento de un estudio centrado en entender cual sería el tamaño óptimo de la superficie-volumen del objetivo para diferentes distancias entre el usuario y el objetivo.

La estimación del apuntamiento realizada usa el ojo dominante de una persona ya que éste influye en como una persona es capaz de ver las cosas. En esta Tesis, todos los usuarios eran diestros, por lo que para las pruebas solo se ha empleado la mano derecha. Se podría verificar con personas zurdas, aunque a priori, no debe existir ninguna discrepancia. Al final hay que formar el vector con los datos que la persona se encuentre más cómoda y adaptada a ellos. En este sentido el sistema debería ser parametrizado para adaptarse de mejor modo a la persona que lo usa y añadirlo como configurable; como ya sucede con el ojo y la mano que se usa.

Con el objetivo de mejorar de la precisión y exactitud es importante ocuparse de problemas como el temblor de la mano y la respuesta de exactitud de la señalización humana. Este segundo aspecto se puede abordar diseñando un método para adaptar mejor el algoritmo de estimación del vector de apuntamiento a la posición de apuntamiento específica de un usuario (sabiendo que este aspecto es muy diferente de un usuario a otro).

Una mejora sería adaptar el sistema a la forma de apuntar del usuario, es decir, personalizar el sistema al propio apuntamiento del usuario. De esta forma, no sería necesario que el usuario aprendiese a *apuntar* en el contexto del sistema de apuntamiento.

Se debería implementar un sistema para simplificar el proceso de calibración del sistema en el entorno y conseguir un despliegue con mayor facilidad, es decir, que el dispositivo se pueda mover a otro entorno sin mucha dificultad. En definitiva, para hacerlo de manera efectiva, aún se necesita una metodología que acelere el despliegue de espacios inteligentes de apuntamiento interactivos.

Como se ha comentado existe una relación entre la altura del punto a detectar y la distancia de este al dispositivo pudiéndose modelado con una regresión lineal; Sin embargo, luego esta fórmula de relación no se usa. Se

podría plantear como trabajo futuro realizar una mejora usando la distancia que separa el dispositivo sensor del usuario.

Como se ha comentado anteriormente, la referencia visual proporcionada a los usuarios en la interacción dentro del entorno es un aspecto importante en el desarrollo de los sistemas de apuntamiento. El problema es que no en todos los servicios de apuntamiento que se implementen en espacios inteligentes se puede proporcionar una referencia visual. El hecho de proporcionar un *feedback* visual en las actividades de interacción sigue siendo un obstáculo que no parece fácil de resolver. La solución puede variar dependiendo de la categoría del usuario (usuario estándar, discapacitado, ancianos, niños, etc) para el que está diseñado el servicio. Por ejemplo, una solución factible podría ser combinar el apuntamiento con el *feedback* que proporcionase el propio objeto apuntado, por lo que en cada objeto interactivo se podría instalar algún tipo de *feedback* visual (por ejemplo, a través de indicadores con luces led). Otra opción podría ser usar sintetizadores de voz para informar al usuario sobre el punto seleccionado, aunque, por ejemplo en entornos multiusuario esto no sería posible. También se podría proporcionar un *feedback* si se emplea un dispositivo adicional, por ejemplo, equipar al usuario con un dispositivo con un sistema de vibración. Por ejemplo, [73] analiza el efecto del *feedback* con tecnología táctil y la distracción visual en el rendimiento de las tareas de señalización para un entorno virtual en 3D. Por otra parte, [71] explora la posibilidad de añadir una nueva señal para proporcionar el movimiento predictivo para terminar el proceso de interacción. En general, es necesario investigar con mayor detenimiento todas las opciones disponibles para que no agreguen mayor complejidad al sistema.

La temática de los servicios puede ser muy variada y se puede orientar hacia lo que se desee. Este tipo de técnicas es válido, como se ha probado, tanto para entornos en dos dimensiones como en tres. Uno de los aspectos a mejorar es la visibilidad que el dispositivo tiene del usuario. Para ello es necesario colocarlo de forma adecuada, y con un ángulo de visión que abarque el mayor rango posible. Sin embargo, en algunas situaciones no se podrán salvar ciertas oclusiones y habría que estudiar como solventarlas para mejorar la respuesta del sistema. Este sistema sería viable para usar en entorno de personas con problemas de movilidad, ancianos, rehabilitación o juegos.

Actualmente el sistema detecta a un único usuario. El dispositivo Kinect es capaz de capturar la información de hasta un máximo de seis usuarios, por lo que, sería posible ampliar el sistema para que detecte hasta un máximo de seis usuarios. En esta misma línea se puede hacer que el sistema, aparte de detectar a más de un usuario, sea realmente multiusuario. El término *multi-*

usuario debe entenderse no solo para captar su información sino también para interactuar con el sistema. Por ejemplo, en un entorno público en el que se permita que varias personas interaccionen a la vez con el entorno inteligente. Esto implica encontrar los inconvenientes en el proceso de configuración del sistema de apuntamiento dentro de un entorno multiusuario.

Anexo A

Cuestionario de evaluación de la experiencia de usuario

Arm-pointing vs. gaze-pointing in smart spaces

Name:

1. Analysis of the interaction performance

1.1. Gaze pointing

1.1.1. Comfort analysis

	1	2	3	4	5	6	7
1 Required actuation force	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	Very low					Very high	
2 Interaction smoothness	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	Very smooth					Very tough	
3 Interaction effort	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	Very low					Very high	
4 Accuracy	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	Very inaccurate					Very accurate	
5 Speed of use	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	Not suitable					Very suitable	
6 General comfort	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	Very uncomfortable					Very comfortable	
7 Overall performance of the input device	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	Very good					Very bad	
8 Neck fatigue	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	Very high					Very low	

9 Eye fatigue

Very high				Very low		

1.1.2. Questions

How was it to <u>understand</u> the interaction mechanism?
How was it to <u>use</u> the interaction mechanism?
Did you experience any <u>improvement on your skills</u> with the control device along the repetitions?
Please comment <u>your feelings</u> when using the interaction device.

1.1.3. Comparison with visual reference or without visual reference

Compared with finish the task with visual reference (red ball), complete without visual reference is:

1	extremely easier	
2	much easier	
3	easier	
4	equal	

5	more difficult	
6	much more difficult	
7	extremely more difficult	

Why?

What method do you prefer?

a	With visual reference	
B	Without visual reference	

Rate its preference (As preferred method selected versus unselected?)

1	2	3	4	5	6	7

little much

Why?

1.2. Arm pointing

1.2.1. Comfort evaluation:

	1	2	3	4	5	6	7
1 Required actuation force	☐	☐	☐	☐	☐	☐	☐
	Very low			Very high			
2 Interaction smoothness	☐	☐	☐	☐	☐	☐	☐
	Very smooth			Very rough			
3 Interaction efforts	☐	☐	☐	☐	☐	☐	☐
	Very low			Very high			
4 Accuracy	☐	☐	☐	☐	☐	☐	☐
	Very innacurate			Very accurate			
5 Speed of use	☐	☐	☐	☐	☐	☐	☐
	No suitable			suitable			
6 General comfort	☐	☐	☐	☐	☐	☐	☐
	Very uncomfortable			Very comfortable			
7 Overall performance of the input device	☐	☐	☐	☐	☐	☐	☐
	Very difficult			Very easy			
8 Neck fatigue	☐	☐	☐	☐	☐	☐	☐
	Very high			Very low			
9 Fingers fatigue	☐	☐	☐	☐	☐	☐	☐
	Very high			Very low			
10 Wrist fatigue	☐	☐	☐	☐	☐	☐	☐
	Very high			Very low			

11 Arm fatigue

Very high				Very low		

12 Shoulder fatigue

Very high				Very low		

1.2.2. Questions:

How was it to <u>understand</u> the interaction mechanism?
How was it to <u>use</u> the interaction mechanism?
Did you experience any <u>improvement on your skills</u> with the control device along the repetitions?
Please comment <u>your feelings</u> when using the interaction device.

1.2.3. Comparison with visual reference or without visual reference

Compared with finish the task with visual reference (red ball), complete without visual reference is:

1	extremely easier	
2	much easier	
3	easier	
4	equal	
5	more difficult	
6	much more difficult	
7	extremely more difficult	

Why?

What method do you prefer?

a	With visual reference	
B	Without visual reference	

Rate its preference (As preferred method selected versus unselected?)

1	2	3	4	5	6	7

little much

Why?

2. Smart space interaction

2.1. Gaze pointing

2.1.1. Standard Usability Score Questionnaire

		strongly disagree				strongly agree
1	I think that I would like to use this system frequently					
2	I find the system unnecessarily complex					
3	I think the system was easy to use					
4	I think that I would need the support of a technical person to be able to use this system					
5	I find the various functions in this system were well integrated					
6	I think there is too much inconsistency in this system					
7	I think that most people would learn to use this system very quickly					
8	I find the system very cumbersome to use					
9	I feel very confident using the system					
10	I need to learn a lot of things before I could get going with this system					

1 2 3 4 5

2.1.2. User experience questionnaire

For the assessment of the pointing system, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Example:

attractive	☐	☒	☐	☐	☐	☐	☐	unattractive
------------	-----------------------	----------------------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	--------------

This response would mean that you rate the application as more attractive than unattractive. Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression. Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. Please remember: there is no wrong or right answer!

Please assess the gesture identification system now by ticking one circle per line.

		1	2	3	4	5	6	7	
1	annoying	☐	☐	☐	☐	☐	☐	☐	enjoyable
2	not understandable	☐	☐	☐	☐	☐	☐	☐	understandable
3	creative	☐	☐	☐	☐	☐	☐	☐	dull
4	easy to learn	☐	☐	☐	☐	☐	☐	☐	difficult to learn
5	valuable	☐	☐	☐	☐	☐	☐	☐	inferior
6	boring	☐	☐	☐	☐	☐	☐	☐	exciting
7	not interesting	☐	☐	☐	☐	☐	☐	☐	interesting
8	unpredictable	☐	☐	☐	☐	☐	☐	☐	predictable
9	fast	☐	☐	☐	☐	☐	☐	☐	slow
10	inventive	☐	☐	☐	☐	☐	☐	☐	conventional
11	obstructive	☐	☐	☐	☐	☐	☐	☐	supportive
12	good	☐	☐	☐	☐	☐	☐	☐	bad
13	complicated	☐	☐	☐	☐	☐	☐	☐	easy
14	unlikable	☐	☐	☐	☐	☐	☐	☐	pleasing
15	usual	☐	☐	☐	☐	☐	☐	☐	leading edge
16	unpleasant	☐	☐	☐	☐	☐	☐	☐	pleasant
17	secure	☐	☐	☐	☐	☐	☐	☐	not secure
18	motivating	☐	☐	☐	☐	☐	☐	☐	demotivating
19	meets expectations	☐	☐	☐	☐	☐	☐	☐	does not meet expectations
20	inefficient	☐	☐	☐	☐	☐	☐	☐	efficient
21	clear	☐	☐	☐	☐	☐	☐	☐	confusing
22	impractical	☐	☐	☐	☐	☐	☐	☐	practical
23	organized	☐	☐	☐	☐	☐	☐	☐	cluttered
24	attractive	☐	☐	☐	☐	☐	☐	☐	unattractive
25	friendly	☐	☐	☐	☐	☐	☐	☐	unfriendly
26	conservative	☐	☐	☐	☐	☐	☐	☐	innovative

2.1.3. Comparison with visual reference or without visual reference

Compared with finish the task with visual reference (red ball), complete without visual reference is:

1	extremely easier	
2	much easier	

3	easier	
4	equal	
5	more difficult	
6	much more difficult	
7	extremely more difficult	

Why?

What method do you prefer?

a	With visual reference	
B	Without visual reference	

Rate its preference (As preferred method selected versus unselected?)

1	2	3	4	5	6	7

little much

Why?

2.2. Arm pointing

Interact 10 times, at least once each object.

After the end of task, please complete the following questionnaire:

2.2.1. Standard Usability Score Questionnaire

		strongly disagree				strongly agree
1	I think that I would like to use this system frequently					
2	I find the system unnecessarily complex					
3	I think the system was easy to use					
4	I think that I would need the support of a technical person to be able to use this system					
5	I find the various functions in this system were well integrated					
6	I think there was too much inconsistency in this system					
7	I think that most people would learn to use this system very quickly					
8	I find the system very cumbersome to use					
9	I feel very confident using the system					
10	I need to learn a lot of things before I could get going with this system					

1 2 3 4 5

2.2.2. User experience questionnaire

For the assessment of the pointing system, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Example:

attractive	○	⊗	○	○	○	○	○	unattractive
------------	---	---	---	---	---	---	---	--------------

This response would mean that you rate the application as more attractive than unattractive. Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression. Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. Please remember: there is no wrong or right answer!

Please assess the gesture identification system now by ticking one circle per line.

		1	2	3	4	5	6	7	
1	annoying	☐	☐	☐	☐	☐	☐	☐	enjoyable
2	not understandable	☐	☐	☐	☐	☐	☐	☐	understandable
3	creative	☐	☐	☐	☐	☐	☐	☐	dull
4	easy to learn	☐	☐	☐	☐	☐	☐	☐	difficult to learn
5	valuable	☐	☐	☐	☐	☐	☐	☐	inferior
6	boring	☐	☐	☐	☐	☐	☐	☐	exciting
7	not interesting	☐	☐	☐	☐	☐	☐	☐	interesting
8	unpredictable	☐	☐	☐	☐	☐	☐	☐	predictable
9	fast	☐	☐	☐	☐	☐	☐	☐	slow
10	inventive	☐	☐	☐	☐	☐	☐	☐	conventional
11	obstructive	☐	☐	☐	☐	☐	☐	☐	supportive
12	good	☐	☐	☐	☐	☐	☐	☐	bad
13	complicated	☐	☐	☐	☐	☐	☐	☐	easy
14	unlikable	☐	☐	☐	☐	☐	☐	☐	pleasing
15	usual	☐	☐	☐	☐	☐	☐	☐	leading edge
16	unpleasant	☐	☐	☐	☐	☐	☐	☐	pleasant
17	secure	☐	☐	☐	☐	☐	☐	☐	not secure
18	motivating	☐	☐	☐	☐	☐	☐	☐	demotivating
19	meets expectations	☐	☐	☐	☐	☐	☐	☐	does not meet expectations
20	inefficient	☐	☐	☐	☐	☐	☐	☐	efficient
21	clear	☐	☐	☐	☐	☐	☐	☐	confusing
22	impractical	☐	☐	☐	☐	☐	☐	☐	practical
23	organized	☐	☐	☐	☐	☐	☐	☐	cluttered
24	attractive	☐	☐	☐	☐	☐	☐	☐	unattractive
25	friendly	☐	☐	☐	☐	☐	☐	☐	unfriendly
26	conservative	☐	☐	☐	☐	☐	☐	☐	innovative

2.2.3. Comparison with visual reference or without visual reference

Compared with finish the task with visual reference (red ball), complete without visual reference is:

1	extremely easier	
2	much easier	
3	easier	

4	equal	
5	more difficult	
6	much more difficult	
7	extremely more difficult	

Why?

What method do you prefer?

a	With visual reference	
B	Without visual reference	

Rate its preference (As preferred method selected versus unselected?)

1	2	3	4	5	6	7

little much

Why?

3. Comparison of both methods

Between the two interaction methods, which one do you prefer? Why?

Would you use the preferred interaction method in your daily life? Please justify your choice.

Would you use a pointing interaction system (arm or gaze-based) combined with speech recognition to broaden its functionalities? (Example: pointing at the lamp while saying the order to complete: 'on')

Additional comments welcome.

Anexo B

Cuestionarios de evaluación de usuarios para las pruebas de videowall

Pointing in smart spaces (videowall)

Nombre:

Ojo dominante:

(Despues de la prueba de precision y cercanía)

Ha mejorado su destreza con el sistema tras las sucesivas iteraciones:

Comfort analysis

	1	2	3	4	5	6	7
1 Required actuation force	☐	☐	☐	☐	☐	☐	☐
	Very low				Very high		
2 Interaction smoothness	☐	☐	☐	☐	☐	☐	☐
	Very smooth				Very tough		
3 Interaction effort	☐	☐	☐	☐	☐	☐	☐
	Very low				Very high		
4 Accuracy	☐	☐	☐	☐	☐	☐	☐
	Very inaccurate				Very accurate		
5 Speed of use	☐	☐	☐	☐	☐	☐	☐
	Not suitable				Very suitable		
6 General comfort	☐	☐	☐	☐	☐	☐	☐
	Very uncomfortable				Very comfortable		
7 Overall performance of the input device	☐	☐	☐	☐	☐	☐	☐

		Very good				Very bad
8	Neck fatigue	☐	☐	☐	☐	☐
		Very high				Very low
9	Arm fatigue	☐	☐	☐	☐	☐
		Very high				Very low

[Standard Usability Score Questionnaire](#)

			strongly disagree			strongly agree
1	I think that I would like to use this system frequently	☐	☐	☐	☐	☐
2	I find the system unnecessarily complex	☐	☐	☐	☐	☐
3	I think the system was easy to use	☐	☐	☐	☐	☐
4	I think that I would need the support of a technical person to be able to use this system	☐	☐	☐	☐	☐
5	I find the various functions in this system were well integrated	☐	☐	☐	☐	☐
6	I think there is too much inconsistency in this system	☐	☐	☐	☐	☐
7	I think that most people would learn to use this system very quickly	☐	☐	☐	☐	☐
8	I find the system very cumbersome to use	☐	☐	☐	☐	☐
9	I feel very confident using the system	☐	☐	☐	☐	☐
10	I need to learn a lot of things before I could get going with this system	☐	☐	☐	☐	☐
		1	2	3	4	5

User experience questionnaire

For the assessment of the pointing system, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Example:

attractive	☐	☒	☐	☐	☐	☐	☐	unattractive
------------	-----------------------	----------------------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	--------------

This response would mean that you rate the application as more attractive than unattractive. Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression. Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. Please remember: there is no wrong or right answer!

Please assess the gesture identification system now by ticking one circle per line.

		1	2	3	4	5	6	7	
1	annoying	☐	☐	☐	☐	☐	☐	☐	enjoyable
2	not understandable	☐	☐	☐	☐	☐	☐	☐	understandable
3	creative	☐	☐	☐	☐	☐	☐	☐	dull
4	easy to learn	☐	☐	☐	☐	☐	☐	☐	difficult to learn
5	valuable	☐	☐	☐	☐	☐	☐	☐	inferior
6	boring	☐	☐	☐	☐	☐	☐	☐	exciting
7	not interesting	☐	☐	☐	☐	☐	☐	☐	interesting
8	unpredictable	☐	☐	☐	☐	☐	☐	☐	predictable
9	fast	☐	☐	☐	☐	☐	☐	☐	slow
10	inventive	☐	☐	☐	☐	☐	☐	☐	conventional
11	obstructive	☐	☐	☐	☐	☐	☐	☐	supportive
12	good	☐	☐	☐	☐	☐	☐	☐	bad
13	complicated	☐	☐	☐	☐	☐	☐	☐	easy
14	unlikable	☐	☐	☐	☐	☐	☐	☐	pleasing
15	usual	☐	☐	☐	☐	☐	☐	☐	leading edge
16	unpleasant	☐	☐	☐	☐	☐	☐	☐	pleasant
17	secure	☐	☐	☐	☐	☐	☐	☐	not secure
18	motivating	☐	☐	☐	☐	☐	☐	☐	demotivating
19	meets expectations	☐	☐	☐	☐	☐	☐	☐	does not meet expectations
20	inefficient	☐	☐	☐	☐	☐	☐	☐	efficient
21	clear	☐	☐	☐	☐	☐	☐	☐	confusing

22	impractical	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	practical
23	organized	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	cluttered
24	attractive	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	unattractive
25	friendly	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	unfriendly
26	conservative	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	innovative

[Additional questions](#)

What selection method do you prefer?	
Gesture other hand	☐
Voice recognition	☐
Dwell time	☐
Why?	
How was it to <u>understand</u> the pointing?	
How was it to <u>use</u> the interaction service?	

Did you experience any <u>improvement on your skills</u> with the control device along the test?
Please comment <u>your feelings</u> when using this pointing interaction.
Any comment is welcome

Anexo C

Modelo de autorización de los participantes de la prueba

Modelo de autorización seguido de las firmas de todos los participantes de la prueba.

Consentimiento informado

Evaluación de un sistema de apuntamiento para aplicaciones de espacios inteligentes

INVESTIGADOR que realiza el estudio: David Gómez Cordero.

GPDS, SSR, ETSIT, UPM

Email: david.gomez@grpss.ssr.upm.es

Propósito del estudio: El objetivo de las pruebas es medir la precisión que consigue un sistema de apuntamiento basado en Kinect, además de recoger algunos datos cualitativos y de experiencia de usuario.

Procedimiento: La prueba consta de dos fases. En la primera fase, los participantes completarán una serie de iteraciones de apuntamiento. En ella, deben seleccionar una serie de objetos virtuales que aparecen proyectados sobre la pared. Esto se repetirá en cuatro posiciones diferentes del entorno. En la segunda fase los participantes utilizarán un servicio real implementado con el sistema de apuntamiento que se está estudiando. Este servicio consiste en seleccionar e interactuar con las 9 pantallas que forman un “videowall”. Finalmente, el participante contestará a un cuestionario para evaluar la experiencia de usuario.

Riesgos: Es posible que se fatigue durante el transcurso de su participación en el estudio. Se le darán oportunidades para descansar, y pausas adicionales si son necesarias. No existen riesgos asociados con la participación en el estudio. Si las tareas se convierten en algo angustioso para usted, se dará por terminado el estudio inmediatamente.

Beneficios: Se espera que los resultados del estudio revelen principios de diseño de sistemas de apuntamiento en espacios inteligentes, que beneficiarán el avance y la aceptación del nuevo modelo de interacción basado en apuntamiento.

Alternativas a participación: La participación en este estudio es voluntaria. Usted es libre de retirarse o dejar de participar en cualquier momento.

Coste y compensación: La participación en este estudio no implicará ningún coste para usted y no hay prevista compensación económica.

Confidencialidad: Toda la información recogida en el estudio se mantendrá de forma estrictamente confidencial. Sus datos se anonimizarán y no se realizarán publicaciones o informes donde se incluya su identificación explícitamente. El material multimedia recogido podrá ser utilizado con objetivos de publicación y difusión, para demostrar el alcance de las pruebas, siempre bajo premisas éticas.

Si acepta participar en este estudio bajo las condiciones descritas, por favor firme a continuación.

Por la presente, yo, _____
declaro que he leído y entendido la información contenida en este formulario.

Firma del participante

Fecha

Investigador responsable del estudio

Fecha

Si tiene alguna consulta sobre este estudio contacte con david.gomez@grpss.ssr.upm.es

Glosario

Aceptación de la tecnología Aceptar la tecnología por parte de un usuario implica que éste considera esa dicha tecnología aceptable para su uso con un fin determinado..

Actuador Es un dispositivo que recibe una orden y en función de ella activa un elemento..

Apuntamiento basado en señalización de la mirada Es el apuntamiento que realiza una persona con su mirada. La detección de este apuntamiento puede realizarse con información del posicionamiento de las pupilas de los usuarios o con la información de su posición y orientación de la cara para conocer donde está mirando..

Apuntamiento basado en señalización manual Es el apuntamiento que realiza una persona con el dedo índice de su mano para mostrar interés hacia algo y que, habitualmente, lleva asociado el movimiento de su brazo..

Calibración del sistema Permite referenciar todas las posiciones o localizaciones que se empleen de acuerdo a un sistema de referencias global único. De este modo las posiciones del usuario, la localización de los dispositivos detectores y los objetos pueden ser referenciados en el mismo sistema de coordenadas. .

Calibración del usuario Tiene como objetivo corregir de forma personalizada la estimación de apuntamiento realizada por el sistema. Para ello es necesario realizar un proceso previo..

Calm technology Es un término que se refiere a la tecnología que informa pero no demanda atención de las personas. No necesita que el usuario este constantemente pendiente de ello y sólo centrará su atención cuando necesite algo concreto..

Confort Comodidad de uso por parte del usuario a la hora de emplear un sistema de interacción..

COTs Dispositivos comerciales listos para usar, *Commercial Off-The-Shelf* como son el sensor Microsoft Kinect o una bombilla inteligente..

Eficacia Un sistema de interacción es eficaz cuando la respuesta del sistema es la adecuada a lo que espera el usuario y de la forma que el desea, por ejemplo, en un tiempo concreto..

Eficiencia Idoneidad para resolver las tareas del usuario..

Entorno inteligente Es un espacio dotado de sensores y activadores que permiten a las personas interactuar con los objetos situados en el entorno. .

Exactitud Distancia del centro de masas del objeto apuntado al punto real estimado..

Experiencia de usuario (*User Experience*) UX Describe las sensaciones y sentimientos que tiene una persona al interactuar con un sistema.

Fatiga Molestia o cansancio del usuario ocasionado por un esfuerzo realizado. .

Fiabilidad También se puede usar el término de confianza como sinónimo en este caso y puede definirse como la sensación de control por parte del usuario. .

Interacción natural Es la interacción que emplea interfaces naturales de usuario. .

Interfaces naturales de usuario, (*Natural User Interfaces, NUIs*) Las interfaces de usuario naturales están basadas en la interacción entre personas tradicional sin que la tecnología que lo hace capaz sea visible al usuario El reconocimiento de gestos o de voz es un ejemplo de este tipo de interfaces..

ISO 9241-9 Norma ISO 9241-9:2000 de la Organización Internacional de Estandarización (*International Organization for Standardization* que recoge los requisitos ergonómicos para el trabajo de oficina con terminales de pantalla visual; concretamente, la parte 9 hace referencia a dispositivos de entrada que no son teclado. .

Ley de Fitts La Ley de Fitts permite modelar la interacción mediante apuntamiento y medir dificultad de un usuario a la hora de apuntar. .

Microsoft Kinect Es un dispositivo sensor de la compañía Microsoft que incorpora una cámara RGB, un sensor de profundidad y un micrófono. Permite rastrear cuerpos de personas, capturar sus movimientos y realizar reconocimiento facial, de voz o gestos..

Objetos inteligentes Objetos dotados de cierta tecnología, conectados a la red y situados dentro de un entorno inteligente. Permiten que las personas interactúen con ellos. Por ejemplo, una lámpara inteligente permite su activación o desactivación por órdenes mediante voz. .

Precisión Dispersión de los valores estimados en varias estimaciones de apuntamiento. A mayor desviación de estos datos, menor precisión, y al contrario, se consigue una mayor precisión con una menor desviación de los datos..

Prototipado Es la acción de construir un prototipo..

Prototipo Un prototipo es el primer modelo que se crea para la construcción de un producto final(tangible o no tangible) . Este a su vez nos permite verificar el diseño y confirmar que cumple los requisitos iniciales..

Rendimiento El rendimiento del sistema se mide con el parámetro TP (*throughput*) planteado en la Ley de Fitts o mediante el tiempo de ejecución. En el caso del parámetro TP, la Ley de Fitts combina las variables de la propia Ley como son el índice de dificultad y el tiempo de movimiento..

Sensor Dispositivo capaz de recoger información del entorno físico. Hay cientos de tipos según la información que recogen, como por ejemplo, información de temperatura, movimiento, de luz, etc. En su caso, la mayoría de los sensores disponen de un controlador para manejar la información captada..

Usabilidad Calidad del sistema de apuntamiento por la cual el usuario final lo empleará y lo encontrará adecuado a sus necesidades..

Videowall Conjunto de pantallas de visualización de contenidos que pueden actuar como si fuesen una única o de modo independiente..

Bibliografía

- [1] M. Weiser and J. S. Brown, “Designing calm technology,” *PowerGrid Journal*, vol. 1, no. 1, pp. 75–85, 1996.
- [2] A. M. Bernardos, L. Bergesio, J. Iglesias, and J. R. Casar, “Meccano: a mobile-enabled configuration framework to coordinate and augment networks of smart objects.” *J. UCS*, vol. 19, no. 17, pp. 2503–2525, 2013.
- [3] D. A. Norman, “Natural user interfaces are not natural,” *interactions*, vol. 17, no. 3, pp. 6–10, 2010.
- [4] A. Fernández, L. Bergesio, A. M. Bernardos, J. A. Besada, and J. R. Casar, “A kinect-based system to enable interaction by pointing in smart spaces,” in *Sensors Applications Symposium (SAS)*, 2015 IEEE. IEEE, 2015, pp. 1–6.
- [5] G. Imai, “Gestures: Body language and nonverbal communication,” Retrieved Oct, 2005.
- [6] S. F. Neggers and H. Bekkering, “Gaze anchoring to a pointing target is present during the entire pointing movement and is driven by a non-visual signal,” *Journal of Neurophysiology*, vol. 86, no. 2, pp. 961–970, 2001.
- [7] J. Brooke et al., “Sus-a quick and dirty usability scale,” *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.
- [8] M. Rauschenberger, S. Olschner, M. P. Cota, M. Schrepp, and J. Thomaschewski, “Measurement of user experience: A spanish language version of the user experience questionnaire (ueq),” in *Information Systems and Technologies (CISTI)*, 2012 7th Iberian Conference on. IEEE, 2012, pp. 1–6.

- [9] X. Xu and R. W. McGorry, “The validity of the first and second generation microsoft kinect™ for identifying joint center locations during static postures,” *Applied ergonomics*, vol. 49, pp. 47–54, 2015.
- [10] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, and A. El Saddik, “Evaluating and improving the depth accuracy of kinect for windows v2,” *IEEE Sensors Journal*, vol. 15, no. 8, pp. 4275–4285, 2015.
- [11] H. Kim, Y. Kim, and E. C. Lee, “Method for user interface of large displays using arm pointing and finger counting gesture recognition,” *TheScientificWorldJournal*, vol. 2014, p. 683045, 09 2014.
- [12] P. Jing and G. Ye-peng, “Human-computer interaction using pointing gesture based on an adaptive virtual touch screen,” *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 6, no. 4, pp. 81–91, 2013.
- [13] C. P. Quintero, R. T. Fomena, A. Shademan, N. Wolleb, T. Dick, and M. Jagersand, “Sepa: Selecting by pointing as an intuitive human-robot command interface,” in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1166–1171.
- [14] Y. Lai, C. Wang, Y. Li, S. S. Ge, and D. Huang, “3d pointing gesture recognition for human-robot interaction,” in *2016 Chinese Control and Decision Conference (CCDC)*, May 2016, pp. 4959–4964.
- [15] J. Shingu, P. Chiu, S. Kratz, J. Vaughan, and D. Kimber, “Depth based shadow pointing interface for public displays,” in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, ser. *UIST ’16 Adjunct*. New York, NY, USA: ACM, 2016, pp. 79–80. [Online]. Available: <http://doi.acm.org/10.1145/2984751.2985710>
- [16] A. M. Bernardos, Í. Marquínez, D. Gómez, J. A. Besada, and J. R. Casar, “A system for multimodal interaction with kinect-enabled virtual windows,” in *Intelligent Technologies for Interactive Entertainment (INTETAIN), 2015 7th International Conference on*. IEEE, 2015, pp. 42–46.
- [17] N. Dhingra, E. Valli, and A. Kunz, “Recognition and localisation of pointing gestures using a rgb-d camera,” 2020.
- [18] M. Tolgyessy, M. Dekan, F. Duchoň, J. Rodina, P. Hubinský, and L. Chovanec, “Foundations of visual linear human–robot interaction via pointing gesture navigation,” *International Journal of Social*

- Robotics, vol. 9, no. 4, pp. 509–523, Sep 2017. [Online]. Available: <https://doi.org/10.1007/s12369-017-0408-9>
- [19] H. Wu, “Hydrascope: Interaction and application of wall sized display,” Master’s thesis, EECS Department, University of California, Berkeley, Jun 2012. [Online]. Available: <http://www2.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-149.html>
- [20] A. M. Bernardos, A. Muñoz, L. Bergesio, J. A. Besada, and J. R. Casar, “A multimodal interaction system for big displays,” in Proceedings of the 6th ACM International Symposium on Pervasive Displays, ser. PerDis ’17. New York, NY, USA: ACM, 2017, pp. 28:1–28:2. [Online]. Available: <http://doi.acm.org/10.1145/3078810.3084353>
- [21] M. Henschke, T. Gedeon, and R. Jones, “Extending the index finger is worse than sitting: Posture and minimal objects in 3d pointing interfaces,” in 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), Dec. 2013, pp. 797–802.
- [22] K. Watanabe, Y. Miyake, N. Nakamichi, T. Yamada, and T. Ozeki, “Remote touch pointing for smart tv interaction,” in 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE). IEEE, 2014, pp. 232–235.
- [23] S. S. Raza Abidi, M. Williams, and B. Johnston, “Human pointing as a robot directive,” in Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction. IEEE Press, 2013, pp. 67–68.
- [24] H. Cheng, T. Shin, and J. Tanaka, Facial Tracking-Assisted Hand Pointing Technique for Wall-Sized Displays. Cham: Springer International Publishing, 2016, pp. 191–201. [Online]. Available: https://doi.org/10.1007/978-3-319-39862-4_18
- [25] L. E. Sibert and R. J. K. Jacob, “Evaluation of Eye Gaze Interaction,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI ’00. New York, NY, USA: ACM, 2000, pp. 281–288. [Online]. Available: <http://doi.acm.org/10.1145/332040.332445>
- [26] N.-K. Chuan and A. Sivaji, “Combining eye gaze and hand tracking for pointer control in hci: Developing a more robust and accurate interaction system for pointer positioning and clicking,” in Humanities, Science and Engineering (CHUSER), 2012 IEEE Colloquium on. IEEE, 2012, pp. 172–176.

- [27] Y. Gizatdinova, O. Špakov, O. Tuisku, M. Turk, and V. Surakka, “Gaze and head pointing for hands-free text entry: Applicability to ultra-small virtual keyboards,” in Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, ser. ETRA ’18. New York, NY, USA: ACM, 2018, pp. 14:1–14:9. [Online]. Available: <http://doi.acm.org/10.1145/3204493.3204539>
- [28] X. Xiong, Z. Liu, Q. Cai, and Z. Zhang, “Eye gaze tracking using an rgbd camera: A comparison with a rgb solution,” in Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, ser. UbiComp ’14 Adjunct. New York, NY, USA: ACM, 2014, pp. 1113–1121. [Online]. Available: <http://doi.acm.org/10.1145/2638728.2641694>
- [29] K. Katsuragawa, K. Pietroszek, J. R. Wallace, and E. Lank, “Watchpoint: Freehand pointing with a smartwatch in a ubiquitous display environment,” in Proceedings of the International Working Conference on Advanced Visual Interfaces, ser. AVI ’16. New York, NY, USA: ACM, 2016, pp. 128–135. [Online]. Available: <http://doi.acm.org/10.1145/2909132.2909263>
- [30] J. Abbasov, T. Horak, and R. Dachsel, “Smartwatch-based pointing interaction,” in Mensch und Computer 2018 - Tagungsband, R. Dachsel and G. Weber, Eds. Bonn: Gesellschaft für Informatik e.V., 2018.
- [31] R. A. Bolt, Put-that-there: Voice and gesture at the graphics interface. ACM, 1980, vol. 14, no. 3.
- [32] A. Alanwar, M. Alzantot, B.-J. Ho, P. Martin, and M. Srivastava, “Selecon: Scalable iot device selection and control using hand gestures,” in Proceedings of the Second International Conference on Internet-of-Things Design and Implementation. ACM, 2017, pp. 47–58.
- [33] X. Wang, “Design and evaluation of novel gesture-centric interaction models in smart spaces,” Ph.D. dissertation, Telecomunicacion, 2016. [Online]. Available: <http://oa.upm.es/43500/>
- [34] C. Andujar, P. Brunet, A. Vinacua, M. Vico, and J. Diaz, “A multi-projector cave system with commodity hardware and gesture-based interaction,” ViRVIG - Grupo de Investigación en Visualización, Realidad Virtual e Interacción Gráfica, Tech. Rep., Jan 2017. [Online]. Available: <http://hdl.handle.net/2117/99668>

- [35] W. A. König, J. Gerken, S. Dierdorf, and H. Reiterer, “Adaptive Pointing – Design and Evaluation of a Precision Enhancing Technique for Absolute Pointing Devices,” in *Human-Computer Interaction – INTERACT 2009*, ser. *Lecture Notes in Computer Science*, T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, and M. Winckler, Eds. Springer Berlin Heidelberg, 2009, no. 5726, pp. 658–671. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-03655-2_73
- [36] D. W. Hansen and Q. Ji, “In the eye of the beholder: A survey of models for eyes and gaze,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478–500, March 2010.
- [37] Y. Rossetti, B. Tadary, and C. Prablanc, “Optimal contributions of head and eye positions to spatial accuracy in man tested by visually directed pointing,” *Experimental brain research*, vol. 97, no. 3, pp. 487–496, 1994.
- [38] R. Stiefelhagen and J. Zhu, “Head orientation and gaze direction in meetings,” in *CHI’02 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2002, pp. 858–859.
- [39] R. Newman, Y. Matsumoto, S. Rougeaux, and A. Zelinsky, “Real-time stereo tracking for head pose and gaze estimation,” in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, 2000, pp. 122–128.
- [40] R. Valenti, Z. Yucel, and T. Gevers, “Robustifying eye center localization by head pose cues,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 612–618.
- [41] R. Valenti, N. Sebe, and T. Gevers, “Combining head pose and eye location information for gaze estimation,” *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 802–815, Feb 2012.
- [42] H. Carnahan and R. G. Marteniuk, “The temporal organization of hand, eye, and head movements during reaching and pointing,” *Journal of Motor Behavior*, vol. 23, no. 2, pp. 109–119, 1991.
- [43] R. Pappu and P. A. Beardsley, “A qualitative approach to classifying gaze direction,” in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, Apr 1998, pp. 160–165.
- [44] J. Wu and M. M. Trivedi, “A two-stage head pose estimation framework and evaluation,” *Pattern Recogn.*, vol. 41, no. 3, pp. 1138–1158, Mar.

2008. [Online]. Available: <http://dx.doi.org/10.1016/j.patcog.2007.07.017>

- [45] V. Rantanen, J. Verho, J. Lekkala, O. Tuisku, V. Surakka, and T. Vanhala, “The effect of clicking by smiling on the accuracy of head-mounted gaze tracking,” in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '12. New York, NY, USA: ACM, 2012, pp. 345–348. [Online]. Available: <http://doi.acm.org/10.1145/2168556.2168633>
- [46] J. Cuaresma and I. S. MacKenzie, “Fittsface: Exploring navigation and selection methods for facial tracking,” in *Universal Access in Human-Computer Interaction. Designing Novel Interactions*, M. Antona and C. Stephanidis, Eds. Cham: Springer International Publishing, 2017, pp. 403–416.
- [47] D. Miniotas, O. Špakov, I. Tugoy, and I. S. MacKenzie, “Speech-augmented eye gaze interaction with small closely spaced targets,” in *Proceedings of the 2006 symposium on Eye tracking research & applications*. ACM, 2006, pp. 67–72.
- [48] C. J. Lin, S.-H. Ho, and Y.-J. Chen, “An investigation of pointing postures in a 3d stereoscopic environment,” *Applied ergonomics*, vol. 48, pp. 154–163, 2015.
- [49] N. Cournia, J. D. Smith, and A. T. Duchowski, “Gaze- vs. handbased pointing in virtual environments,” in *In CHI '03: CHI '03 extended abstracts on Human factors in computing systems*. ACM Press, 2003, pp. 772–773.
- [50] S. Zhai, C. Morimoto, and S. Ihde, “Manual and gaze input cascaded (magic) pointing,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '99. New York, NY, USA: ACM, 1999, pp. 246–253. [Online]. Available: <http://doi.acm.org/10.1145/302979.303053>
- [51] G. Kouroupetroglou, A. Pino, A. Balmpakakis, D. Chalastanis, V. Golematis, N. Ioannou, and I. Koutsoumpas, “Using wiimote for 2d and 3d pointing tasks: Gesture performance evaluation,” in *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*, ser. *Lecture Notes in Computer Science*, E. Efthimiou, G. Kouroupetroglou, and S.-E. Fotinea, Eds. Springer

- Berlin Heidelberg, 2012, vol. 7206, pp. 13–23. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-34182-3_2
- [52] A. Pino, E. Tzemis, N. Ioannou, and G. Kouroupetroglou, “Using Kinect for 2d and 3d Pointing Tasks: Performance Evaluation,” in *Human-Computer Interaction. Interaction Modalities and Techniques*, ser. *Lecture Notes in Computer Science*, M. Kurosu, Ed. Springer Berlin Heidelberg, 2013, no. 8007, pp. 358–367. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-39330-3_38
 - [53] H. Fürntratt and H. Neuschmied, “Evaluating pointing accuracy on kinect v2 sensor,” in *International Conference on Multimedia and Human-Computer Interaction (MHCI)*, 2014, pp. 124–1.
 - [54] A. Zaranek, B. Ramoul, H. F. Yu, Y. Yao, and R. J. Teather, “Performance of Modern Gaming Input Devices in First-person Shooter Target Acquisition,” in *Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, ser. *CHI EA '14*. New York, NY, USA: ACM, 2014, pp. 1495–1500. [Online]. Available: <http://doi.acm.org/10.1145/2559206.2581291>
 - [55] M. Schwaller, S. Brunner, and D. Lalanne, “Two handed mid-air gestural hci: Point+ command,” in *International Conference on Human-Computer Interaction*. Springer, 2013, pp. 388–397.
 - [56] A. Biskupski, A. R. Fender, T. M. Feuchtner, M. Karsten, and J. D. Willaredt, “Drunken ed: a balance game for public large screen displays,” in *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2014, pp. 289–292.
 - [57] L. Gallo, A. P. Placitelli, and M. Ciampi, “Controller-free exploration of medical image data: Experiencing the kinect,” in *Computer-based medical systems (CBMS), 2011 24th international symposium on*. IEEE, 2011, pp. 1–6.
 - [58] C. Peng, J. T. Hansberger, L. Cao, and V. A. Shanthakumar, “Hand gesture controls for image categorization in immersive virtual environments,” in *2017 IEEE Virtual Reality (VR)*, March 2017, pp. 331–332.
 - [59] G. Drossis, C. Birliraki, G. Margetis, and C. Stephanidis, “Immersive 3d environment for data centre monitoring based on gesture based interaction,” in *International Conference on Human-Computer Interaction*. Springer, 2017, pp. 103–108.

- [60] I. O. for Standardization, “Ergonomic requirements for office work with visual display terminals (vdts) – part 9: Requirements for non-keyboard input devices,” International Organization for Standardization, Standard, Feb. 2000.
- [61] A. M. Bernardos, D. Gómez, and J. R. Casar, “A comparison of head pose and deictic pointing interaction methods for smart environments,” *International Journal of Human-Computer Interaction*, vol. 32, no. 4, pp. 325–351, 2016. [Online]. Available: <http://dx.doi.org/10.1080/10447318.2016.1142054>
- [62] P. M. Fitts, “The information capacity of the human motor system in controlling the amplitude of movement.” *Journal of experimental psychology*, vol. 47, no. 6, p. 381, 1954.
- [63] R. W. Soukoreff and I. S. MacKenzie, “Towards a Standard for Pointing Device Evaluation, Perspectives on 27 Years of Fitts’ Law Research in HCI,” *Int. J. Hum.-Comput. Stud.*, vol. 61, no. 6, pp. 751–789, Dec. 2004. [Online]. Available: <http://dx.doi.org/10.1016/j.ijhcs.2004.09.001>
- [64] I. S. MacKenzie, “Fitts’ law as a research and design tool in human-computer interaction,” *Human-computer interaction*, vol. 7, no. 1, pp. 91–139, 1992.
- [65] S. K. Card, W. K. English, and B. J. Burr, “Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a crt,” *Ergonomics*, vol. 21, no. 8, pp. 601–613, 1978.
- [66] I. S. MacKenzie and R. J. Teather, “Fittstilt: The application of fitts’ law to tilt-based interaction,” in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, ser. NordiCHI ’12. New York, NY, USA: ACM, 2012, pp. 568–577. [Online]. Available: <http://doi.acm.org/10.1145/2399016.2399103>
- [67] K. A. F. Mora and J.-M. Odobez, “Gaze estimation from multimodal kinect data,” in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2012, pp. 25–30.
- [68] N. Bevan, “What is the difference between the purpose of usability and user experience evaluation methods,” in *Proceedings of the Workshop UXEM*, vol. 9, 2009, pp. 1–4.

- [69] A. Murata and H. Iwase, “Extending fitts’ law to a three-dimensional pointing task,” *Human movement science*, vol. 20, no. 6, pp. 791–805, 2001.
- [70] L. J. Cronbach, “Coefficient alpha and the internal structure of tests,” *psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.
- [71] K. Cho, J.-H. Lee, B.-T. Lee, and E. Park, “Effects of feedforward in in-air remote pointing,” *International Journal of Human-Computer Interaction*, vol. 31, no. 2, pp. 89–100, 2015.
- [72] J. Nielsen and T. K. Landauer, “A mathematical model of the finding of usability problems,” in *Proceedings of the INTERACT ’93 and CHI ’93 Conference on Human Factors in Computing Systems*, ser. CHI ’93. New York, NY, USA: ACM, 1993, pp. 206–213. [Online]. Available: <http://doi.acm.org/10.1145/169059.169166>
- [73] B. Corbett, C. Nam, and T. Yamaguchi, “The effects of haptic feedback and visual distraction on pointing task performance,” *International Journal of Human-Computer Interaction*, vol. 32, p. 151006145548004, 10 2015.