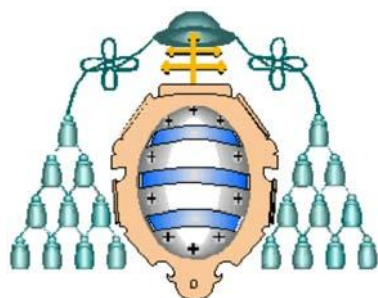


UNIVERSIDAD DE OVIEDO
DEPARTAMENTO DE INFORMÁTICA
ÁREA DE INGENIERÍA TELEMÁTICA



TESIS DOCTORAL

**EVALUACIÓN Y DISEÑO DE MÉTODOS
EFECTIVOS DE INSERCIÓN DE TEXTO EN
ENTORNOS DE TELEVISIÓN DIGITAL
INTERACTIVA**

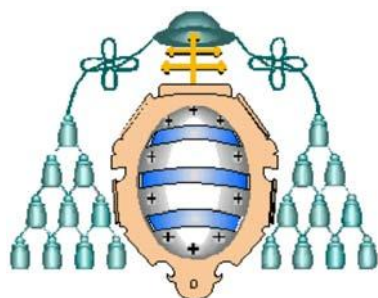
AUTORA: AURORA BARRERO LÓPEZ

DIRECTOR: DAVID MELENDI PALACIO

**Tesis Doctoral presentada en la Universidad de Oviedo para la
obtención del título de Doctor por la Universidad de Oviedo**

Gijón, julio de 2017

UNIVERSIDAD DE OVIEDO
DEPARTAMENTO DE INFORMÁTICA
ÁREA DE INGENIERÍA TELEMÁTICA



TESIS DOCTORAL

**EVALUACIÓN Y DISEÑO DE MÉTODOS
EFECTIVOS DE INSERCIÓN DE TEXTO EN
ENTORNOS DE TELEVISIÓN DIGITAL
INTERACTIVA**

AUTORA: AURORA BARRERO LÓPEZ

DIRECTOR: DAVID MELENDI PALACIO

**Tesis Doctoral presentada en la Universidad de Oviedo para la
obtención del título de Doctor por la Universidad de Oviedo**

Gijón, julio de 2017



RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: Evaluación y diseño de métodos efectivos de inserción de texto en entornos de TV Digital Interactiva.	Inglés: Evaluation and design of effective methods to insert text in Interactive Digital TV environments.
2.- Autor	
Nombre: Aurora Barrero López	DNI/Pasaporte/NIE: -A
Programa de Doctorado: Ingeniería Informática	
Órgano responsable: Departamento de Informática	

RESUMEN (en español)

Desde que el acceso a internet se ha convertido en un servicio al alcance de la mayoría de hogares, la televisión digital interactiva (IDTV) y sus aplicaciones están jugando un papel destacado en la sociedad actual. Las SmartTV están ocupado un lugar predominante como dispositivo multimedia de entretenimiento en el hogar. Este nuevo tipo de televisión permite que los usuarios accedan a servicios de una forma muy similar a como lo hacen con sus ordenadores y móviles.

Sin embargo, el principal dispositivo de interacción del que disponemos para interactuar con el televisor es el mando a distancia tradicional, que aunque es apropiado para realizar interacciones simples en el contexto de la IDTV, dado que no está diseñado para escribir texto, necesitaría una modernización para mejorar la experiencia de los usuarios que necesiten realizar tareas más avanzadas.

El objetivo principal de esta tesis es contribuir a determinar cuál es la combinación óptima de método y dispositivo de inserción de texto para aplicaciones de IDTV teniendo en cuenta diferentes perfiles de usuario y distintos contextos de uso.

En primer lugar, se desarrolló una metodología que nos permite establecer un proceso para realizar tests de usuario y pruebas subjetivas en el ámbito de la inserción de texto en IDTV, abarcando todas las etapas de una experimentación: desde la definición de métricas, pasando por el desarrollo de prototipos y realización de las pruebas hasta concluir con el análisis de resultados.

En segundo lugar, se realizó un amplio estudio sobre métodos de inserción de texto para aplicaciones de la IDTV utilizando como dispositivos de interacción diversos modelos de mandos a distancia, con diferente disposición de teclas, y con otros dispositivos alternativos como pueden ser teclados, dispositivos apuntadores o touchpads. Como métodos de escritura se utilizaron teclados virtuales (alfabético, genético, Qwerty) y métodos tipo móvil Multitap, T9 y 2-key. El entorno de pruebas se diseñó simulando un salón como el que el participante tendría en su propia casa. El estudio se realizó con 240 usuarios reales, con diferentes perfiles, rango de edad, hábitos y conocimiento en tecnología, que tuvieron que escribir una serie de frases seleccionadas previamente. Se analizó la velocidad de escritura, las tasas de error, los perfiles de aprendizaje, y las impresiones subjetivas, teniendo siempre en cuenta las características particulares de los usuarios.

En tercer lugar, en este estudio analizamos los errores cometidos por 82 usuarios en un experimento en el que escribieron 7,395 frases. Teniendo en cuenta las características particulares de los usuarios, se clasificaron los errores y analizaron sus principales causas. En cuanto al análisis de errores según el perfil de los participantes, las personas mayores tuvieron valores promedio más altos en casi todos los tipos de errores, sin embargo, las diferencias con el resto de los usuarios fueron estadísticamente significativas sólo en el caso de errores diagonal *proximity*.



En relación al análisis de mandos a distancia, los resultados muestran que cuando los usuarios tienen que introducir textos sencillos, el método con el mejor rendimiento es Multitap, a pesar de que la tasa de error es mayor que con los teclados virtuales. Sin embargo, al escribir textos complejos, los teclados virtuales presentan una velocidad de escritura igual o superior que Multitap con una tasa de error significativamente más baja. Además, cuando combinamos teclados virtuales con ciertos botones en el mando a distancia para acciones frecuentes mejoramos considerablemente el rendimiento. Cabe destacar que la forma y localización de las teclas en los mandos a distancia no afecta significativamente el rendimiento. Teniendo en cuenta estas consideraciones, si se quiere lograr un compromiso entre rapidez y precisión, el teclado virtual elegido sería el Genético, que además ha recogido muy buenas opiniones entre los participantes. El análisis de los dispositivos alternativos al mando a distancia indica que el teclado tradicional obtiene los mejores resultados en términos de velocidad de escritura. Sin embargo, el número de errores en condiciones reales es muy alto porque los usuarios tienen problemas al presionar más de una tecla a la vez mientras sostienen el teclado. Otro resultado interesante es que las impresiones subjetivas no siempre coinciden con los resultados empíricos: a pesar de la baja velocidad del touchpad, obtuvimos retroalimentación bastante positiva de más de la mitad de los usuarios que declararon explícitamente que se trataba de un dispositivo simple y fácil de usar.

Finalmente, usando los datos y resultados obtenidos en las experimentaciones anteriores, en este trabajo se presenta un sistema hardware de interacción, basado en un touchpad al que denominamos TVPad, que persigue mejorar la experiencia de los usuarios en entornos de IDTV pero conservando el confort del uso tradicional de la TV. Este sistema se ha pensado para alcanzar las capacidades de interacción avanzadas que los usuarios demandan para este tipo de aplicaciones, como son acciones de ampliación y reducción o la selección de elementos o texto en pantalla.

RESUMEN (en Inglés)

With internet access at home becoming a mainstream service, the interactive digital television (IDTV) and its applications are playing a leading role in today's society. SmartTVs are growing into an important multimedia device for home entertainment. This new type of television allows users to access services in a way similar to mobile phones or computers.

However, the main interaction device available to interact with the TV is still the traditional remote control which may be suitable for the IDTV context to carry out simple interactions, but since it was not designed to write text, it requires further development to improve the experience of the users who need to perform advanced tasks.

This Ph.D. Thesis is aimed at contributing to determine the optimal combination of method and device to introduce text in IDTV applications, taking into account different user profiles and scenarios of use.

Firstly, we developed a methodology that allowed us to establish a process to run users tests and subjective assessment in the field of text entry for IDTV, covering all the phases in the experimentation: from defining metrics to developing prototypes and carrying out tests, and ending with the analysis of the results.

Secondly, we carried out a wide research about text entry methods for IDTV applications, using as interaction devices a variety of remote control models with different key dispositions, and other alternative devices like keyboards, pointing devices or touchpads. As writing methods we use virtual keyboards (alphabetic, genetic, Qwerty) and mobile-type methods like Multitap, T9 and 2-key. The test environment was designed simulating a living room similar to that which subjects would have in their homes. The study was performed with 240 real users, with different profiles, age range, habits and tech knowledge. The users had to write a series of previously selected sentences. We analyzed writing speed, error rates, learning profile, subjective opinions, always keeping in mind user specific characteristics.



Thirdly, we analyzed in the study the errors committed by 82 users in an experiment in which they wrote 7,395 sentences. These errors were classified and their main causes analyzed, taking into account the particular characteristics of the users. Concerning the error analysis by participant profile, older people had on average higher rates in almost every type of error, although differences with the rest of the users were statistically significant only in the cases of diagonal proximity errors.

Regarding the analysis of remote controls, results show that when simple texts need to be written, the method with the best performance is Multitap, despite having a higher error rate than virtual keyboards. However, when writing complex texts, virtual keyboards present an equal or superior writing speed to Multitap with a notably lower error rate. It should be pointed out that performance is not significantly affected by the shape of the remote control or the position of the keys. However, when virtual keyboards and certain remote control buttons are combined to carry out frequent actions, performance is considerably improved. Bearing in mind these factors, if we want to achieve a compromise between speed and accuracy, we should select Genetic as the virtual keyboard, which also received very good opinions from participants.

The analysis of alternative devices to the remote control indicates that conventional keyboards obtain the best results in terms of speed writing. Nevertheless, the number of errors under real conditions is very high due to the fact that users have problems when pressing more than one key at the same time while holding the keyboard. Another result worth mentioning is that subjective opinions do not always match empirical results: regardless of the low speed of the touchpad, the feedback from more than half of the users was fairly positive, and they stated that the device was straightforward and easy to use.

Finally, in this thesis we present a hardware interaction system, based on a touchpad, which we have named TVPad. This device aims to improve user experience for IDTV context but preserving the comfort of traditional use of TV. This system was designed to achieve advanced interaction abilities that users demand for this type of applications, such as zoom in and out or screen text and element selection.

**SR. DIRECTOR DE DEPARTAMENTO DE INFORMÁTICA
SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN INGENIERÍA
INFORMÁTICA**

Agradecimientos

En primer lugar a David Melendi Palacio, mi director de tesis, por la dedicación y el apoyo que ha brindado a este proyecto, y cuyas ideas y sugerencias han sido imprescindibles para que este trabajo llegara a buen puerto.

A mis amigos, por estar ahí en todo momento para brindarme su consejo tanto profesional como personal. Y por enseñarme a afrontar los obstáculos con alegría disfrutando del tiempo libre que muchas veces ni siquiera tenía.

A mis padres y hermanos, por enseñarme que la perseverancia y el esfuerzo son el camino para lograr cualquier objetivo que me marque.

A todos los participantes de las experimentaciones, porque sin su tiempo, colaboración y enorme paciencia este proyecto no hubiera sido posible.

Y por último a todas las personas que han aportado su granito de arena, en particular a los integrantes del Grupo de Investigación DMMS de la Universidad de Oviedo cuya ayuda, apoyo y consejo han sido indispensables.

A todos.

GRACIAS.

Aurora Barrero López, julio 2017.

Resumen

Desde que el acceso a internet se ha convertido en un servicio al alcance de la mayoría de hogares, la televisión digital interactiva (IDTV) y sus aplicaciones están jugando un papel destacado en la sociedad actual. Las SmartTV están ocupando un lugar predominante como dispositivo multimedia de entretenimiento en el hogar. Este nuevo tipo de televisión permite que los usuarios accedan a servicios de una forma muy similar a como lo hacen con sus ordenadores y móviles.

Sin embargo, el principal dispositivo de interacción del que disponemos para interactuar con el televisor es el mando a distancia tradicional, que aunque es apropiado para realizar interacciones simples en el contexto de la IDTV, dado que no está diseñado para escribir texto, necesitaría una modernización para mejorar la experiencia de los usuarios que necesiten realizar tareas más avanzadas.

El objetivo principal de esta tesis es contribuir a determinar cuál es la combinación óptima de método y dispositivo de inserción de texto para aplicaciones de IDTV teniendo en cuenta diferentes perfiles de usuario y distintos contextos de uso.

En primer lugar, se desarrolló una metodología que nos permite establecer un proceso para realizar tests de usuario y pruebas subjetivas en el ámbito de la inserción de texto en IDTV, abarcando todas las etapas de una experimentación: desde la definición de métricas, pasando por el desarrollo de prototipos y realización de las pruebas hasta concluir con el análisis de resultados.

En segundo lugar, se realizó un amplio estudio sobre métodos de inserción de texto para aplicaciones de la IDTV utilizando como dispositivos de interacción diversos modelos de mandos a distancia, con diferente disposición de teclas, y con otros dispositivos alternativos como pueden ser teclados, dispositivos apuntadores o touchpads. Como métodos de escritura se utilizaron teclados virtuales (alfabético, genético, Qwerty) y métodos tipo móvil Multitap, T9 y 2-key. El entorno de pruebas se diseñó simulando un salón como el que el participante tendría en su propia casa. El estudio se realizó con 240 usuarios reales, con diferentes perfiles, rango de edad, hábitos y conocimiento en tecnología, que tuvieron que escribir una serie de frases seleccionadas previamente. Se analizó la velocidad de escritura, las tasas de error, los perfiles de aprendizaje, y las impresiones subjetivas, teniendo siempre en cuenta las características particulares de los usuarios.

En tercer lugar, en este estudio analizamos los errores cometidos por 82 usuarios en un experimento en el que escribieron 7,395 frases. Teniendo en cuenta las

características particulares de los usuarios, se clasificaron los errores y analizaron sus principales causas. En cuanto al análisis de errores según el perfil de los participantes, las personas mayores tuvieron valores promedio más altos en casi todos los tipos de errores, sin embargo, las diferencias con el resto de los usuarios fueron estadísticamente significativas sólo en el caso de errores diagonal *proximity*.

En relación al análisis de mandos a distancia, los resultados muestran que cuando los usuarios tienen que introducir textos sencillos, el método con el mejor rendimiento es Multitap, a pesar de que la tasa de error es mayor que con los teclados virtuales. Sin embargo, al escribir textos complejos, los teclados virtuales presentan una velocidad de escritura igual o superior que Multitap con una tasa de error significativamente más baja. Además, cuando combinamos teclados virtuales con ciertos botones en el mando a distancia para acciones frecuentes mejoramos considerablemente el rendimiento. Cabe destacar que la forma y localización de las teclas en los mandos a distancia no afecta significativamente el rendimiento. Teniendo en cuenta estas consideraciones, si se quiere lograr un compromiso entre rapidez y precisión, el teclado virtual elegido sería el Genético, que además ha recogido muy buenas opiniones entre los participantes. El análisis de los dispositivos alternativos al mando a distancia indica que el teclado tradicional obtiene los mejores resultados en términos de velocidad de escritura. Sin embargo, el número de errores en condiciones reales es muy alto porque los usuarios tienen problemas al presionar más de una tecla a la vez mientras sostienen el teclado. Otro resultado interesante es que las impresiones subjetivas no siempre coinciden con los resultados empíricos: a pesar de la baja velocidad del touchpad, obtuvimos retroalimentación bastante positiva de más de la mitad de los usuarios que declararon explícitamente que se trataba de un dispositivo simple y fácil de usar.

Finalmente, usando los datos y resultados obtenidos en las experimentaciones anteriores, en este trabajo se presenta un sistema hardware de interacción, basado en un touchpad al que denominamos TVPad, que persigue mejorar la experiencia de los usuarios en entornos de IDTV pero conservando el confort del uso tradicional de la TV. Este sistema se ha pensado para alcanzar las capacidades de interacción avanzadas que los usuarios demandan para este tipo de aplicaciones, como son acciones de ampliación y reducción o la selección de elementos o texto en pantalla.

Abstract

With internet access at home becoming a mainstream service, the interactive digital television (IDTV) and its applications are playing a leading role in today's society. SmartTVs are growing into an important multimedia device for home entertainment. This new type of television allows users to access services in a way similar to mobile phones or computers.

However, the main interaction device available to interact with the TV is still the traditional remote control which may be suitable for the IDTV context to carry out simple interactions, but since it was not designed to write text, it requires further development to improve the experience of the users who need to perform advanced tasks.

This Ph.D. Thesis is aimed at contributing to determine the optimal combination of method and device to introduce text in IDTV applications, taking into account different user profiles and scenarios of use.

Firstly, we developed a methodology that allowed us to establish a process to run users tests and subjective assessment in the field of text entry for IDTV, covering all the phases in the experimentation: from defining metrics to developing prototypes and carrying out tests, and ending with the analysis of the results.

Secondly, we carried out a wide research about text entry methods for IDTV applications, using as interaction devices a variety of remote control models with different key dispositions, and other alternative devices like keyboards, pointing devices or touchpads. As writing methods we use virtual keyboards (alphabetic, genetic, Qwerty) and mobile-type methods like Multitap, T9 and 2-key. The test environment was designed simulating a living room similar to that which subjects would have in their homes. The study was performed with 240 real users, with different profiles, age range, habits and tech knowledge. The users had to write a series of previously selected sentences. We analyzed writing speed, error rates, learning profile, subjective opinions, always keeping in mind user specific characteristics.

Thirdly, we analyzed in the study the errors committed by 82 users in an experiment in which they wrote 7,395 sentences. These errors were classified and their main causes analyzed, taking into account the particular characteristics of the users. Concerning the error analysis by participant profile, older people had on average higher rates in almost every type of error, although differences with the rest of the users were statistically significant only in the cases of diagonal proximity errors.

Regarding the analysis of remote controls, results show that when simple texts need to be written, the method with the best performance is Multitap, despite having a higher error rate than virtual keyboards. However, when writing complex texts, virtual keyboards present an equal or superior writing speed to Multitap with a notably lower error rate. It should be pointed out that performance is not significantly affected by the shape of the remote control or the position of the keys. However, when virtual keyboards and certain remote control buttons are combined to carry out frequent actions, performance is considerably improved. Bearing in mind these factors, if we want to achieve a compromise between speed and accuracy, we should select Genetic as the virtual keyboard, which also received very good opinions from participants.

The analysis of alternative devices to the remote control indicates that conventional keyboards obtain the best results in terms of speed writing. Nevertheless, the number of errors under real conditions is very high due to the fact that users have problems when pressing more than one key at the same time while holding the keyboard. Another result worth mentioning is that subjective opinions do not always match empirical results: regardless of the low speed of the touchpad, the feedback from more than half of the users was fairly positive, and they stated that the device was straightforward and easy to use.

Finally, in this thesis we present a hardware interaction system, based on a touchpad, which we have named TVPad. This device aims to improve user experience for IDTV context but preserving the comfort of traditional use of TV. This system was designed to achieve advanced interaction abilities that users demand for this type of applications, such as zoom in and out or screen text and element selection

INDICE

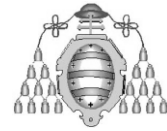
1	Introducción	1
2	Objetivos	5
3	Antecedentes	7
4	Discusión de resultados	17
4.1	Metodología de Test de Usuario y Pruebas Subjetivas	17
4.2	Experimentaciones	20
4.2.1	<i>Fases</i>	20
4.2.2	<i>Resultados de las experimentaciones</i>	28
4.3	Clasificación errores de escritura en IDTV	34
4.4	TVPad: Mecanismo efectivo de inserción de texto	39
4.4.1	<i>Diseño del dispositivo de interacción</i>	39
4.4.2	<i>Diseño del interfaz de usuario</i>	40
5	Conclusiones y Trabajo Futuro	43
5.1	Conclusiones	43
5.2	Trabajo Futuro	46
6	Compendio de publicaciones	49
6.1	An empirical investigation into text input methods for interactive digital television application	51
6.2	An Empirical Investigation Into Typing Errors in Interactive Digital Television Applications	75
6.3	A research on typing methods for interactive Digital Television Applications	93
6.4	Evaluation of Text Entry Methods for Interactive Digital Television Applications with Devices Alternative to Conventional Remote Controls	105
7	Informe sobre la calidad de las publicaciones	119
8	Otras Publicaciones de la Tesis	121
8.1	Metodología	123
8.2	Sistemas de introducción de texto en aplicaciones de TV interactiva	133
8.3	TVpad: Un novedoso sistema de interacción con la televisión	153
9	Bibliografía	165

FIGURAS

Figura 1 Fases y tareas de la metodología en notación BPMN	19
Figura 2 Multitap Keyboard	21
Figura 3 QWERTY Keyboard	21
Figura 4 Square Alphabetic Keyboard	21
Figura 5 Genetic Keyboard	21
Figura 6 Interfaz del Teclado Multitap-T9	22
Figura 7 Teclado virtual Qwerty caracteres en minúsculas	22
Figura 8 Teclado virtual Genético caracteres en mayúsculas	22
Figura 9 Teclado Virtual Genético de caracteres en minúsculas.....	23
Figura 10 Teclado virtual Qwerty y Genético caracteres especiales.....	23
Figura 11 Mandos a Firefly, Golden Interstar y Avermedia.Fase 4.	23
Figura 12 Mando Firefly. Teclas mayúsculas y caracteres especiales y teclas de espacio y borrar.	24
Figura 13 Teclas mayúsculas, caracteres especiales y teclas de espacio y borrar. Golden Interstar.....	24
Figura 14 Teclas mayúsculas, caracteres especiales y teclas de espacio y borrar. Avermedia.	25
Figura 15 Maxi Keyboard. Logitech.....	25
Figura 16 Mini Keyboard. Logitech	25
Figura 17 Modified Touchpad. Apple Magic Trackpad	26
Figura 18 Gyroscopic Remote Control. Fujitsu Air Command Plus. Entorno de Experimentación	26
Figura 19 Disposición real del experimento.	27
Figura 20 Interface de usuario de la aplicación utilizada en los experimentos.	27
Figura 21 Velocidad de escritura en las cinco primeras sesiones de la Fase 1.	29
Figura 22 Tasa de error en las cinco primeras sesiones de la Fase 1.	30
Figura 23 Apple Magic Trackpad y Fujitsu Air Command Plus.....	31
Figura 24 Velocidad de escritura en las tres sesiones de la Fase 5.....	32
Figura 25 Análisis y clasificación de los errores más comunes en las tres sesiones de la Fase 5.	33
Figura 26 Proporción de frases con errores versus frases sin errores.	35
Figura 27 Prototipo del dispositivo de interacción propuesto TVPad.....	40
Figura 28 Diseño conceptual del interfaz de usuario TVPad.	41

TABLAS

Tabla 1 Experimentación y fases	20
Tabla 2 Resumen de los resultados del análisis de errores	37

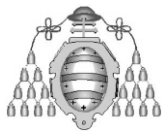


1 Introducción

En los últimos años la Televisión Digital Interactiva (IDTV) ha ido incrementando su popularidad hasta ocupar un papel importante en nuestra sociedad actual y evolucionar hacia lo que hoy conocemos como smartTV. Éste se ha convertido en un dispositivo destacado para el entretenimiento en el hogar [1] ya que aproxima la experiencia de usuario (UX) de los telespectadores a la de los usuarios de un ordenador conectado a Internet.

Debido a esta popularidad de la smartTV, el mercado de la IDTV está en auge ya que la mayoría de los proveedores de contenidos han apostado por una convergencia entre la WEB y la televisión (TV). Esta situación, unida al hecho de que el número de usuarios está en continuo crecimiento, ha provocado un clima de competitividad elevado entre los diferentes proveedores que ofrecen sus servicios multimedia interactivos mediante una smartTV o un set-top-box (STB). Este nuevo modelo de TV permite a un telespectador acceder a aplicaciones interactivas para buscar información, para fines de comunicación o educativos, o simplemente para entretenerse o usar una de las aplicaciones y servicios que se pueden instalar en la SmartTV [2]. Los telespectadores han pasado de tener un rol pasivo a tener un rol cada vez más activo, adquiriendo incluso la capacidad de crear, editar, compartir y controlar contenidos [3].

El dispositivo más común para interactuar con la TV sigue siendo el mando a distancia tradicional, aunque hay otros dispositivos disponibles en el mercado que pueden ser usados en este contexto. El mando a distancia es adecuado en el contexto IDTV para llevar a cabo interacciones simples como desplazarse por los menús utilizando un conjunto reducido de teclas. El problema surge cuando las acciones que quieren llevar a cabo los usuarios son más avanzadas y necesitan un tipo de interacción más compleja, como escribir usuarios y contraseñas de servicios, realizar búsquedas o interactuar con otros usuarios. Dado que los mandos a distancia convencionales no fueron diseñados para escribir texto y además existe una falta de estandarización [4] en los mismos, insertar texto en una SmartTV puede convertirse en una tarea relativamente difícil para algunos usuarios [5] y en una experiencia que definitivamente necesita mejorarse. Dado que la escritura de texto en aplicaciones de IDTV supone un problema real para los



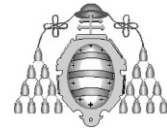
usuarios, nace la necesidad de encontrar métodos efectivos de escritura de texto para entornos IDTV.

En esta tesis, hemos llevado a cabo un estudio para evaluar los distintos métodos de inserción de texto con varios dispositivos entre los que se encuentran tanto los mandos a distancia convencionales que están presentes en los hogares, como los nuevos dispositivos que podríamos denominar “competidores” y que son de fácil acceso en el mercado, como teclados, touchpads y dispositivos giroscópicos.

Debido al elevado número de pruebas que fue necesario hacer para el estudio, decidimos desarrollar una metodología [6] que nos permitiera establecer un proceso más ordenado para realizar los tests de usuario y las pruebas subjetivas. En esta experimentación participaron un grupo heterogéneo de 240 usuarios reales, con distintos perfiles y diferente background IT, que tuvieron que escribir una serie de frases predefinidas en un entorno de pruebas que simulaba un salón de un hogar con el fin de que la experiencia fuera lo más realista posible.

Los métodos de inserción de texto que se utilizaron fueron 3 teclados virtuales: QWERTY, ALFABÉTICO y GENÉTICO, además de sus versiones optimizadas, y 3 mecanismos que provienen de los dispositivos móviles: Multitap, T9 y 2-key. Como dispositivos se utilizaron 3 mandos a distancia diferentes con distintas disposiciones de teclas, y como dispositivos competidores al mando a distancia se utilizaron un teclado de tamaño normal, un mini teclado, un dispositivo apuntador giroscópico tipo Wii y un touchpad modificado.

En los experimentos se analizaron las velocidades de entrada, las tasas de error, los perfiles de aprendizaje y las impresiones subjetivas, teniendo en cuenta las características particulares de los participantes. Se intentó responder a preguntas tales como qué método es el más rápido, cuál es el método con el que los usuarios cometen menos errores, cuál es el mejor dispositivo de interacción, cómo afecta la edad rendimiento, cómo mejora el rendimiento el uso de atajos o teclas rápidas, cómo un conjunto extendido de caracteres disminuye el rendimiento, cómo el diseño y disposición de teclas del mando a distancia afecta el rendimiento, cuál es la combinación de método y dispositivo óptima, cuál es la tipología de los errores más comunes en la inserción de texto en IDTV y cuáles son sus causas principales.



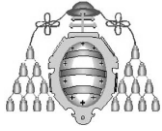
Nuestros resultados muestran, por ejemplo, que Multitap es un buen método para textos simples. Es entre 12% y 34% más rápido que el teclado virtual más rápido, dependiendo de la edad del usuario. Sin embargo, cuando los textos complejos necesitan ser escritos, los teclados virtuales presentan las mismas o incluso mejores velocidades de escritura (QWERTY es un 13% más rápido) y con tasas de error significativamente menores.

Respecto a los dispositivos competidores al mando a distancia los resultados muestran que el método más rápido es el teclado estándar, pero muchos usuarios informan de molestias y problemas bajo condiciones de poca iluminación. Además, los índices de error son considerablemente más altos con ambos teclados cuando es necesario utilizar las teclas modificadoras. Los resultados obtenidos con el mando giroscópico y el touchpad son similares. Sin embargo, mientras que los usuarios se quejan de los problemas de fatiga con el primero, sus comentarios sobre este último son muy positivos. También hemos observado que la edad es un factor importante que afecta el rendimiento de los usuarios con este tipo de dispositivos.

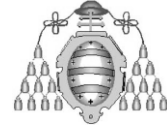
Los resultados muestran que los errores más frecuentes son errores de proximidad en todos los métodos evaluados. Aparecen, dependiendo del método de inserción de texto, entre el 13,75% y el 32,31% de las frases analizadas. Además, la edad y los background son aspectos importantes que afectan el número y el tipo de errores.

Finalmente, se ha diseñado un dispositivo touchpad denominado TVPad [7] que, según los resultados obtenidos, podría mejorar la experiencia de los usuarios de aplicaciones IDTV dado que mantiene una alta capacidad de interacción al mismo tiempo que proporciona el confort necesario para ver la TV en el salón de un hogar.

El problema de la introducción de texto en IDTV ha despertado el interés y ha aumentado el número de estudios que lo abordan en los últimos años. A diferencia de otros trabajos, en esta tesis hemos intentando realizar un estudio lo más realista posible, lo que significó seleccionar un grupo de usuarios con perfiles diversos y en un número significativo. Por otro lado, también nos esforzamos en replicar durante los experimentos el entorno que el usuario tendría en su casa. Esta tesis analiza la interacción humana con el entorno IDTV consiguiendo unos datos muy interesantes para los investigadores en este campo y para los diseñadores de aplicaciones IDTV.



Esta tesis se presenta como un compendio de publicaciones, estando organizada como sigue: En el capítulo 2 se plantean sus objetivos. El capítulo 3 presenta los antecedentes sobre las temáticas que vertebran esta tesis. Los resultados obtenidos y las contribuciones aportadas se describen en el capítulo 4. Las conclusiones y trabajo futuro se comentan en el capítulo 5. El capítulo 6 contiene los artículos que forman el compendio de publicaciones, mientras que el informe sobre su impacto se detalla en el capítulo 7, y cierra la exposición el capítulo 8, con las publicaciones adicionales presentadas en congresos.



2 Objetivos

En los últimos años, el estudio y comparación de métodos de inserción de texto es un campo que ha generado mucha investigación. Sin embargo, los avances más recientes se han centrado en dispositivos móviles como smartphones o tablets. Aunque algunas soluciones pueden ser aplicadas al ámbito de la IDTV, todavía hay muchas cuestiones por considerar.

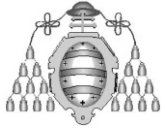
En primer lugar, el dispositivo más extendido es el mando a distancia que no ha sido diseñado para escribir texto [5], y que tiene una serie de características inherentes a su uso tradicional distintas a las de un teclado, lápiz o pantalla táctil. A esto hay que añadir que la actual falta de estandarización [4] en los mandos a distancia tampoco facilita esta tarea.

En segundo lugar, las teclas y la pantalla están en dispositivos separados, de modo que los usuarios no pueden mirar al dispositivo de interacción y a la pantalla al mismo tiempo.

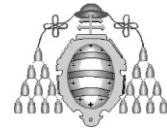
Y en tercer lugar, hay otros tipos dispositivos como los teclados externos o algunos dispositivos apuntadores, pero ni son muy populares ni su diseño los hace cómodos para su utilización desde el sofá.

El objetivo general de esta tesis es contribuir a determinar cuál es la combinación óptima de método y dispositivo de inserción de texto para aplicaciones de televisión digital interactiva teniendo en cuenta diferentes perfiles de usuario y distintos contextos de uso. Para alcanzar este objetivo proponemos los siguientes objetivos específicos:

1. Analizar métodos de inserción de texto aplicables a un mando a distancia, como pueden ser los teclados virtuales o los métodos heredados de los teclados de los teléfonos móviles.
2. Diseñar y aplicar una metodología para evaluar la usabilidad de los métodos de introducción de texto con usuarios reales.
3. Identificar y evaluar optimizaciones en los métodos de inserción de texto aplicables a un mando a distancia que permitan mejorar su rendimiento y/o experiencia de usuario.



4. Identificar entornos y analizar métodos específicos de IDTV , tales como la escritura de otros tipos de caracteres como los números, los símbolos de puntuación o otros especiales como "/" o "@" o ":".
5. Revisar el impacto del formato y disposición de las teclas en distintos mandos a distancia.
6. Identificar dispositivos de interacción competidores al mando a distancia como tabletas, móviles, mandos giroscópicos o teclados de dimensiones reducidas; y analizar los métodos de inserción de texto específicos de cada uno de ellos. Además, resulta de interés la comparación con el resto de métodos evaluados.
7. Identificar y evaluar optimizaciones en los métodos de inserción de texto para dispositivos competidores.
8. Diseñar métricas de rendimiento, de cara a realizar las evaluaciones y a disponer de un conjunto de datos que permita la consecuente extracción de conclusiones. Éstas tendrán que evaluar tanto aspectos subjetivos (facilidad, sencillez, comodidad, etc.) como aspectos objetivos (palabras por minuto, tasa de errores, etc.)
9. Diseñar e implementar plataformas de experimentación que, para los dispositivos y métodos planteados en la investigación, permitan interactuar con los usuarios que participen en los experimentos y capturar los datos que posteriormente serán analizados.
10. Planificar y ejecutar los test de usuario y pruebas subjetivas de las experimentaciones.
11. Realizar estudios de rendimiento según las métricas y plataformas diseñadas utilizando usuarios reales.
12. Detectar e identificar los errores más comunes en la inserción de texto para IDTV.
13. Cuantificar y clasificar errores detectados en la inserción de texto para IDTV.
14. Proponer mejoras para los diferentes métodos de inserción de texto compatibles con la IDTV.
15. Diseñar mecanismos efectivos de inserción de texto en entornos de Televisión Digital Interactiva.

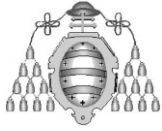


3 Antecedentes

El estudio de diferentes métodos de introducción de texto es un campo amplio donde se han realizado muchas investigaciones, aunque la mayoría de las obras se han centrado en dispositivos móviles [8]. Sin embargo, en la reciente literatura sobre *Human-Computer Interaction* (HCI) se puede encontrar una variedad de estudios con diferentes dispositivos y métodos de entrada de texto que permiten interactuar en el contexto de IDTV. Algunos de estos dispositivos son diferentes a los mandos a distancia tradicionales, como por ejemplo, teclados, pantallas táctiles, micrófonos o dispositivos apuntadores.

En el campo de los dispositivos móviles hay varios estudios sobre teclados virtuales con resultados que pueden ser aplicables parcialmente en servicios de IDTV. Los teclados virtuales son representaciones en pantalla de un teclado sobre el que el usuario se puede desplazar con un conjunto reducido de teclas, acompañadas de una tecla de confirmación.

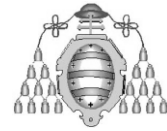
Bellman y MacKenzie [9] comparan dos tipos de teclados virtuales. El primero de ellos es el conocido QWERTY, mientras que el segundo es la disposición optimizada FOCL o *Fluctuating Optimal Character Layout*. Este método consiste en cambiar la posición de las letras en el teclado de forma que las letras con mayor probabilidad de ser seleccionadas tras una letra determinada se pintan junto a ésta. Para comparar estos dos métodos se realizó un estudio con once personas seleccionadas de entre los alumnos de los autores. Cada participante tenía que escribir frases con los métodos en diez sesiones de quince minutos. Las frases se seleccionaban de artículos publicados en Internet. Los resultados muestran que no había diferencias significativas entre los métodos, aunque los participantes reconocían que no podían utilizar de forma óptima los teclados FOCL. Por otro lado, Zhai et al [10] presentan un nuevo teclado virtual. Según los autores, las disposiciones de teclas existentes, y entre ellas la QWERTY, están pensadas para ser utilizadas con ambas manos de forma que no son adecuadas para dispositivos como teléfonos móviles o tabletas. Para incrementar la velocidad de escritura en estos dispositivos, los autores crearon la disposición de teclas Metrópolis, con la que llegan a obtener una velocidad de escritura un 40% más rápida que con QWERTY. Brewbaker [11] persigue el mismo objetivo, pero utilizando un algoritmo genético que le permite personalizar las dimensiones y el idioma del teclado. Los resultados obtenidos por



Brewbaker son similares a los de Zhai et al [10], pero por desgracia estos teclados no fueron probados con usuarios reales, quedando la duda de si serán de su agrado o no.

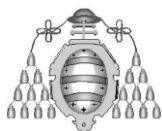
Además de los teclados virtuales, existen otros mecanismos de escritura en dispositivos móviles que podrían ser de aplicación en el campo de la IDTV. Por ejemplo, Silfverberg et al [12] comparan tres métodos de escritura basados en el teclado de los teléfonos móviles: el típico método de escritura de SMS conocido como multi-press (con tiempo límite y con botón siguiente), otro que llaman Two-Key, que consiste en que el usuario selecciona un grupo de caracteres mediante una pulsación y luego el carácter deseado con una segunda pulsación, y el método predictivo T9. La evaluación de estos métodos se realizó con 12 voluntarios con edades entre los 23 y los 47 años, casi todos usuarios frecuentes de teléfonos móviles. Los resultados muestran que el mejor rendimiento se obtiene con T9. De forma similar, Butts y Cockburn [13] repiten el experimento (sin considerar el T9) en una evaluación con 8 personas graduadas en Informática. En este caso los mejores resultados se obtienen con el multi-press con botón siguiente y los peores con el Two-Key. Por otro lado, MacKenzie et al [14] prueban otra técnica de desambiguación para la escritura de texto en teléfonos móviles basada en probabilidades de prefijos, que denominaron LetterWise. Esta técnica pretende ayudar a los usuarios proporcionándoles un mecanismo de desambiguación usando una base de datos con probabilidades de prefijos en vez de un mecanismo basado en un diccionario como el T9. Para comparar estos dos métodos, llevaron a cabo un estudio con veinte participantes. A cambio de una compensación económica, estos participantes llevaron a cabo veinte sesiones de entre 25 y 30 minutos en las que tenían que escribir frases de un conjunto preparado por los autores. Las velocidades de escritura fueron similares en la primera sesión, pero la última sesión mostró una diferencia significativa entre los métodos. No obstante, no pudieron establecer una diferencia significativa entre las tasas de error. MacKenzie y Tanaka-Ishii [15] presentan diferentes organizaciones de teclados y distintos métodos de desambiguación. Obtienen las mediciones características de teclados de teléfonos ambiguos y, de forma metodológica, explican cómo evaluar y comparar teclados utilizando el número de pulsaciones por carácter y por palabra (KSPC/KSPW) llegando a desarrollar un modelo.

Además de los teclados virtuales y los métodos basados en la utilización de un conjunto reducido de teclas, existen otras técnicas denominadas concurrentes que



persiguen combinar dos métodos de forma simultánea. Aunque hay varios estudios de esta naturaleza, los resultados de Widgor y Balakrishnan [16] son especialmente interesantes. Llevan a cabo un estudio comparativo sobre las mejoras en la velocidad de escritura mediante este tipo de técnicas. En este caso, añaden tres teclas adicionales a un teclado de teléfono móvil clásico, de forma que los usuarios tienen que combinarlas con una tecla numérica en caso de que deseen escribir un número o una letra. Realizaron un experimento con quince usuarios con experiencia en la utilización de teléfonos móviles y detectaron un incremento del rendimiento significativo. De forma similar, Oniszczyk y MacKenzie [17] presentan una técnica nueva llamada RollPad. Es un teclado de 12 teclas en las que las pulsaciones tienen que acompañarse de movimientos. Por ejemplo, para escribir la letra "a" el usuario tiene que presionar la tecla 2 y mover el dedo a la izquierda. Comparan este método con el tradicional método Multitap un experimento con 14 participantes con edades comprendidas entre los 18 y los 35 años. Aunque los resultados no muestran ninguna diferencia significativa en las velocidades y las tasas de error, el número de pulsaciones por carácter disminuye con RollPad, suponiendo un esfuerzo físico menor pero a costa de una mayor carga cognitiva.

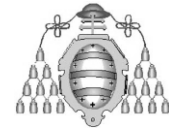
En los últimos años, los dispositivos de juego también han sido estudiados para comprobar si las características de sus periféricos podían influir sobre las velocidades de escritura. MacKenzie et al [18] comparan velocidades de escritura y tasas de error entre un Apple iPhone y una consola Nintendo Wii utilizando un teclado virtual en ambos casos. 16 personas con una media de edad cercana a los 25 años participaron en un experimento en el que el objetivo era introducir tres frases en varios bloques utilizando ambas técnicas. Cada sesión rondaba los 45 minutos de duración. Los resultados muestran una diferencia significativa entre las de escritura obteniendo mejores resultados con el iPhone (con 18,5 palabras por minuto frente a las 9,2 de la Wii). Los autores atribuyen esta diferencia a la gran concentración necesaria para escribir texto con la Wii en comparación con el iPhone. También encuentran una diferencia significativa entre las tasas de error, obteniendo más errores con el iPhone que con la Wii (7,7% frente a 2,8%). Esto se explica por el "problema del dedo gordo" que consiste en la dificultad de seleccionar la tecla adecuada cuando su tamaño o el espacio entre teclas es muy pequeño [19]. Esta línea fue seguida igualmente por otros autores como Jones et al. [20] o Castellucci y MacKenzie [21]. Por un lado, Jones et al. [20] evaluaron el rendimiento de un dispositivo Wii mote con dos técnicas de entrada basadas en acelerómetro: Tri-center



y Matrix. Participaron del estudio 15 usuarios de 19 a 28 años con experiencia en controladores basados en acelerómetros. La velocidad de entrada de texto de la interfaz de Matrix fue de 3,7 WPM y la interfaz de Tri-centter fue de 3,3 WPM. Por otro lado, MacKenzie et al [22] llevaron a cabo un estudio empírico con 16 participantes usando un iPhone de Apple y un Wii mote. Basados en un teclado virtual convencional, los autores compararon velocidades de entrada de texto y tasas de error. Los participantes introdujeron texto a 18,5 WPM con el iPhone y a 9,2 WPM con el Wii mote. Los resultados sugirieron que usar una pantalla táctil para ingresar texto era mucho más rápido que usar un puntero remoto como un mando Wii.

En el caso de los dispositivos touchpad, ha habido pocos estudios aplicados en el contexto de las aplicaciones IDTV. Choi et al. [23] realizaron un estudio comparando un touchpad "clickable" con un joystick "clickable" (un controlador de Xbox 360). Doce estudiantes universitarios participaron en la primera parte del estudio. Tenían una edad media de 25,4 y todos tenían algo de experiencia con los touchpads. En la segunda parte del estudio mejoraron el diseño del touchpad gracias a la retroalimentación obtenida en la primera parte. En la segunda parte, los participantes tuvieron que realizar tres tareas (gestos numéricos, selección de programas y entrada de texto) utilizando el dispositivo touchpad bajo dos condiciones diferentes (Hover-Tracking y No-Hover-Tracking). Solamente 11 de los 12 estudiantes iniciales participaron en esta parte del estudio. Los resultados concluyeron que no había diferencia en las velocidades de entrada de texto entre las condiciones y que la velocidad máxima que podían alcanzar era alrededor de 8,5 WPM. Más recientemente, Young y Mi choi [24], probaron un nuevo prototipo para la inserción de texto en SmartTV combinando la interacción entre un Touchpad y un teclado físico. Realizaron un estudio con 20 estudiantes universitarios para comparar este prototipo con un mando a distancia, un TouchPad y un teclado físico. Los participantes hicieron una tarea de inserción de texto y una tarea de edición de texto en cada uno de los 4 dispositivos. Los resultados indicaron que la combinación de un teclado virtual con el touchpad para insertar y editar texto puede conducir a una entrada y edición de texto más rápida.

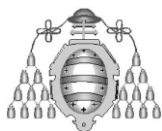
Por otro lado, Költringer et al [25] hicieron pruebas con la Microsoft Xbox 360. Compararon una nueva técnica llamada TwoStick con un teclado virtual QWERTY. Esta técnica se basa en una rejilla de 3x3 celdas en la que cada celda se subdivide en otras 9



celdas más pequeñas. El teclado final permite escribir hasta 56 caracteres diferentes incluyendo letras, números y caracteres especiales. Las teclas de espacio en blanco y borrado se han asignado en cada zona a las teclas centrales (izquierda y derecha). La tecla central de cada zona no se utiliza. Para moverse de una zona a otra los usuarios deben utilizar el joystick izquierdo y para moverse dentro de una zona deben utilizar el joystick derecho. En el experimento, 8 usuarios de edades entre los 22 y los 29 años tenían que completar 20 sesiones en las que escribían frases con cada método. En la sesión final el método TwoStick consigue ser significativamente más rápido que el QWERTY, con 14,87 WPM frente a 12,9. Incluso los resultados muestran diferencias significativas entre las tasas de error en la primera sesión, pero no en la última. En el caso del método TwoStick la causa más frecuente de error fue una temporización incorrecta de los movimientos de los joysticks. Esto ocurría frecuentemente cuando los usuarios empezaban a mover el joystick de la subzona (izquierdo) antes de situarse en la zona correcta (joystick derecho).

De forma similar, Wobbrock et al [26] diseñaron un método de escritura basado en un joystick llamado EdgeWrite. Este método utiliza gestos para seleccionar símbolos en un alfabeto. Los autores compararon este método con un teclado virtual con disposición Alfabética y con el DateStamp, ampliamente utilizado para escribir las iniciales en los juegos de arcade tradicionales. 18 usuarios con una edad media de 21,5 años participaron en los experimentos. Los resultados muestran que EdgeWrite fue al menos 1,5 veces más rápido que el teclado virtual y 2,4 veces más rápido que DateStamp.

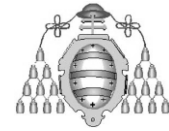
En el campo específico de los servicios de IDTV no hay muchos trabajos publicados relacionados con la escritura de texto. La primera referencia puede ser el trabajo de Iatrino y Modeo [5]. Estos autores identifican dos formas de usar un mando a distancia para escribir: utilizar los cursores y la tecla OK o usar un conjunto reducido de teclas como en los teléfonos móviles. Evalúan ambas posibilidades en un experimento grupal. Se hacen pruebas con tres métodos: Multipress (el método SMS), Multipress con una retroalimentación visual y un teclado virtual QWERTY. 36 personas participaron en el experimento teniendo que llevar a cabo dos tareas con cada método. La primera tarea consistía en escribir una dirección de correo electrónico, mientras que en la segunda había que escribir una frase corta en italiano. Los autores concluyen que el mejor método es el Multipress, destacando múltiples problemas relacionados con la internacionalización.



Ingmarsson [27] presenta una nueva técnica llamada TNT similar al ya comentado TwoStick [25]. El sistema se basa en una rejilla de 3x3 en la que cada celda se subdivide en otras 9 celdas menores. Cada una de las celdas menores tiene un carácter, alcanzando un total de 81 posibilidades. Para escribir un carácter el usuario debe seleccionar una de las celdas principales utilizando el teclado numérico del mando y luego seleccionar el carácter que desea escribir con una nueva pulsación. De esta forma, cada carácter es accesible con sólo dos pulsaciones. Cinco personas pagadas con edades entre los 27 y los 32 años probaron el sistema durante 10 sesiones de 45 minutos para escribir una novela corta en sueco. Los resultados muestran velocidades comparables o superiores a la escritura manual en una PDA o un método Multipress. Los usuarios destacaban su gran sencillez.

Por otra parte, Geleijnse et al [28] compararon tres técnicas para mandos a distancia (Multitap, T9 y un teclado virtual) con un teclado convencional QWERTY (con y sin autocompletado). En el experimento participaron 22 personas con edades entre los 21 y los 32 años. Tenían que buscar vídeos en Youtube escribiendo pares "artista-pista" con cada uno de los métodos. A pesar de que los autores no especifican la experiencia de los usuarios con las nuevas tecnologías, no es de sorprender que los resultados muestren una diferencia significativa entre las técnicas del mando a distancia y el teclado convencional. Lo que es sorprendente, es que los autores no encontraron ninguna diferencia entre los métodos utilizados con el mando a distancia. Los autores concluyeron que, bajo las condiciones del experimento, "las encuestas no han demostrado ningún indicio de que los usuarios no acepten un teclado en un salón". Esta conclusión contrasta con la de Orbist et al [29] que abogan por sistemas basados en voz. Los resultados de su estudio etnográfico indican que es poco realista centrarse en estudios dependientes de la popularidad de periféricos para televisión distintos de los convencionales. Un enfoque diferente al problema de escribir texto sería usar un teclado físico convencional. De hecho, Marshall, Foster y Jack [30] afirmaron que ésta era la mejor solución para la entrada de texto IDTV. Por otra parte, otros autores como Lee y Schmidt [31] argumentan que es mejor evitar trasladar el uso que se hace del teclado en el escritorio porque un teclado convencional es difícil de manejar en muchas posturas en el sofá.

Aoki et al. [32] diseñaron un método de entrada de texto para mandos a distancia de TV llamado Twist&Tap. El método utiliza una combinación de operaciones que son fáciles

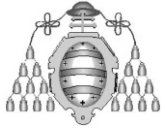


de recordar: Multitap de las cuatro teclas direccionales y cinco niveles de giro de la muñeca sosteniendo el mando a distancia en sentido horario o antihorario. Se evaluó a un conjunto de 14 mujeres diestras que fueron seleccionadas de una agencia de trabajo temporal. Su rango de edad estaba comprendido entre 24 y 33 años. Los experimentos confirman que Twist&Tap ofrece una mayor velocidad de entrada de texto que con el Multitap después de sólo 3 horas de uso del mando.

Gargi y Gossweiler [33] presentan un nuevo sistema predictivo diseñado para mejorar la velocidad de escritura en teclados virtuales: QuickSuggest. Esta técnica muestra un anillo que rodea al carácter seleccionado y muestra los cuatro caracteres que le siguen con más frecuencia. Cuando se escribe un carácter el anillo aparece y si el usuario iba a escribir uno de los caracteres que se sugieren solamente tiene que seleccionarlo y presionar OK. El cursor se mueve a la tecla correspondiente. Si el carácter deseado no es ninguno de los sugeridos, el usuario debe moverse a la posición del teclado en el que se encuentra el carácter deseado. Este método requiere un mínimo de dos pulsaciones por carácter. Los autores realizaron la evaluación en dos pasos: en primer lugar llevaron a cabo un estudio teórico del método y a continuación un estudio con 10 participantes. Los autores emplean la métrica de pulsaciones por carácter (KSPC) y calculan el tiempo requerido por los usuarios en la segunda fase del estudio.

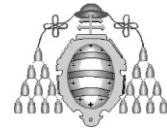
Sporka et al [34] comparan el ya comentado TNT [27] con un nuevo método llamado TwiceTap. Este método sigue la misma filosofía que TNT pero, además de permitir que los usuarios escriban caracteres sueltos, les permite escribir bloques de caracteres frecuentemente utilizados (n-gramas). 18 usuarios remunerados con una media de 22,7 años de edad participaron en los experimentos. Aunque no hubo diferencias significativas entre los métodos, los usuarios se decantaron por TwiceTap principalmente por su similitud con Multitap y por la posibilidad de seleccionar los n-gramas.

Por otro lado, no existen trabajos que realicen un estudio detallado de los tipos de errores de inserción de texto para el contexto específico de IDTV, simplemente comparan diferentes métodos de entrada teniendo en cuenta el número de errores y no su naturaleza específica. Existen algunos trabajos de investigación fuera del contexto de IDTV en los que se proporciona información detallada sobre los errores cometidos por los usuarios reales. Este puede ser el caso de MacKenzie y Soukoreff [8], quienes afirmaron que un análisis completo implica determinar qué tipo de errores ocurren y por



qué. Identificaron cuatro tipos básicos de errores: los usuarios introducen un carácter incorrecto (sustitución), los usuarios omiten un carácter (omisión), los usuarios añaden un carácter extra (inserción) o los usuarios intercambian caracteres vecinos (transposición). Sin embargo, el objetivo del artículo no fue incluir resultados empíricos. Además, Chen et al. [35] presentaron un estudio que investiga los errores de usuarios que acceden a través del móvil mientras escriben y apuntan al mismo tiempo. Por tanto las conclusiones de trabajos de investigación basados en experimentos con usuarios reales como Trewin y Pain [36], Chen et al. [35], o Henze et al. [37] no son válidos en este contexto.

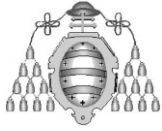
Finalmente, hay algunos estudios que tratan de resolver el problema de la escritura de texto utilizando otras técnicas. La mayoría de ellos se basan en reconocimiento de voz, superficies interactivas o gestos a mano alzada. Este es el caso de Vega-Oliveros et al. [38] que exploraron una entrada de texto multimodal en IDTV combinando el reconocimiento de voz, un teclado de celda y un teclado virtual. Además, Rick [39] estudió la eficiencia de la introducción de texto mediante el uso de superficies interactivas, como un tablero táctil, en el que los usuarios se mueven a través de las letras de una palabra en un teclado virtual sin levantar el dedo. Del mismo modo, Varcholik et al. [40] presentaron dos estudios en los que se utilizó un teclado QWERTY virtual en una plataforma multitáctil. Además, Ren y O'Neill [41] investigaron la entrada de texto usando gestos a mano alzada capturados con un sistema de sensores usando tres técnicas de selección y dos diseños: QWERTY y Dual-Circle. Más recientemente, Ren et al. [42] investigaron el diseño de la interacción gestual de mano alzada para TV interactiva basándose en dos tareas comunes para el usuario: selección de menú e inserción de texto. Estudiaron muchos factores de diseño, como la retroalimentación táctil / de audio, la profundidad de los gestos, el tamaño y la posición de los elementos de menú, así como el diseño de teclados virtuales y métodos de selección de teclas. Sus estudios sugieren un conjunto de hallazgos interesantes y directrices de diseño para la interacción gestual de mano alzada. Como se puede apreciar, la escritura en entornos de IDTV es un problema muy estudiado en los últimos años. Sin embargo, son pocos los estudios con resultados obtenidos de usuarios reales. Muchos trabajos presentan conclusiones basadas en los modelos de predicción de Fitts [43] u otros análisis teóricos. Su rendimiento con usuarios reales es desconocido. Algunos trabajos realizan evaluaciones con un conjunto reducido de usuarios o con características muy



homogéneas: generalmente jóvenes con formación técnica. Otros trabajos estudian los mismos o métodos similares pero presentan resultados contradictorios. Además, hay cuestiones importantes que no han sido estudiadas en profundidad, como el impacto de un conjunto complejo de símbolos o la forma y ubicación de las teclas del mando a distancia. Consideramos que la cuestión de escribir texto en aplicaciones de IDTV con un mando a distancia necesita un estudio en profundidad. En nuestra opinión, el estudio que llevamos a cabo arroja nuevos resultados. En las primeras fases, hemos analizado varios tipos de diseños de teclado virtual y otros métodos, algunos de ellos complementados con diferentes mejoras, como técnicas de desambiguación o sistemas de sugerencia. En estas fases, hemos realizado las evaluaciones utilizando frases generales y caracteres específicos y textos. Además, hemos evaluado cómo la forma y la ubicación de las teclas en los mandos a distancia afectan el rendimiento del usuario. La última característica importante en esta fase de nuestro estudio es el número de usuarios y las variaciones en sus características: edad, sexo, nivel de conocimiento y hábitos de uso de tecnología.

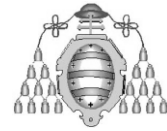
Por otro lado, hay destacar que otros autores han diseñado métodos a medida para solucionar el problema de la inserción de texto en IDTV, pero no están disponibles en el mercado o requieren un entrenamiento específico. Además, la mayoría de los autores presentan estudios con un número reducido de usuarios y normalmente de características homogéneas. Y para los tests que usan en el experimento normalmente usan frases cortas en inglés y no son necesarios símbolos complejos. Creemos que es necesario evaluar métodos de inserción de texto con otros dispositivos alternativos al mando a distancia y que puedan ser de fácil acceso a los usuarios en el entorno de su casa como son un teclado tradicional, un mini Teclado, un dispositivo giroscópico (mando tipo Wii) y un Touchpad. También, hemos realizado las experimentaciones con un número significativo de usuarios reales con características heterogéneas y con un background diferente. Además, hemos utilizado un entorno de pruebas equivalente al que un usuario tendría en su propia casa.

No existen en la literatura trabajos que realicen un estudio detallado de los tipos de errores de inserción de texto para el contexto específico de IDTV. Los existentes comparan diferentes métodos de entrada teniendo en cuenta el número de errores y no su naturaleza específica. Creemos que es importante conocer la tipología de los errores cometidos por los usuarios para encontrar soluciones efectivas para reducir o incluso



eliminar los errores tipográficos en este entorno y así mejorar la calidad de la experiencia de los usuarios. De ahí que otra de las aportaciones de esta tesis sea el análisis y tipificación de los errores más comunes en métodos de inserción de texto en IDTV.

Y por último, destacar que a veces los dispositivos de interacción con la TV necesitan capacidades de interacción avanzadas, además de la inserción de texto, como pueden ser la capacidad de: ejecución de acciones de ampliación y reducción, la selección de elementos en pantalla y/o de fragmentos de texto, el uso del portapapeles o el accionamiento de un control habilitado en pantalla. Por eso, como parte final de la investigación, se presenta TVPad, un dispositivo hardware de interacción con la televisión basado en un touchpad que persigue mejorar la experiencia de los usuarios en entornos IDTV.



4 Discusión de resultados

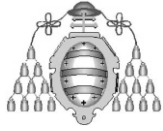
En este capítulo se presentan los resultados más destacados obtenidos durante la tesis, los cuales están divididos en 4 secciones. En primer lugar, en la sección 4.1 se describe la metodología desarrollada para establecer un proceso para realizar tests de usuario y pruebas subjetivas en el ámbito de la inserción de texto en IDTV. A continuación, en la sección 4.2 se detallan la evaluación y los resultados obtenidos de las 5 fases del estudio realizado con los distintos mecanismos de inserción de texto a usuarios reales con diferentes perfiles. En la sección 4.3 se realiza un análisis de las frases introducidas por los usuarios de las experimentaciones anteriores para identificar, cuantificar y clasificar los errores más comunes en la inserción de texto para IDTV. Por último, en la sección 4.4, se describe el diseño TVPad, un nuevo mecanismo de interacción elaborado a partir de los resultados obtenidos en las experimentaciones de la sección anterior.

4.1 Metodología de Test de Usuario y Pruebas Subjetivas

Aunque en los últimos años se ha producido una importante proliferación de tests de usuario y evaluaciones subjetivas, no existen muchos trabajos en los que se trate el tema específico de la escritura de texto en aplicaciones de IDTV. Esto es debido a la problemática que se genera a la hora de abordar, de manera eficiente, el elevado número de pruebas que es necesario hacer por cada tipo de experimentación, puesto que el tiempo y el coste de algunos recursos suele ser elevado. De ahí que haya surgido la necesidad de encontrar un proceso que facilite la ejecución de las tareas necesarias a la hora de realizar los tests, definir el entorno de pruebas, diseñar la interfaz de la aplicación y analizar los resultados obtenidos.

Como parte de este trabajo se presenta una metodología ágil e intuitiva, cuya aplicación será medir tanto la usabilidad como la satisfacción de los usuarios en métodos de inserción de texto en IDTV. Se tendrán en cuenta cuestiones importantes como el diseño de la interfaz, las características del medio de interacción a utilizar y el entorno.

Se establece un método que abarca todas las etapas de una experimentación, desde la definición de métricas y objetivos, al diseño de la misma, el desarrollo de prototipos y



la realización de las pruebas de usuario, finalizando con la recogida y validación de datos y el análisis de resultados.

Esta metodología está basada en una serie de fases, subdivididas a su vez en una serie de tareas. En el diagrama de la Figura 1 se muestra un flujo de trabajo formal de las fases y tareas propuestas en notación Business Process Model and Notacion (BPMN).

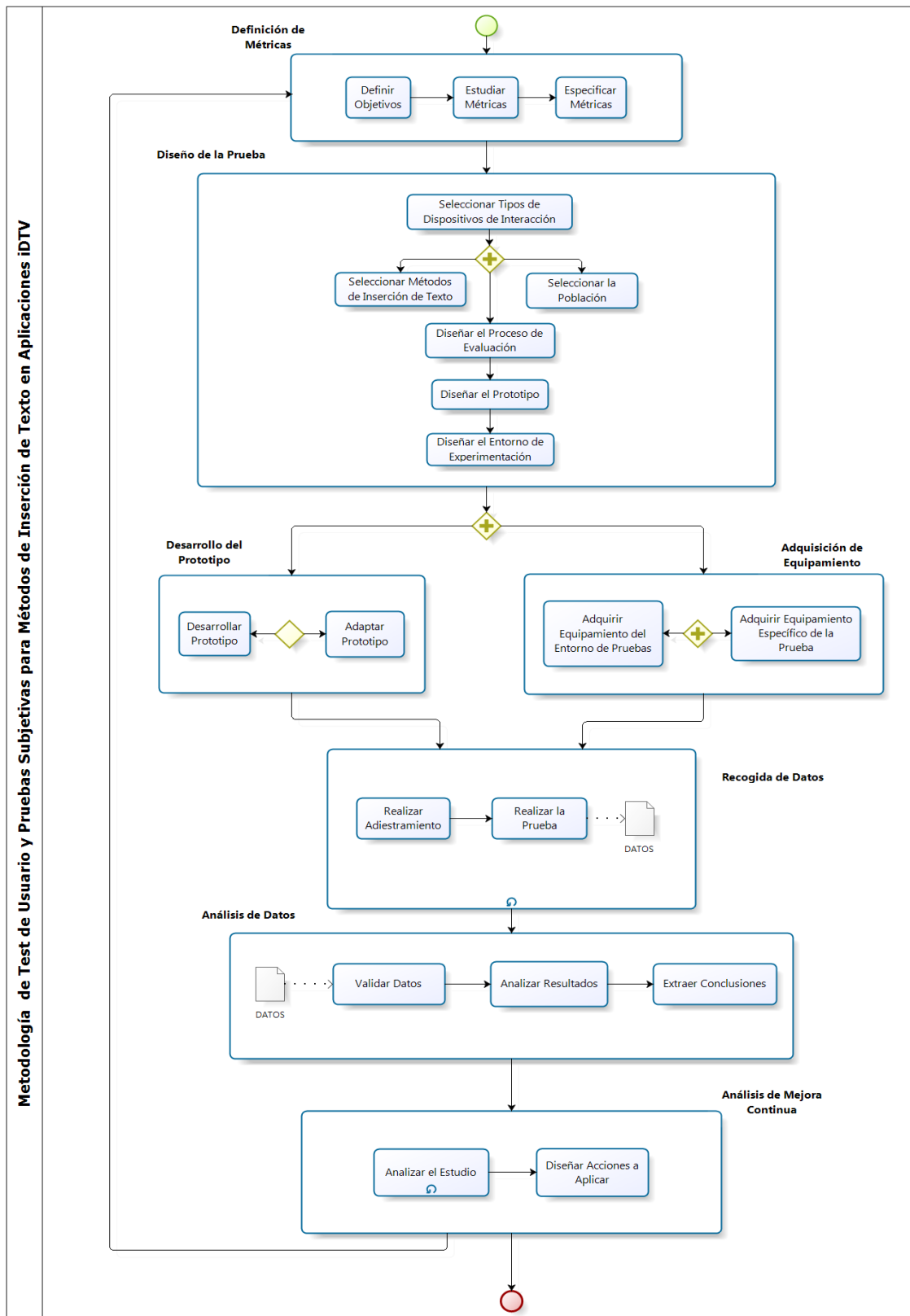
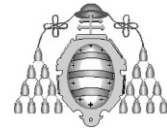
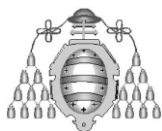


Figura 1 Fases y tareas de la metodología en notación BPMN



4.2 Experimentaciones

Para alcanzar los objetivos planteados en la tesis, se han realizado una serie de experimentaciones en las que se convocó a 240 personas que tenían que hacer pruebas de escritura de texto en una aplicación desarrollada al efecto.

4.2.1 Fases

La experimentación se estructuró en cinco fases, para cada una de las cuales, se aplicó la metodología diseñada, descrita en el apartado anterior, y que se recogen en la Tabla 1 que se muestra a continuación:

FASE	USUARIOS	MÉTODOS	DISPOSITIVO	SESIONES	JUEGO DE CARACTERES	TEXTO
1	57	QWERTY, ALFABÉTICO, GENÉTICO, MULTITAP	SNAPSTREAM FIREFLY	5	BÁSICO	5 FRASES DE CORPUS
2	48	QWERTY, GENÉTICO, T9, 2-KEY	SNAPSTREAM FIREFLY	2	BÁSICO	5 FRASES DE CORPUS
3	42	QWERTY, GENÉTICO Y MULTITAP MODIFICADOS	SNAPSTREAM FIREFLY	1	COMPLEJO	5 CAMPOS CON DATOS PERSONALES + TEXTO COMÚN
4	42	QWERTY, GENÉTICO Y MULTITAP MODIFICADOS	GOLDEN INTERSTAR, AVERMEDIA RMKS	1	COMPLEJO	5 CAMPOS CON DATOS PERSONALES + TEXTO COMÚN
5	52	QWERTY, GENÉTICO	TECLADOS LOGITECH K400 Y DÍNOVO MINI, APPLE MAGIC TRACKPAD Y FUJITSU AIR COMMAND PLUS	3	COMPLEJO	2 FRASES DE CORPUS Y 3 CAMPOS CON DATOS PROPUESTOS

Tabla 1 Experimentación y fases

4.2.1.1 Fase 1: Experimentación con métodos básicos de inserción de texto.

Experimentar con métodos populares y compatibles con mandos a distancia convencionales. En concreto, se utilizaron teclados virtuales con distintas disposiciones de teclas y el método Multitap (Figura 2). Las disposiciones para los teclados virtuales fueron QWERTY (Figura 3), Alfabética (Figura 4) y un teclado Genético (Figura 5) generado a partir del trabajo de Brewbaker [11].

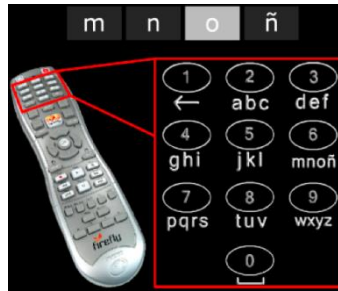
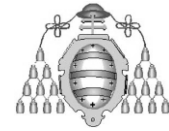


Figura 2 Multitap Keyboard



Figura 3 QWERTY Keyboard



Figura 4 Square Alphabetic Keyboard

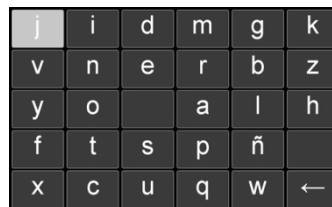


Figura 5 Genetic Keyboard

4.2.1.2 Fase 2: Experimentación con métodos optimizados.

Identificar y experimentar con optimizaciones de los métodos utilizados en la Fase 1. En base a los resultados y la retroalimentación de los usuarios de la Fase 1, se plantearon una serie de optimizaciones para los métodos empleados. Entre estas optimizaciones, está la utilización de algunos botones del mando a distancia para el borrado de caracteres y la escritura de espacios en blanco para los teclados virtuales. Igualmente, se eligieron como optimizaciones del Multitap de la fase 1, el T9 (Figura 6) y el 2-Key.

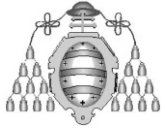


Figura 6 Interfaz del Teclado Multitap-T9

4.2.1.3 Fase 3: Experimentación en entornos específicos de IDTV.

En las fases anteriores se utilizaban textos sencillos, pero ahora se pretenden utilizar textos más complejos que reflejen situaciones reales de interacciones de usuarios con aplicaciones de IDTV, como introducir datos típicos como son el email, fecha de nacimiento, DNI, dirección postal o incluso introducir una URL. Por eso, para que el experimento fuese lo más real posible en esta fase, se usaron textos que incluían caracteres en minúsculas, caracteres en mayúsculas, acentos, tildes y caracteres especiales. Para ello, fue necesario incorporar estos símbolos en los métodos, tal y como se muestra en la Figura 7, Figura 8, Figura 9 y Figura 10.

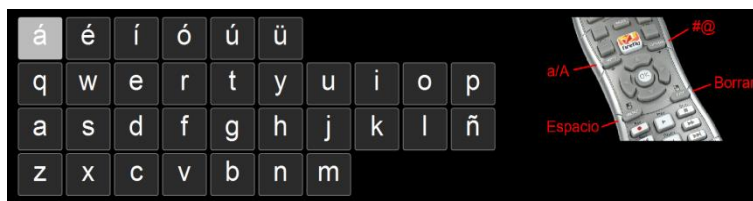


Figura 7 Teclado virtual Qwerty caracteres en minúsculas

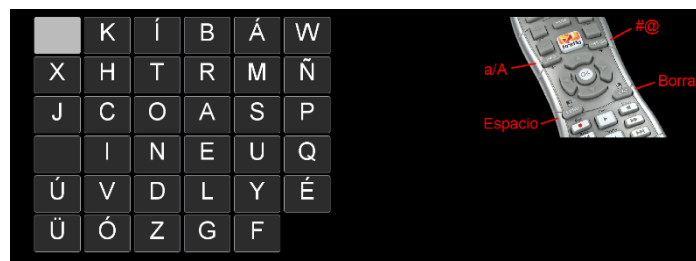


Figura 8 Teclado virtual Genético caracteres en mayúsculas

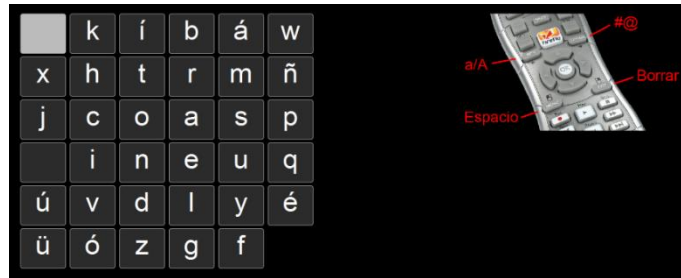
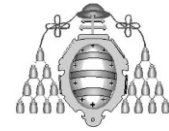


Figura 9 Teclado Virtual Genético de caracteres en minúsculas

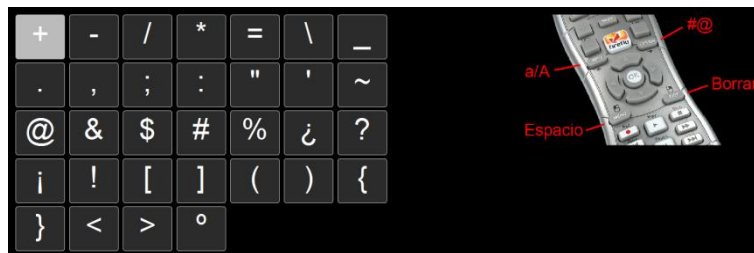


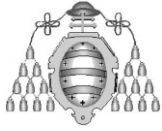
Figura 10 Teclado virtual Qwerty y Genético caracteres especiales

4.2.1.4 Fase 4: Experimentación con mandos heterogéneos.

Dado que no hay un modelo de mando a distancia estándar [4], es necesario evaluar el impacto que tiene la forma o la disposición de teclas que utiliza cada mando. Para comparar distintos modelos repetimos las pruebas de la Fase 3 con dos nuevos modelos de mando, para contrastar los resultados de esta fase con los de la fase anterior. Los mandos utilizados se muestran en la Figura 11.



Figura 11 Mandos a Firefly, Golden Interstar y Avermedia.Fase 4.



El mando a distancia del centro (Golden Interstar) de la Figura 11 tiene los números en la parte superior izquierda y las flechas y la tecla OK en el centro a la derecha. Los botones de colores son usados para las acciones especiales de cambiar al teclado en mayúsculas, al teclado de caracteres especiales, así como para borrar o introducir un espacio en blanco (Figura 13). Esto contrasta con el mando a distancia original usado en las fases anteriores ya que la ubicación de los botones es distinta, y las teclas utilizadas para las acciones especiales están alrededor de las teclas de flecha como podemos ver en la Figura 12.



Figura 12 Mando Firefly. Teclas mayúsculas y caracteres especiales y teclas de espacio y borrar.

El mando a distancia de la derecha de la Figura 11 tiene los números en la parte inferior y estos no tienen etiquetado de letras en ellos, las flechas y las teclas OK están en la parte superior y no dispone de las 4 teclas de color. Las teclas de las acciones especiales se encuentran alrededor de las flechas (Figura 14). Este mando contrasta de nuevo con los otros mandos a distancia en la forma (es más pequeño que los anteriores), en la disposición de las teclas y en el etiquetado de letras en las teclas numéricas

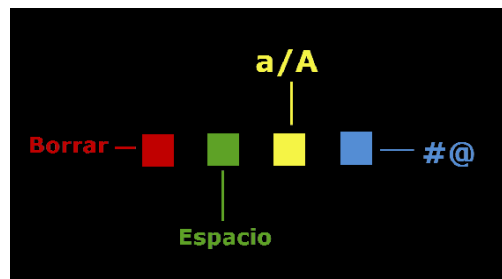


Figura 13 Teclas mayúsculas, caracteres especiales y teclas de espacio y borrar. Golden Interstar.

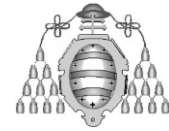


Figura 14 Teclas mayúsculas, caracteres especiales y teclas de espacio y borrar. Avermedia.

4.2.1.5 Fase 5: Experimentación con dispositivos competidores.

Se realizó una comparación de periféricos distintos del mando a distancia convencional. Por ello, en esta nueva fase la experimentación se realiza utilizando 4 nuevos periféricos. Se trata de un teclado convencional (Figura 15), un teclado de dimensiones reducidas (Figura 16), un touchpad (Figura 17) y un dispositivo apuntador similar al "wiimote" (

Figura 18). El touchpad y el dispositivo apuntador, mostrados en la Fig. 6, actúan sobre un teclado virtual con disposiciones QWERTY y Genética, a los que se han añadido símbolos complejos como sucedía en la Fase 3 de la investigación.



Figura 15 Maxi Keyboard. Logitech



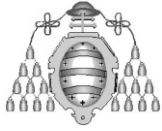


Figura 16 Mini Keyboard. Logitech



Figura 17 Modified Touchpad. Apple Magic Trackpad



Figura 18 Gyroscopic Remote Control. Fujitsu Air Command Plus. Entorno de Experimentación

Los experimentos de las 5 fases se realizaron en una sala intentando que la experiencia del usuario fuera lo más parecida posible a la del salón de su casa. En la sala se ubicó un sofá a una distancia de 2 metros de una televisión de 32 pulgadas (Figura 19). La televisión estaba conectada a un PC con Windows. Al PC se conectaban los dispositivos necesarios en cada prueba.

En el PC se ejecutaba una aplicación Adobe® AIR® diseñada para los experimentos. Tal y como se muestra en la Figura 20, la aplicación se ejecuta a pantalla completa con fondo negro. En función de la prueba, el texto a escribir se pinta en la parte superior. En la zona central un campo de texto muestra lo que escribe el usuario. La zona inferior se utiliza para pintar información de apoyo a la escritura.

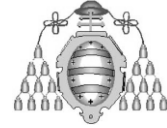


Figura 19 Disposición real del experimento.

La aplicación ejecutaba un ciclo de prueba con cada método de escritura. El orden de los métodos se alteraba para cada usuario para prevenir una posible influencia en los resultados. Dentro de cada ciclo el participante escribía una serie de cadenas que se le proponían o que correspondían a datos que él o ella conocía. Cuando se proponía una cadena se dejaban unos segundos para que el usuario la leyese. Esa iteración finalizaba cuando la cadena escrita por el usuario coincidía con la propuesta. Cuando no se proponía una cadena, la iteración finalizaba cuando el usuario indicaba que había terminado. No había límite de tiempo para escribir cada cadena. Entre cada iteración y cada ciclo se dejaban unos instantes para que el usuario se adaptase al cambio.

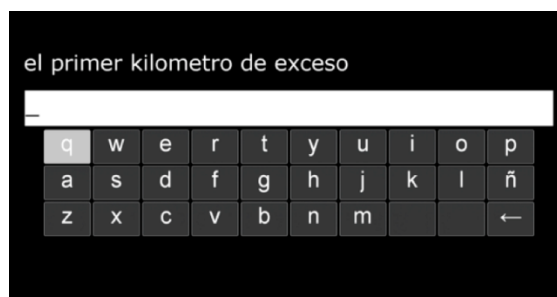
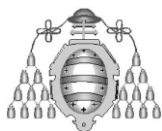


Figura 20 Interface de usuario de la aplicación utilizada en los experimentos.

Para proponer los textos, se creó un corpus de frases cortas extraídas de medios de comunicación populares en España, de forma similar a Bellman y MacKenzie [9].



La aplicación almacenaba datos en un fichero XML sobre cada uno de los eventos que registraba, incluyendo una marca de tiempo. Quedaba constancia de todas las pulsaciones que realizaban los usuarios, lo que permite calcular, entre otros:

- La velocidad de escritura obtenida con cada método.
- El número de errores producidos y su naturaleza, mediante el análisis de las correcciones realizadas.
- El aprendizaje con la experiencia, vista la evolución de la velocidad y de los errores.

Al final de la prueba se proporcionaban unos cuestionarios a los usuarios para obtener información subjetiva. Unas escalas Likert (de 0 a 4) [44] permitían al usuario valorar la facilidad de uso, la velocidad de escritura y la satisfacción general con cada uno de los métodos.

4.2.2 Resultados de las experimentaciones

4.2.2.1 Resultados de las Fases 1 a la 4

Tal y como se muestra en la Tabla 1, todos los experimentos de estas fases se han llevado a cabo con mandos convencionales.

En cuanto a la velocidad de escritura, tal y como se observa en la Figura 21, el método más rápido es Multitap con diferencias estadísticamente significativas con el resto ($p < 0,001$) [45]. De entre los teclados virtuales, el mejor es el Genético, pero sin apreciar diferencias significativas con el resto de métodos. Además se observa que los usuarios mejoran su rendimiento en general con la experiencia.

Al introducir las optimizaciones, se observa que el rendimiento mejora significativamente en QWERTY ($M=38,39$, $SD=6,75$) y T9 ($M=58,59$, $SD=18,81$). La mejora del Genético no es significativa ($M=39,84$, $SD=8,34$). En el caso particular del 2-key ($M=32,34$, $SD=8,39$), lo que se obtiene es una reducción en la velocidad estadísticamente significativa (37,17% peor que con Multitap). Los resultados de T9 y 2-Key siguen las pautas en [12] y [13].

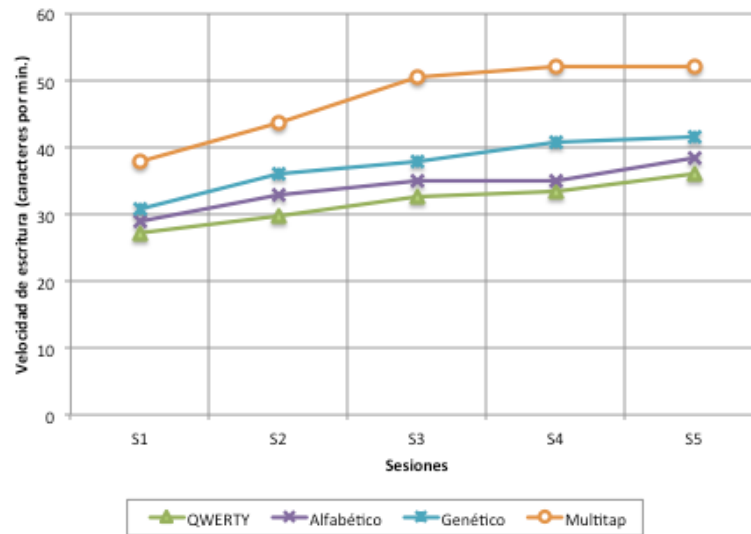
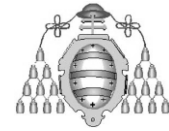
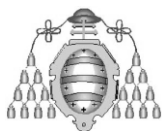


Figura 21 Velocidad de escritura en las cinco primeras sesiones de la Fase 1.

No obstante, cuando no se proponen textos a los usuarios y se introducen símbolos más complejos, los resultados cambian completamente (Fase 3 y Fase 4). En este caso el método más rápido es QWERTY ($M=30,66$, $SD=6,71$) seguido de Multitap ($M=27,07$, $SD=6,92$) y del Genético ($M=26,97$, $SD=6,49$), pero las diferencias entre métodos no son significativas. También se observa que el rendimiento es peor que el obtenido en fases anteriores con una disminución de hasta un 48,09% en el caso del método Multitap. Finalmente, en la comparación entre mandos no se obtienen resultados estadísticamente significativos, lo que nos lleva a concluir que la disposición de teclas y la forma del mando no afecta al rendimiento.

En lo que respecta a los errores de escritura, los resultados de la Fase 1 muestran que el método con la mayor tasa de errores es Multitap, tal y como se observa en la Figura 22. Este método se sitúa muy por encima de los teclados virtuales, presentando diferencias estadísticamente significativas ($p < 1e-05$). También se observa un notable descenso en el número de errores con la experiencia de los usuarios.

Con la introducción de las optimizaciones en la Fase 2, observamos que el método con más errores es T9 ($M=20,92$, $SD=18,7$), con diferencias significativas con el resto ($p < 0,001$). El Genético es el mejor ($M=1,63$, $SD=1,5$), sin diferencias significativas con QWERTY ($M=3,2$, $SD=2,31$) pero sí con 2-key ($M=7,41$, $SD=7,06$) ($p = 0,028$). Comparando estos resultados con los de la fase anterior, vemos algunas diferencias



importantes y estadísticamente significativas en el método T9. La tasa de error de los usuarios que participaron en ambas fases se incrementa un 204% en el T9 respecto al Multitap, mientras que si realizamos la comparación con usuarios que no participaron en la Fase 1, este empeoramiento es del 124,87%.

La incorporación de símbolos complejos junto a la exigencia de escribir textos no propuestos de la Fase 3 incrementa las tasas de error notablemente. La peor tasa se obtiene con Multitap ($M=16,86$, $SD=9,71$), que presenta diferencias significativas con el resto de métodos ($p < 0,001$). Le siguen el método Genético ($M=5,49$, $SD=5,65$) y QWERTY ($M=4,85$, $SD=3,51$). En cuanto a la comparación entre mandos, no se observan diferencias estadísticamente significativas, tal y como sucedía con la velocidad.

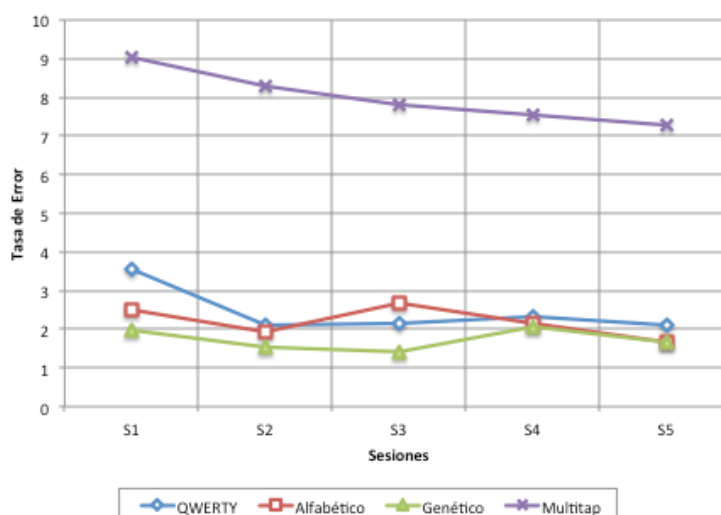
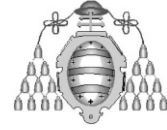


Figura 22 Tasa de error en las cinco primeras sesiones de la Fase 1.

La tipología de los errores es algo que también se ha estudiado en detalle [46]. Además del elevado número de errores, nuestros resultados muestran que en función del tipo de método utilizado hay determinados errores que se cometen más frecuentemente. Por ejemplo, Multitap presenta un número notable de ocasiones en las que el usuario tarda demasiado tiempo en detectar un error ($M=0,728$, $SD=0,415$), lo que claramente indica que el usuario mira al mando a distancia en lugar de a la pantalla. La mayoría de estos errores se producen por un número erróneo de pulsaciones al seleccionar el símbolo que se desea ($M=0,307$, $SD=0,296$). Por otro lado, los errores de proximidad son de los más frecuentes (selección de teclas/símbolos adyacentes). Estos



problemas van desde una media de 0,725 errores por frase en 2-Key (al seleccionar un grupo de símbolos) hasta 0,158 en el Alfabético (adyacencias horizontales o verticales).

4.2.2.2 Resultados de la Fase 5

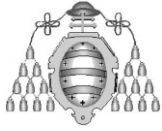
En las fases anteriores previas a la Fase 5 se hizo un estudio pormenorizado de distintos métodos que son susceptibles de ser utilizados en un entorno de IDTV, considerando que el mando a distancia es el dispositivo de referencia. No obstante, pese a las afirmaciones de otros autores [21], no es descartable que el mando tradicional sea paulatinamente sustituido por otros tipos de dispositivo.

Tal y como se muestra en la Tabla 1, en la Fase 5 se realizaron experimentos utilizando 4 nuevos periféricos: un teclado convencional, un teclado de dimensiones reducidas, un touchpad y un dispositivo apuntador similar al "wiimote". El touchpad y el dispositivo apuntador, mostrados en la Figura 23, actúan sobre un teclado virtual con disposiciones QWERTY y Genética, a los que se han añadido símbolos complejos como sucedía en la Fase 3 de la investigación.



Figura 23 Apple Magic Trackpad y Fujitsu Air Command Plus.

Si observamos los resultados obtenidos que se muestran en la Figura 24, el método más rápido es el teclado convencional en todas las sesiones de esta fase de la investigación, ya que su velocidad es muy superior a la del resto de métodos. En el otro extremo se situaría el dispositivo apuntador, que obtiene los peores resultados con la disposición Genética También se observa una disminución en la velocidad en la tercera sesión, debido a la introducción de símbolos complejos. La caída más pronunciada se



produce en los teclados físicos con una disminución promedio del 48,14%. No obstante, se observa que esta caída no es tan pronunciada en el resto de métodos, con una disminución media del 20,6%.

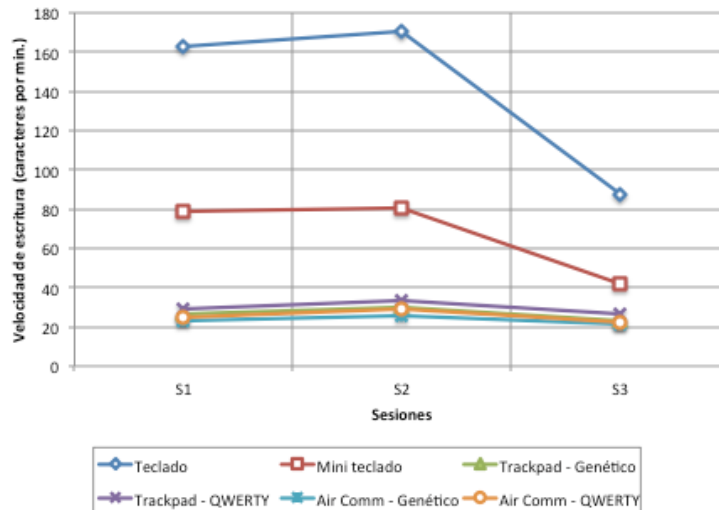
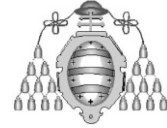


Figura 24 Velocidad de escritura en las tres sesiones de la Fase 5.

El problema de la velocidad en los teclados físicos se debe a la necesidad de pulsar más de una tecla simultáneamente para escribir determinados símbolos, sin disponer de una superficie sobre la que apoyar el teclado. Esto quedó también reflejado en las encuestas de los usuarios en las que el 34,66% considera que este método es incómodo para usar en el contexto de visionado de televisión. En el caso del touchpad y el dispositivo apuntador, los mejores resultados se obtienen con el touchpad y la disposición QWERTY, con una velocidad media de 33,285 caracteres por minuto sin símbolos complejos y una media de 26,597 cuando se utilizan estos símbolos.

En cuanto a los errores, en la Figura 25 se observa que la tasa de error disminuye en la segunda sesión, salvo en el caso de los teclados físicos. Es decir, con un poco de experiencia los usuarios mejoran con el touchpad y el dispositivo apuntador pero no con los teclados físicos. Cuando se introducen los símbolos complejos en la tercera sesión, la tasa de error crece en general, pero de forma más pronunciada en los teclados físicos. A la vista de los resultados, podemos afirmar que el touchpad es el mejor método desde el punto de vista de la tasa de error. Sin símbolos complejos los mejores resultados se



obtienen con la disposición Genética con una tasa promedio del 6,093%. Cuando los símbolos son complejos, QWERTY es la mejor disposición con una tasa del 11,109%.

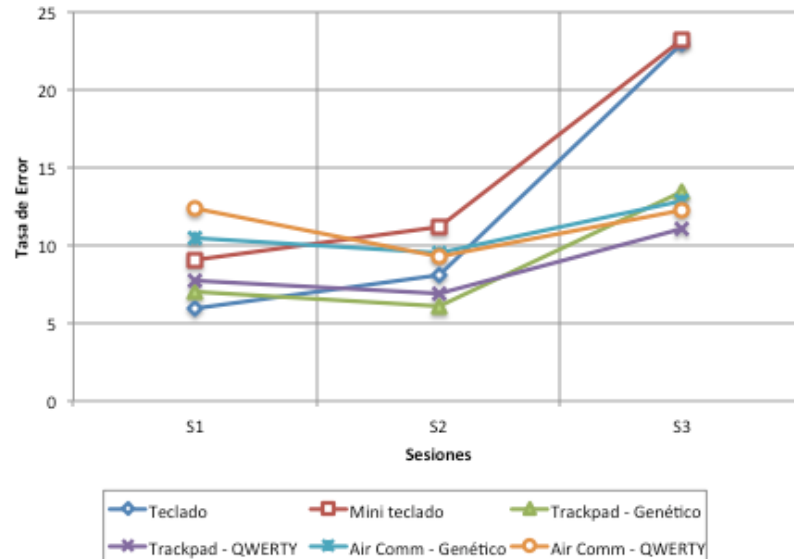
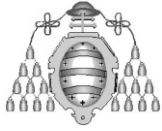


Figura 25 Análisis y clasificación de los errores más comunes en las tres sesiones de la Fase 5.

4.2.2.3 Resultados de las valoraciones subjetivas

También, hemos considerado relevantes las valoraciones subjetivas de los usuarios del experimento. Aquí trasladamos algunos comentarios:

- De forma generalizada los usuarios consideran que el método más rápido de escritura de texto es el teclado convencional. No obstante, un 36% de los usuarios indicaron explícitamente que era incómodo de utilizar en el contexto de visionado de la televisión. Otros usuarios proporcionaron comentarios sobre la falta de iluminación, que dificulta ver las teclas del teclado. Finalmente, algunos usuarios que no utilizan el ordenador frecuentemente indicaron problemas a la hora de escribir símbolos que necesitan la pulsación simultánea de varias teclas (un 18% de los usuarios).
- Los teclados de reducidas dimensiones presentaron problemas adicionales a los teclados convencionales debido a su tamaño. Un 45% de los usuarios



proporcionaron comentarios relativos a las dimensiones del teclado, la reorganización de las teclas y/o la dimensión de cada tecla.

- Un 46% de los usuarios indicaron que habían sufrido problemas de fatiga durante la prueba. Todos ellos coinciden en el cansancio que produce la utilización del dispositivo de tipo *wiimote*.
- Un 56% de los usuarios indica explícitamente que el *touchpad* es un dispositivo muy sencillo y cómodo de utilizar, llegando a afirmar que “es el mejor método para personas que no saben mecanografía” o que “es el mejor método de todos como mando”

Basándonos en las pruebas que hemos realizado, creemos que un touchpad puede ser un dispositivo totalmente adecuado para interactuar con una televisión. Obviamente, no es el mejor dispositivo para escribir texto, pero consideramos que es adecuado para reemplazar al mando a distancia convencional.

4.3 Clasificación errores de escritura en IDTV

Con el objetivo de detectar e identificar los errores más comunes en la inserción de texto para IDTV se analizaron los datos recogidos en las Fases de la 1 a la 4, durante las que se utilizaron varios métodos de inserción de texto compatibles con aplicaciones de la IDTV.

Dichos datos corresponden a los errores cometidos por 82 usuarios que escribieron 7.395 frases.

Debido a la organización de los experimentos y otros problemas con los usuarios, con cada uno de los métodos se escribió un número diferente de oraciones. Se escribieron 2.017 oraciones con QWERTY, 1.397 con el Alfabético, 2.001 con el Genético, 1.375 con el Multitap y 605 con 2-key.

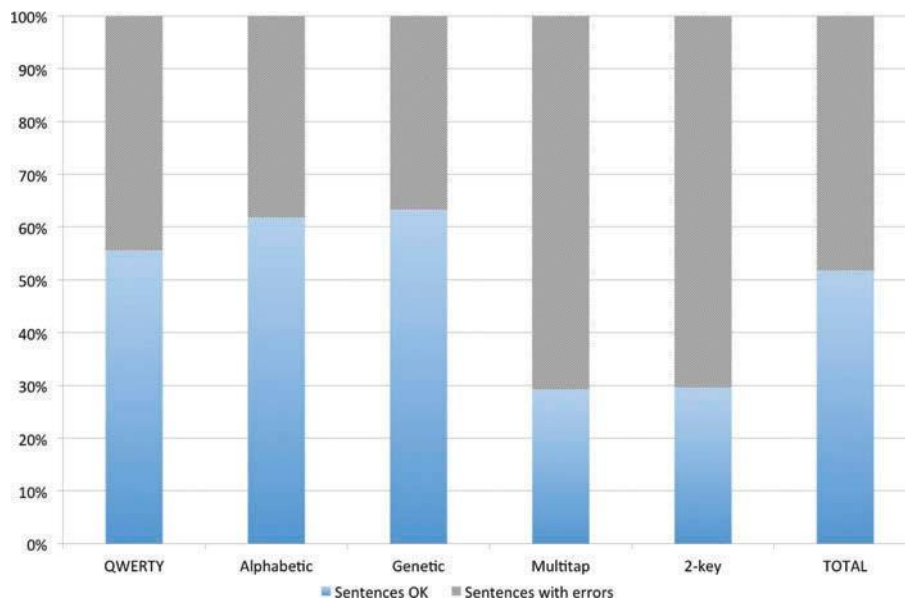
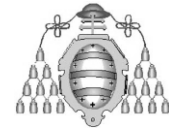
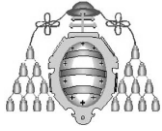


Figura 26 Proporción de frases con errores versus frases sin errores.

Consideramos que se produce un error cada vez que un usuario debe eliminar algo. Teniendo en cuenta esta premisa, el 97,26% de los usuarios cometieron al menos un error y el promedio de errores por frase es cercano a 1 ($M = 0,961$, $SD = 0,443$). Esto demuestra que escribir con un mando a distancia en una televisión puede ser considerado como una tarea difícil. Como se muestra en la Figura 26, los usuarios cometen más errores con los métodos diseñados para móviles (Multitap y 2-key) que con los teclados virtuales. De hecho, con Multitap o 2-key, el 98,63% de los usuarios cometieron al menos un error y el promedio de errores por frase es alto ($M = 1,889$, $SD = 1,191$ para Multitap y $M = 2,459$, $SD = 1,389$ para 2-Key).

Con los métodos de tipo móvil se detectó al menos un error en el 70% de las frases. La situación mejora notablemente con los teclados de tipo virtual. Aunque el 95,94% de los usuarios del estudio cometieron al menos un error, el promedio de errores por frase fue mucho menor que en el caso anterior ($M = 0,619$, $SD = 0,357$ para QWERTY, $M = 0,575$, $SD = 0,343$ para el genético y $M = 0,516$, $SD = 0,327$ para el Alfabético). Si comparamos los resultados obtenidos con cada uno de los métodos, pueden existir diferencias significativas según el test ANOVA de un factor (*one-way ANOVA*), $F(4, 370) = 74,793$, $p < 0,05$. El test post hoc Tukey con un intervalo de confianza de 95% confirma que existen diferencias significativas entre varios métodos. 2-key es el método con el



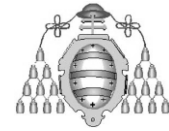
mayor número de errores por frase ($p < 0,008$). Multitap es mejor que el método 2-key ($p = 0,008$) pero peor que el resto ($p < 0,05$). Finalmente, QWERTY es ligeramente peor que el método Alfabético ($p = 0,043$).

Durante el análisis de las frases con errores, hemos identificado los siguientes tipos de error:

- **Missed symbols.** Al escribir una frase, hemos identificado que muchos usuarios se olvidan de escribir un símbolo y siguen escribiendo los símbolos que van a continuación. Hemos considerado los espacios en blanco de forma separada del resto de los símbolos porque se trata de una situación especial: después de escribir una palabra, los usuarios empiezan a escribir la que sigue inmediatamente sin introducir el espacio en blanco.
- **Mistaken deletions.** A veces los usuarios eliminan un carácter correcto y tienen que escribirlo de nuevo.
- **Late identified.** Hemos identificado que una cantidad considerable de errores se producen porque los usuarios continúan escribiendo después de que se ha cometido un error, y una vez que se dan cuenta del error, tienen que borrar muchos símbolos que eran correctos.
- **Bounce.** Estos errores se producen cuando un símbolo se escribe más veces de lo necesario.
- **Transposition.** Estos errores ocurren cuando dos símbolos adyacentes se escriben en orden inverso.
- **Proximity.** Estos errores se producen cuando se escribe un símbolo incorrecto debido a algún tipo de proximidad con el símbolo correcto.

Todos estos tipos de error son comunes a todos los métodos de inserción de texto antes mencionados. Pero también hay otros errores que son específicos de algunos métodos y se enumeran a continuación:

- **Wrong finger bounces with Multitap.** En Multitap, si se pulsa una tecla determinada más o menos veces de lo necesario, se escribe un símbolo incorrecto. Tenemos en cuenta estos errores sólo cuando se escribe un símbolo incorrecto de la tecla correcta.



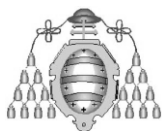
- **Wrong group selection for 2-key.** Éstos se producen cuando un usuario selecciona un grupo incorrecto de símbolos en 2-Key.
- **Wrong symbol selection for 2-key.** Éstos se producen cuando un usuario selecciona el grupo correcto de símbolos pero un carácter incorrecto.

Hemos seguido parcialmente la semántica en Trewin y Pain [36] para dar nombres a estos errores. Su definición se ha adaptado al contexto del experimento.

% USUARIOS CON ERRORES	ERRORES POR FRASE	DEPENDENCIAS	
MISSED BLANKS	90.41	$M = 0.074, SD = 0.064$	METHOD
MISSED SYMBOLS	97.26	$M = 0.110, SD = 0.064$	METHOD, BACKGROUND
MISTAKEN DELETIONS	79.45	$M = 0.025, SD = 0.030$	METHOD
LATE IDENTIFICATION	95.89	$M = 0.214, SD = 0.118$	METHOD
BOUNCE ERRORS	89.00	$M = 0.049, SD = 0.036$	METHOD
TRANSPOSITION ERRORS	13.69	$M = 0.001, SD = 0.005$	—
PROXIMITY ERRORS (VIRTUAL KEYBOARDS)	97.26	$M = 0.158, SD = 0.086$	AGE, BACKGROUND
PROXIMITY ERRORS (MULTITAP)	95.00	$M = 0.418, SD = 0.489$	NONE
WRONG FINGER BOUNCES (MULTITAP)	88.33	$M = 0.307, SD = 0.296$	BACKGROUND
WRONG GROUP	92.18	$M = 0.488, SD = 0.390$	BACKGROUND
WRONG SYMBOL	93.75	$M = 0.460, SD = 0.300$	BACKGROUND

Tabla 2 Resumen de los resultados del análisis de errores

Como puede verse en la Tabla 2, los métodos de tipo móvil generalmente producen un mayor número de errores que los teclados virtuales. [5] Iatrino y Modeo afirman que si comparamos televisores con dispositivos móviles o superficies táctiles, uno de los principales problemas de la TV es el hecho de que las teclas y la pantalla están en diferentes dispositivos, lo que hace imposible para los usuarios ver el dispositivo de interacción y la pantalla al mismo tiempo. Hemos sido capaces de demostrar que esto sigue siendo cierto. El notable número de errores *Late identification* de Multitap ($M = 0,728, SD = 0,415$) es claramente una consecuencia del usuario que mira el mando a distancia mientras escribe, en lugar de mirar al televisor. Cuando cometen un error, son conscientes de ello demasiado tarde, por lo que tienen que borrar muchos símbolos correctos para volver al lugar donde se cometió el error. Esto no sucede con los otros métodos, porque los usuarios tienen que mirar a la pantalla del televisor. Incluso con 2-key, los usuarios tienen que mirar la pantalla para hacer una selección.



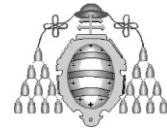
Una posible mejora podría ser incorporar algún tipo de feedback de entrada que permita al usuario saber lo que está escribiendo sin mirar a la televisión [47]. Por ejemplo, una opción puede ser que el mando a distancia o la televisión informe al usuario del símbolo escrito cada vez que escribe algo. Otra técnica menos disruptiva puede ser una pantalla en el mando a distancia. En general, esto es aplicable para métodos en los que los símbolos disponibles no están en la pantalla (por ejemplo, cuando se controla la televisión con un smartphone).

Hemos observado que los textos escritos tanto Multitap y 2-key provocan un número significativo de errores tipo *proximity*, alcanzando un promedio de 0,725 errores por frase al seleccionar un grupo en el método 2-key. Este problema se produce principalmente debido al tamaño de las teclas. Los mandos a distancia actuales, debido a su tamaño, no cumplen con el tamaño de tecla eficaz e ideal de aproximadamente 18 mm cuadrados [48]. El problema antes mencionado de tener las teclas y la pantalla en diferentes dispositivos significa que si el usuario está viendo la televisión en lugar de mirar el mando a distancia, es muy fácil presionar una tecla equivocada porque están muy juntas. Este problema se puede minimizar aumentando la distancia entre los botones del mando a distancia, como indican Siek et al. [19].

En el caso de teclados virtuales, el número de errores *late identification* es mucho menor. Esto es una consecuencia de que los usuarios miran la pantalla todo el tiempo. Los usuarios no necesitan mirar el mando a distancia porque el conjunto de teclas es reducido y su posición hace que estos métodos sean muy intuitivos. Si cometen un error, son conscientes de ello inmediatamente. Sin embargo, los errores más frecuentes en los teclados virtuales son *proximity* (el usuario elige un símbolo adyacente al correcto), generalmente como consecuencia de querer escribir más rápido. La situación se puede mejorar mediante el uso de delays en las pulsaciones de las teclas.

En cuanto a la edad de los usuarios, las personas mayores tuvieron valores promedio más altos en casi todos los tipos de errores. Las diferencias con el resto de los usuarios fueron estadísticamente significativas sólo en el caso de errores *diagonal proximity*.

Por último, también cabe destacar que no hemos encontrado ninguna diferencia estadísticamente significativa entre los usuarios masculinos y femeninos o entre usuarios zurdos y diestros.



4.4 TVPad: Mecanismo efectivo de inserción de texto

El diseño de un mecanismo de inserción de texto efectivo para IDTV se sustenta sobre la idea de que, en general, el mecanismo más versátil y sencillo de utilizar para interactuar con la televisión es mediante algún tipo de dispositivo apuntador. No obstante, la utilización de un dispositivo apuntador para controlar la televisión necesita un cambio en el interface que se muestra a los usuarios en pantalla. Por ello, en esta sección se hace una propuesta tanto del interface a mostrar en pantalla, como del dispositivo de interacción en sí mismo.

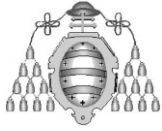
Con los resultados de las experimentaciones de las fases anteriores se realizó el diseño de un dispositivo de interacción llamado TVPad [7]. Se diseñó un mecanismo efectivo de inserción de texto en IDTV, es decir, diseñamos la combinación óptima de método y dispositivo que conjuga la eficiencia en la escritura con su uso cómodo por parte de un televidente.

4.4.1 Diseño del dispositivo de interacción

De los dispositivos de interacción apuntadores disponibles en el mercado, hemos seleccionado el touchpad como base para nuestro diseño. Además de los resultados empíricos de los experimentos realizados en las fases anteriores [45] [49], se han tenido en cuenta otras consideraciones para tomar esta decisión. Por ejemplo, el hecho de que un 56% de los usuarios de la fase 5 considere explícitamente que el touchpad es un dispositivo muy sencillo y cómodo de utilizar, llegando a afirmar que "es el mejor método para personas que no saben mecanografía" o que "es el mejor método de todos como mando"[50]. Por todo lo anterior, consideramos que es adecuado para reemplazar al mando a distancia convencional.

El TVPad tendría las siguientes características:

- Dimensiones similares a las del Magic Trackpad de Apple y desarrollado de forma que pueda ser utilizado sin necesidad de un apoyo (Figura 17).



- Tal y como se observa en el prototipo de la Figura 27, la mayoría de la superficie se destinaría a la interacción táctil.
- Permitirá el acceso a las funciones más frecuentes, proponiéndose incorporar una serie de botones:
 - Encendido y apagado del televisor.
 - Control de volumen.
 - Cambio de programa o desplazamiento por una lista.
 - Cambio de modo de funcionamiento del televisor.
 - Retroceso o salir.
- Se propone ubicar en el frontal del dispositivo un interruptor para apagar el dispositivo.

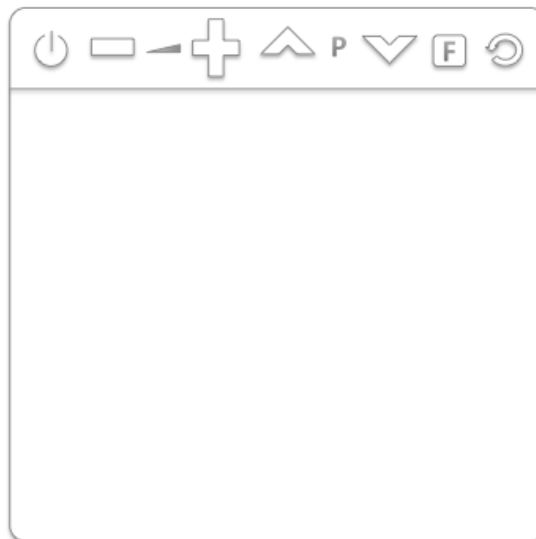


Figura 27 Prototipo del dispositivo de interacción propuesto TVPad.

4.4.2 Diseño del interfaz de usuario

Por otro lado, en la Figura 28 se muestra un diseño conceptual de cómo sería el interfaz de usuario en modo smartTV.

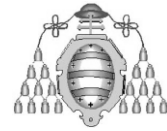


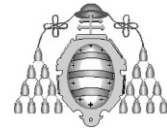
Figura 28 Diseño conceptual del interfaz de usuario TVPad.

En este diseño se incluyen 4 zonas de interacción en pantalla:

- Zona de control: acceso a la guía electrónica, configuración del televisor, etc
- Zona de programa: permite realizar acciones sobre el programa como activar el teletexto, cambiar de idioma o activar/seleccionar los subtítulos.
- Zona de teclado: permite introducir el número del canal.
- Zona de volumen

Con independencia de estos aspectos, en el modo de visionado estándar se plantea la posibilidad de realizar algunos gestos sobre el TVPad para las siguientes acciones:

- Ampliar o reducir la imagen, con un movimiento de pinza en diagonal.
- Subir o bajar el volumen, con un movimiento de desplazamiento vertical con dos dedos.
- Cambiar de modo de funcionamiento, con un movimiento de desplazamiento horizontal con dos dedos.



5 Conclusiones y Trabajo Futuro

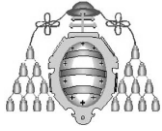
5.1 Conclusiones

El principal objetivo de este trabajo era evaluar el rendimiento obtenido con distintos métodos de escritura de texto susceptibles de ser utilizados en servicios de televisión digital interactiva.

Una de las aportaciones principales de este trabajo es haber realizado experimentaciones sobre métodos y dispositivos de inserción de texto en entornos IDTV, monitorizando a 240 participantes, con diferentes perfiles y rango de edad, en un contexto similar al de sus hogares. Este es un aspecto que contrasta respecto a otros estudios en los que seleccionan un grupo o muy reducido de personas o con características muy similares: comúnmente jóvenes y con experiencia en nuevas tecnologías.

Previamente a dichas experimentaciones se diseñó una metodología, para avanzar en la interacción humana en el ámbito de los métodos de inserción de texto para IDTV. Su aplicación será medir tanto la usabilidad, como la satisfacción de los usuarios, y abarcará todas las etapas necesarias para la realización de los test de usuario y pruebas subjetivas: desde la definición de métricas, pasando por el desarrollo de prototipos y realización de las pruebas hasta concluir con el análisis de resultados. Los resultados han mostrado la existencia de diferencias significativas entre los métodos y dispositivos en lo que se refiere a velocidad de escritura, tasa de error e impresiones generales de los usuarios.

En las primeras fases de la experimentación, se ha tenido en cuenta que el dispositivo mayoritariamente disponible en estos servicios es el mando a distancia, por lo que los métodos se han seleccionado acorde a esta situación. Por otro lado, hemos visto cómo los métodos más rápidos provenían del mundo de los dispositivos móviles (Multitap y T9), pero a costa de unas tasas de error altísimas. La mayoría de los errores se producen porque los usuarios no se dan cuenta de que han cometido un error y continúan escribiendo. Esta tasa de error se incrementa notablemente cuando los textos incorporan



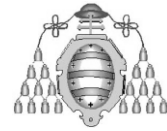
símbolos que podríamos denominar complejos (signos de puntuación, tildes, etc.) Dado que estos símbolos son necesarios en un contexto real, los teclados virtuales son la opción más recomendada. Con las diferentes disposiciones de símbolos que hemos probado no hemos obtenido diferencias significativas, pero la retroalimentación de los usuarios inclina la balanza hacia las disposiciones QWERTY y Genética.

Se ha comprobado que la edad es un aspecto muy importante a tener en cuenta, ya que a medida que ésta aumenta disminuye la rapidez de escritura. Además, usar habitualmente el ordenador potencia la velocidad de entrada de datos. Por el lado contrario, ha quedado de manifiesto que el sexo de los usuarios y el hecho de ser diestros o zurdos no fue relevante.

Se obtuvieron igualmente algunas conclusiones sorprendentes, como que el uso del sistema predictivo T9 o el envío de mensajes SMS no mejoraban la rapidez de escritura, sino que por el contrario, el uso del T9 aumentaba el número de errores cometidos utilizando dicho sistema.

En cuanto a los métodos evaluados, el más rápido ha sido el Multitap simple, aunque es también el de mayor tasa de error. Si se quiere lograr un compromiso entre rapidez y precisión, el teclado elegido sería el Genético que ha recogido muy buenas opiniones entre los participantes, los cuales notaron la progresión entre sesiones. El problema de este último método es la necesidad de un período de aprendizaje para que los usuarios consigan familiarizarse con la disposición de los caracteres en el teclado. El método 2-Key, aunque obtuvo la mejor progresión entre sesiones, ha resultado ser el más lento y por consiguiente, el más duramente criticado por los participantes del experimento.

Para que el experimento fuese lo más real posible, en otra fase posterior se consideró además de caracteres en minúsculas, caracteres en mayúsculas, acentos, tildes y caracteres especiales; ya que cualquier interacción de un usuario en una aplicación IDTV real tendría que introducir datos típicos como son el email, fecha de nacimiento, DNI, dirección postal o incluso introducir una URL, y que son una combinación de los caracteres comentados anteriormente. Esto influyó decisivamente en los resultados, provocando una disminución de velocidad y un aumento de tasa de error, debido a la



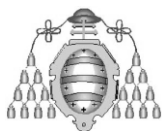
complejidad añadida, ya que el participante tiene que cambiar a la hora de introducir las frases entre el teclado en minúsculas, mayúsculas y caracteres especiales y números.

Por otro lado, tampoco hemos encontrado diferencias significativas cuando se altera la forma del mando a distancia o la disposición de sus teclas. Los resultados de los análisis realizados para detectar la combinación mando-método óptima no han sido concluyentes, es decir, el tipo de mando no ha influido en los resultados estudiados de los métodos de entrada de texto para este experimento.

Otra contribución de esta tesis derivada de los resultados obtenidos de los experimentos con otros dispositivos, es el desarrollo de una propuesta de lo que en nuestra opinión podría ser el dispositivo ideal. En términos objetivos de rendimiento, los mejores resultados se obtienen con un teclado convencional, a costa de unas elevadas tasas de error cuando se introducen símbolos complejos. No obstante, los usuarios manifiestan claramente que este periférico no es cómodo para ser utilizado en un entorno de visionado de televisión. Hemos visto que con la experiencia los usuarios mejoran su rendimiento con otros periféricos, acercándose al del teclado convencional. Por ello, objetiva y subjetivamente el mejor dispositivo es el touchpad. Tomando este periférico como referencia, el estudio concluye con un diseño de un touchpad adaptado a la televisión que hemos denominado TVPad. Al diseño del periférico incorporamos una propuesta de interface gráfica de usuario ajustada al modo de visionado de televisión.

Entre las aportaciones de esta tesis, está la identificación, análisis y clasificación de los errores más comunes en inserción de texto para IDTV. Para ello hemos analizado un conjunto de 7.395 frases escritas por 82 usuarios para identificar, cuantificar y proponer mejoras para varios métodos de inserción de texto compatibles con el contexto IDTV. Además del considerable número de usuarios, también es importante considerar que en general presentan características heterogéneas: edad, sexo, nivel de conocimiento y antecedentes tecnológicos.

Nuestros resultados muestran que, en general, el número de errores por frase es mucho menor en los teclados virtuales que en los métodos móviles. En cuanto a la edad de los usuarios, las personas mayores tuvieron valores promedio más altos en casi todos los tipos de errores, pero las diferencias con el resto de los usuarios fueron



estadísticamente significativas sólo en el caso de errores diagonal proximity. Asimismo, también cabe destacar que no hemos encontrado ninguna diferencia estadísticamente significativa entre los usuarios masculinos y femeninos o entre usuarios zurdos y diestros.

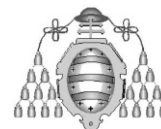
Por último, en las pruebas realizadas en el marco del proyecto, se ha constatado la importancia de elegir un método de escritura adecuado, dado que esta elección impacta notablemente sobre la velocidad de escritura y el número de errores que cometen los usuarios. Pese a que la elección de este método depende en gran medida de los fabricantes de televisores o set-top-box, no es menos cierto que los resultados de este proyecto son directamente aplicables por los desarrolladores de aplicaciones, que sin duda tendrán un peso específico en el sector de los servicios de IDTV.

5.2 Trabajo Futuro

Siguiendo esta línea de investigación, un trabajo futuro se basaría principalmente en experimentos con dispositivos nuevos o modificados. Nos gustaría comprobar el rendimiento obtenido con los dispositivos que han sido diseñados para otros fines pero que pueden utilizarse para interactuar con televisores, como es el caso de los smartphones o tablets. La razón principal es que estos dispositivos ya están disponibles en la mayoría de hogares, por lo que su uso en este contexto es totalmente factible.

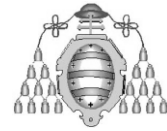
Además, el hecho de conseguir buenos resultados con velocidades de inserción de texto con teclados físicos, en esta nueva ampliación del estudio nos gustaría experimentar aún más para ver si podemos reducir las tasas de error. Por ejemplo, podríamos usar las StickyKeys en lugar de las teclas modificadoras (Ctrl/Alt/Fn/Shift) o experimentar con un teclado retroiluminado de tamaño normal. Además, mediante la incorporación de otros equipos, queremos medir otros factores que pueden influir en la experiencia de los usuarios como los niveles de estrés, la actividad cerebral o el efecto sobre la vista, que nos permitirían sacar conclusiones aún más interesantes.

Otro trabajo futuro será realizar una implementación real de TVPad tanto del interface de usuario como del dispositivo hardware. Esta implementación real nos permitirá realizar pruebas de usabilidad con usuarios reales, dado que las pruebas que hemos hecho hasta



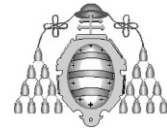
el momento se han limitado a tareas de escritura de texto. Estas pruebas de usuario permitirán probar algunas hipótesis planteadas en este estudio.

Por último, la metodología desarrollada permite organizar estudios con usuarios reales de forma ordenada en el ámbito de la inserción de texto, pero sería interesante realizar mejoras para que fuese más genérica y su ámbito de aplicación sea cualquier tipo de aplicaciones en IDTV con o sin inserción de texto.



6 Compendio de publicaciones

En este capítulo se adjuntan las publicaciones que conforman la tesis doctoral.



6.1 An empirical investigation into text input methods for interactive digital television application

A. Barrero, D. Melendi, X. G. Pañeda, R. García, y S. Cabrero, «An empirical investigation into text input methods for interactive digital television applications», *International Journal of Human-Computer Interaction*, vol. 30, n.o 4, pp. 321–341, 2014. DOI: 10.1080/10447318.2013.858461

An Empirical Investigation Into Text Input Methods for Interactive Digital Television Applications

Aurora Barrero, David Melendi, Xabiel G. Pañeda, Roberto García, and Sergio Cabrero

University of Oviedo, Asturias, Spain

Nowadays there is a huge market emerging in the interactive digital TV realm. In this context, we need new and effective methods of user interaction, as the main interaction device is still the classical remote control. Remote controls are especially problematic when it comes to writing text, something needed in most applications. Thus, we have carried out an empirical investigation to find effective methods of text entry with remote controls. We analyze several methods by performing experiments based on a methodology in which a heterogeneous set of real users carries out several sequential tasks in an incremental process. We analyze entry speeds, error rates, learning profiles, and subjective impressions, taking into account the particular characteristics of the users. Our results show, for instance, that Multitap is a good method for simple texts. It is between 12% and 34% faster than the fastest virtual keyboard, depending on the age of the user. Nevertheless, when complex texts need to be written, virtual keyboards present the same or even better writing speeds (QWERTY is 13% faster) and with significant lower error rates (Multitap is 347% worse than QWERTY). We consider that our results are very interesting for researchers, designers of TV applications, and hardware vendors.

1. INTRODUCTION

Recent innovations in the Internet protocol television (IPTV) field (Spira, 2011) and regulation changes (Congressional Record, 2009; European Union, 2005) have created a new scenario in the television realm in which content providers are encouraged to offer advanced services to be competitive. Most of these advanced services require interaction capabilities that are possible nowadays thanks to the proliferation of home Internet connections (Brandtzæg, Heim, & Karahasanović, 2011), the improvement of set-top-box and TV devices and the general evolution of telecommunication networks. The Internet and TV have merged, forming services such as IPTV or Internet TV (Spira, 2011). However, if complex applications are to be

developed, we also have to design effective methods of user interaction, taking into account that the main interaction device available is still the remote control. This is a problem when it comes to writing text, as remote controls have not been designed for this purpose. Thus, if we need the users to introduce text on interactive digital TV (IDTV) applications, we have to find effective methods of text entry with a remote control considering aspects such as the experience of the users or their age (Taveira & Choi, 2009).

The main goal of this article is to provide a detailed study on text entry methods that may be used with commonly available technology that most people will probably already have in their homes. Thus, several methods are analyzed by performing experiments with a heterogeneous set of real users in an incremental process based on several sequential tasks. Some preliminary results published in Perrinet et al. (2011) are now improved with new results, more text entry methods, and usage contexts. We conclude answering questions such as which method is the fastest, which method is the one that leads to fewer user mistakes, how age affects performance, how performance improves with shortcuts or fast keys, how an extended set of characters diminishes performance, or how the layout of the remote control affects performance.

Our article presents an important set of contributions compared with previous work. We have analyzed a considerable number of input methods, some of them complemented with several improvements, such as disambiguation or suggestion mechanisms. We have also performed the evaluations under general contexts, also taking into account the special needs of other contexts and issues such as internationalization. Moreover, we have evaluated how the shape and the location of the keys in remote controls affect user performance. Finally, an important number of users participate in our study with a heterogeneous set of characteristics: age, gender, knowledge level, and technology usage habits.

The rest of the article is organized as follows. In Section 2 previous work is revised. The plan of the experiments is defined in Section 3. The evaluation and the obtained results are detailed in Section 4. Finally, conclusions and future work are presented in Section 5.

Address correspondence to David Melendi, University of Oviedo, 2.7.11 Edificio Polivalente, Campus de Viesques, s/n, Xixón, Asturias 33203, Spain. E-mail: melendi@uniovi.es

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.

2. RELATED WORK

The study and comparison of different text input methods is a field where a lot of research has been carried out. Nevertheless, most recent advances have been focused on mobile devices such as smartphones or tablet devices. Although some of these advances may be applied to the IDTV realm, there are still many open issues as stated by [Iatrino and Modeo \(2006\)](#). First, the main interaction device is a remote control and not a keyboard, stylus, or touch screen. Nowadays this still holds true. Second, comparing TVs and mobile devices, the keys and the screen are in different devices, so users cannot look at the interaction device and the screen at the same time. Thus, interactive digital television services offer new challenges to be solved.

Commonly available remote controls can be used in only two possible ways: using cursors and OK keys over a virtual keyboard or using a reduced set of keys. Regarding virtual keyboards, in the field of mobile devices there have been several publications with interesting results for the IDTV realm. [Bellman and MacKenzie \(1998\)](#) compared two types of virtual keyboards. The first type is the well-known QWERTY layout, and the second is an optimized alternative named FOCL. FOCL proposes placing letters in the keyboard surrounding a predetermined letter depending on their probability of following this letter. Although the results show no significant differences between each method, FOCL can be considered as a feasible improvement for IDTV context. [Zhai, Hunter, and Smith \(2000\)](#) presented a new virtual keyboard layout called Metropolis, 40% faster than QWERTY, to improve writing performance in certain devices. The authors stated that existing layouts such as QWERTY have been designed to be used with both hands, so they are not appropriate for these devices. [Brewbaker \(2008\)](#) pursued the same goal but using a genetic algorithm to customize the size and language of the keyboard. The results of [Brewbaker \(2008\)](#) and [Zhai et al. \(2000\)](#) are similar, but their methods were not tested with real users. Although it is unknown whether these methods are good for real users, the algorithmic approach toward the design of a new layout can be interesting for IDTV applications.

Also, there are writing mechanisms for mobile devices based on a reduced set of keys suitable for IDTV applications. For instance, [Silfverberg, MacKenzie, and Korhonen \(2000\)](#) compared three methods: the multipress method used to write Short Message Service (SMS) messages (sometimes called Multitap) either with timeout or next key to switch between letters, a method based on two keystrokes called 2-key (first users select a group of letters and afterwards the desired letter), and the predictive method T9 (based on the theory that each combination of keystrokes produces a single word or a reduced set of words). These methods are evaluated with 12 users between 23 and 47 years of age, and the results show that the best results are obtained with T9 and the worst with 2-key. Similarly, [Butts and Cockburn \(2002\)](#) repeated the experiment of [Silfverberg et al. \(2000\)](#) but without the T9 method. Eight advanced users obtained the best results with the multipress method with next

key and the worst with 2-key. [MacKenzie, Kober, Smith, Jones, and Skepner \(2001\)](#) presented a different disambiguation technique based on prefix probabilities called LetterWise. This method is compared with T9 in an experiment with 20 paid users, who participated in 20 sessions of 25 to 30 min. Users had to write sentences from a set defined by the authors. Whereas in the first session the results between methods were similar, in the last sessions there were some differences in performance. Nevertheless, they are unable to establish a significant difference between error rates for each of the methods. [Nesbat \(2003\)](#) presented a novel text entry system (MessagEase) for small electronic devices with a unique keyboard based on letter frequency and positional matrix. This keyboard is applicable to hard-key devices with a limited number of keys. The text entry system presented here provides full text entry (full ASCII 220) and is adaptable for any language. Based on the application of Fitts's law, this system is determined to be 67% faster than the QWERTY soft keyboard and 31% faster than multipress, but it is unknown whether it is good for users, as the results are only theoretical.

Apart from virtual keyboards and methods based on a reduced set of keys, there are other methods called "concurrent techniques" designed to combine different methods at the same time. Although there are several papers where these techniques are used, the results of [Wigdor and Balakrishnan \(2004\)](#) are especially interesting. They conducted a comparative experiment in which they add three additional keys to a conventional mobile phone keyset. Users have to combine these additional keys with regular keys depending if they want to write a number or a letter. They performed an experiment with 15 experienced users and detected a significant improvement in performance. These concurrent techniques can also be applied to an IDTV environment, for instance, when a greater set of symbols is needed.

The aforementioned papers mainly come from the field of mobile devices, but there are also a few noteworthy studies in which the issue of writing text in IDTV applications is treated. [Iatrino and Modeo \(2006\)](#) compared multipress (the SMS method), multipress with visual feedback, and a QWERTY virtual keyboard layout. Thirty-six people participated in an experiment where each user had to write an e-mail address and a short sentence in Italian. Results show that the best method is multipress. Also, the authors mention important problems with internationalization. [Ingmarsson, Dinka, and Zhai \(2004\)](#) presented a new technique called TNT, similar to TwoStick ([Költringer, Isokoski, & Grechenig, 2007](#)). Each character is accessible with two keystrokes with a similar approach to 2-key ([Silfverberg et al., 2000](#)). Five people between 27 and 32 years of age participated in the experiment during 10 sessions of 45 min to write a short novel in Swedish. Results show speeds comparable to or faster than manual writing in a personal digital assistant or the multipress method. The most valuable feature for the users of the experiment was the simplicity of the method. [Geleijnse, Aliakseyeu, and Sarroukh \(2009\)](#) compared Multitap

(multipress), T9, and a QWERTY virtual keyboard layout with the speed of a conventional QWERTY keyboard. The goal was to search for author-track pairs in YouTube. Twenty-two people between 21 and 32 years of age participated in the experiment. The authors conclude that the conventional keyboard is faster than remote control methods and that there are no significant differences between remote control methods, in contrast to previous works. Their final conclusion is also surprising: “User tests have not shown any indications that users do not accept the keyboard in a living room-like setting” p. 148. This conclusion contrasts with [Orbist, Bernhaupt, and Tscheligi \(2008\)](#), who consider that voice recognition systems may be a better solution. Their results show that it is not feasible to rely on the popularity of external peripherals for televisions different from those conventional. More recently, [Gargi and Gossweiler \(2010\)](#) presented a new predictive system designed to improve writing speed in virtual keyboards: QuickSuggest. This method shows a ring surrounding a given character, with the four characters that most probably follow it. If a user writes a symbol, he or she only has to select the next character with a cursor key and OK. Then the system moves automatically to the written character and it starts over again. The authors perform a theoretical performance study and an experiment with 10 real users. [Sporka, Polacek, and Slavik \(2012\)](#) compared TNT ([Ingmarsson et al., 2004](#)) and a new method named TwiceTap. This method has the same philosophy as TNT but, apart from allowing users to type single characters, it also allows them to write frequent blocks of characters (n-grams). Eighteen paid users, with an average age of 22.7, participated in the experiment. There were no significant differences between the methods, but users preferred TwiceTap mostly because of its similarity to Multitap and the availability of n-grams.

2.1. Discussion and Problem Description

As mentioned previously, there is little research on text writing methods for IDTV applications if we consider that the main interaction device is still the remote control.

In recent years there have been other papers, apart from those mentioned previously, aiming to solve the problem of writing text. Nevertheless these studies rely on interaction devices that have nothing to do with a remote control. This is the case of [Wobbrock, Myers, and Aung \(2004\)](#) and [Költringer et al. \(2007\)](#) studying joystick text entry methods; [Oniszczak and MacKenzie \(2004\)](#) combining keystrokes with finger movements in RollPad; [Orbist et al. \(2008\)](#) and [Vega-Oliveros, Pedrosa, Pimentel, and De Mattos Fortes \(2010\)](#) with voice recognition; [MacKenzie, Lopez, and Castelluci \(2009\)](#) and [Aoki, Maeda, Watanabe, Kobayashi, and Abe \(2010\)](#) with pointing devices; [Rick \(2010\)](#) and [Varcholik, LaViola, and Hughes \(2012\)](#) with mechanisms of writing text on interactive surfaces; or [Choi, Han, Lee, Lee, and Lee \(2011\)](#) with an ad-hoc-designed remote control equipped with a touchpad. Although the results of all these previous works are very interesting and may be applied to IDTV environments in the

future, they are not eligible with current massively available TV technologies, as the main interaction device is still the conventional remote control.

As we have seen, there are only a few studies about text entry in TV environments with methods suitable for remote controls. Also, there are even fewer studies with results obtained from real users. Many papers present conclusions based on Fitts’s prediction models ([MacKenzie, 1991](#)) or other theoretical analyses. Their performance with real users is unknown. Other papers perform evaluations with a reduced set of users or with very homogeneous characteristics: usually young people with a technical background. Other papers study the same or similar methods but present contradictory results. Furthermore, there are important issues that have not been studied in depth, such as the impact of a complex set of symbols, the shape of the remote control, or the location of the keys. We consider that the issue of writing text on IDTV applications with a remote control needs further study.

In our opinion, our article presents new results regarding text input issues in the IDTV realm. First, we have analysed several types of virtual keyboard layouts and other methods, some of them complemented with different improvements, such as disambiguation techniques or suggestion systems. Second, we have performed the evaluations using both general sentences and specific characters and texts. Third, we have evaluated how the shape and the location of the keys in remote controls affect user performance. The last important characteristic of our study is the number of users and the variations in their characteristics: age, gender, level of knowledge, and technology usage habits.

3. TEST PLAN

3.1. Goals and Metrics

The main goal of the experiments was to compare several input methods for IDTV applications. To achieve this goal, various metrics were initially chosen:

- Measured entry speed: in characters per minute. We considered the time taken to write a certain sentence, including the time spent in writing the desired sentence, but also the time spent in deleting mistaken characters and rewriting the correct ones.
- Measured error rate: percentage of mistaken characters written. To calculate error rates we counted the total amount of characters written by the user and compared it with the number of characters in the desired sentence.
- Measured learning curve: improvement (or not) of users’ performance.
- Subjective impression of ease of use (0 = *very difficult* to 4 = *very easy*).
- Subjective impression of speed of use (0 = *very slow* to 4 = *very fast*).
- Subjective impression of user satisfaction (0 = *unsatisfied* to 4 = *very satisfied*).

We have focused the analyses on entry speed and error rates because if we observe previous work, we can clearly see that all the authors use them as basic metrics to compare text entry methods. Nevertheless, as shown in MacKenzie et al. (2001), the values of these basic metrics evolve when users participate in several sessions in the experiments. Thus, it is also important to study how entry speed and error rates vary with experience.

Apart from empirical data, it is also important to collect subjective information. Previous work shows that the opinions of the users are very valuable and that these do not always coincide with empirical performance data. For instance, Sporka et al. (2012) showed no empirical differences between methods but a clear preference of the users toward TwiceTap. We have chosen speed of use to compare the impression of the users with the measured speed. We have also selected user satisfaction as a good indicator of the general impression of the users, as in Iatrino and Modeo (2006). Finally, we have chosen ease of use because this is one of the most valuable features for users in previous work (Ingmarsson et al., 2004).

3.2. Subjects

In total, 96 users participated in the experiments. These participants were as heterogeneous as possible, with ages ranging between 20 and 64, different technological skills, gender, and laterality. The participants did not receive any money or any other kind of compensation. Some of their characteristics are shown in Table 1.

Not all the users participated in all the tasks, so further details are provided in the description of each task. We grouped users according to their age as follows: users belonging to the “mobile phone generation,” ages 18 to 30 (young); users belonging to the “computer generation,” ages 31 to 45 (adult); and “pre-pc” users, older than 45 (older). We have focused the analyses on age because changes in perceptual and motor skill capabilities that accompany the aging process bring implications for the design of human–computer interfaces (Taveira & Choi, 2009). This is aligned with the differences reported by Siek, Rogers, and Connelly (2005). The rest of the data in Table 1 are descriptive. Nevertheless, as we expect users to behave differently depending on their attitude toward technology, we have taken

into account the habits of the users using the information gathered in several questionnaires. Finally, we have also considered gender and laterality, because these are also common types of groups of demographic segments in usability.

3.3. Apparatus

Hardware. During the experiment, each participant was left alone in a room, to avoid distractions. To create the most realistic situation possible, users sat as if they were in their own living room. An armchair was placed in front of a 32-in. LCD television of dimensions 814 × 599 mm placed on top of a table around 1 m in height. According to the recommendations of the television manufacturer, the armchair was placed at a distance of 2 m (the optimal distance varies from 1.5 to 3 m for this type of display).

The television was connected to a PC with Microsoft Windows XP running an application designed to carry out the tests. We also connected several remote controls to the PC equipped with USB RF or infrared receivers, which incorporated the most common keys of conventional television remote controls. This decision was based on the fact that one of our design goals was to investigate text entry methods that may be used with commonly available technology that most people would probably already have in their homes. In most of the tests we used a SnapStream Firefly PC remote control, shown in Figure 1.

Software. We developed an Adobe AIR application to carry out the tests. This application, shown in Figure 1, had a black full-screen layout and one or more text input fields to allow the users to write. Our fonts were Verdana 65 point for the proposed sentences and the input fields, and Arial 60 points for the keyboard layouts. Depending on the test, the text to be written was placed on top of these input fields.

The application was developed to process keyboard events. To map remote control strokes into keyboard events, we used the EventGhost software. No forced time lags between keystrokes were introduced by the system, and users were allowed to delete symbols.

The application showed a 10-s counter at the beginning of the experiment and between different input methods. This counter was also used before writing each sentence, allowing users to

TABLE 1
Characteristics of the Participants

Age	Total	Gender		Field (Profession or Studies)		Education		
		M	F	IT	Non-IT	Elementary	High School	University
Young	57	32	25	39	18	0	7	50
Adult	27	17	10	14	13	0	5	22
Elderly	12	8	4	1	11	5	1	6

Note. IT = information technology.

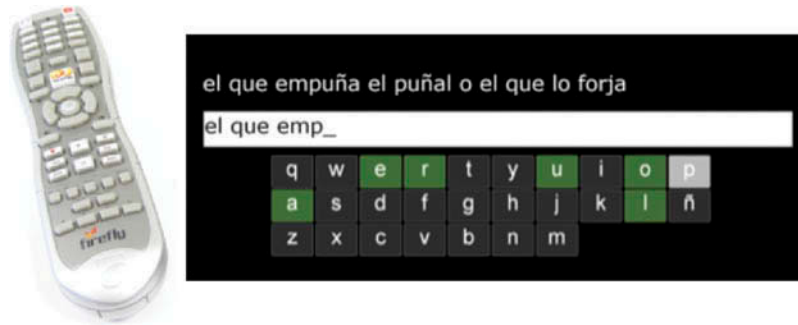


FIG. 1. Remote control and testbed application.

read sentences in advance. Once the user was allowed to write, the application checked whether the sentence was written completely. When a sentence had been completed, the application switched to a new sentence or to a different input method.

When virtual keyboards were used, the keyboard layout was placed on the bottom part of the screen, and users had to use the cursors and the OK button to move around the layout. We also included the following improving mechanisms:

- Users could move faster by keeping a cursor button pressed. This reduced their physical load. Based on pilot tests, we adjusted both the start of the auto-repeat function and its repetition rate to 0.2 s.
- The borders of the layout were wrapped around both vertically and horizontally. For instance, users could go from the top part to the bottom by pressing the up cursor and vice versa.
- We implemented a suggestion system inspired by the LetterWise disambiguation method (MacKenzie et al., 2001) and similar to those in FOCL (Bellman & MacKenzie, 1998) and QuickSuggest (Gargi & Gossweiler, 2010). Based on what a user wrote, the system changed the color of the six letters most likely to follow what was already written, as shown in Figure 1. To provide these recommendations, the system loaded a dictionary of the Spanish language (RAE – Royal Spanish Academy, 2001), and each time a user added a letter to a word, the system searched for all the words containing the string the user had written. With this list of words, the system computed the most frequent letters after the current string and displayed the six most probable with a green background.

When mobile-like techniques were used, the bottom part of the screen was used to show help information, as in Figure 2. The keys in the application were placed in accordance with the European standard ETSI ES 202 130 V2.1.2 (2007-09). To assist users when they pressed the same key several times, the current letter was highlighted.

The application stored information of the activity of the users in XML log files. In these files the application registered when a certain key in the remote control was pressed, information about

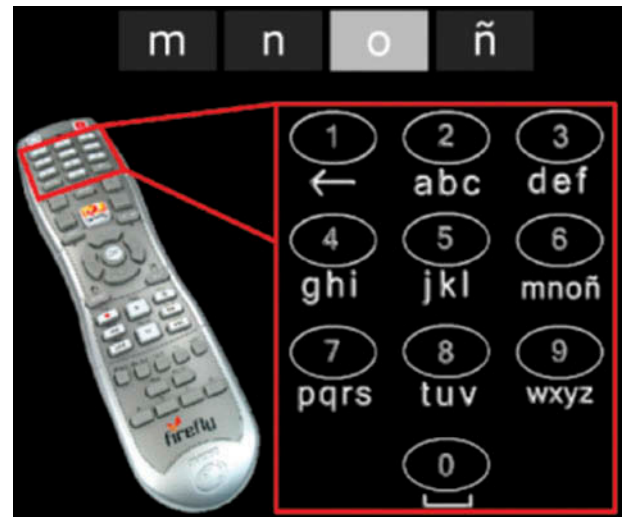


FIG. 2. Implementation of a mobile-like method in our application.

the answers to questionnaires and information about both the sentences proposed by the system and the sentences written by users. Further details of the format of these files may be found in Perrinet et al. (2011).

3.4. Text Input Methods

During the experiment, we compared the performance of several text input methods suitable for interactive TV applications. These methods can be classified into two groups: virtual keyboards and mobile-like methods.

- Virtual keyboards:
 - QWERTY: This is the traditional keyboard layout, broadly used in previous work.
 - Square alphabetic—or simply alphabetic: This is based on the alphabet, the position of the keys in alphabetical order.
 - Genetic: We used a genetic algorithm to generate a new keyboard layout in order to improve the efficiency of the users. Using a modern version of Don

Quixote as a reference text for Spanish, the algorithm positioned the keys so that the most used letters appear in the center of the keyboard. The resulting layout is shown in Figure 3. The details and cost model of this algorithm can be found in Brewbaker (2008).

- Modified QWERTY: This was QWERTY with an additional row of symbols including special vowels as shown in Figure 4. Also, four additional buttons in the remote control were used. One allowed the user to switch to a layout with capital letters, another allowed the user to switch to another layout with special symbols (Figure 5), and two more buttons allowed the user to write blank spaces and delete.
- Modified Genetic: Using the genetic algorithm we generated a new layout but including special vowels. The result is shown in Figure 6. Also, four additional buttons in the remote control were used as in the



FIG. 3. Genetic keyboard layout.

modified QWERTY layout, with exactly the same purpose.

- Mobile-like methods:
 - Multitap: Based on the mobile phone system, this method uses number keys to write text as shown in Figure 2. To differentiate between multiple strokes corresponding to a single letter and strokes between successive letters, we use a threshold of 1 s as in a regular phone.
 - T9: This is a predictive system based on the idea of pressing one single key for each letter of the word the user wants to write (Silfverberg et al., 2000). For example, if we consider the relation between the keys and the letters in Figure 2, the word “this” can be written by pressing 8, 4, 4, and 7. In most cases, the word that appears after pressing a sequence of keys is the desired one. When there is more than one possibility matching the sequence of keystrokes, users must select one of the available options.
 - 2-key: This is a system designed to reach any letter with only two keystrokes (Silfverberg et al., 2000). Using numeric keys, a first keystroke is used to select a group of letters and a second keystroke is used to select the desired letter.
 - Modified Multitap: In this version of the Multitap method, new characters were inserted in the keyboard with special vowels as shown in Figure 7. Also, we used a concurrent technique to change the functionality of the keyboard as in Wigdor and Balakrishnan (2004). A button allowed the user to switch to capital letters and numbers. Another button allowed the user to insert special characters.



FIG. 4. Modified QWERTY layout.



FIG. 5. Special characters layout.



FIG. 6. Modified genetic layout.

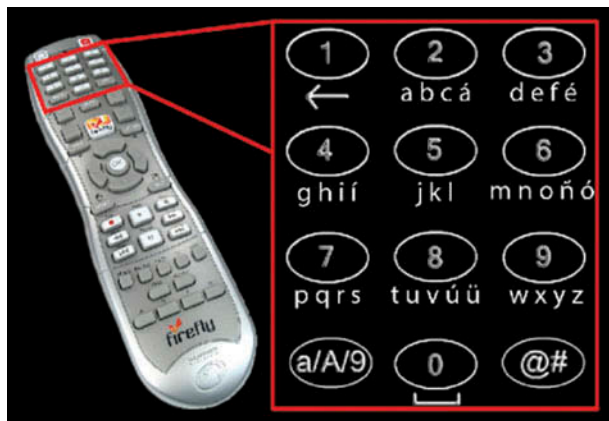


FIG. 7. Modified multitap layout.

Regarding virtual keyboard layouts, we have chosen QWERTY because it is already known by a great number of people familiar with computers and because it is the main layout used in previous work. This method was designed placing many adjacent letter pairs (digraphs) on the opposite sides of the keyboard, facilitating the frequent alternation of the left and right hand. But, on a virtual keyboard, the polarizing common digraphs mean that the user has to move back and forth more frequently and over greater distances than necessary (Zhai et al., 2000). Nevertheless, it is still currently the most popular layout. The Alphabetic layout was considered because it is supposed to be the “easiest” in the sense that anyone who knows the alphabet would be able to use it without a learning process. Thus, nonexpert users would find it very easy to use according to Zhai et al. (2000). We have used the Genetic layout, which had never been used in Spanish, as a more experimental option and obtained very promising results. We conducted a theoretical analysis in order to check the efficiency of this layout. Measuring the minimum number of keystrokes needed to write Don Quixote, this method is 45% faster than QWERTY for a proficient user. Finally, we have used the modified versions of QWERTY and the Genetic layouts to include symbols that are nowadays common in applications and symbols required in languages different to English.

On the other hand, mobile-like methods were chosen mainly inspired by Silfverberg et al. (2000). We consider that Multitap is interesting for the IDTV domain. Although it is progressively disappearing as modern mobile phones incorporate physical keyboards or touch screens, Multitap was extremely popular among young users not so long ago. Thus, the number of expert users of this method is still huge. This situation combined with the fact that previous works show that it is more efficient than virtual keyboards (Iatrino & Modeo, 2006) are the reasons why we considered this method for the evaluation. Also, T9 is a promising method according to the results in Silfverberg et al. (2000), surpassing the results of Multitap. We have chosen 2-key because, in theory, it is faster than Multitap for Spanish if we compare its two keystrokes with a calculated weighted average of 2.23 Multitap keystrokes per letter for the words in the official Spanish dictionary (RAE – Royal Spanish Academy, 2001). Finally, as in the case of virtual keyboards, we have a version of Multitap with an enhanced number of symbols to be used in real IDTV applications and languages different to English.

3.5. General Procedure

The experiment consisted of four tasks, each of them requiring the users to attend one or several sessions in which the same text entry methods were used. A summary of the plan of the experiment is shown in Table 2, and a brief description of each of the tasks follows:

- In the first task, we started by testing popular methods of text entry with a predefined set of sentences: QWERTY, Alphabetic, Genetic, and Multitap.
- In the second task we included several improvements, both proposed by users during the first task or gathered from previous works such as Silfverberg et al. (2000). The methods were QWERTY, Genetic, T9, and 2-key.
- In the third task we performed tests with personalized sentences (not predefined) and nonalphanumeric characters, using the modified versions of QWERTY, Genetic, and Multitap. Previous works such as Ingmarsson et al. (2004) or Iatrino and Modeo

TABLE 2
Summary of the Plan of the Experiment

Task	Users	Methods	Remote Controls	Sessions	Set of Characters	Texts
1	57	QWERTY, Alphabetic, Genetic, Multitap	SnapStream Firefly	5	Basic	5 sentences of corpus
2	48	QWERTY, Genetic, T9, 2-key	SnapStream Firefly	2	Basic	5 sentences of corpus
3	42	Modified QWERTY, Modified Genetic, Modified Multitap	SnapStream Firefly	1	Complex	5 fields with personal data and 1 common text
4	42	Modified QWERTY, Modified Genetic, Modified Multitap	Golden Interstar, AverMedia RMKS	1	Complex	5 fields with personal data and 1 common text

(2006) reported internationalization problems and, partially, problems with special symbols (@, commas, etc.). We considered this task important due to these problems and bearing in mind the deployment of IDTV applications in contexts such as web browsing or e-commerce.

- In the final task we performed tests with different remote controls to see whether the device itself influences performance. We carried out this task due to the lack of standard remote control layouts (Nielsen, 2012), as only general recommendations exist (European Broadcasting Union, 2006).

All these tasks were not designed a priori. When we first performed the experiments published in Perrinet et al. (2011), we discovered other problems that needed to be taken into account. First, the impact of certain optimizations, analysed in the second task. Second, the usage of special symbols to accommodate specific contexts, analyzed in the third task. Third, the influence of the design of the remote control, analyzed in the final task. These problems were further analysed in the aforementioned tasks, conforming the full experiment presented here.

As shown in Table 2, in some of these tasks we planned several sessions to measure how users learn. We tried to confirm the relation between experience and performance (MacKenzie et al., 2001). During the first session of each task, we gave the participants a brief explanation and an example of each input method. No instructions were given regarding how to handle the remote control, that is, with the left hand, with the right hand, loosely, tightly, and so on. Thus, we expected them to use the remote as they usually do. The methods were assigned to participants in a counterbalanced order to neutralize learning effects (analyses of variance [ANOVAs] for test order showed no significant differences).

In the first and second tasks we provided users with the texts to write. We created a small corpus with 50 representative short sentences in Spanish gathered from local newspapers. The sentences had an average of five words, and they included

only numbers and lowercase letters from the English alphabet, with the single exception of the ñ symbol.

We used several questionnaires to gather users' details and subjective information:

- During their first session, users had to fill in a questionnaire with details of their gender, age, profession, level and type of studies, habits of using television, mobile phone and computer, and whether they were left- or right-handed.
- In all of the sessions users had to complete a final questionnaire providing us with some feedback. They had to comment on each method, for example, if they had the impression they were improving, and rate three Likert scales (0–4): ease of use, entry speed, and global satisfaction.

Finally, to extract the conclusions of the experiment we performed several statistical analyses. The main goal was to determine if there were real differences between the means of two or more groups of variables. We used the most common tests for this type of analysis, according to Crawley (2007). These tests depended on the assumptions of normality and homoscedasticity of the data. First, we checked normality with Shapiro–Wilk tests and homoscedasticity with Bartlett tests. When data met both normality and homoscedasticity, we used one-way ANOVA tests to compare the data. If homoscedasticity failed, then we used Kruskal–Wallis tests. Kruskal–Wallis tests were also used in situations with a strong failure in normality (p values in normality tests over .05). Finally, if differences existed, we used Tukey tests with a confidence coefficient of 95% to perform pairwise comparisons.

4. EVALUATION OF TEXT INPUT METHODS

4.1. First Task: Evaluation of Main Input Methods

In this task we compared the performance of four text entry methods suitable for IDTV applications: QWERTY, Alphabetic

and Genetic virtual keyboards, and Multitap. As previously stated, these methods were chosen after an analysis of previous work.

To track the progression of the participants, users had to complete five sessions on 5 consecutive days. In each of these sessions, they had to write five random sentences of the corpus with all the methods. The particular characteristics of the participants are shown in Table 3.

Some preliminary results of this task were published in Perrinet et al. (2011) with fewer users than in the present study.

Analysis of writing speed and learning profiles. The general results show that the fastest method is Multitap, whereas the slowest is QWERTY. The average writing speed in characters per minute for each method and Sessions 1 to 5 is shown in Figure 8. It is noticeable that Multitap is much faster than the rest in all sessions. Also, performance regularly improves with experience for all the methods. We have calculated how this progression may be in future sessions using the power law of learning (Ritter & Schooler, 2002) with the results shown in Figure 8. In this figure we can see a great progression in

Multitap. For virtual keyboards the Genetic layout presents a slightly better tendency with similar results for QWERTY and Alphabetic. Moreover, our initial assumption was that QWERTY does not require a learning period, but learning does in fact occur as in the case of Clarkson, Clawson, Lyons, and Starner (2005).

To check whether the differences between each method have any statistical significance, we have performed pairwise comparisons with Tukey tests. Basically, Tukey applies simultaneously to the set of all pairwise comparisons $\{\mu_i - \mu_j\}$. Confidence intervals including 0 are not significantly different, and all the other pairs are significantly different. Confidence intervals greater than 0 represent greater values in the i set than in the j set, whereas confidence intervals lower than 0 mean the opposite. These comparisons present differences between the methods in all the sessions, but they are statistically significant only in the case of Multitap ($p < .001$), which is the fastest method. This coincides with Iatrino and Modeo (2006), but is in clear contrast with Geleijnse et al. (2009). For example, Figure 9 shows the results of these comparisons for Session 5. Genetic is the best virtual keyboard, QWERTY

TABLE 3
Characteristics of the Participants in the First Task

Age	Total	Gender		Field (Profession or Studies)		Education		
		M	F	IT	Non IT	Elementary	High School	University
Young	27	15	12	20	7	0	3	24
Adult	19	12	7	12	7	0	2	17
Older	11	7	4	0	11	5	1	5

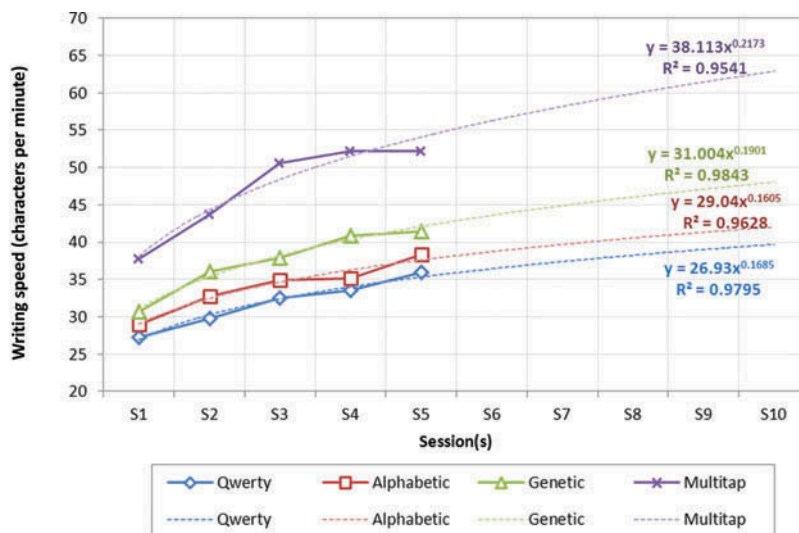


FIG. 8. Average writing speed per method and session and prediction of improvement with further experience.

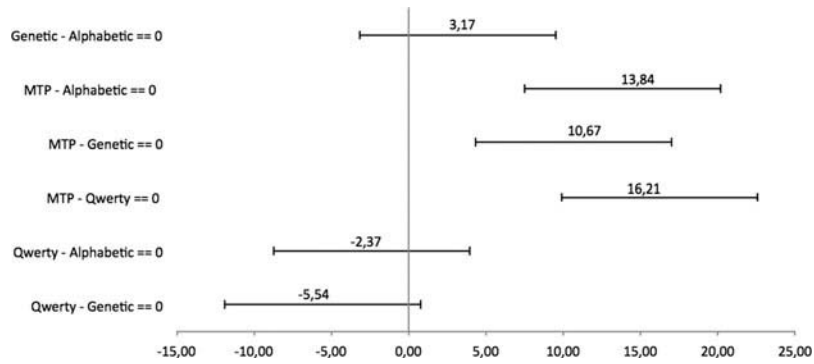


FIG. 9. Pairwise comparisons of average writing speed in Session 5 (95% family-wise confidence level). Note. MTP = Multitap.

being the worst, but this cannot be considered statistically significant.

If we compare empirical data with the habits reported by the users, we see that users who reported sending more than one SMS message per day are faster than users who send messages less frequently, but only for Multitap. The average speed is 58.87 characters per minute for frequent SMS users versus 46.60 for the rest of the users. Also, the participants who use computers frequently are faster in all the methods. Users with more than 12 hr a week of computer usage show statistically significant differences with the rest of the users in all the virtual keyboards ($p < .02$). In the case of Multitap, this difference is statistically significant only if we compare users with more than 12 hr a week with users with less than 3 hr a week ($p = .007$). Although users with weekly computer usage between 3 and 12 hr are faster in general than users with less than 3 hr, this is not statistically significant according to Tukey tests. Habits reported by users regarding their usage of TV do not seem to affect performance.

In the results we have seen that age is a major factor affecting performance, as shown in Figure 10. In general, the younger the user, the faster he or she writes. Pairwise comparisons of average speeds show that young users and adult users are faster than older users with a statistically significant difference ($p < .001$). Also, although young users are faster than adult users,

this difference is statistically significant only for Multitap ($p = .0043$) and Alphabetic ($p = .02$).

Finally, ANOVA tests show us that there are no statistically significant differences between male and female users and left versus right-handed users.

Analysis of error rates and learning profiles. The general results show that the text input method with the highest error rate is Multitap, in contrast with virtual keyboards that present much lower error rates as shown in Figure 11. Multitap is worse than the rest in all the sessions. Moreover, in general for all the methods, error rates decrease with experience. Using the power law of learning we have calculated how this progression may be in future sessions with the results shown in Figure 11. In this figure we can see a great progression in Multitap and similar results for all the virtual keyboards.

Pairwise comparisons performed with Tukey tests show that only the differences between Multitap and the virtual keyboards are statistically significant making the Multitap the worst method considering error rates. For instance, in Session 5 we obtain a $p < 1e-05$ if we compare Multitap ($M = 7.26$, $SD = 5.59$) with Genetic ($M = 1.68$, $SD = 1.37$), QWERTY ($M = 2.11$, $SD = 2.27$), and Alphabetic ($M = 1.67$, $SD = 1.72$). In general, Genetic is the best method but with results similar to those of QWERTY and Alphabetic.

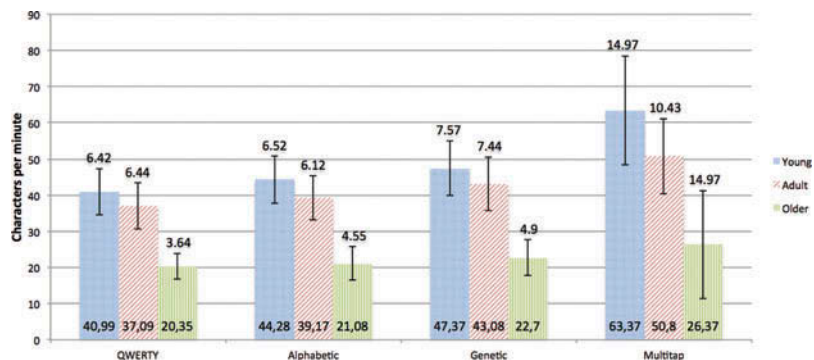


FIG. 10. Mean and standard deviation values of writing speed per method and age.

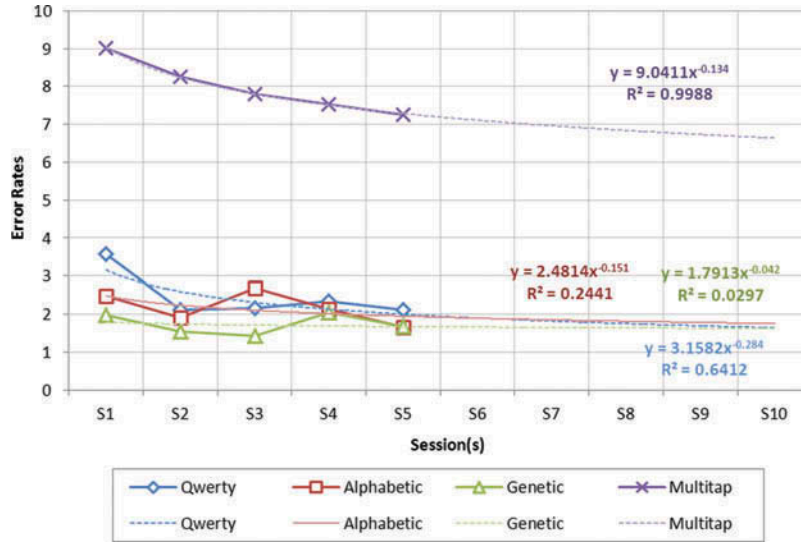


FIG. 11. Average error rates per method and session and prediction of improvement with experience.

In the results we have seen that age also affects performance. Pairwise comparisons of average error rates show that older users commit more errors than young and adult users with a statistically significant difference for all the methods ($p < .04$) but for the Alphabetic. There are no significant differences between young and adult users. Figure 12 shows 95% family-wise confidence intervals for the methods with statistical significant differences.

If we compare the empirical data with the habits reported by the users, we see that there are no differences in error rates for users with different habits toward SMS or TV services. Nevertheless, error rates depend to some extent on computer usage habits but only for QWERTY and Genetic. In general, the higher the computer usage, the lower the error rates are for these methods. This is clearer for users with more than 12 hr a week of computer usage, who present lower error rates than the rest of the users with statistically significant differences ($p < .047$).

Finally, the results present no differences between male and female users and left- versus right-handed users.

Subjective impressions. As mentioned previously, users had to fill in questionnaires in each session reporting their degree of satisfaction and their impressions about speed and ease of use. Figure 13 shows the mean scores of these values in Sessions 1 and 5. In the figure we can see how users improve their scores as sessions pass by. Not only do they feel faster with experience (with the single exception of Alphabetic), which coincides with empirical data, but they also feel more confident in general with all the methods.

If we compare subjective scores with empirical data, we can see that slight differences exist. First, if we use empirical data to rank the methods according to their speed, we obtain in all the sessions the same classification, as shown in Figure 8. On the other hand, if we rank the methods according to subjective impressions, we see how users change their minds as the experiment evolves. For instance, users feel that the fastest method in the first session is Multitap, which coincides with empirical data, but in the last sessions they feel that the fastest method is the Genetic. Second, the subjective impressions do not coincide

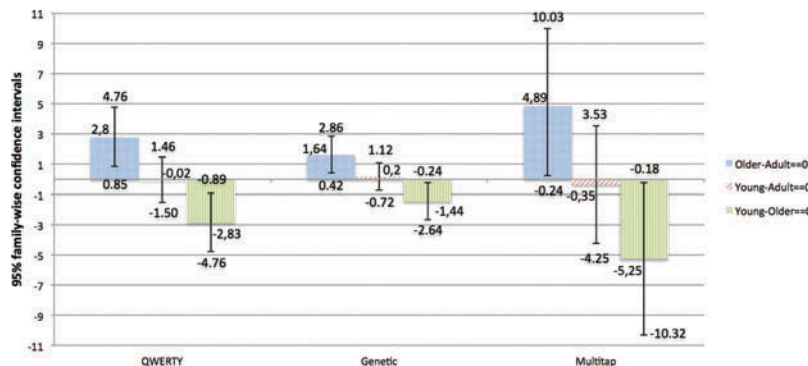


FIG. 12. 95% family-wise confidence intervals for error rates in methods with significant differences.

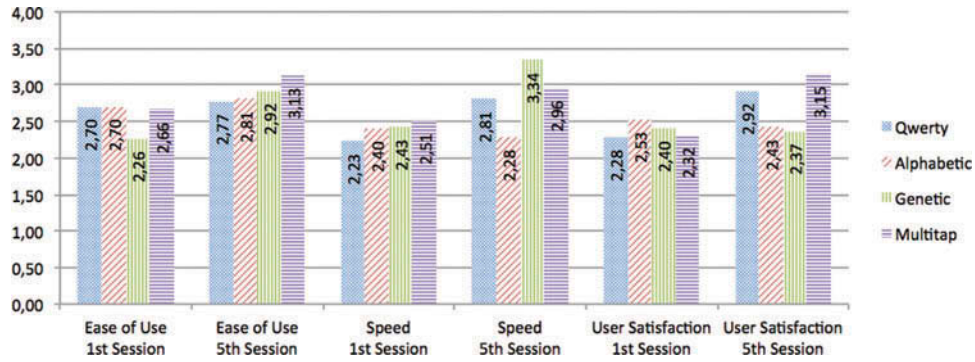


FIG. 13. Mean scores of subjective impressions per method in Sessions 1 and 5 (0 = *unsatisfied* to 4 = *very satisfied*).

with empirical data. For example, at the end of the experiment users found that the slowest method was the Alphabetic and the fastest the Genetic, which contrasts with empirical data that shows Multitap the fastest and QWERTY the slowest. Third, although empirical data reflect a good performance of the Genetic method, this is the worst method according to user satisfaction at the end of the experiment. It is considered to be easy and fast, but users were not satisfied perhaps due to the confusing position of the letters. Finally and surprisingly, users considered the Genetic to be the fastest method, but they were much more satisfied with Multitap, even though this is the worst method according to empirical error rates.

4.2. Second Task: Evaluation of Optimized Methods

In this task, only three of the four methods used in the previous phase were used: QWERTY, Genetic, and Multitap. Alphabetic was discarded because it did not show any significant differences with QWERTY in the previous task. The methods were modified to include a set of optimizations. In the previous task, many users highlighted that in virtual keyboards they had to cross the layout to insert a blank space or to delete a symbol. They suggested using keys of the remote for these actions, and in this task we tried to check whether this improves performance. Also, two enhancements of Multitap taken from previous work were considered: T9 and 2-key (Silverberg et al., 2000).

Each user carried out two sessions on 2 consecutive days. With regard to the previous task, we reduced the number of sessions, as we were more interested in the effect of the improvements than in the analysis of learning patterns. In every session, users had to write five random sentences of the corpus with all the methods.

The particular characteristics of the participants are shown in Table 4. Some of the users who participated in this task also participated in the first task. Twelve of these users were in the group young, seven in adult, and four in older. This situation will be taken into account, as their previous experience may affect the results.

Analysis of writing speed. If we compare the performance of all the methods, pairwise comparisons performed with Tukey tests present clear differences between them. These comparisons are statistically significant ($p < .001$) in the case of T9, which is the fastest method ($M = 58.59, SD = 18.81$). Also, comparisons are statistically significant in the case of 2-key ($p < .05$), which is the slowest method ($M = 32.34, SD = 8.39$). Genetic ($M = 39.84, SD = 8.34$) is slightly faster than QWERTY ($M = 38.39, SD = 6.75$), but the difference is not statistically significant. Comparing these results with those gathered in the first task, we are able to detect whether the optimizations have been effective. To reach more interesting conclusions, we have different remarks for users who participated in the first task and users who did not. The results are shown in Table 5. For repeating users, we have compared the

TABLE 4
Characteristics of the Participants in the Second Task

	Total	Gender		Field (Profession or Studies)		Education		
		M	F	IT	Non IT	Elementary	High School	University
Age								
Young	33	21	12	26	7	0	3	30
Adult	10	7	3	7	3	0	2	8
Older	5	3	2	1	4	2	1	2

TABLE 5
Differences Between the First Task and the Second Task

	Repeating Users		New Users	
	Task 1 Session 5	Task 2 Session 2	Task 1 Session 2	Task 2 Session 2
QWERTY	38.56	38.05	29.74	38.62
Genetic	44.34	39.43	36.06	40.10
Multitap – T9	52.60	56.13	43.71	60.20
Multitap – 2-key	52.60	33.04	43.71	31.88

Note. Units: characters per minute.

speed in the last session of each task. For new users, we have compared the speed of the users of the first task in the second session of that task with the speed of new users in the second task in their second session, as all of them had the same degree of experience.

For repeating users pairwise comparisons show that there are no statistically significant differences for QWERTY, Genetic, and T9. Nevertheless, the results obtained with 2-key are 37.17 % worse than with the standard Multitap, with a statistically significant difference. In the case of new users, they improve performance with the optimizations for QWERTY ($p < .001$), Genetic, and T9 ($p < .001$), although in the case of Genetic the difference is not statistically significant. Again, the performance with 2-key is worse than with the standard Multitap ($p < .001$). In conclusion, the results show that new users improve faster with the optimizations than without them for all the methods but the 2-key. Also, a little experience greatly improves performance, reaching similar results within a reduced period of time. It is also interesting to see that the results of Multitap, T9, and 2-key methods coincide with those in [Silfverberg et al. \(2000\)](#) and in [Butts and Cockburn \(2002\)](#), but they are in contrast with

those in [Ingmarsson et al. \(2004\)](#); TNT having a two-keystroke approach similar to 2-key) and [Geleijnse et al. \(2009\)](#).

If we compare the empirical data with the habits reported by the users, although there are some differences in speed depending on the usage of the SMS service, computers, TV, or T9 in mobile phones, these differences are not statistically significant. Mean values are shown in [Table 6](#). It is very interesting to see that the results obtained with the T9 method by frequent T9 users ($M = 65.86$ characters per minute) are not particularly better than those obtained by other users ($M = 57.08$ characters per minute).

In the results we have seen that age is a major factor affecting performance, as shown in [Figure 14](#). Again, the general criterion is that the younger the user, the faster he or she writes. Pairwise comparisons of average speeds show that young users and adult users are faster than older users with a statistically significant difference in all the methods ($p < .023$) but the 2-key. In the case of the 2-key, the difference between adult and older users is not significant. Also, although young users are faster than adult users, this difference is statistically significant only for T9 ($p < .015$) and 2-key ($p < .01$).

Finally, ANOVA tests show us that there are no statistically significant differences between male and female users and left-versus right-handed users.

Analysis of error rates. In general, virtual keyboards show much lower error rates than the other methods. Tukey tests show that T9 is the worst method ($M = 20.92$, $SD = 18.7$), with statistically significant differences with the rest of the methods ($p < .001$). Genetic is the best method ($M = 1.63$, $SD = 1.5$) with a significant difference with 2-key ($M = 7.41$, $SD = 7.06$; $p = .028$) but not with QWERTY ($M = 3.2$, $SD = 2.31$). Finally, the difference between QWERTY and 2-key is not significant.

We can compare these results with those of the first task, with different conclusions for users who participated in the first task and users who did not ([Table 7](#)). For repeating users, we

TABLE 6
Writing Speed Depending on User Habits

	QWERTY	Genetic	T9	2-Key
SMS messages per day	<1: 38.62	<1: 40.11	<1: 58.58	<1: 33.60
	1–5: 35.97	1–5: 37.58	1–5: 55.03	1–5: 28.47
	>5: 40.74	>5: 42.91	>5: 59.68	>5: 39.69
Hours of computer per week	<3: 38.46	<3: 40.69	<3: 57.51	<3: 41.15
	3–12: 36.32	3–12: 36.69	3–12: 60.35	3–12: 32.23
	>12: 38.05	>12: 39.72	>12: 57.20	>12: 31.89
Hours of TV per week	<7: 37.91	<7: 38.96	<7: 57.72	<7: 30.77
	7–21: 38.18	7–21: 40.40	7–21: 57.64	7–21: 33.44
	>21: 34.12	>21: 30.80	>21: 54.20	>21: 28.87
Usage of T9	No: 37.93	No: 39.66	No: 57.08	No: 31.15
	Yes: 36.71	Yes: 36.94	Yes: 65.86	Yes: 35.26

Note. Units: characters per minute. SMS = Short Text Messages.

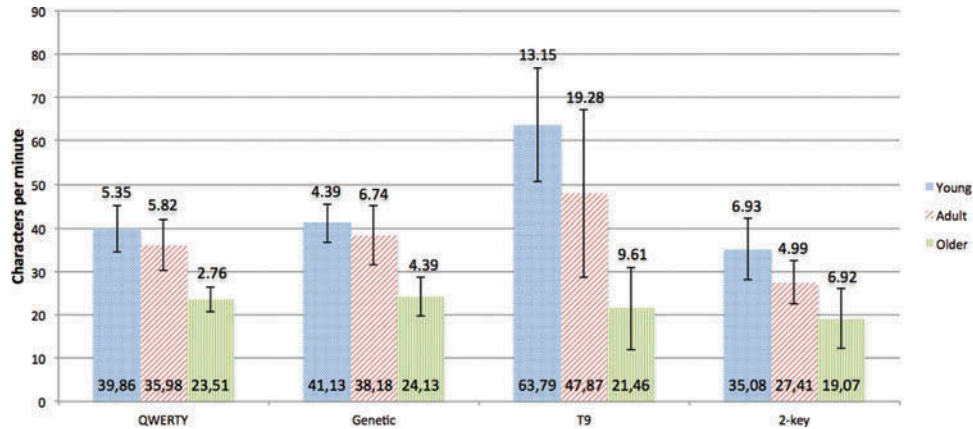


FIG. 14. Mean and standard deviation values of writing speed per method and age.

TABLE 7
Differences in Error Rates Between the First Task and the Second Task

	Repeating Users		New Users	
	Task 1 Session 5	Task 2 Session 2	Task 1 Session 2	Task 2 Session 2
QWERTY	2.05	2.91	2.11	3.39
Genetic	1.49	1.94	1.55	1.50
Multitap – T9	8.04	24.47	8.27	18.60
Multitap – 2-key	8.04	7.58	8.27	7.31

have compared error rates in the last session. For new users, we have compared error rates in the second session, as all of them had the same degree of experience. Pairwise comparisons show that the only statistically significant difference for repeating users is with T9, with an increase of 204% in error rates. Furthermore, for new users the difference is statistically significant for QWERTY ($p < .001$), with an increase of 60.18%, and T9 ($p < .001$), with an increase of 124.87%.

Although there are slight differences in error rates depending on the age of the users or their habits, pairwise comparisons show that these are not statistically significant. Also, there are no differences between male and female users and left- versus right-handed users.

4.3. Third Task: Evaluation of Methods With Special Characters and Specific Information

In the previous tasks we considered only numbers and lowercase letters. However, if applications need to be developed in a global environment, new symbols are needed. First, there are symbols needed to solve internationalization issues as happened in Iatrino and Modeo (2006) or Ingmarsson et al. (2004). Second, certain contexts such as web browsing, messaging, e-commerce, or e-government require special symbols

(@, slashes, commas). Thus, tildes, accents and special symbols need to be included somehow if we would like to work under realistic conditions. To accommodate these situations, the input methods chosen for this task are the modified versions of QWERTY, Genetic, and Multitap.

To ensure the use of these special characters, the users had to fill in a form with their personal information including name, ID number, date of birth, address, and e-mail. All the users also had to write a common piece of text: an Internet internationalized resource identifier (this IRI is the same for all the users). An example follows:

- Name and surnames: Xurde Rodríguez Fonticiella
- Personal identification code: 12345678X
- Date of birth: 20/02/2002
- Home address: C/ My Address, nº 11
- E-mail address: myemail.address@email.com
- Internet IRI: http://es.wikipedia.org/wiki/Teclado_informática

As the system did not provide the texts to be written, we followed a different approach to calculate entry speed and error rates. Entry speed was calculated once the form was completed by using the number of characters in the form and the total time taken by the user. To calculate error rates we counted the characters deleted by the users.

This experiment took place on a single day, during a session of about 35 min. We performed only one session, as the goal was to analyze the increase in the complexity of the methods. The particular characteristics of the participants are shown in Table 8. Although some of the users who participated in this task also participated in the first and/or second tasks, we have not considered this situation in the analyses because the design of this task differs greatly from the design of the previous tasks. Thus, we consider any previous experience irrelevant for the results of this task.

Analysis of writing speed. In this case, the fastest method is QWERTY ($M = 30.66$, $SD = 6.71$) followed by Multitap ($M = 27.07$, $SD = 6.92$), and Genetic ($M = 26.97$, $SD = 6.49$).

TABLE 8
Characteristics of the Participants in the Third and Fourth Tasks

Age	Total	Gender		Field (Profession or Studies)		Education		
		M	F	IT	Non IT	Elementary	High School	University
Young	24	17	7	20	4	0	2	22
Adult	13	9	4	9	4	0	1	12
Older	5	3	2	1	4	2	1	2

Nevertheless, an ANOVA test shows that there are no significant differences between the methods ($p = .058$). If we compare these results with those gathered in previous tasks, we can see that they are much worse. If we compare the results of virtual keyboards in Tasks 2 and 3 (optimizations are used in both tasks) with QWERTY, the speed decreases 20.15%, whereas with Genetic the decrease is 32.30%. Moreover, if we compare the results of Multitap in Tasks 1 and 3 (T9 and 2-key are not used), speed is now 48.09% worse. Clearly, there is an increase of complexity in usability that has a great impact on performance.

If we compare the empirical data with the habits reported by the users, the differences in speed depending on the usage of SMS, computers or TV are not statistically significant. Mean values are shown in Table 9. The only exception is Multitap: Users who send more than five SMS messages per day are faster than the rest of the users ($p < .045$).

Again, we have seen that age affects performance. As in previous tasks, the younger the user, the faster he or she writes. Pairwise comparisons show that young and adult users are faster than older users with a statistically significant difference in all the methods ($p < .01$). Furthermore, the difference between young and adult users for Multitap is also significant ($p = .018$), which coincides with the results of previous tasks.

TABLE 9
Writing Speed Depending on User Habits

	QWERTY	Genetic	Multitap
SMS messages per day	<1: 30.30	<1: 27.28	<1: 27.38
	1–5: 28.61	1–5: 26.17	1–5: 23.81
	>5: 37.86	>5: 30.39	>5: 36.97
Hours of computer per week	≤12: 30.28	≤12: 25.85	≤12: 26.34
	>12: 30.39	>12: 27.24	>12: 26.93
Hours of TV per week	≤7: 28.26	≤7: 26.35	≤7: 23.44
	>7: 31.50	>7: 27.56	>7: 28.71

Note. Units: characters per minute.

We have not found any statistically significant differences between male and female users and left- versus right-handed users.

Analysis of error rates. As in previous tasks, Multitap produces higher error rates than virtual keyboards. Tukey tests show that Multitap is the worst method ($M = 16.86$, $SD = 9.71$), with statistically significant differences with other methods ($p < .001$). QWERTY ($M = 4.85$, $SD = 3.51$) is slightly better than Genetic ($M = 5.49$, $SD = 5.65$), but the differences are not significant.

If we compare these results with those gathered in previous tasks, we can see that error rates are much worse. If we compare the results of virtual keyboards in Tasks 2 and 3 with QWERTY error rates, these increase 48%. With Genetic they increase 211%. Also, if we compare the results of Multitap in Tasks 1 and 3, error rates increase 163%. Again, the increase of complexity in usability has a great impact on performance.

Although there are slight differences in error rates depending on the age of the users, pairwise comparisons show that these are not statistically significant. Also, we have not found any significant differences if we compare the methods with users' habits. The only exception is Genetic: Users who use computers more than 12 hr a week have lower error rates than users who do not (4.67% vs. 12.85%). Finally, the results also show that there are no differences between male and female users and left- versus right-handed users.

4.4. Fourth Task: Evaluation of Different Remote Controls

In this case, users had to use remote controls with different keysets. As pointed out by Nielsen (2012), the differences in the design of remote controls due to the lack of any standardization constitute a huge usability problem. Thus, the goal of this task is to discover if changes in the shape of the remote control or the location of keys affect user performance. In this experiment two additional remote controls were used, as shown in Figure 15.

The remote control on the left of the figure had been used in the rest of the tasks. It is a SnapStream Firefly remote control, as mentioned previously, and is referred to as "remote #1." We chose the remote control in the center of the image because the number and arrow sets are placed to one side. The numbers are



FIG. 15. Remote controls used in the fourth task.

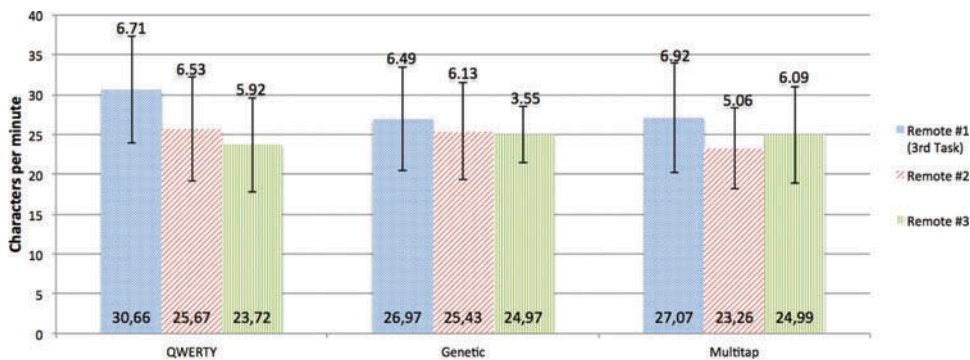


FIG. 16. Mean and standard deviation values of writing speed per method.

situated on the top-left part and the arrows and OK keys on the center-right part. It is a Golden Interstar remote control and will be referred to as “remote #2.” We chose the remote control on the right of the image to check whether the size was important or not, as it is notably smaller than the rest. Also, this remote allowed us to check the impact of reversing the order of its elements, as the numbers are on the bottom-part and the arrows and OK keys on the top part. This is an AverMedia RMKS remote control and is referred to as “remote #3.”

To compare the results with those gathered in previous tasks, we repeated the third task but using the new remote controls. This allows us to reuse the data gathered in the third task by comparing it with the data obtained with remote controls #2 and #3. Obviously, this time the form had to be filled in twice, each time with each of the new remote controls. To make data comparable, we used exactly the same group of people in the third and fourth tasks, so their particular characteristics are shown in Table 8.

Analysis of writing speed. In this case, the results are slightly worse than in the third task, as shown in Figure 16. If we

consider independently the results of each remote control, there are differences in speed between each of the methods but they are not statistically significant according to ANOVA tests ($p > .05$). The same happened in the third task with remote #1.

If we perform pairwise comparisons using all the information gathered in the third and fourth tasks, we obtain the best and worst pair remote-method. As shown in Figure 17, the best results are obtained with remote #1 when QWERTY is used, with statistically significant differences in some cases (p ranging from less than 0.001 to 0.04 in those cases). On the other hand, the worst results are obtained with remote #2 and Multitap, but the differences are not significant according to Tukey tests.

Also, in the same pairwise comparison we can see whether the shape or location of the keys affect performance. For this purpose, we can analyze each method independently by comparing its pairs with all the remote controls (e.g., compare all the pairs of QWERTY with each other). In Figure 17 we can see that the only statistically significant difference is with remotes #1 and #3 with QWERTY ($p = .003$). Thus, our results show

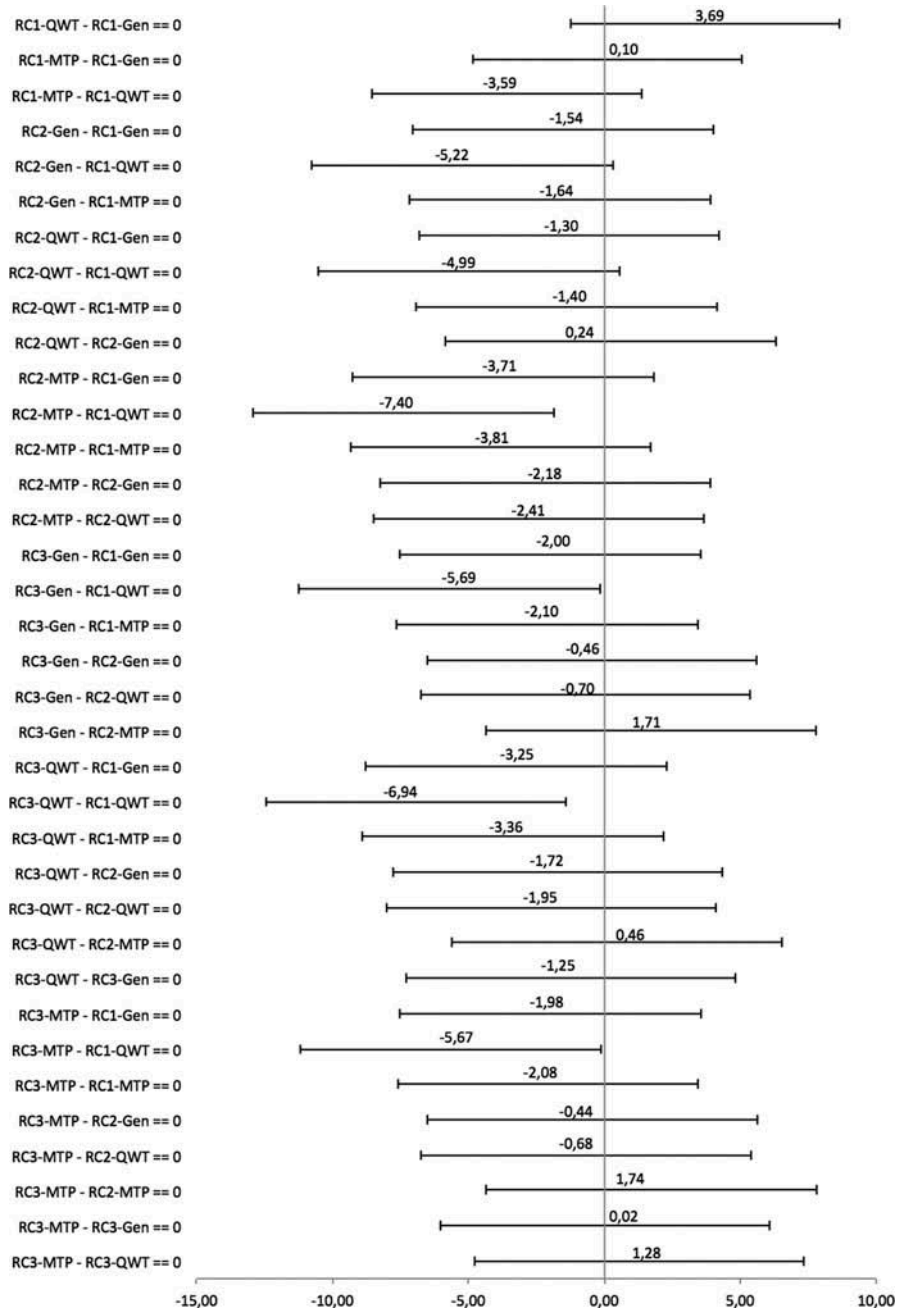


FIG. 17. Pairwise comparisons of writing speeds for each method and remote control (95% family-wise confidence level). Note. MTP = Multitap.

that, in general, the shape and location of the keys does not produce significant differences in speed.

Analysis of error rates. As in previous tasks, error rates registered with Multitap are much higher than with the rest of the methods in every remote control, as shown in Figure 18. Furthermore, the difference between Multitap and the rest of the methods is statistically significant in remote controls #2 and #3 as shown in Figure 19 ($p < .001$). The same happened in the third task with remote control #1.

As with writing speed, if we perform pairwise comparisons using all the data of the third and fourth tasks, we obtain the best and worst pair remote-method. As shown in Figure 19, the best results are obtained with Genetic and remote #2. The differences are statistically significant with Multitap for all the remote controls, as stated previously. On the other hand, the worst results are obtained with remote #1 and Multitap with statistically significant differences with virtual keyboards for all the remote controls according to Tukey tests.

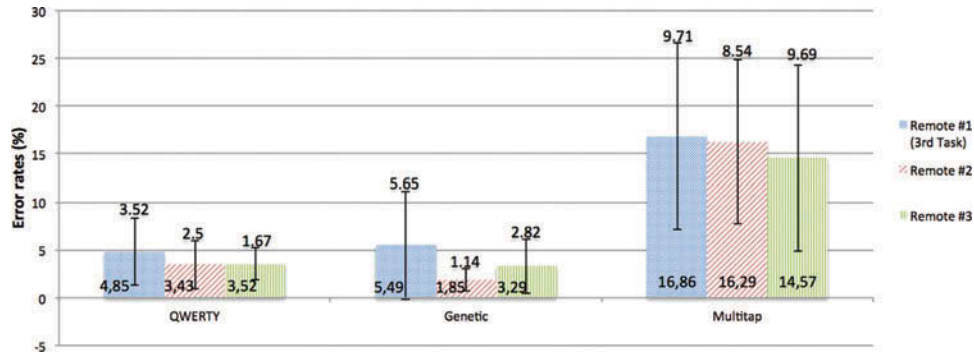


FIG. 18. Mean and standard deviation values of error rates per method.

As with writing speed, we can see if the shape or location of the keys affects error rates in the same pairwise comparison. In Figure 19 we can see that there are no significant differences if we analyze each method independently. Our results show that, in general, the shape and location of the keys does not produce significant differences in error rates.

5. CONCLUSIONS AND FUTURE WORK

Our article deals with human interaction in the IDTV realm with an important set of contributions. We have analyzed several types of virtual keyboard layouts and other methods, some of them complemented with different improvements, such as disambiguation techniques or suggestion systems. We have performed the evaluations using both general sentences and specific characters and texts, attending to issues such as internationalization or usage in particular contexts. We have evaluated how the shape and the location of the keys in remote controls affect user performance. Also, we present results gathered with a considerable number of real users with heterogeneous characteristics: age, gender, level of knowledge, and technology usage habits. These contributions are especially important if we consider that there is little research on text writing methods for IDTV applications with remote controls.

Our results show that, when simple texts need to be written, the method with the best performance is Multitap, despite the fact that this method produces higher error rates than virtual keyboards. Nevertheless, when writing complex texts, virtual keyboards present the same or even better writing speeds than Multitap and with significantly lower error rates. Using concurrent techniques has been fairly positive as stated by Wigdor and Balakrishnan (2004). When we combine virtual keyboards with certain buttons in the remote control for frequent actions we improve performance considerably. A logical conclusion is that users improve with experience, but we have seen that each method evolves differently: Multitap and the Genetic virtual layout show the best progressions. Another interesting conclusion is that subjective impressions do not always coincide with empirical results, and we think that the opinion of users can be as important or more than real performance. Furthermore, our results show that the shape and location of the keys in remote

controls does not significantly affect performance. Another general conclusion is that user habits may have an impact on performance in some cases, but not on a general basis. Although performance is not affected by certain user characteristics such as gender or hand orientation, age is a major factor affecting performance: the younger the user, the faster he or she is. Performance tends to be similar between mobile-like methods and virtual keyboards for older users. Another interesting conclusion is that certain optimizations are not as efficient as they were intended to be, which is confirmed by the results obtained with methods such as T9, 2-key or Alphabetic virtual keyboard layouts.

Some recommendations for developers of IDTV applications follow:

- It is a good idea to provide multiple methods so users can choose the method they are most comfortable with. Performance is important, but we have seen that subjective opinions may indicate something totally different.
- If only one method can be used, there are better and worse methods depending on the context of the application. If simple tests need to be written, Multitap or T9 are good options. On the other hand, if special symbols are required virtual keyboards are better, as a wider range of users would be satisfied. In the latter case, the Genetic is a good option as performance improves greatly with experience.
- Use concurrent techniques to take advantage of the rest of the buttons of the remote control. For instance, spare buttons can be used for frequent actions such as deleting characters or writing blank spaces.
- Apparently, the design of the remote control does not have an impact on performance, but remote controls should be designed with enough distance between buttons to avoid fat-finger problems (Siek et al., 2005).

One of our design goals was to investigate text entry methods that may be used with commonly available technology that most people will probably already have in their homes. Thus, future work will be mainly based on experiments with devices different to conventional remote controls.

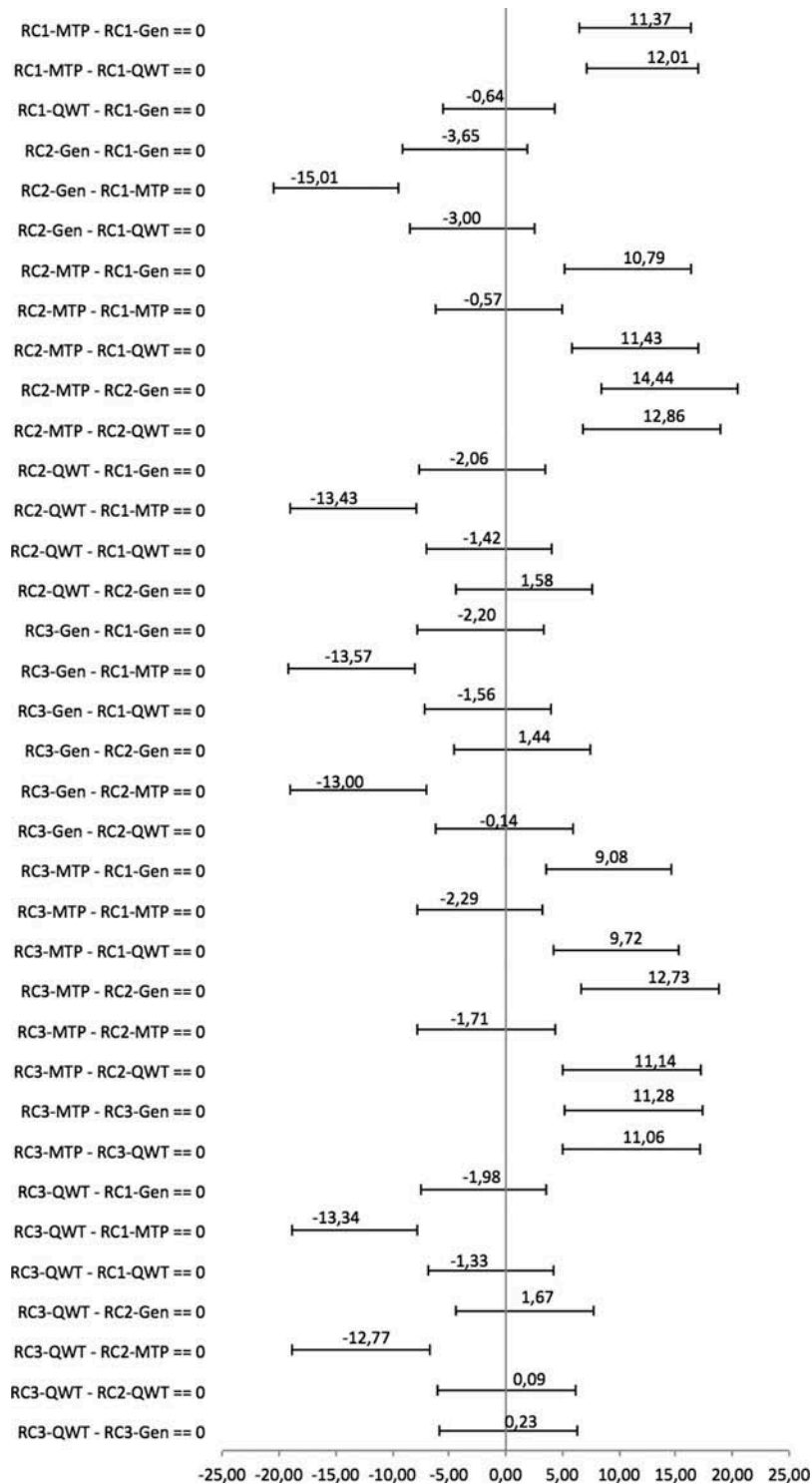


FIG. 19. Pairwise comparisons of error rates for each method and remote control (95% family-wise confidence level).

Although articles such as [Orbist et al. \(2008\)](#), recommend not to rely on these devices, and the results of some experiments such as those in [MacKenzie et al. \(2009\)](#) may not

be as were expected, it is increasingly evident that the popularity of devices to interact with televisions different to the conventional remote controls will increase in the near future.

Thus, experiments may be undertaken to compare the performance of the evaluated methods with devices such as gyroscopic remote controls (pointing devices), mini-keyboards, touchpads, tablets, smartphones, or multitouch devices. Apart from writing speed and accuracy, we would also like to measure other factors that may influence the experience of users. By incorporating further equipment, we would like to measure stress levels, brain activity, eyesight effects, and similar parameters, which may allow us to draw even more interesting conclusions.

FUNDING

This work was partially supported by the University of Oviedo and the Principality of Asturias through the project SV-PA-13- ECOEMP-75 and by the cable operator Telecable de Asturias S.A.U. through the project “Diseño de sistema para gestión y provisión de aplicaciones/servicios de televisión digital interactive.”

REFERENCES

- Aoki, R., Maeda, A., Watanabe, T., Kobayashi, M., & Abe, M. (2010). Twist tap: Text entry for TV remotes using easy-to-learn wrist motion and key operation. *IEEE Transactions on Consumer Electronics*, *56*, 161–168.
- Bellman, T., & MacKenzie, I.S. (1998). A probabilistic character layout strategy for mobile text entry. *Proceedings of Graphics Interface '98*, 168–176.
- Brandtzæg, P. B., Heim, J., & Karahasanović, A. (2011). Understanding the new digital divide—A typology of Internet users in Europe. *International Journal of Human-Computer Studies*, *69*, 123–138.
- Brewbaker, C. R. (2008). *Optimizing stylus keyboard layouts with a genetic algorithm: Customization and internationalization*. Department of Computer Science, Iowa State University, Ames.
- Butts, L., & Cockburn, A. (2002). An evaluation of mobile phone text input methods. *Journal of Australian Computer Science Communications*, *24*, 55–59.
- Choi, S., Han, J., Lee, G., Lee, N., & Lee, W. (2011). RemoteTouch: Touchscreen-like interaction in the TV viewing environment. *Proceedings of CHI 2011*, 393–402.
- Clarkson, E., Clawson, J., Lyons, K., & Starner, T. (2005). An empirical study of typing rates on mini-QWERTY keyboards. *Proceedings of CHI '05*, 1288–1291.
- Congressional Record, Vol. 155. (2009). DTV Delay Act, Public Law 111-4 Feb. 11, 2009.
- Crawley, M. J. (2007). *The R book*. Chichester, England: Wiley & Sons.
- European Broadcasting Union. (2006). Digital Video Broadcasting (DVB); Multimedia Home Platform (MHP) Specification 1.0.3. ETSI specification ETSI ES 201 812 V1.1.2.
- European Telecommunications Standards Institute (ETSI) ES 202 130 V2.1.2, 2007-09, Human Factors (HF); User Interface; Character repertoires, ordering rules and assignments to the 12-key telephone keypad. http://portal.etsi.org/stfs/STF_HomePages/STF300/es_202130v020102p.zip
- European Union. (2005). *Communication from the Commission to the Council, the European Parliament, the European Economic and Social committee and the Committee of the Regions on accelerating the transition from analogue to digital broadcasting* {SEC(2005)661}.
- Gargi, U., & Gossweiler, R. (2010). QuickSuggest: Character prediction on web appliances. *Proceedings of WWW 2010*, 1249–1252.
- Geleijnse, G., Aliakseyeu, D., & Sarroukh, E. (2009). Comparing text entry methods for interactive television applications. *Proceedings of EuroITV'09*, 145–148.
- Iatrino, A., & Modeo, S. (2006). Text editing in digital terrestrial television: A comparison of three interfaces. *Proceedings of EuroITV'06*, 224–241.
- Ingmarsson, M., Dinka, D., & Zhai, S. (2004). TNT—A numeric keypad based text input method. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 639–646.
- Költringer, T., Isokoski, P., & Grechenig, T. (2007). TwoStick: Writing with a game controller. *Proceedings of Graphics Interface 2007*, 103–110.
- MacKenzie, I. S. (1991). *Fitts' law as a performance model in human-computer interaction* (Unpublished doctoral dissertation). University of Toronto, Toronto, Ontario, Canada.
- MacKenzie, I. S., Kober, H., Smith, D., Jones, T., & Skepner, E. (2001). LetterWise: Prefix-based disambiguation for mobile text input. *Proceedings of the ACM Symposium on UIST*, 111–120.
- MacKenzie, I. S., Lopez, M. H., & Castellucci, S. (2009). Text entry with the Apple iPhone and the Nintendo Wii. *Proceedings of CHI2009*.
- Nesbat, S. B. (2003). A system for fast, full-text entry for small electronic devices. *Proceedings of the 5th International Conference on Multimodal Interfaces*, 4–11.
- Nielsen, J. (2012). *Remote control anarchy*. Available from <http://www.useit.com/alertbox/20040607.html>
- Oniszcak, A., & MacKenzie, I. S. (2004). A comparison of two input methods for keypads on mobile devices. *Proceedings of the Third Nordic Conference on Human-Computer Interaction*, 101–104.
- Orbist, M., Bernhaupt, R., & Tscheligi, M. (2008). Interactive TV for the home: An ethnographic study on users' requirements and experiences. *International Journal of Human-Computer Interaction*, *24*, 174–196.
- Perrinet, J., Pañeda, X. G., Cabrero, S., Melendi, D., García, R., & García, V. (2011). Evaluation of virtual keyboards for interactive digital television applications. *International Journal of Human-Computer Interaction*, *27*, 703–728.
- RAE - Royal Spanish Academy. (2001). *Spanish language dictionary* (22nd ed.). Available from <http://www.rae.es/rae.html>
- Rick, J. (2010). Performance optimizations of virtual keyboards for stroke-based text entry on a touch-based tabletop. *Proceedings of the 23rd Annual ACM Symposium on UIST*, 77–86.
- Ritter, F. E., & Schooler, L. J. (2002). The learning curve. In *International encyclopedia of the social and behavioral sciences* (pp. 8602–8605). Amsterdam, the Netherlands: Pergamon. Available from <http://www.iesbs.com/>
- Siek, K. A., Rogers, Y., & Connelly, K. H. (2005). Fat finger worries: How older and younger users physically interact with PDAs. *Proceedings of Interact 2005*, 267–280.
- Silfverberg, M., MacKenzie, I. S., & Korhonen, P. (2000). Predicting text entry speed on mobile phones. *Proceedings of the CHI2000 Conference*, 9–16.
- Spira, J. B. (2011). Internet TV: Almost ready for prime time [Tools & Toys]. *IEEE Spectrum*, *48*, 24–26.
- Sporcka, A. J., Polacek, O., & Slavik, P. (2012). Comparison of two text entry methods on interactive TV. *Proceedings of the 10th European Conference on Interactive TV and Video*, 49–52.
- Taveira, A. D., & Choi, S. D. (2009). Review study of computer input devices and older users. *International Journal of Human-Computer Interaction*, *25*, 455–474.
- Varcholik, P. D., LaViola, J. J., Jr., & Hughes, C. E. (2012). Establishing a baseline for text entry for a multi-touch virtual keyboard. *International Journal of Human-Computer Studies*, *70*, 657–672.
- Vega-Oliveros, D. A., Pedrosa, D. C., Pimentel, M. G. C., & De Mattos Fortes, R. P. (2010). An approach based on multiple text input modes for interactive digital TV applications. *Proceedings of the 28th ACM International Conference on Design of Communication*. 191–198.
- Wigdor, D., & Balakrishnan, R. (2004). A comparison of consecutive and concurrent input text entry techniques for mobile phones. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 81–88.
- Wobbrock, J. O., Myers, B. A., & Aung, H. H. (2004). Writing with a joystick: A comparison of date stamp, selection keyboard, and EdgeWrite. *Proceedings of Graphics Interface '04*, 1–8.
- Zhai, S., Hunter, M., & Smith, B. A. (2000). The Metropolis keyboard—An exploration of quantitative techniques for virtual keyboard design. *Proceedings of ACM Symposium on User Interface Software and Technology*, 119–128.

ABOUT THE AUTHORS

Aurora Barrero is a Computer Science Engineer from the University of Oviedo and a Ph.D. student with an interest in the area of multimedia systems and services, content distribution networks, and interactive digital TV services. She is a Research Engineer in the Department of Computer Science of the University of Oviedo.

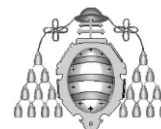
David Melendi is a Computer Science Engineer with a Ph.D. from the University of Oviedo and an interest in multimedia systems and services, content distribution networks, idTV services, and mobile ad hoc networks. He is an Associate Professor at the University of Oviedo and a member of the W3C.

Xabiel G. Pañeda is a Computer Science Engineer with a Ph.D. from the University of Oviedo and an interest in multimedia systems and services, content distribution networks,

idTV services, and mobile ad hoc networks. He is an Associate Professor at the University of Oviedo and a member of the W3C.

Roberto García is a Telecommunications Engineer from The Technical University of Madrid with a Ph.D. from the University of Oviedo and an interest in telecommunication networks and services, applied to performance analysis, modelling, and simulation of systems and services. He is an Associate Professor at the University of Oviedo.

Sergio Cabrero is a Telecommunications Engineer from the University of Oviedo and a Ph.D. student with an interest in the area of telecommunication networks, interactive digital TV services, multimedia services, and mobile ad hoc networks. He is a Teaching Assistant in the Department of Computer Science of the University of Oviedo.



6.2 An Empirical Investigation Into Typing Errors in Interactive Digital Television Applications

A. Barrero, D. Melendi, X. G. Pañeda, R. García, y S. Cabrero, «An Empirical Investigation Into Typing Errors in Interactive Digital Television Applications», *International Journal of Human-Computer Interaction*, vol. 31, n.o 3, pp. 210–225, 2015. DOI: 10.1080/10447318.2014.994195

An Empirical Investigation Into Typing Errors in Interactive Digital Television Applications

Aurora Barrero, David Melendi, Xabiel G. Pañeda, Roberto García, and Sergio Cabrero

Department of Informatics, University of Oviedo, Asturias, Spain

Remote controls are especially problematic for writing text, something necessary in many interactive digital TV applications (IDTV). In the context of research aimed at finding effective entry methods for IDTV applications, an empirical investigation was carried out and is presented in this article. The study analyzed the errors committed by 82 users in an experiment in which they wrote 7,395 sentences. Several typing methods suitable for IDTV contexts were used for experimentation. There were 3,562 of the sentences that registered at least one error. These errors were classified and their main causes analyzed, the particular characteristics of the users taken into account. The results show that the most frequent errors are proximity mistakes in all the evaluated methods. They appear in between 13.75% and 32.31% of the sentences, depending on the input method. Also, age and background are major aspects affecting the number and type of mistakes. The results could be the basis of changes in the design of conventional remote controls or in the design of advanced techniques for error recognition and correction.

1. INTRODUCTION

These days, fully interactive digital television (IDTV) applications are possible thanks to recent innovations in the Internet Protocol TV field. These innovations have encouraged content providers to offer advanced services in order to be competitive. Most of these advanced services require interaction capabilities, which are possible now thanks to the proliferation of home Internet connections (Brandtzæg, Heim, & Karahasanovic, 2011), the improvement of set-top-box and TV devices, and the evolution of telecommunication networks. The Internet and TV have merged, forming services such as Internet Protocol TV (Spira, 2011): a current, massively deployed Internet of Things in the Home service (Brown et al, 2013). However, if complex applications are to be developed, effective methods of user interaction have to be designed, taking into account that the main interaction device available is still the remote control. This is a problem when it comes to writing text, as remote controls were not designed for this purpose. Thus, if we need the

Address correspondence to David Melendi, 2.7.11 Edificio Polivalente, Campus de Viesques, s/n, Xixón, 33203 Asturias, Spain. E-mail: melendi@uniovi.es

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.

users to introduce text on IDTV applications, we have to find effective methods of text entry with a remote control, considering aspects such as speed, accuracy, the opinion of the users, their background, and their age (Taveira & Choi, 2009).

To find effective typing methods for IDTV applications, we carried out empirical research on several text entry methods suitable for this context. The main goal of this research is to analyze text entry methods that may be used with commonly available technology that most people will probably already have in their homes. Although the general results of our research are available in Barrero, Melendi, Pañeda, García, and Cabrero (2014), we have gone beyond these results and now present empirical information about the errors made by a set of heterogeneous users. These users participated in an experiment in which they had to type several sentences with a set of writing methods using a conventional television remote control. The results show the most common errors for each of the evaluated methods and the influence of the particular characteristics of the users. This knowledge may take us further than the raw comparison of input methods. For example, our results show that, in general, the most frequent errors are proximity errors in all methods (users writing a symbol adjacent to the desired one). We have also seen that age and background influence to some extent the type and number of errors made, but not gender or laterality. Finally, several improvements to these typing methods are proposed in order to reduce the number of errors, and thus enhance the experience of the users.

The rest of the article is organized as follows. In section 2, previous work is revisited. The plan of the experiments to be carried out and the subjects who participated in the experiments are described in section 3. Results are detailed in section 4. Finally, conclusions and future work are presented in section 5.

2. RELATED WORK

The study and comparison of text-typing methods is a field in which much research has been carried out. Nevertheless, most recent advances have been focused on mobile devices such as smartphones or tablet devices. Although some solutions may be transferred from these domains to IDTV, there are still many open issues as stated by Iatrino and Modeo (2007). Although several years have passed since Iatrino and Modeo made this

statement, it still holds true. First, the main interaction device is still a conventional remote control. The interaction of conventional remote controls is mainly performed with number, arrow, and OK keys, and this may impose severe constraints for IDTV applications. Second, if we compare TVs with mobile devices or tactile surfaces, we must take into account the fact that the keys and the screen are in different devices, making it impossible for users to look at the interaction device and the screen at the same time. In the rest of this section we only revise some previous work either addressing the issue of writing text in the specific context of IDTV or studying errors in detail. For a general survey on typing methods, please see Lewis, Commarford, Kennedy, and Sadowski (2008).

For the specific context of IDTV, there are only a few noteworthy studies in which the issue of writing text is treated. Still, none of them performs a detailed study of typing errors. They simply compare different input methods taking into account the number of errors and not their specific nature, as part of a performance evaluation. Such is the case of Ingmarsson, Dinka, and Zhai (2004), who presented a new technique called TNT. TNT works by letting the user press two numeric keys to produce a letter on the screen. The first key press selects a group, and the second selects a member in that group. Each character is accessible with two keystrokes. Five people between ages 27 and 32 participated in the experiment to write a short novel. Results show speeds comparable to or faster than manual writing in a PDA or the Multitap method. Multitap was designed for mobile phones and is based on having several symbols assigned to each key. With multiple keystrokes, the user cycles through the symbols for that key. Køltringer, Ngo Van, and Grechenig (2007) presented a study comparing an alphabetical selection keyboard to a Multitap selection keyboard using a game controller as input device. Ten volunteer users with an average age of 27.3 participated in the experiment. They analyzed writing speed and the number of corrected and uncorrected errors. Geleijnse, Aliakseyeu, and Sarroukh (2009) compared Multitap, T9, and a QWERTY virtual keyboard layout, with the speed of a conventional QWERTY keyboard (with the same keyset of Multitap, T9 is a predictive method that allows words to be written by a single keystroke for each of its letters). They also introduced a recommendation system. Twenty-two people between 21 and 32 years of age participated in the experiment. In terms of the time spent on the tasks, the authors concluded that a conventional keyboard using their recommendation system is the most efficient. With their recommendation system and Multitap or the virtual keyboard, completion times are comparable with a conventional keyboard without intelligence. The worst results are obtained with T9. Orbist, Bernhaupt, and Tscheligi (2008) stated that a voice-controlled IDTV would probably be the most important gateway between humans without PC experience and IDTV. Furthermore, Vega-Oliveros, Pedrosa, Pimentel, and De Mattos Fortes (2010) explored a multimodal text input on IDTV combining speech input, Multitap, and a virtual keyboard. Using questionnaires and usability tests, they highlight

the importance of using complementary text input modes in order to satisfy the needs of different users. The experiments in Perrinet et al. (2011) were further developed in Barrero et al. (2014). Ninety-six people between 20 and 64 years of age participated in the experiments to analyze writing speed, learning profiles, and the global number of errors using several virtual keyboards—Multitap, 2-key, and T9 with different remote controls. The same as TNT, 2-key is based on a first keystroke to choose a group of symbols and a second keystroke for the selection of a symbol (Silfverberg, MacKenzie, & Korhonen, 2000). Choi, Han, Lee, Lee, and Lee (2011) performed experiments with an ad hoc–designed remote control equipped with a touchpad. Twelve university students with an average age of 25.7 participated in the experiments. They measured entry speed, error rates, and the participant’s task workload. Sporka, Polacek, and Slavik (2012) compared TNT and a new method called TwiceTap. This method has the same philosophy as TNT but, apart from allowing users to type single characters, it allows them to write frequent blocks of characters (n-grams). Eighteen paid users, with an average age of 22.7, participated in the experiment. There were no significant differences between the methods, but users preferred TwiceTap mostly because of its similarity to Multitap and the availability of n-grams. It is true that Iatrino and Modeo (2007) paid some attention to the problems suffered by the users during their experiments, but not to the empirical information of the errors they committed. In conclusion, to our knowledge there are no previous research works providing detailed empirical information about the errors made by IDTV users.

There are some research works outside the IDTV context in which detailed information about errors made by real users is provided. This can be the case of MacKenzie and Soukoreff (2002), who stated that a complete analysis involves determining what kind of errors occurs and why. They identified four basic types of errors: users entering an incorrect character (substitution), users omitting a character (omission), users adding an extra character (insertion), or users swapping neighboring characters (transposition). Nevertheless, the aim of the article was not to include empirical results. Also, Chen, Yesilada, and Harper (2010) presented a user study that investigates input errors of mobile web users while both typing and pointing.

The study reruns the experiments in Trewin and Pain (1999), but in a different context, and identified six types of typing errors: long key press errors (a key is pressed too long and generates unwanted copies), bounce errors (a key is unintentionally pressed more than once and generates unwanted copies), missing key errors (a target letter is not typed in), transposition errors (two characters adjacent to each other are typed in reverse order), additional key errors (a key adjacent to the target key is unintentionally pressed), and key ambiguity errors (a participant cannot distinguish different letters on the same key). Some of the errors identified in Trewin and Pain (1999) are not considered because they are not valid for the context of the experiment. Nineteen people between 19 and 44 years of age participated

in an experiment in which they had to perform a set of typing and pointing tasks with a PDA equipped with a small keyboard. The number of errors detected is relatively low, key ambiguity errors being the most frequent (average of 9.33 errors per user). They suggested that key ambiguity errors are closely related to users' familiarity with small keyboards and experiences in text entry. Also, Henze, Rukzio, and Boll (2012) study typing problems when using virtual keyboards on mobile devices. For this purpose they developed a typing game that recorded how users touched a standard Android keyboard. Nevertheless, they did not analyze the types of errors made by the users. They focused on analyzing the distribution of touch events by computing the distances between hit locations and the center of each key.

Also outside the IDTV domain, it is worth mentioning previous work aimed at studying accessibility issues, as in the case of Trewin and Pain (1999), where the authors perform an empirical investigation into typing and pointing errors caused by motor disabilities. Twenty-six users 25 to 72 years of age participated in experiments where a conventional keyboard and a mouse were used. Regarding typing errors, they identify seven types of errors: long key press errors (a key is pressed too long and generates unwanted copies), additional key errors (a key adjacent to the target key is activated), missing errors (the intended key is missed), dropping errors (user fails to press two keys simultaneously), bounce errors (a key is unintentionally pressed more than once), remote errors (a key not adjacent to the intended key is pressed) and transposition errors (two keys are transposed). In some cases, the number of errors detected is considerable, long key press errors being the most frequent (2,610 of 3,201). They pointed out that the main problem is to find the ideal key repeat delay for each user. Although the comparison group made no key press length errors, pressing keys for longer than the default key repeat delay was the most common problem for users with physical disabilities. They performed several tests to measure the lengths of the key presses of the users. Users from the comparison group did the tests with a repeat delay of 10 ticks (one tick = 1/60 seconds), other users had a repeat delay of 24 ticks, and the rest of the users did the tests with the key repeat switched off. Eventually, they used the obtained values to infer the number of errors that would have occurred had each subject been using the default key repeat delay of a Macintosh at the time (16 ticks). The average key

press length of the users of the comparison group was five ticks, which is consistent with Lewis, Potosnak, and Magyar (1997). But among the main group of users, the lengths ranged from four to 20 ticks, with some users reporting serious problems to press keys quickly. They concluded that key repeat settings are probably the most important issue in keyboard accessibility.

As can be seen from the literature discussed previously, there is little research into the types of problems that users experience when typing text. Furthermore, this is even scarcer in the IDTV domain. As stated previously, the main interaction device is still a conventional remote control. Thus, the conclusions of research works based on experiments with real users such as Trewin and Pain (1999), Chen et al. (2010), or Henze et al. (2012) are not valid in this context. Nevertheless, we think that it is important to know the errors made by the users in order to find effective solutions to reduce or even eliminate typing mistakes in this environment. Although some solutions may be transferred across different domains, the constraints imposed by the design of conventional remote controls need to be taken into account. In this article we present an empirical study that aims to identify the types of errors committed by users of an IDTV application and to propose solutions to improve their current quality of experience.

3. EXPERIMENTAL DESIGN

Data were gathered from 82 users who were asked to write several sentences in an application using a remote control. They were required to attend several sessions in which the same text entry methods were used.

3.1. Description of the Population

The users who participated in the experiments were as heterogeneous as possible, with ages ranging between 20 and 64 and different technological skills, gender, and laterality. They did not receive money or any other kind of compensation. Some of their characteristics are shown in Table 1.

We grouped users according to their age because changes in perceptual and motor skill capabilities that accompany the aging process bring important implications for the design of human-computer interfaces (Taveira & Choi, 2009). This is

TABLE 1
Characteristics of the Participants in the Experiment

Age	Total	Gender		Field (Profession or Studies)		Education		
		M	F	IT	Non-IT	Elementary	High School	University
Young	48	27	21	34	14	0	5	43
Adult	22	14	8	12	10	0	4	18
Older	12	8	4	1	11	5	1	6

Note. IT = information technology.

aligned with the differences reported by Siek, Rogers, and Connelly (2005). Several users belonged to the “mobile phone generation,” between 18 and 30 years of age (young); other users belonged to the “computer generation,” between 31 and 45 years of age (adult); and finally “pre-pc” users older than 45 (older). In the analyses, we use the term “background” to refer to both the profession and the studies. In general, the members of each group had a different attitude toward technology due to their personal development contexts.

3.2. Text Input Methods

Users had to write with several text input methods suitable for IDTV applications. The description of these methods follows:

- QWERTY virtual keyboard: This is probably currently the most popular layout. It was designed placing many adjacent letter pairs (digraphs) on the opposite sides of the keyboard. This facilitates the frequent alternation of the left and right hand but, on a virtual keyboard, the polarizing common digraphs mean that the user has to move back and forth more frequently and over greater distances than necessary (Zhai, Hunter, & Smith, 2000).
- Square alphabetic virtual keyboard: In this layout the keys are placed in alphabetical order. It is supposed to be the “easiest” in the sense that anyone who knows the alphabet would be able to use it without difficulty. Nevertheless, previous work reports usability problems (Norman & Fisher, 1982) and a considerably worse performance if compared with QWERTY (Lewis, LaLomia, & Kennedy, 1999).
- Genetic virtual keyboard: We used a genetic algorithm to generate a new keyboard layout in order to improve the efficiency of traditional keyboards. Using a modern version of Don Quixote as a reference text for Spanish, the algorithm positioned the keys so that the most used letters appear in the center of the keyboard. The result is shown in Figure 1. The details and cost model of this algorithm can be found in Brewbaker (2008). The idea is to achieve higher text entry speeds



FIG. 1. Genetic keyboard layout.

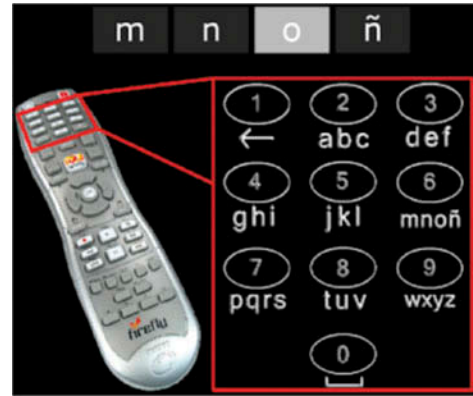


FIG. 2. Multitap method.

by reducing keystrokes, as in Bellman and MacKenzie (1998). We conducted a theoretical analysis to compare this layout with the QWERTY layout. If we consider the minimum number of keystrokes needed to write Don Quixote in Spanish, this method requires 45% less keystrokes than the QWERTY layout.

- Multitap: Based on mobile phones, this method uses number keys to write text as shown in Figure 2. As these keys are available in almost every remote control, it is possible to use this mechanism for IDTV systems. This, combined with the fact that previous works show that it is more efficient than virtual keyboards (Iatrino & Modeo, 2007), is the reason why we considered this method for the evaluation. To differentiate between multiple strokes corresponding to a single letter and strokes between successive letters, we use a threshold of 1 s as in a conventional phone.
- 2-key: This is a system designed to reach any letter with only two keystrokes (Silfverberg et al., 2000). Using numeric keys, a first keystroke is used to select a group of letters and a second keystroke is used to select the desired letter. In the example shown in Figure 3, if the user would like to write the symbol “t,” he or she would have to press number key 8 and then 1. 2-key is slightly faster than Multitap for Spanish if we compare two keystrokes with a calculated weighted average of 2.23 Multitap keystrokes per letter for the words in the official Spanish dictionary (RAE – Royal Spanish Academy, 2001).

To neutralize learning effects, the methods were assigned to participants in a counterbalanced order (analyses of variance for test order showed no significant differences).

3.3. Apparatus

To carry out the experiment, we installed a 32-in. LCD television with dimensions 814 × 599 mm placed on top of a table of around 1 m high. The equipment was placed in a quiet private

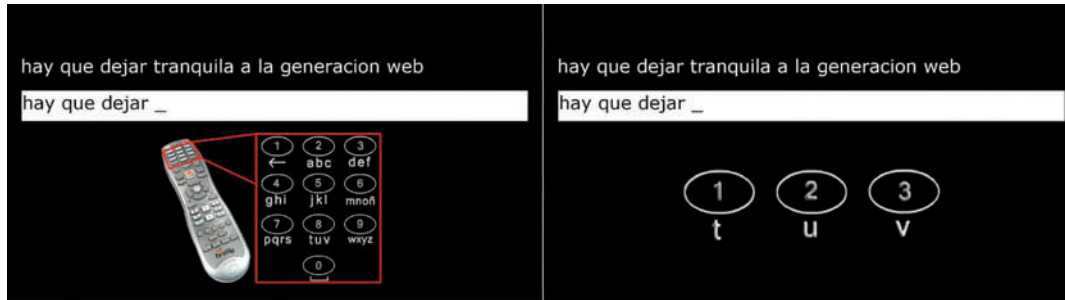


FIG. 3. 2-key method.

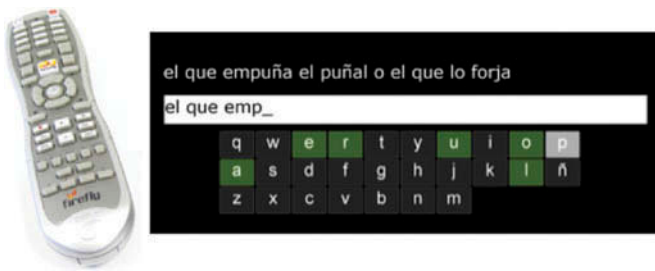


FIG. 4. Remote control and testbed application.

office with an illumination of 200 lx as recommended in ITU-T (2009). According to the recommendations of the television manufacturer, an armchair was placed at a distance of 2 m.

We connected the television to a PC with the Microsoft Windows XP operating system. We also connected to the PC a SnapStream Firefly PC remote control through an USB radio frequency receiver. This remote, shown in Figure 4, incorporated the most common keys of conventional television remote controls: numbers, OK, and cursors. It is important to take into account that one of our design goals was to investigate typing methods that may be used with commonly available technology that most people would probably already have in their homes.

The PC had an Adobe® AIR® application designed to carry out the tests. As shown in Figure 4, our design is similar to the application of Millet, Asfour, and Lewis (2009). Our application had a black full-screen layout and one or more text input fields to allow the users to write. The text to be written was placed on top of an input field. Our fonts were Verdana 65 points for the proposed sentences and the input field, and Arial 60 points for keyboard layouts.

The application was developed to process keyboard events, so it was fully usable with a standard keyboard. Then we mapped remote control strokes into keyboard events by using the EventGhost software. The system has no auto-repeat feature (MacKenzie, 2002), and no forced time lags between keystrokes were introduced.

The application showed a 10-s counter at the beginning of the experiment and between different input methods. This

allowed the users to take a look at what they were about to find on screen: mainly the structure of virtual keyboards and the correspondence between the symbols and the keys on the remote for Multitap and 2-key. This 10-s counter was also used before writing each sentence, allowing the user to read the sentence in advance. After this 10-s counter, the application allowed the user to write until the written sentence coincided exactly with the proposed sentence. To handle mistakes, the application provided users with mechanisms to delete characters (a key in the remote or a symbol on screen depending on the method). Once a sentence was written, the application showed a new sentence or switched to a different input method.

When virtual keyboards were used, the keyboard layout was placed on the bottom part of the screen as shown in Figure 1. Users had to use the cursors of the remote control to move around the keyboard layout and the “OK” button to perform an action. We used a wraparound cursor feature as in Millet, Asfour, and Lewis (2008). On the other hand, when mobile-like techniques were used, the bottom part of the screen was used to show help information, as in Figures 2 and 3. The keys in the application were placed in accordance with the European standard ETSI ES 202 130 V2.1.2 (2007-09).

The application stored information of the activity of the users in XML log files. In these files the application registered the following events:

- Start of a typing method
- Start of a sentence
- Timestamp and type of keystroke
- Timestamp and type of symbol written
- Answers to questionnaires

3.4. Procedure

At the beginning of the first session, the participants were given a brief explanation on the general structure of the experiment and an example of each of the typing methods to be used. They were not allowed to try these methods before the experiment. Also, no instructions were given to users regarding how to handle the remote control, that is, with the left hand, with the right hand, with both hands, loosely, tightly, and so on. Thus,

we expected them to use the remote control as they usually do. Also, in this session users had to fill in a questionnaire with their personal details (gender, age, profession, level and type of studies, habits of using television, etc.). This information allowed us to refine our conclusions by contrasting users' characteristics with empirical data.

Users were left alone in a room in order to avoid distractions. We wanted the experiment to be as realistic as possible, so each user sat in front of a television as if the user were in his or her own living room and had no time constraints to complete the experiment. Users were provided with the texts to write. For this purpose, we created a small corpus with 50 short sentences representative of the Spanish language gathered from local newspapers. We followed a similar approach to MacKenzie and Soukoreff (2003) but for the Spanish language. The sentences had an average of five words, and they included only numbers and lowercase letters from the English alphabet, with the single exception of the Spanish ñ character. Two examples follow "móviles que hacen de todo" and "la idea es ponerselo fácil."

After each session, users had to complete a final questionnaire providing us with some feedback.

The experiment was divided into two phases, each having a different number of sessions and using different methods. We had two phases, because the complete experiment had a considerable amount of goals that recommended doing this. For instance, in the first phase we wanted to analyze learning patterns with a set of methods. Nevertheless, this was not a goal of the second phase, in which we wanted to include additional typing methods. In the first phase, five sessions were carried out on 5 consecutive days. In these sessions the methods used were QWERTY, Alphabetic, Genetic, and Multitap. In the second phase, users had to attend two sessions and perform tests with QWERTY, Genetic, and 2-key. In each session, users had to write five sentences with each of the methods. Other results of these experiments can be found in Barrero et al. (2014), including speed, error rates, learning profiles, and subjective opinions of the users.

To extract the conclusions of the experiment we performed several statistical analyses. The main goal was to determine if there were real differences between the means of two or more groups of variables. We used the most common tests for this type of analysis, according to Crawley (2007). These tests depended on the assumptions of normality and homoscedasticity of the data. First, we checked normality with Shapiro-Wilk tests and homoscedasticity with Bartlett tests. When data met both normality and homoscedasticity we used one-way analysis of variance (ANOVA) tests to compare the data. If homoscedasticity failed, then we used Kruskal-Wallis tests. Kruskal-Wallis tests were also used in situations with a strong failure in normality (p values in normality tests over .05). Finally, if differences existed, we used student's t tests or Tukey tests with a confidence coefficient of 95% to perform pairwise comparisons.

4. RESULTS

During the experiment we gathered 7,395 sentences written by the users with the aforementioned typing methods. Due to the organization of the experiments and other issues with the users, a different number of sentences was written with each of the methods. There were 2,017 sentences written with QWERTY, 1,397 with the Alphabetic, 2,001 with the Genetic, 1,375 with Multitap, and 605 with 2-key.

We consider that an error is produced every time a user must delete something. Under these circumstances, 97.26% of the users made at least one error, and the average number of errors per sentence is close to 1 ($M = 0.961$, $SD = 0.443$). This shows that writing with a remote control on a television can be considered as a difficult task. As shown in Figure 5, users make more mistakes with mobile-designed methods (Multitap and 2-key) than with virtual keyboards. In fact, with Multitap or 2-key, 98.63% of the users made at least one error, and the average number of errors per sentence is high ($M = 1.889$, $SD = 1.191$ for Multitap and $M = 2.459$, $SD = 1.389$ for 2-key). As shown in Figure 5, with mobile-like methods one error was at least detected in 70% of the sentences. The situation improves notably with virtual keyboards. Although 95.94% of the users made at least one mistake, the average number of errors per sentence was much lower than in the previous case ($M = 0.619$, $SD = 0.357$ for QWERTY, $M = 0.575$, $SD = 0.343$ for the Genetic, and $M = 0.516$, $SD = 0.327$ for the Alphabetic). If we compare the results obtained with each of the methods, significant differences might exist according to one-way ANOVA tests, $F(4, 370) = 74.793$, $p < .05$. A 95% confidence level post hoc Tukey test confirms that there are significant differences between several methods. 2-key is the method with the highest number of errors per sentence ($p < .008$). Multitap is better than 2-key ($p = .008$) but worse than the rest ($p < .05$). Finally, QWERTY is slightly worse than the Alphabetic method ($p = .043$).

During the analysis of the sentences with errors, we have identified the following types of error:

- Missed symbols. When writing a sentence, we have identified that many users forget to write a given symbol and keep on writing the symbols that follow. We have considered separately blank spaces and the rest of the symbols because this is a special situation: After writing a given word, users start writing the one that follows immediately.
- Mistaken deletions. Sometimes users delete a correct character and, thus, have to write it again.
- Late identified errors. We have identified that a considerable amount of errors are produced because users continue writing after an error has been made, and once they have noticed the error, they have to delete many symbols that were correct.
- Bounce errors. These errors are produced when a given symbol is written more times than necessary.

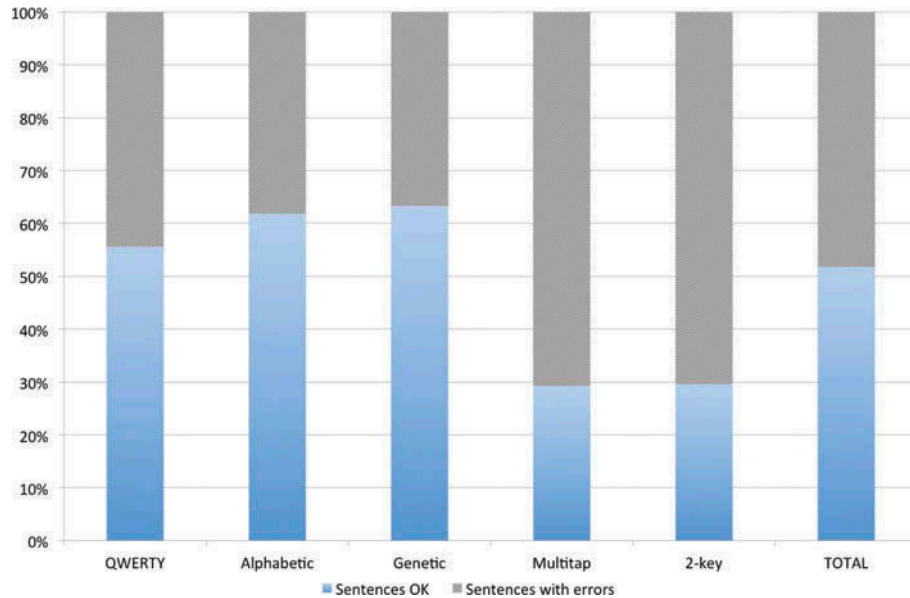


FIG. 5. Proportion of sentences with errors versus sentences without them.

- Transposition errors. These errors occur when two adjacent symbols are typed in reverse order.
- Proximity errors. These errors are produced when an incorrect symbol is written due to some kind of proximity with the correct symbol.

All these error types are common to all the aforementioned typing methods. But there are also other errors that are specific to a given typing method, which are listed next:

- Wrong finger bounces with Multitap. In Multitap, if a given key is pressed more or less times than necessary an incorrect symbol is written. We take into account these errors only when an incorrect symbol from the same key of the correct one is written.
- Wrong group selection for 2-key. These are produced when a user selects an incorrect group of symbols in 2-key.
- Wrong symbol selection for 2-key. These are produced when a user selects the correct group of symbols but an incorrect character.

We have partially followed the semantics in Trewin and Pain (1999) to give names to these errors. Their definition has been adapted to the context of the experiment.

A summary of the results is shown in Table 2.

4.1. Missed Symbols

When writing the sentences, some users occasionally tend to omit symbols. This effect may be due to many reasons including an off-target user aim, insufficient force when pressing a

given key, or just a momentary lapse. It is interesting to see how Trewin and Pain (1999) claimed that this was one of the most frequent errors in their control group (users with no disabilities). We have analyzed these missed symbols differentiating between blank spaces and other characters. Tables 3 and 4 show the most relevant results of the analyses for missed blank spaces and missed symbols, respectively.

As we can see in Table 3, blank spaces have been forgotten by 90.41% of the users at least once during the experiments. There were 49.54% of these errors made during the first session of the users; thus we can consider that they are caused by the experiment itself. In fact the general average number of errors per sentence is low. Users want to write the words in the proposed sentence, and they forget the first blank space in the majority of the cases. If we compare the results obtained with each of the methods, significant differences might exist according to one-way ANOVA tests, $F(4, 370) = 19.378$, $p < .05$. A 95% confidence level post hoc Tukey test shows that significant differences exist between virtual keyboards and mobile-like methods ($p < .0002$). Furthermore, significant differences exist between the Genetic and other virtual keyboards ($p < .034$) and between Multitap and 2-key ($p = .003$). Thus, we can conclude that this error is more common in virtual keyboards than in mobile-designed methods, being Genetic the best virtual keyboard and Multitap the best method in general.

To check the influence of the personal characteristics of the users with the number of errors of this type, we have performed several analyses. Regarding the age of the users, no significant differences exist between groups according to one-way ANOVA tests, $F(2, 82) = 0.601$, $p = .55$. The same happens with their

TABLE 2
Summary of the Results of the Experiment

	% Users With Errors	Errors per Sentence	Dependencies
Missed blanks	90.41	$M = 0.074, SD = 0.064$	Method
Missed symbols	97.26	$M = 0.110, SD = 0.064$	Method, Background
Mistaken deletions	79.45	$M = 0.025, SD = 0.030$	Method
Late identification	95.89	$M = 0.214, SD = 0.118$	Method
Bounce errors	89.00	$M = 0.049, SD = 0.036$	Method
Transposition errors	13.69	$M = 0.001, SD = 0.005$	—
Proximity errors (Virtual keyboards)	97.26	$M = 0.158, SD = 0.086$	Age, Background
Proximity errors (Multitap)	95.00	$M = 0.418, SD = 0.489$	None
Wrong finger bounces (Multitap)	88.33	$M = 0.307, SD = 0.296$	Background
Wrong group	92.18	$M = 0.488, SD = 0.390$	Background
Wrong symbol	93.75	$M = 0.460, SD = 0.300$	Background

TABLE 3
Results of the Analysis of Missed Blank Spaces

	% of Users With Errors	Errors per Sentence
General	90.41	$M = 0.074, SD = 0.064$
Results per method		
QWERTY	80.82	$M = 0.114, SD = 0.110$
Genetic	75.00	$M = 0.073, SD = 0.089$
Alphabetic	77.05	$M = 0.105, SD = 0.113$
Multitap	16.67	$M = 0.006, SD = 0.015$
2-key	23.44	$M = 0.026, SD = 0.054$
Results per age group		
Young	85.00	$M = 0.068, SD = 0.069$
Adult	95.45	$M = 0.072, SD = 0.055$
Older	100.00	$M = 0.094, SD = 0.062$
Results per background		
IT	88.09	$M = 0.068, SD = 0.073$
Non-IT	93.55	$M = 0.079, SD = 0.050$

Note. IT = information technology.

TABLE 4
Results of the Analysis of Missed Symbols

	% of Users With Errors	Errors per Sentence
General	97.26	$M = 0.110, SD = 0.064$
Results per method		
QWERTY	84.93	$M = 0.090, SD = 0.073$
Genetic	84.93	$M = 0.097, SD = 0.080$
Alphabetic	63.01	$M = 0.092, SD = 0.099$
Multitap	63.01	$M = 0.075, SD = 0.073$
2-key	75.34	$M = 0.290, SD = 0.238$
Results per age group		
Young	95.24	$M = 0.097, SD = 0.060$
Adult	100.00	$M = 0.113, SD = 0.067$
Older	100.00	$M = 0.151, SD = 0.058$
Results per background		
IT	97.60	$M = 0.096, SD = 0.056$
Non-IT	96.77	$M = 0.127, SD = 0.069$

Note. IT = information technology.

background according to one-way ANOVA tests, $F(1, 82) = 0.567, p = .453$.

If we consider different symbols than the blank space, a correct symbol has been missed by 97.26% of the users at least once, as shown in Table 4. The average number of errors per sentence is slightly higher than with blank spaces. On this occasion, we have not detected any influence on the number of the session, but depending on the method users made more or fewer mistakes of this type. We can see that this error is relatively common in the case of the 2-key method. In fact, significant differences between methods might exist according to one-way ANOVA tests, $F(4, 370) = 31.961, p < .05$. A 95% confidence level post hoc Tukey test shows that significant differences exist between 2-key and the rest of the methods ($p < .05$). A symbol

was missed in 24.46% of the sentences written with 2-key. Further analyses show that, in the particular case of this method, 75.49% of these errors were produced when users pressed an adjacent key to the correct one. If we ignore these situations, analyzed later in the article, the real percentage of users who with missed symbols with 2-key decreases to 30.14% ($M = 0.12, SD = 0.205$ errors per sentence).

If we analyze the influence of age with the number of errors of this type, no significant differences exist according to one-way ANOVA tests, $F(2, 82) = 2.819, p = .066$. In the case of the background of the users, significant differences might exist according to one-way ANOVA tests, $F(1, 82) = 4.421, p = .039$. A 95% confidence level *T* test shows that users with an IT background make less mistakes of this type than non-IT users

($p = .019$). We have not found any statistically significant difference between male and female users or between left-handed and right-handed users.

4.2. Mistaken Deletions

Users sometimes delete a character that is correct. This occurred 146 times during the experiments. As shown in Table 5, 79.45% of the users made at least one error of this type, but the average number of mistaken deletions per sentence is low.

If we analyze the results obtained with the different methods, significant differences between them might exist according to one-way ANOVA tests, $F(4, 370) = 12.242, p < .05$. A 95% confidence level post hoc Tukey test shows that significant differences exist between 2-key and the rest of the methods ($p < .05$) and between the Alphabetic and the rest of the methods ($p < .019$) but the Genetic. Nevertheless, we can observe that this error is not frequent. This is also confirmed if we analyze the number of sentences with mistaken deletions. This error appeared in 2.18% of the sentences written with QWERTY, 1.00% with the Alphabetic, 0.95% with the Genetic, 2.04% with Multitap, and 6.78% with 2-key. We have analyzed whether this effect was caused by an excessive number of delete actions while correcting other mistakes, but these cases were not significant.

Given the poor relevance of this error, we have not performed further analyses.

4.3. Late Identification

As reported by Iatrino and Modeo (2007), if we compare televisions with other devices, the keys and the screen are in different places, so users cannot look at the interaction device and the screen at the same time. This produces problems when a mistake is made and the user is not aware of it. The user continues writing and after a while notices that an error has been made. Due to the sequential order of writing text, the user then has to delete a lot of correct information. Table 6 shows the most relevant results of the analyses for late identification errors.

TABLE 5
Results of the Analysis of Mistaken Deletions

	% of Users With Errors	Errors per Sentence
General	79.45	$M = 0.025, SD = 0.030$
Results per method		
QWERTY	38.00	$M = 0.025, SD = 0.057$
Genetic	15.27	$M = 0.013, SD = 0.040$
Alphabetic	16.39	$M = 0.009, SD = 0.022$
Multitap	31.00	$M = 0.020, SD = 0.034$
2-key	46.00	$M = 0.068, SD = 0.088$

TABLE 6
Results of the Analysis of Late Identification Errors

	% of Users With Errors	Errors per Sentence
General	95.89	$M = 0.214, SD = 0.118$
Results per method		
QWERTY	82.19	$M = 0.134, SD = 0.161$
Genetic	98.63	$M = 0.075, SD = 0.090$
Alphabetic	54.10	$M = 0.056, SD = 0.108$
Multitap	100.00	$M = 0.728, SD = 0.415$
2-key	81.25	$M = 0.280, SD = 0.253$
Results per age group		
Young	92.85	$M = 0.199, SD = 0.108$
Adult	100.00	$M = 0.227, SD = 0.137$
Older	100.00	$M = 0.250, SD = 0.112$
Results per background		
IT	95.23	$M = 0.191, SD = 0.103$
Non-IT	96.77	$M = 0.245, SD = 0.131$

This occurred 1,640 times during the experiment, which is a considerable amount. According to one-way ANOVA tests, significant differences between methods might exist, $F(4, 370) = 89.945, p < .05$. A 95% confidence level post hoc Tukey test shows that significant differences exist between all the methods ($p < .003$), with the single exception of the Genetic compared with the Alphabetic ($p = .143$). The number of sentences with this error is also considerably higher with mobile-like methods (45.16% of the sentences with Multitap and 22.15% with 2-key) than with virtual keyboards (8.87% of the sentences written with QWERTY, 4.66% with the Alphabetic and 5.60% with the Genetic). The fact that virtual keyboards require the user to look at the screen makes this error less frequent with these methods. On the other hand, it is clear that users do not look at the screen when using a mobile like method, and thus this error tends to be very frequent (especially with Multitap).

We have also checked the influence of the personal characteristics of the users with the number of errors of this type. If we consider age, no significant differences exist according to one-way ANOVA tests, $F(2, 82) = 0.856, p = .429$. The same happens with the background of the users, as no significant differences exist according to one-way ANOVA tests, $F(2, 82) = 3.872, p = .053$. We have not found any statistically significant difference between male and female users or between left-handed and right-handed users.

4.4. Bounce Errors

Bounce errors are produced when a correct symbol is written more times than necessary. This happened 394 times. As shown in Table 7, 89% of the users performed at least a bounce error during the experiments, but the average number of errors per sentence is relatively low.

TABLE 7
Results of the Analysis of Bounce Errors

	% of Users With Errors	Errors per Sentence
General	89.00	$M = 0.049, SD = 0.036$
Results per method		
QWERTY	71.23	$M = 0.051, SD = 0.048$
Genetic	69.44	$M = 0.066, SD = 0.069$
Alphabetic	57.38	$M = 0.054, SD = 0.061$
Multitap	36.66	$M = 0.019, SD = 0.028$
2-key	29.68	$M = 0.046, SD = 0.083$

In general, a higher number of users perform this error with virtual keyboards than with mobile-like methods. Actually, significant differences between methods might exist according to one-way ANOVA tests, $F(4, 370) = 4.946, p = .0007$. A 95% confidence level post hoc Tukey test shows that significant differences exist between Multitap and the rest of the methods ($p < .011$). In the case of virtual keyboards, these errors are produced when users unintentionally press the OK key of the remote more than once. In the case of Multitap, 37.04% of these errors were made when entering the blank space. The key for the blank space does not have any other symbols, so it does not have any delay. Thus, it is easy to perform a double-click in this key. Also for Multitap, 44.44% of these errors were caused by wrong finger bounces, which are analyzed later in the article. For 2-key we have found that these errors are a collateral effect of 2-key specific features, which are analyzed later in the article.

If we consider the number of sentences with this error, it was detected in 5.45% of the sentences written with QWERTY, 5.37% with the Alphabetic, 6.75% with the Genetic, 1.96% with Multitap, and 4.46% with 2-key. As we can see, this error is not frequent, so no further analyses have been made.

4.5. Transposition Errors

Sometimes users write two adjacent symbols in reverse order, for example, “on” instead of “no.” This happened only 12 times, and only 13.69% of the users performed one or more mistakes of this type ($M = 0.001, SD = 0.005$ errors per sentence). The error was not registered with the QWERTY method, and it happened only once with the Alphabetic and the Genetic, eight times with Multitap, and twice with 2-key.

Due to the poor relevance of this error, we have not made any detailed analyses.

4.6. Proximity Errors in Virtual Keyboards

Proximity errors are produced when an incorrect symbol is written instead of a certain symbol, and they are both adjacent in the layout projected on the screen.

In virtual keyboard methods, this happened 1,240 times. Of the users, 97.26% performed at least one error of this type ($M = 0.158, SD = 0.086$ errors per sentence). We have differentiated between diagonal and cross proximity errors. In the case of cross proximity (75.48% of proximity errors), we consider symbols adjacent to the correct character either horizontally or vertically. The results of their analysis are shown in Table 8. On the other hand, diagonal proximity errors are produced when users press a key adjacent to the desired one, but diagonally. Their results are shown in Table 9.

For cross proximity errors, no significant differences between methods exist according to one-way ANOVA tests, $F(2, 232) = 1.54, p = .216$. Furthermore, to check the influence of the personal characteristics of the users with the number

TABLE 8
Results of the Analysis of Cross Proximity Errors

	% of Users With Errors	Errors per Sentence
General	95.89	$M = 0.119, SD = 0.066$
Results per method		
QWERTY	91.78	$M = 0.158, SD = 0.103$
Genetic	93.05	$M = 0.183, SD = 0.121$
Alphabetic	85.25	$M = 0.149, SD = 0.120$
Results per age group		
Young	92.85	$M = 0.113, SD = 0.074$
Adult	100.00	$M = 0.124, SD = 0.055$
Older	100.00	$M = 0.133, SD = 0.057$
Results per background		
IT	95.23	$M = 0.104, SD = 0.061$
Non-IT	96.77	$M = 0.137, SD = 0.069$

Note. IT = information technology.

TABLE 9
Results of the Analysis of Diagonal Proximity Errors

	% of Users With Errors	Errors per Sentence
General	87.67	$M = 0.039, SD = 0.033$
Results per method		
QWERTY	68.49	$M = 0.046, SD = 0.046$
Genetic	76.38	$M = 0.066, SD = 0.070$
Alphabetic	62.29	$M = 0.049, SD = 0.052$
Results per age group		
Young	80.95	$M = 0.037, SD = 0.033$
Adult	95.45	$M = 0.032, SD = 0.017$
Older	100.00	$M = 0.065, SD = 0.055$
Results per background		
IT	86.04	$M = 0.034, SD = 0.021$
Non-IT	90.32	$M = 0.046, SD = 0.044$

of cross proximity errors, we have performed several analyses. If we consider age, no significant differences exist according to one-way ANOVA tests, $F(2, 82) = 0.446$, $p = .641$. Nevertheless, if we consider the background of the users, significant differences might exist according to one-way ANOVA tests, $F(1, 82) = 4.55$, $p = .036$. A 95% confidence level T test shows that users with an IT background make fewer mistakes of this type than non-IT users ($p = .018$).

Furthermore, 87.67% of the users performed at least one diagonal proximity error, but the frequency is lower than in the previous case. No significant differences between methods exist according to one-way ANOVA tests, $F(2, 232) = 2.66$, $p = .072$. Again, we have checked the influence of the characteristics of the users with the number of diagonal proximity errors. According to one-way ANOVA tests, significant differences exist between age groups, $F(2, 82) = 3.336$, $p = .041$. A 95% confidence level post hoc Tukey test shows that older users perform more diagonal proximity errors than younger users ($p < .025$). Regarding the background of the users, no significant differences exist according to one-way ANOVA tests, $F(1, 82) = 2.44$, $p = .122$. We have not found any statistically significant difference between male and female users or between left-handed and right-handed users.

4.7. Wrong Finger Bounces With Multitap

The design of the Multitap method leads to a specific error that arises when a user bounces a given key more or fewer times than needed. For example, as shown in Figure 2, if the user would like to write the symbol “o,” he or she should press number key 6 three times. If he or she presses the number key more or fewer times, an incorrect symbol is written. Depending on the number of keystrokes, this does not produce an error, because the same symbol appears again in the sequence. For instance, if the user would like to write symbol “o,” he or she should press number key 6 three times, but “o” can be written again with four additional bounces. We have considered errors only when an incorrect symbol is written.

This error happened 432 times, and the most relevant results are shown in Table 10. Of the users, 88.33% made this error at least once, and the average number of errors per sentence is relatively high. If we separate fewer-keystroke and more-keystroke errors, the desired key was pressed fewer times than necessary—170 times. On the other hand, a more than necessary number of keystrokes was made 262 times.

To check the influence of personal characteristics on the number of errors of this type, we have performed several analyses. In general, no significant differences exist between age groups according to one-way ANOVA tests, $F(2, 68) = 0.302$, $p = .74$. If we only consider less-keystroke errors, no significant differences exist between age groups according to one-way ANOVA tests, $F(2, 68) = 0.798$, $p = .455$. The same happens if we only consider more-keystroke errors, $F(2, 68) = 0.109$, $p = .897$. In general, if we analyze the background of the users,

TABLE 10
Results of the Analysis of Wrong Finger Bounces With Multitap

	% of Users With Errors	Errors per Sentence
Globally	88.33	$M = 0.307$, $SD = 0.296$
Results per age group		
Young	84.37	$M = 0.280$, $SD = 0.351$
Adult	95.00	$M = 0.327$, $SD = 0.208$
Older	91.67	$M = 0.361$, $SD = 0.266$
Results per background		
IT	85.29	$M = 0.273$, $SD = 0.312$
Non-IT	92.31	$M = 0.35$, $SD = 0.274$
Less keystrokes	80.00	$M = 0.121$, $SD = 0.111$
Results per age group		
Young	71.87	$M = 0.106$, $SD = 0.126$
Adult	95.00	$M = 0.128$, $SD = 0.081$
Older	83.33	$M = 0.160$, $SD = 0.113$
Results per background		
IT	76.47	$M = 0.094$, $SD = 0.099$
Non-IT	84.61	$M = 0.156$, $SD = 0.119$
More keystrokes	78.33	$M = 0.186$, $SD = 0.210$
Results per age group		
Young	69.69	$M = 0.174$, $SD = 0.244$
Adult	90.00	$M = 0.199$, $SD = 0.167$
Older	91.67	$M = 0.201$, $SD = 0.176$
Results per background		
IT	74.28	$M = 0.179$, $SD = 0.234$
Non-IT	84.61	$M = 0.194$, $SD = 0.177$

Note. IT = information technology.

no significant differences exist according to one-way ANOVA tests, $F(1, 68) = 0.986$, $p = .325$. In the case of less-keystroke errors, significant differences might exist according to one-way ANOVA tests, $F(1, 68) = 4.861$, $p = .031$. A 95% confidence level T test shows that users with an IT background make fewer mistakes of this type than non-IT users ($p = .015$). Finally, for more-keystroke errors, no significant differences between background groups exist according to one-way ANOVA tests, $F(1, 68) = 0.071$, $p = .791$. We have not found any statistically significant difference between male and female users or between left-handed and right-handed users.

4.8. Proximity Errors With Multitap

We have separated proximity errors in virtual keyboards and Multitap, as they follow very different mechanics. A proximity error in a virtual keyboard means that a user has moved over the layout to a wrong position and pressed OK, whereas on Multitap it means that a user has pressed a key on the remote close to the proper key. This also affects performance, because when a user presses the wrong key on Multitap, he or she has to wait until

TABLE 11

Results of the Analysis of Proximity Errors With Multitap

	% of Users With Errors	Errors per Sentence
General	95	$M = 0.418, SD = 0.489$
Results per age group		
Young	96.87	$M = 0.378, SD = 0.331$
Adult	100.00	$M = 0.459, SD = 0.673$
Older	83.33	$M = 0.478, SD = 0.537$
Results per background		
IT	100.00	$M = 0.440, SD = 0.541$
Non-IT	88.46	$M = 0.389, SD = 0.421$

Note. IT = information technology.

the end of the timer in order to delete the symbol, creating a delay in writing speed.

In Multitap, a key close to the correct one was pressed 585 times. The most significant results are shown in Table 11.

We have checked the influence of personal characteristics on the number of proximity errors using several analyses. In the case of age groups, no significant differences exist according to one-way ANOVA tests, $F(2, 68) = 0.234, p = .791$. Regarding the background of the users, no significant differences exist according to one-way ANOVA tests, $F(1, 68) = 0.158, p = .692$. We have not found any statistically significant difference between male and female users or between left-handed and right-handed users.

4.9. Group and Symbol Selection Errors in 2-key

2-key is a system designed to reach any letter with only two keystrokes. Using numeric keys, a first keystroke is used to select a group of symbols and a second keystroke is used to select the desired symbol. Apart from what has already been mentioned, 2-key errors are mainly produced due to the selection of either an incorrect group or an incorrect symbol within a correct group. In the example of Figure 3, if the user would like to write the symbol “t” and he or she presses a key different than number key 8, a group selection error is produced. In the same example, even if the user pressed number key 8, if he or she pressed afterward a key different to number key 1, an incorrect symbol is written and this second type of error is produced. In both cases, the problem can be caused by the proximity of the pressed key to the correct key. These situations are analyzed in this section.

An incorrect group of characters was chosen 298 times. As shown in Table 12, 92.18% of the users made this error at least once, and the average number of errors per sentence is high. Of the errors, 90.93% were caused by pressing an adjacent key to the correct one, that is, the adjacent group.

TABLE 12

Results of the Analysis of Group Selection Errors in 2-key

	% of Users With Errors	Errors per Sentence
Globally	92.18	$M = 0.488, SD = 0.390$
Results per age group		
Young	85.29	$M = 0.403, SD = 0.433$
Adult	100.00	$M = 0.533, SD = 0.272$
Older	100.00	$M = 0.725, SD = 0.399$
Results per background		
IT	91.89	$M = 0.442, SD = 0.423$
Non-IT	92.86	$M = 0.546, SD = 0.342$
Adjacent group	90.62	$M = 0.446, SD = 0.373$
Results per age group		
Young	82.86	$M = 0.365, SD = 0.400$
Adult	100.00	$M = 0.488, SD = 0.289$
Older	100.00	$M = 0.675, SD = 0.391$
Results per background		
IT	88.88	$M = 0.398, SD = 0.412$
Non-IT	92.85	$M = 0.507, SD = 0.315$

Note. IT = information technology.

We have performed several analyses to check the influence of personal characteristics on the number of wrong selections. In general, in the case of age groups, no significant differences exist according to one-way ANOVA tests, $F(2, 72) = 2.55, p = .086$. This also happens if we specifically treat errors performed when selecting the adjacent group, $F(2, 72) = 2.56, p = .085$. Regarding the background of the users, in general, no significant differences exist according to one-way ANOVA tests, $F(1, 72) = 1.116, p = .295$. This also happens when the adjacent group was chosen: no significant differences exist according to one-way ANOVA tests, $F(1, 72) = 1.345, p = .250$. We have not found any statistically significant difference between male and female users or between left-handed and right-handed users.

Even when the group was correct, the wrong symbol was chosen 282 times. As shown in Table 13, 93.75% of the users made this error at least once, and the average number of errors per sentence is also high. Of these errors, 77.74% were caused by pressing an adjacent key to the correct one.

Again, we have performed several analyses to check the influence of personal characteristics on the number of wrong symbol selections. In general, if we consider age groups no significant differences exist according to one-way ANOVA tests, $F(2, 72) = 0.265, p = .768$. This also happens if we specifically treat errors performed when selecting the adjacent symbol. No significant differences exist according to one-way ANOVA tests, $F(2, 72) = 0.514, p = .6$. Regarding the background of the users, in general, no significant differences exist according to one-way ANOVA tests, $F(1, 72) = 2.662, p = .108$. Nevertheless, if we consider cases when the adjacent symbol was chosen, significant differences might exist

TABLE 13
Results of the Analysis of Symbol Selection Errors in 2-key

	% of Users With Errors	Errors per Sentence
Globally	93.75	$M = 0.46, SD = 0.300$
Results per age group		
Young	91.17	$M = 0.468, SD = 0.343$
Adult	95.45	$M = 0.474, SD = 0.265$
Older	100.00	$M = 0.387, SD = 0.181$
Results per background		
IT	94.44	$M = 0.513, SD = 0.313$
Non-IT	92.86	$M = 0.392, SD = 0.270$
Adjacent symbol	92.18	$M = 0.350, SD = 0.260$
Results per age group		
Young	88.23	$M = 0.347, SD = 0.281$
Adult	95.45	$M = 0.383, SD = 0.248$
Older	100.00	$M = 0.275, SD = 0.183$
Results per background		
IT	94.44	$M = 0.407, SD = 0.278$
Non-IT	89.28	$M = 0.277, SD = 0.213$

Note. IT = information technology.

between background groups according to one-way ANOVA tests, $F(1, 72) = 4.194, p = .045$. In fact, a 95% confidence level T test shows that users with an IT background make more mistakes of this type than non-IT users ($p = .022$). We have not found any statistically significant difference between male and female users or between left-handed and right-handed users.

5. DISCUSSION

As can be seen, mobile-like methods produce in general a higher number of errors than virtual keyboards. Iatrino and Modeo (2007) stated that if we compare TVs with mobile devices or tactile surfaces, one of the main problems of TV is the fact that the keys and the screen are in different devices, making it impossible for users to look at the interaction device and the screen at the same time. We have been able to prove that this still holds true. The remarkable number of late identification errors of Multitap ($M = 0.728, SD = 0.415$) is clearly a consequence of the user looking at the remote control while typing, instead of looking at the television. When an error is made, they are aware of it too late, so they have to delete many correct symbols in order to return to the place where the error was made. This does not happen with the rest of the methods, because users have to look at the screen of the television. Even with 2-key, users have to look at the screen to make a selection. A possible improvement could be to incorporate some kind of input feedback to let the user know what he or she is writing without looking at the television (Roe, Muto, & Blake, 1984). For instance, an option could be the remote or the television informing the user of the symbol written every time he or she

writes something. Another less disruptive technique may be a display in the remote control. In general, this is applicable for methods in which the available symbols are not on the screen (e.g., when controlling the television with a smartphone).

Both Multitap and 2-key also suffer a remarkable number of proximity errors, that is, users clicking on a key adjacent to the correct one. Proximity errors reach an average of 0.725 errors per sentence when selecting a group in 2-key. This problem is mainly produced because of the size of the keys. Current remote controls, because of their size, do not meet the ideal effective key size of about 18 mm square (Lewis et al., 1997). The aforementioned problem of having the keys and the screen in different devices means that if the user is looking at the television instead of looking at the remote, it is very easy to press a wrong key because they are very close to each other. This problem can be minimized by increasing the distance between the buttons of the remote control, as stated by Siek et al. (2005).

Multitap also registers a relatively high number of wrong finger bounces. Taking into account that we have only considered errors when an incorrect symbol is written, users choose the correct key but perform more or fewer keystrokes than necessary. This is also a consequence of where the focus of the users is. Previous recommendations are valid to improve this particular type of error. Nevertheless, if we consider that blank spaces had their own key and that 37.04% of bounce errors with Multitap were made when entering the blank space, everything indicates that keystroke delays need further improvement. A forced delay between successive keystrokes would also help to reduce the number of these errors.

2-key also has some particular problems. Symbol selection errors reach a remarkable average of 0.46 errors per sentence. This can be produced due to the fact that when a user chooses a group of symbols, he or she must necessarily write a character to continue writing. Thus, if a user chooses a wrong group, he or she will also perform a symbol selection error. In fact, 38.66% of symbol selection errors are produced after the selection of a wrong group. Thus, it would be interesting to allow users to cancel a symbol once an incorrect group has been chosen. If an additional key is used to roll back the selection of a group, we can reduce the number of errors. This key could be a button on the remote, an on-screen symbol, or both.

In the case of virtual keyboards, the number of late identification errors is much lower. This is a consequence of users looking at the screen all the time. Users do not need to look at the remote, because the reduced set of keys and their position makes these methods very intuitive. If they perform a mistake, they are aware of it immediately. Nevertheless, the most frequent errors in virtual keyboards are proximity errors, that is, users choosing a symbol adjacent to the correct one. Because no forced time lags between keystrokes were introduced, it is very easy for a user who wants to write fast to go beyond the desired key. The situation may be improved by using keystroke delays.

There are other errors that can be reduced by applying spell-checking techniques, as suggested by Chen et al. (2010). This can be the case of missed symbols, bounce errors, or transposition errors. Nevertheless, in this context a new approach must be followed, as the interaction device is a remote control. A good approach would be to use the fastext color keys of the remote to choose between four possible words. These words can be shown in the bottom part of the screen with a colored symbol representing the button to press in the remote control.

In general, users with an IT background are more proficient than users without. In the majority of the cases, they make fewer mistakes than non-IT users. Nevertheless, there is an exception with symbol selection errors with 2-key. In this case IT users make more mistakes, but the reason for this is not clear enough to reach a valid conclusion.

Regarding the age of the users, older subjects had higher average values than the rest of the users in almost all the types of errors. Nevertheless, these differences were statistically significant only in the case of diagonal proximity errors. Thus, although it is known that age is an important factor in finger precision, we cannot say that this is problematic when we study specific errors in detail.

Finally, it is also noteworthy that we have not found any statistically significant difference between male and female users or between left-handed and right-handed users.

6. CONCLUSIONS AND FUTURE WORK

In this article we have analyzed a set of 7,395 sentences written by 82 users to identify, quantify, and propose improvements for several input methods compatible with the IDTV context. Apart from the considerable number of users, it is also important to consider that by and large they present heterogeneous characteristics: age, gender, level of knowledge, and technology background.

Our results show that, in general, the number of errors per sentence is much lower in virtual keyboards than in mobile-like methods. There is a great difference between having 2.459 errors per sentence in a method such as 2-key and having 0.516 errors per sentence with an Alphabetic virtual keyboard. A great number of errors per sentence can be extremely frustrating for users, so we think that this is something that must be taken into account very seriously. Also, we have seen that methods such as 2-key and Multitap usually concentrate errors of different types in the same sentence. This is why their general averages of errors per sentence are so high.

Moreover, we have seen that depending on the input method, users tend to make certain mistakes more frequently. For instance, Multitap registers a remarkable number of late identification errors, with an average of 0.728 errors per sentence. This is clearly a consequence of the user looking at the remote control while typing instead of looking at the television. Furthermore, virtual keyboards suffer a high rate of cross-proximity errors, with an average of 0.158 errors per

sentence. We have also seen that age and background influence to some extent the type and number of errors made. For instance, diagonal-proximity errors in virtual keyboards are especially significant for older users, and this also happens with cross-proximity errors and users without an IT background. We provide a set of recommendations in order to improve these results and, thus, reduce the number of errors committed by the users. It is important to take into account that one of our goals was to investigate text entry methods that may be used with commonly available technology that most people will probably already have in their homes. Thus, these recommendations have been provided with this constraint in mind. For example, we propose to allow users to cancel a symbol once an incorrect group has been chosen with 2-key, or to include a display in the remote or some sort of feedback to reduce late identification errors. This would reduce the number of wrong symbols written considerably.

Future work will be based mainly on experiments with devices different to conventional remote controls such as touchpads, mini keyboards, or wiimote-like remotes. Although publications such as Orbist et al. (2008) recommend not to rely on these devices, and the results of some experiments such as those in MacKenzie, Lopez, and Castelluci (2009) may not be as was expected, it is increasingly evident that the popularity of devices to interact with televisions different to the conventional remote controls will increase in the near future (Anido, Valladares, Fernandez-Iglesias, Rivas, & Gomez, 2013). Thus, experiments may be undertaken to detect the most frequent errors made with these alternative methods. We would also like to introduce environmental issues during the experiments such as different levels of illumination or noise. The inclusion of these issues will allow us to check how the performance of the users is affected and, thus, draw even more interesting results. We would also like to perform new experiments to check whether the improvements suggested in this article are effective. This includes a study of the parameters controlling the behavior of input methods, similarly to MacKenzie and Felzer (2010), to check whether these parameters influence the number and type of errors. For instance, how does the Multitap threshold affect the type and number of errors? Or, does a time lag between keystrokes improve proximity errors in virtual keyboards? The final goal would be to design an adaptive method in order to adjust these values to the characteristics of each user. For example, different thresholds for Multitap can be used depending on the age of the user.

REFERENCES

- Anido, L. E., Valladares, S. M., Fernandez-Iglesias, M. J., Rivas, C., & Gomez, M. (2013). Adapted interfaces and interactive electronic devices for the smart home. *Proceedings of 8th IEEE International Conference on Computer Science & Education*, 472–477.
- Barrero, A., Melendi, D., Pañeda, X.G., García, R., & Cabrero, S. (2014). An empirical investigation into text input methods for interactive digital television applications. *International Journal of Human-Computer Interaction*, 30, 321–341.

- Bellman, T., & MacKenzie, I. S. (1998). A probabilistic character layout strategy for mobile text entry. *Proceedings of Graphics Interface '98*, 168–176.
- Brandtzæg, P. B., Heim, J., & Karahasanovic, A. (2011). Understanding the new digital divide—A typology of Internet users in Europe. *International Journal of Human-Computer Studies*, 69, 123–138.
- Brewbaker, C. R. (2008). *Optimizing stylus keyboard layouts with a genetic algorithm: Customization and internationalization*. Technical report. Iowa, USA: Department of Computer Science, Iowa State University.
- Brown, M., Coughlan, T., Lawson, G., Goulden, M., Houghton, R. J., & Mortier, R. (2013). Exploring interpretations of data from the Internet of things in the home. *Interacting with Computers*, 25, 204–217.
- Chen, T., Yesilada, Y., & Harper, S. (2010). What input errors do you experience? Typing and pointing errors of mobile Web users. *International Journal of Human-Computer Studies*, 68, 138–157.
- Choi, S., Han, J., Lee, G., Lee, N., & Lee, W. (2011). RemoteTouch: Touch-screen-like interaction in the TV viewing environment. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 393–402.
- Crawley, M. J. (2007). *The R book*. Chichester, England: Wiley & Sons.
- European Telecommunications Standards Institute ES 202 130 V2.1.2, 2007-09, Human Factors (HF); User Interface; Character repertoires, ordering rules and assignments to the 12-key telephone keypad.
- Geleijnse, G., Aliakseyeu, D., & Sarroukh E. (2009). Comparing text entry methods for interactive television applications. *Proceedings of the 7th European Conference on Interactive TV and Video*, 145–148.
- Henze, N., Rukzio, E., & Boll, S. (2012). Observational and experimental investigation of typing behaviour using virtual keyboards on mobile devices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2659–2668.
- Iatrino, A., & Modeo, S. (2007). Text editing in digital terrestrial television: A comparison of three interfaces. In G. Lekakos, K. Chorianopoulos, & G. Doukidis (Eds.), *Interactive digital television: Technologies and applications* (pp. 224–241). Hershey, PA: Idea Group.
- Ingmarsson, M., Dinka, D., & Zhai, S. (2004). TNT—A numeric keypad based text input method. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 639–646.
- ITU-T Rec. BT.500-12. 2009, Methodology for the subjective assessment of the quality of television pictures.
- Költringer, T., Ngo Van, M., & Grechenig, T. (2007). Game controller text entry with alphabetic and multi-tap selection keyboards. *Extended Abstracts Proceedings of the 2007 Conference on Human Factors in Computing Systems*, 2513–2518.
- Lewis, J. R., Commarford, P. M., Kennedy, P. J., & Sadowski, W. J. (2008). Handheld electronic devices. In C. Melody Carswell (Ed.), *Reviews of Human factors and ergonomics* (Vol. 4, pp. 105–148). Santa Monica, CA: Human Factors and Ergonomics Society.
- Lewis, J. R., LaLomia, M. J., & Kennedy, P. J. (1999). Evaluation of typing key layouts for stylus input. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*, 420–424.
- Lewis, J. R., Potosnak, K. M., & Magyar, R. L. (1997). Keys and keyboards. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction* (pp. 1285–1315). Amsterdam, the Netherlands: Elsevier.
- MacKenzie, I. S. (2002). Mobile text entry using three keys. *Proceedings of NordiCHI 2002*, 27–34.
- MacKenzie, I. S., & Felzer, T. (2010). SAK: Scanning ambiguous keyboard for efficient one-key text entry. *ACM Transactions on Computer-Human Interaction*, 17, 1–39.
- MacKenzie, I. S., Lopez, M. H., & Castelluci, S. (2009). Text entry with the Apple iPhone and the NintendoWii. Retrieved from www.malchevic.com/papers/MobileHCI2009.pdf (Accessed January 7, 2015)
- MacKenzie, I. S., & Soukoreff, R. W. (2002). Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction*, 17, 147–198.
- MacKenzie, I. S., & Soukoreff, R. W. (2003). Phrase sets for evaluating text entry techniques. *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*, 754–755.
- Millet, B., Asfour, S., & Lewis, J. R. (2008). Designing a hybrid layout for a five-key text entry technique. *Proceedings of the 20th Annual International Occupational Ergonomics and Safety Conference*, 156–162.
- Millet, B., Asfour, S., & Lewis, J. R. (2009). Selection-based virtual keyboard prototypes and data collection application. *Behavior Research Methods*, 41, 951–956.
- Norman, D. A., & Fisher, D. (1982). Why alphabetic keyboards are not easy to use: Keyboard layout doesn't much matter. *Human Factors*, 24, 509–519.
- Orbist, M., Bernhaupt, R., & Tscheligi, M. (2008). Interactive TV for the home: An ethnographic study on users' requirements and experiences. *International Journal of Human-Computer Interaction*, 24, 174–196.
- Perrinet, J., Pañeda, X. G., Cabrero, S., Melendi, D., García, R., & García, V. (2011). Evaluation of virtual keyboards for interactive digital television applications. *International Journal of Human-Computer Interaction*, 27, 703–728.
- RAE – Royal Spanish Academy. (2001). *Spanish language dictionary* (22nd ed.). Available from <http://www.rae.es/rae.html>.
- Roe, C. J., Muto, W. H., & Blake, T. (1984). Feedback and key discrimination on membrane keypads. *Proceedings of the Human Factors Society 28th Annual Meeting*, 277–281.
- Siek, K. A., Rogers, Y., & Connelly, K. H. (2005). Fat finger worries: How older and younger users physically interact with PDAs. *Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction*, 267–280.
- Silfverberg, M., MacKenzie, I. S., & Korhonen, P. (2000). Predicting text entry speed on mobile phones. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 9–16.
- Spira, J. B. (2011). Internet TV: Almost ready for prime time [Tools & Toys]. *IEEE Spectrum*, 48, 24–26.
- Sporka, A. J., Polacek, O., & Slavik, P. (2012). Comparison of two text entry methods on interactive TV. *Proceedings of the 10th European conference on Interactive TV and Video*, 49–52.
- Taveira, A. D., & Choi, S. D. (2009). Review study of computer input devices and older users. *International Journal of Human-Computer Interaction*, 25, 455–474.
- Trewin, S., & Pain, H. (1999). Keyboard and mouse errors due to motor disabilities. *International Journal of Human Computer Studies*, 50, 109–144.
- Vega-Oliveros, D. A., Pedrosa, D. C., Pimentel, M. G. C., & De Mattos Fortes, R. P. (2010). An approach based on multiple text input modes for interactive digital TV applications. *Proceedings of the 28th ACM International Conference on Design of Communication*, 191–198.
- Zhai, S., Hunter, M., & Smith, B. A. (2000). The Metropolis keyboard—An exploration of quantitative techniques for virtual keyboard design. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 119–128.

ABOUT THE AUTHORS

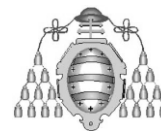
Aurora Barrero is a Computer Science Engineer from the University of Oviedo and a Ph.D. student, with an interest in the area of multimedia systems and services, content distribution networks, and interactive digital TV services. She is a Research Engineer in the Department of Computer Science of the University of Oviedo.

David Melendi is a Computer Science Engineer with a Ph.D. from the University of Oviedo, with an interest in multimedia systems and services, content distribution networks, idTV services, and mobile ad hoc networks. He is an Associate Professor at the University of Oviedo and member of the W3C.

Xabiel G. Pañeda is a Computer Science Engineer with a Ph.D. from the University of Oviedo, with an interest in multimedia systems and services, content distribution networks, idTV services, and mobile ad hoc networks. He is an Associate Professor at the University of Oviedo and member of the W3C.

Roberto García is a Telecommunications Engineer from The Technical University of Madrid with a Ph.D. from the University of Oviedo, with an interest in telecommunication networks and services, applied to performance analysis, modelling, and simulation of systems and services. He is an Associate Professor at the University of Oviedo.

Sergio Cabrero is a Telecommunications Engineer from the University of Oviedo and a Ph.D. student, with an interest in the area of telecommunication networks, interactive digital TV services, multimedia services, and mobile ad hoc networks. He is a Teaching Assistant in the Department of Computer Science of the University of Oviedo.



6.3 A research on typing methods for interactive Digital Television Applications

A. Barrero, D. Melendi, X. G. Paneda, R. Garcia, L. Pozueco, y J. L. Arciniegas, «A research on typing methods for interactive Digital Television Applications», *IEEE Latin America Transactions*, vol. 13, n.º 11, pp. 3612–3620, 2015.

A research on typing methods for interactive Digital Television Applications

A. Barrero, D. Melendi, X. G. Pañeda, R. García, L. Pozueco, J. L. Arciniegas

Abstract— It has been several years now since we started to experience a change in traditional television services. Recently, the digitalization of these services has improved the usage of the available electromagnetic spectrum, increased the quality of the service and increased the number of channels. The next step in this evolution is the transformation of traditional television into a fully interactive service. This change allows users to run applications on the television, similar to those of mobile phones or computers. Nevertheless, the main interaction device delivered with televisions and set-top-boxes is still the traditional remote control. This device is suitable to perform certain actions such as move a pointer or select something, but it has not been designed to carry out more complex actions such as writing text. In this paper we present the main outcomes of a project aimed at researching typing methods for interactive digital television services. During the project we have carried out several experiments with real users, who had to use different typing methods and devices that may be suitable for interactive digital television applications.

Keywords— Interactive Digital Television, Typing, Text, Interaction, HCI.

I. INTRODUCCIÓN

Las ventajas que presentan los servicios de televisión digital sobre los servicios convencionales de televisión son de sobra conocidas. Se realiza un mejor aprovechamiento del medio de transmisión correspondiente para ofrecer a los usuarios más y mejores contenidos audiovisuales. Al mismo tiempo, se da a los proveedores de servicios de televisión la posibilidad de ofrecer servicios alternativos y/o complementarios a la mera difusión de audio/vídeo. Esto ya era posible hace tiempo en el caso de las compañías de cable o satélite, pero el impulso definitivo para el lanzamiento generalizado de servicios de televisión digital interactiva se produjo hace unos pocos años. Entre otros factores, este impulso se debe a la evolución en las tecnologías de acceso a Internet, la proliferación de conexiones en los hogares de los usuarios, y la aparición de nuevos dispositivos set-top-box y televisiones con capacidad de cómputo. Podemos afirmar que la Televisión Digital Interactiva o idTV, se ha convertido en un servicio real de Internet de las cosas en el hogar [1].

En los servicios de idTV un usuario puede interactuar con facilidad utilizando un mando a distancia convencional. Principalmente, lo que hace es desplazarse por menús y realizar la selección de determinados elementos. No obstante, algunos servicios pueden llegar a requerir la escritura de algún

tipo de texto. Este puede ser el caso de algunos juegos interactivos, servicios de banca electrónica, navegación por Internet o servicios de t-learning. Por ejemplo, un usuario puede tener que escribir texto en forma de nombres de usuario, contraseñas, códigos, instrucciones, mensajes, etc. El problema de la escritura de texto es que el mando a distancia convencional no está diseñado para ello. Además, la actual falta de estandarización en este campo tampoco facilita esta tarea [2]. Es cierto que hay otros tipos de dispositivos como los teclados externos o algunos elementos apuntadores, pero ni son muy populares ni su diseño los hace confortables para su utilización desde el sofá.

Dado que la escritura de texto en aplicaciones de idTV supone un problema real para los usuarios, hemos llevado a cabo un proyecto para evaluar distintos métodos de escritura en este entorno. Durante el proyecto se han evaluado distintos métodos compatibles con un mando a distancia, así como otros dispositivos que podríamos denominar “competidores”. Igualmente, se ha diseñado un dispositivo que, según los resultados obtenidos, podría mejorar la experiencia de los usuarios de aplicaciones de idTV. Los resultados más destacables de este proyecto se presentan en este trabajo.

El resto del artículo se ha estructurado como sigue. En la sección 2 se presentan algunos trabajos previos que podrían relacionados con el objetivo del proyecto. En la sección 3 se muestra el plan de trabajo seguido durante el desarrollo del proyecto. En la sección 4 se describe el entorno de experimentación utilizado. En la sección 5 se describen algunos de los resultados más relevantes, para concluir con la sección 6 en la que se presentan las conclusiones del proyecto.

II. TRABAJOS PREVIOS

El estudio de diferentes métodos escritura es un campo que ha generado muchos resultados de investigación. No obstante, la mayoría de los últimos avances corresponde al mundo de los dispositivos móviles. Aunque es posible reutilizar algunos de estos diseños en entornos de idTV, hay dos aspectos clave que deben tenerse en cuenta [3]. En primer lugar, el dispositivo más extendido es el mando a distancia, que tiene una serie de características inherentes a su uso tradicional distintas a las de un teclado, lápiz o pantalla táctil. En segundo lugar, las teclas y la pantalla están en dispositivos separados, de modo que los usuarios no pueden mirar al dispositivo de interacción y a la pantalla al mismo tiempo. Pese a ello, se comentan algunas referencias que han sido interesantes desde el punto de vista del proyecto.

A. Barrero, D. Melendi, X.G. Pañeda, R. García, L. Pozueco, Universidad de Oviedo, {barreroaurora, melendi, xabiel, garciaroberto, pozuecolaura}@uniovi.es

J. L. Arciniegas, Departamento de Telemática, Universidad del Cauca, jlarci@unicauca.edu.co

A. Innovaciones en la escritura en dispositivos móviles

En el campo de los dispositivos móviles hay varios estudios sobre teclados virtuales con resultados que pueden ser aplicables parcialmente en servicios de idTV. Los teclados virtuales son representaciones en pantalla de un teclado sobre el que el usuario se puede desplazar con un conjunto reducido de teclas, acompañadas de una tecla de confirmación.

Bellman y MacKenzie [4] comparan dos tipos de teclados virtuales. El primero de ellos es el conocido QWERTY, mientras que el segundo es la disposición optimizada FOCL o *Fluctuating Optimal Character Layout*. Este método consiste en cambiar la posición de las letras en el teclado de forma que las letras con mayor probabilidad de ser seleccionadas tras una letra determinada se pintan junto a ésta. Para comparar estos dos métodos se realizó un estudio con once personas seleccionadas de entre los alumnos de los autores. Cada participante tenía que escribir frases con los métodos en diez sesiones de quince minutos. Las frases se seleccionaban de artículos publicados en Internet. Los resultados muestran que no había diferencias significativas entre los métodos, aunque los participantes reconocían que no podían utilizar de forma óptima los teclados FOCL. Por otro lado, Zhai et al [5] presentan un nuevo teclado virtual. Según los autores, las disposiciones de teclas existentes, y entre ellas la QWERTY, están pensadas para ser utilizadas con ambas manos de forma que no son adecuadas para dispositivos como teléfonos móviles o tabletas. Para incrementar la velocidad de escritura en estos dispositivos, los autores crearon la disposición de teclas Metrópolis, con la que llegan a obtener una velocidad de escritura un 40% más rápida que con QWERTY. Brewbaker [6] persigue el mismo objetivo, pero utilizando un algoritmo genético que le permite personalizar las dimensiones y el idioma del teclado. Los resultados obtenidos por Brewbaker son similares a los de Zhai et al [5], pero por desgracia estos teclados no fueron probados con usuarios reales, quedando la duda de si serán de su agrado o no.

Además de los teclados virtuales, existen otros mecanismos de escritura en dispositivos móviles que podrían ser de aplicación en el campo de la idTV. Por ejemplo, Silfverberg et al [7] comparan tres métodos de escritura basados en el teclado de los teléfonos móviles: el típico método de escritura de SMS conocido como multi-press (con tiempo límite y con botón siguiente), otro que llaman Two-Key, que consiste en que el usuario selecciona un grupo de caracteres mediante una pulsación y luego el carácter deseado con una segunda pulsación, y el método predictivo T9. La evaluación de estos métodos se realizó con 12 voluntarios con edades entre los 23 y los 47 años, casi todos usuarios frecuentes de teléfonos móviles. Los resultados muestran que el mejor rendimiento se obtiene con T9. De forma similar, Butts y Cockburn [8] repiten el experimento (sin considerar el T9) en una evaluación con 8 personas graduadas en Informática. En este caso los mejores resultados se obtienen con el multi-press con botón siguiente y los peores con el Two-Key. Por otro lado, MacKenzie et al [9] prueban otra técnica de desambiguación para la escritura de texto en teléfonos móviles basada en probabilidades de prefijos, que denominaron LetterWise. Esta

técnica pretende ayudar a los usuarios proporcionándoles un mecanismo de desambiguación usando una base de datos con probabilidades de prefijos en vez de un mecanismo basado en un diccionario como el T9. Para comparar estos dos métodos, llevaron a cabo un estudio con veinte participantes. A cambio de una compensación económica, estos participantes llevaron a cabo veinte sesiones de entre 25 y 30 minutos en las que tenían que escribir frases de un conjunto preparado por los autores. Las velocidades de escritura fueron similares en la primera sesión, pero la última sesión mostró una diferencia significativa entre los métodos. No obstante, no pudieron establecer una diferencia significativa entre las tasas de error. MacKenzie y Tanaka-Ishii [10] presentan diferentes organizaciones de teclados y distintos métodos de desambiguación. Obtienen las mediciones características de teclados de teléfonos ambiguos y, de forma metodológica, explican cómo evaluar y comparar teclados utilizando el número de pulsaciones por carácter y por palabra (KSPC/KSPW) llegando a desarrollar un modelo.

Además de los teclados virtuales y los métodos basados en la utilización de un conjunto reducido de teclas, existen otras técnicas denominadas concurrentes, que persiguen combinar dos métodos de forma simultánea. Aunque hay varios estudios de esta naturaleza, los resultados de Widgor y Balakrishnan [11] son especialmente interesantes. Llevan a cabo un estudio comparativo sobre las mejoras en la velocidad de escritura mediante este tipo de técnicas. En este caso, añaden tres teclas adicionales a un teclado de teléfono móvil clásico, de forma que los usuarios tienen que combinarlas con una tecla numérica en caso de que deseen escribir un número o una letra. Realizaron un experimento con quince usuarios con experiencia en la utilización de teléfonos móviles y detectaron un incremento del rendimiento significativo. De forma similar, Oniszczak y MacKenzie [12] presentan una técnica nueva llamada RollPad. Es un teclado de 12 teclas en las que las pulsaciones tienen que acompañarse de movimientos. Por ejemplo, para escribir la letra “a” el usuario tiene que presionar la tecla 2 y mover el dedo a la izquierda. Comparan este método con el tradicional método Multitap un experimento con 14 participantes con edades comprendidas entre los 18 y los 35 años. Aunque los resultados no muestran ninguna diferencia significativa en las velocidades y las tasas de error, el número de pulsaciones por carácter disminuye con RollPad, suponiendo un esfuerzo físico menor pero a costa de una mayor carga cognitiva.

B. Innovaciones en la escritura en otros dispositivos

En los últimos años, los dispositivos de juego también han sido estudiados para comprobar si las características de sus periféricos podían influir sobre las velocidades de escritura. MacKenzie et al [13] comparan velocidades de escritura y tasas de error entre un Apple iPhone y una consola Nintendo Wii utilizando un teclado virtual en ambos casos. 16 personas con una media de edad cercana a los 25 años participaron en un experimento en el que el objetivo era introducir tres frases en varios bloques utilizando ambas técnicas. Cada sesión rondaba los 45 minutos de duración. Los resultados muestran

una diferencia significativa entre las velocidades de escritura obteniendo mejores resultados con el iPhone (con 18,5 palabras por minuto frente a las 9,2 de la Wii). Los autores atribuyen esta diferencia a la gran concentración necesaria para escribir texto con la Wii en comparación con el iPhone. También encuentran una diferencia significativa entre las tasas de error, obteniendo más errores con el iPhone que con la Wii (7,7% frente a 2,8%). Esto se explica por el “problema del dedo gordo” que consiste en la dificultad de seleccionar la tecla adecuada cuando su tamaño o el espacio entre teclas es muy pequeño [14]. Esta línea fue seguida igualmente por otros autores como Castellucci y MacKenzie [15] o Jones et al. [16].

Por otro lado, Költringer et al [17] hicieron pruebas con la Microsoft Xbox 360. Compararon una nueva técnica llamada TwoStick con un teclado virtual QWERTY. Esta técnica se basa en una rejilla de 3x3 celdas en la que cada celda se subdivide en otras 9 celdas más pequeñas. El teclado final permite escribir hasta 56 caracteres diferentes incluyendo letras, números y caracteres especiales. Las teclas de espacio en blanco y borrado se han asignado en cada zona a las teclas centrales (izquierda y derecha). La tecla central de cada zona no se utiliza. Para moverse de una zona a otra los usuarios deben utilizar el joystick izquierdo y para moverse dentro de una zona deben utilizar el joystick derecho. En el experimento, 8 usuarios de edades entre los 22 y los 29 años tenían que completar 20 sesiones en las que escribían frases con cada método. En la sesión final el método TwoStick consigue ser significativamente más rápido que el QWERTY, con 14,87 palabras por minuto frente a 12,9. Incluso los resultados muestran diferencias significativas entre las tasas de error en la primera sesión pero no en la última. En el caso del método TwoStick la causa más frecuente de error fue una temporización incorrecta de los movimientos de los joysticks. Esto ocurría frecuentemente cuando los usuarios empezaban a mover el joystick de la subzona (izquierdo) antes de situarse en la zona correcta (joystick derecho)

De forma similar, Wobbrock et al [18] diseñaron un método de escritura basado en un joystick llamado EdgeWrite. Este método utiliza gestos para seleccionar símbolos en un alfabeto. Los autores compararon este método con un teclado virtual con disposición Alfabética y con el DateStamp, ampliamente utilizado para escribir las iniciales en los juegos de arcade tradicionales. 18 usuarios de 21,5 años de edad participaron en los experimentos. Los resultados muestran que EdgeWrite fue al menos 1,5 veces más rápido que el teclado virtual y 2,4 veces más rápido que DateStamp.

C. Innovaciones específicas en servicios de idTV

En el campo específico de los servicios de idTV no hay muchos trabajos publicados relacionados con la escritura de texto. La primera referencia puede ser el trabajo de Iatrino y Modeo [3]. Estos autores identifican dos formas de usar un mando a distancia para escribir: utilizar los cursores y la tecla OK o usar un conjunto reducido de teclas como en los teléfonos móviles. Evalúan ambas posibilidades en un experimento grupal. Se hacen pruebas con tres métodos: Multipress (el método SMS), Multipress con una

retroalimentación visual y un teclado virtual QWERTY. 36 personas participaron en el experimento teniendo que llevar a cabo dos tareas con cada método. La primera tarea consistía en escribir una dirección de correo electrónico, mientras que en la segunda había que escribir una frase corta en italiano. Los autores concluyen que el mejor método es el Multipress, destacando múltiples problemas relacionados con la internacionalización.

Ingmarsson et al [19] presentan una nueva técnica llamada TNT similar al ya comentado TwoStick [17]. El sistema se basa en una rejilla de 3x3 en la que cada celda se subdivide en otras 9 celdas menores. Cada una de las celdas menores tiene un carácter, alcanzando un total de 81 posibilidades. Para escribir un carácter el usuario debe seleccionar una de las celdas principales utilizando el teclado numérico del mando y luego seleccionar el carácter que desea escribir con una nueva pulsación. Como se puede observar, cada carácter es accesible con sólo dos pulsaciones. Cinco personas pagadas con edades entre los 27 y los 32 años probaron el sistema durante 10 sesiones de 45 minutos para escribir una novela corta en Sueco. Los resultados muestran velocidades comparables o superiores a la escritura manual en una PDA o un método Multipress. Los usuarios destacaban su gran sencillez.

Por otra parte, Geleijnse et al [20] compararon tres técnicas para mandos a distancia (Multitap, T9 y un teclado virtual) con un teclado convencional QWERTY (con y sin autocompletado). En el experimento participaron 22 personas con edades entre los 21 y los 32 años. Tenían que buscar vídeos en Youtube escribiendo pares “artista-pista” con cada uno de los métodos. A pesar de que los autores no especifican la experiencia de los usuarios con las nuevas tecnologías, no es de sorprender que los resultados muestren una diferencia significativa entre las técnicas del mando a distancia y el teclado convencional. Lo que es sorprendente, es que los autores no encontraron ninguna diferencia entre los métodos utilizados con el mando a distancia. Los autores concluyeron que, bajo las condiciones del experimento, “las encuestas no han demostrado ningún indicio de que los usuarios no acepten un teclado en un salón”. Esta conclusión contrasta con la de Orbist et al [21] que abogan por sistemas basados en voz. Los resultados de su estudio etnográfico indican que es poco realista centrarse en estudios dependientes de la popularidad de periféricos para televisión distintos de los convencionales.

Gargi y Gossweiler [22] presentan un nuevo sistema predictivo diseñado para mejorar la velocidad de escritura en teclados virtuales: QuickSuggest. Esta técnica muestra un anillo circundante al carácter seleccionado, mostrando los cuatro caracteres que le siguen con más frecuencia. Cuando se escribe un carácter el anillo aparece y si el usuario iba a escribir uno de los caracteres que se sugieren solamente tiene que seleccionarlo y presionar OK. El cursor se mueve a la tecla correspondiente. Si el carácter deseado no es ninguno de los sugeridos, el usuario debe moverse a la posición del teclado en el que se encuentra el carácter deseado. Este método requiere un mínimo de dos pulsaciones por carácter. En cuanto a la evaluación, los autores la realizaron en dos pasos. En primer lugar llevaron a cabo un estudio teórico del

método, mientras que en el segundo efectuaron un estudio con 10 participantes. Los autores emplean la métrica de pulsaciones por carácter (KSPC) y calculan el tiempo requerido por los usuarios en la segunda fase del estudio.

Sporka et al [23] comparan el ya comentado TNT [19] con un nuevo método llamado TwiceTap. Este método sigue la misma filosofía que TNT pero, además de permitir a los usuarios escribir caracteres sueltos, les permite escribir bloques de caracteres frecuentemente utilizados (n-gramas). 18 usuarios remunerados con una media de 22,7 años de edad participaron en los experimentos. Aunque no hubo diferencias significativas entre los métodos, los usuarios se decantaron por TwiceTap principalmente por su similitud con Multitap y por la posibilidad de seleccionar los n-gramas.

D. Limitaciones en el estado del arte

Como se puede apreciar, no hay muchos trabajos en los que se trate el tema de la escritura en aplicaciones de idTV. Muchos trabajos se basan en los modelos de predicción de Fitt [24]. Otros confían en técnicas no muy extendidas o de difícil implantación. En otros trabajos se ejecutan evaluaciones con usuarios reales pero con un grupo muy reducido de personas o con características muy similares: comúnmente jóvenes y con experiencia en nuevas tecnologías. Por ello, el proyecto que hemos llevado a cabo realiza avances en el terreno de la usabilidad en servicios de idTV en el campo de la escritura de texto. Para ello, se ha trabajado en dos líneas:

- La identificación de métodos susceptibles de ser utilizados en servicios de idTV.
- La experimentación con un elevado número de usuarios reales con características heterogéneas.

III PLAN DE TRABAJO

Para alcanzar los objetivos planteados en el proyecto, se ha seguido un plan de trabajo estructurado en seis fases:

- Fase #1: Experimentación con métodos populares y compatibles con mandos a distancia convencionales. En concreto, se utilizaron teclados virtuales con distintas disposiciones de teclas y el método Multitap. Las disposiciones para los teclados virtuales fueron QWERTY, Alfabética y un teclado Genético generado a partir del trabajo de Brewbaker [6].
- Fase #2: Identificar y experimentar con optimizaciones de los métodos utilizados en la Fase #1. En base a los resultados y la retroalimentación de los usuarios de la Fase #1, se plantearon una serie de optimizaciones para los métodos empleados. Entre estas optimizaciones, está la utilización de algunos botones del mando a distancia para el borrado de caracteres y la escritura de espacios en blanco para los teclados virtuales. Igualmente, se eligieron T9 y 2-Key como optimizaciones seleccionadas de trabajos anteriores para Multitap.
- Fase #3: Experimentación en entornos específicos. En tareas anteriores se utilizaban textos sencillos, pero ahora se pretenden utilizar textos más complejos, en los que aparecen vocales con tildes y caracteres especiales. Para ello, es necesario incorporar estos símbolos en los métodos, tal y como se muestra en la Fig. 1.

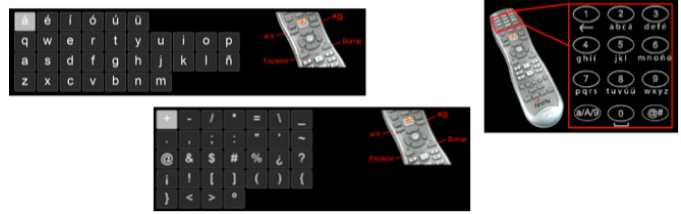


Figura 1. Disposiciones de caracteres especiales. En la parte superior izquierda teclado virtual QWERTY. En la parte inferior central, juego complementario para teclados virtuales. En la parte derecha, juego de símbolos para Multitap.

- Fase #4: Experimentación con mandos heterogéneos. Dado que no hay un modelo de mando a distancia estándar [2], es necesario evaluar el impacto que tiene la forma o la disposición de teclas que utiliza cada mando. Para comparar distintos modelos repetimos las pruebas de la Fase #3 con dos nuevos modelos de mando, para comparar los resultados de esta fase con los de la fase anterior. Los mandos utilizados se muestran en la Fig. 2.
- Fase #5: Experimentación con dispositivos competidores. A pesar de las afirmaciones de algunos autores [21], el mando a distancia sufrirá necesariamente algunas modificaciones para acomodarlo al entorno de idTV o paulatinamente se empezarán a incorporar dispositivos alternativos. Por ello, se pretende realizar una comparación de periféricos distintos del mando a distancia convencional.
- Fase #6: Diseño de nuevos mecanismos de interacción. Con los resultados de las fases anteriores se pretende realizar un diseño de un método y/o dispositivo que conjugue la eficiencia en la escritura con su uso confortable por parte de un televidente.

Tal y como se observa en el plan de trabajo, las cinco primeras fases del proyecto fueron de experimentación propiamente dicha. En ellas se convocaba a personas que tenían que hacer pruebas de escritura de texto en una aplicación desarrollada al efecto. Un resumen de los datos de las fases de experimentación se puede observar en la tabla 1.

IV ENTORNO DE EXPERIMENTACIÓN

Los experimentos se realizaron en una sala intentando que la experiencia del usuario fuera lo más parecida posible a la del salón de su casa. En la sala se ubicó un sofá a una distancia de 2 metros de una televisión de 32 pulgadas. La televisión estaba conectada a un PC con Windows XP. Al PC se conectaban los dispositivos necesarios en cada prueba.



Figura 2. Mandos utilizados en las primeras fases del proyecto.

TABLA I. PLAN DE EXPERIMENTACIÓN

FASE	USUARIOS	MÉTODOS	DISPOSITIVO	SESIONES	JUEGO DE CARACTERES	TEXTO
1	57	QWERTY, ALFABÉTICO, GENÉTICO, MULTITAP	SNAPSTREAM FIREFLY	5	BÁSICO	5 FRASES DE CORPUS
2	48	QWERTY, GENÉTICO, T9, 2-KEY	SNAPSTREAM FIREFLY	2	BÁSICO	5 FRASES DE CORPUS
3	42	QWERTY, GENÉTICO Y MULTITAP MODIFICADOS	SNAPSTREAM FIREFLY	1	COMPLEJO	5 CAMPOS CON DATOS PERSONALES + TEXTO COMÚN
4	42	QWERTY, GENÉTICO Y MULTITAP MODIFICADOS	GOLDEN INTERSTAR, AVERMEDIA RMKS	1	COMPLEJO	5 CAMPOS CON DATOS PERSONALES + TEXTO COMÚN
5	52	QWERTY, GENÉTICO	TECLADOS LOGITECH K400 Y DIÑOVO MINI, APPLE MAGIC TRACKPAD Y FUJITSU AIR COMMAND PLUS	3	COMPLEJO	2 FRASES DE CORPUS Y 3 CAMPOS CON DATOS PROPUESTOS

En el PC se ejecutaba una aplicación Adobe® AIR® diseñada para los experimentos. Tal y como se muestra en la Fig. 3, la aplicación se ejecuta a pantalla completa con fondo negro. En función de la prueba, el texto a escribir se pinta en la parte superior. En la zona central un campo de texto muestra lo que escribe el usuario. La zona inferior se utiliza para pintar información de apoyo a la escritura.

La aplicación ejecutaba un ciclo de prueba con cada método de escritura. El orden de los métodos se alteraba para cada usuario para prevenir una posible influencia en los resultados. Dentro de cada ciclo el participante escribía una serie de cadenas que se le proponían o que correspondían a datos que él o ella conocía. Cuando se proponía una cadena se dejaban unos segundos para que el usuario la leyese. Esa iteración finalizaba cuando la cadena escrita por el usuario coincidía con la propuesta. Cuando no se proponía una cadena, la iteración finalizaba cuando el usuario indicaba que había terminado. No había límite de tiempo para escribir cada cadena. Entre cada iteración y cada ciclo se dejaban unos instantes para que el usuario se adaptase al cambio.

Para proponer los textos, se creó un corpus de frases cortas extraídas de medios de comunicación populares en España, de forma similar a Bellman y MacKenzie [4].

La aplicación almacenaba datos en un fichero XML sobre cada uno de los eventos que registraba, incluyendo una marca de tiempo. Quedaba constancia de todas las pulsaciones que realizaban los usuarios, lo que permite calcular, entre otros:

- La velocidad de escritura obtenida con cada método.
- El número de errores producidos y su naturaleza, mediante el análisis de las correcciones realizadas.
- El aprendizaje con la experiencia, vista la evolución de la velocidad y de los errores.

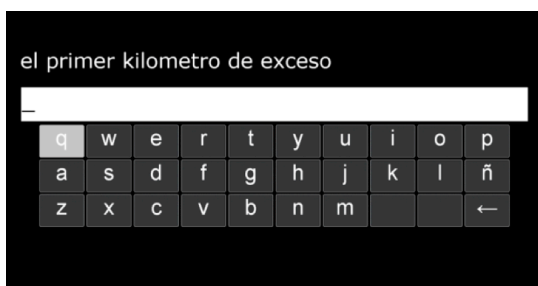


Figura 3. Interface de usuario de la aplicación utilizada en los experimentos.

Al final de la prueba se proporcionaban unos cuestionarios a los usuarios para obtener información subjetiva. Unas escalas Likert (de 0 a 4) permitían al usuario valorar la facilidad de uso, la velocidad de escritura y la satisfacción general con cada uno de los métodos.

V RESULTADOS OBTENIDOS

A. Resultados de las Fases #1 a #4

Tal y como se observa en la tabla 1, todos los experimentos de estas fases se realizaron con los mandos a distancia convencionales que se muestran en la Fig. 2.

En cuanto a la velocidad de escritura, tal y como se observa en la Fig. 4, el método más rápido es Multitap con diferencias estadísticamente significativas con el resto ($p < 0,001$) [25]. De entre los teclados virtuales, el mejor es el Génético, pero sin apreciar diferencias significativas con el resto de métodos. Además se observa que los usuarios mejoran su rendimiento en general con la experiencia.

Al introducir las optimizaciones, se observa que el rendimiento mejora significativamente en QWERTY ($M=38,39$, $SD=6,75$) y T9 ($M=58,59$, $SD=18,81$). La mejora del Génético no es significativa ($M=39,84$, $SD=8,34$). En el caso particular del 2-key ($M=32,34$, $SD=8,39$), lo que se obtiene es una reducción en la velocidad estadísticamente significativa (37,17% peor que con Multitap). Los resultados de T9 y 2-Key siguen las pautas en [7] y [8].

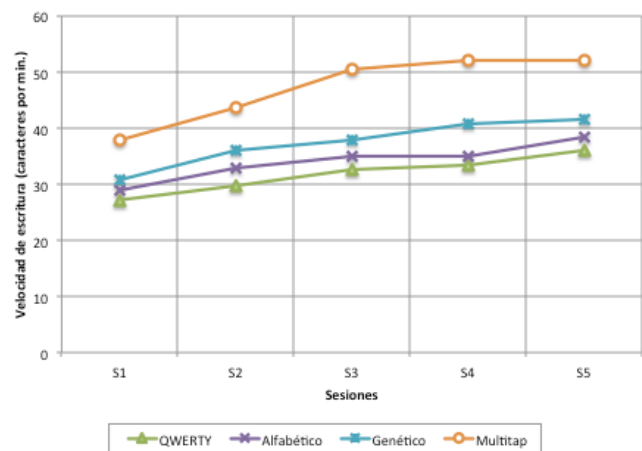


Figura 4. Velocidad de escritura en las cinco primeras sesiones de la Fase #1.

No obstante, cuando no se proponen textos a los usuarios y se introducen símbolos más complejos, los resultados cambian completamente (Fases #3 y #4). En este caso el método más rápido es QWERTY (M=30,66, SD=6,71) seguido de Multitap (M=27,07, SD=6,92) y del Genético (M=26,97, SD=6,49), pero las diferencias entre métodos no son significativas. También se observa que el rendimiento es peor que el obtenido en fases anteriores con una disminución de hasta un 48,09% en el caso de Multitap. Finalmente, en la comparación entre mandos no se obtienen resultados estadísticamente significativos, lo que nos lleva a concluir que la disposición de teclas y la forma del mando no afecta al rendimiento.

En lo que respecta a los errores de escritura, los resultados de la fase #1 muestran que el método con la mayor tasa de errores es Multitap, tal y como se observa en la Fig. 5. Este método se sitúa muy por encima de los teclados virtuales, presentando diferencias estadísticamente significativas ($p < 1e-05$). También se observa un notable descenso en el número de errores con la experiencia de los usuarios.

Con la introducción de las optimizaciones en la fase #2, observamos que el método con más errores es T9 (M=20,92, SD=18,7), con diferencias significativas con el resto ($p < 0,001$). El Genético es el mejor (M=1,63, SD=1,5), sin diferencias significativas con QWERTY (M=3,2, SD=2,31) pero sí con 2-key (M=7,41, SD=7,06) ($p = 0,028$). Comparando estos resultados con los de la fase anterior, vemos algunas diferencias importantes y estadísticamente significativas en el método T9. La tasa de error para los usuarios que participaron en ambas fases se incrementa un 204% en el T9 respecto al Multitap, mientras que si realizamos la comparación con usuarios que no participaron en la fase #1, este empeoramiento es del 124,87%.

La incorporación de símbolos complejos junto a la exigencia de escribir textos no propuestos de la fase #3, incrementa las tasas de error notablemente. La peor tasa se obtiene con Multitap (M=16,86, SD=9,71), que presenta diferencias significativas con el resto de métodos ($p < 0,001$). Le siguen el método Genético (M=5,49, SD=5,65) y QWERTY (M=4,85, SD=3,51). En cuanto a la comparación entre mandos, no se observan diferencias estadísticamente significativas, tal y como sucedía con la velocidad.

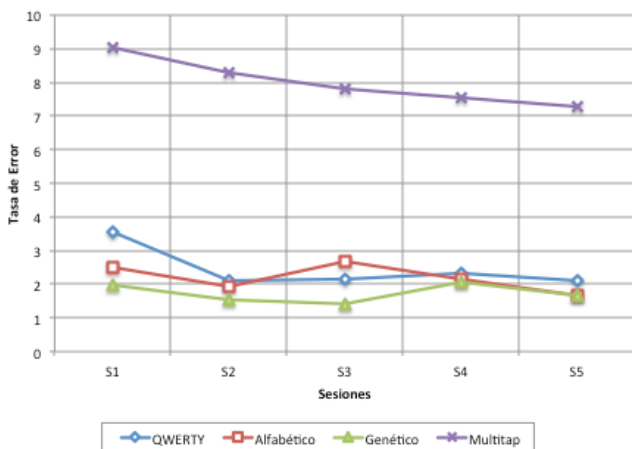


Figura 5. Tasa de error en las cinco primeras sesiones de la Fase #1.

La tipología de los errores es algo que también se ha estudiado en detalle [26]. Además del elevado número de errores, nuestros resultados muestran que en función del tipo de método utilizado hay determinados errores que se cometen más frecuentemente. Por ejemplo, Multitap presenta un número notable de ocasiones en las que el usuario tarda demasiado tiempo en detectar un error (M=0,728, SD=0,415). Esto es una clara consecuencia de un usuario que mira al mando a distancia en vez de a la pantalla. La mayoría de estos errores se producen por un número erróneo de pulsaciones al seleccionar el símbolo que se desea (M=0,307, SD=0,296). Por otro lado, los errores de proximidad son de los más frecuentes (selección de teclas/símbolos adyacentes). Estos problemas van desde una media de 0,725 errores por frase en 2-Key (al seleccionar un grupo de símbolos) hasta 0,158 en el Alfabético (adyacencias horizontales o verticales).

B. Resultados de la Fase #5

En las fases anteriores del proyecto se hizo un estudio pormenorizado de distintos métodos que son susceptibles de ser utilizados en un entorno de idTV, considerando que el mando a distancia es el dispositivo de referencia. No obstante, pese a las afirmaciones de otros autores [21], no es descartable que el mando tradicional sea paulatinamente sustituido por otros tipos de dispositivo. Por ello, en esta nueva fase de la investigación se realizan experimentos utilizando 4 nuevos periféricos, tal y como se muestra en la tabla 1. Se trata de un teclado convencional, un teclado de dimensiones reducidas, un touchpad y un dispositivo apuntador similar al “wiimote”. El touchpad y el dispositivo apuntador, mostrados en la Fig. 6, actúan sobre un teclado virtual con disposiciones QWERTY y Genética, a los que se han añadido símbolos complejos como sucedía en la fase #3 de la investigación.

Si observamos los resultados obtenidos, el método más rápido es el teclado convencional en todas las sesiones de esta fase de la investigación, tal y como se observa en la Fig. 7. La velocidad es muy superior a la del resto de métodos. En el otro extremo se situaría el dispositivo apuntador, que obtiene los peores resultados con la disposición Genética. También se observa una disminución en la velocidad en la tercera sesión, debido a la introducción de símbolos complejos. La caída más pronunciada se produce en los teclados físicos con una disminución promedio del 48,14%. No obstante, se observa que esta caída no es tan pronunciada en el resto de métodos, con una disminución media del 20,6%.



Figura 6. Apple Magic Trackpad y Fujitsu Air Command Plus.

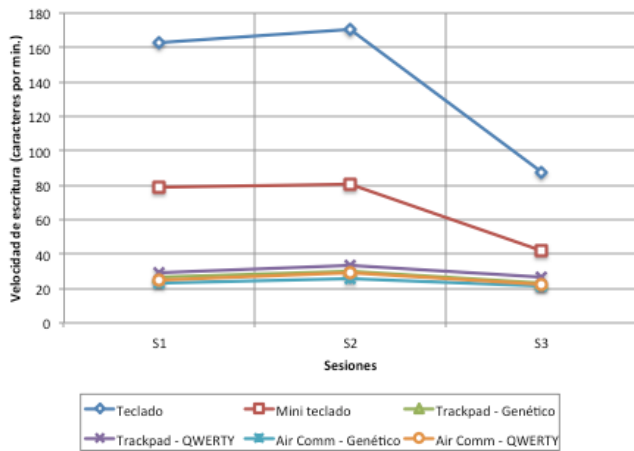


Figura 7. Velocidad de escritura en las tres sesiones de la Fase #5.

El problema de los teclados físicos se debe a la necesidad de pulsar más de una tecla simultáneamente para escribir determinados símbolos, sin una superficie sobre la que apoyar el teclado. Esto quedó también reflejado en las encuestas de los usuarios. Un 34,66% considera que este método es incómodo para usar en el contexto de visionado de televisión. En el caso del touchpad y el dispositivo apuntador, los mejores resultados se obtienen con el touchpad y la disposición QWERTY, con una velocidad media de 33,285 caracteres por minuto sin símbolos complejos y una media de 26,597 cuando se utilizan estos símbolos.

En cuanto a los errores, en la Fig. 8 se observa que la tasa de error disminuye en la segunda sesión, salvo en el caso de los teclados físicos. Es decir, con un poco de experiencia los usuarios mejoran con el touchpad y el dispositivo apuntador pero no con los teclados físicos. Cuando se introducen los símbolos complejos en la tercera sesión, la tasa de error crece en general, pero de forma más pronunciada en los teclados físicos. A la vista de los resultados, podemos afirmar que el touchpad es el mejor método desde el punto de vista de la tasa de error. Sin símbolos complejos los mejores resultados se obtienen con la disposición Genética con una tasa promedio del 6,093%. Cuando los símbolos son complejos, QWERTY es la mejor disposición con una tasa del 11,109%.

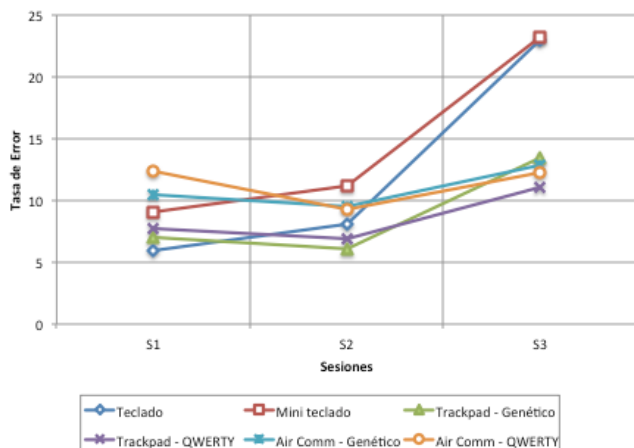


Figura 8. Tasa de error en las tres sesiones de la Fase #5.

C. Resultados de la Fase #6

Tomando como referencia de los resultados obtenidos en las fases anteriores, se realizó el diseño de un dispositivo de interacción llamado TVPad [27]. Este diseño se basa en la idea de que el mecanismo más versátil y sencillo para interactuar con la televisión es un periférico que permita apuntar a alguna zona del televisor. El dispositivo se acompañaría de un interface de usuario adaptado.

De los dispositivos de interacción apuntadores disponibles en el mercado, hemos seleccionado el touchpad como base para nuestro diseño. Además de los resultados empíricos de los experimentos realizados en las fases anteriores, se han tenido en cuenta otras consideraciones para tomar esta decisión. Por ejemplo, el hecho de que un 56% de los usuarios de la fase #5 considere explícitamente que el touchpad es un dispositivo muy sencillo y cómodo de utilizar, llegando a afirmar que “es el mejor método para personas que no saben mecanografía” o que “es el mejor método de todos como mando”. Por todo lo anterior, consideramos que es adecuado para reemplazar al mando a distancia convencional. Nuestra propuesta es que TVPad tenga unas dimensiones similares a las del Magic Trackpad de Apple, además de desarrollarlo de forma que pueda ser utilizado sin necesidad de un apoyo. Tal y como se observa en el prototipo de la Fig. 9, la mayoría de la superficie se destinaría a la interacción táctil. Por otro lado, para permitir el acceso a las funciones más frecuentes, se propone incorporar una serie de botones:

- Encendido y apagado del televisor.
- Control de volumen.
- Cambio de programa o desplazamiento por una lista.
- Cambio de modo de funcionamiento del televisor.
- Retroceso o salir.

Igualmente, se propone ubicar en el frontal del dispositivo un interruptor para apagar el dispositivo.

Por otro lado, en la Fig. 10 se muestra un diseño conceptual de cómo sería el interface de usuario. Con toda probabilidad, el televisor tendrá varios modos de funcionamiento y el cambio se podría controlar desde el TVPad. En un modo “smartTV” el dispositivo podría utilizarse tal y como se hace en un ordenador. Por ello, el diseño conceptual de la Fig. 10 se plantea para dar cobertura al modo de visionado de televisión.

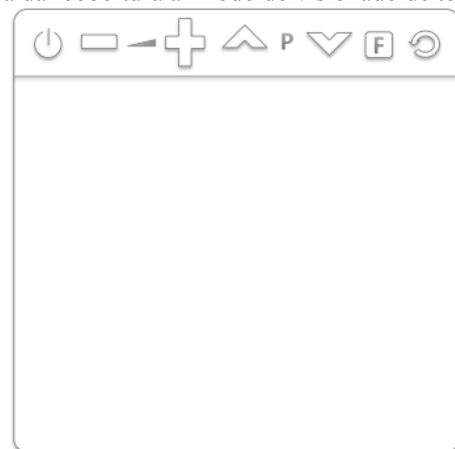


Figura 9. Prototipo del dispositivo de interacción propuesto.



Figura 10. Diseño conceptual del interface de usuario.

En el modo de visionado, el usuario podría utilizar los botones disponibles en el TVPad para las acciones más comunes. Adicionalmente, cuando el usuario toca la superficie táctil del dispositivo, aparecería el cursor en pantalla junto con cuatro recuadros, tal y como se muestra en la Fig. 10. El recuadro superior se utilizaría como zona de control. El del lado izquierdo se utilizaría para pintar un teclado numérico virtual. El del lado derecho se utilizaría para controlar el volumen. Finalmente, el inferior se utilizaría como zona habilitada para el programa que se está visualizando.

En la zona de control, se propone ubicar botones para acceder a la guía electrónica de programación, a la configuración del televisor y otros controles que, en función de las capacidades del televisor, puedan ser necesarios en este modo y que no sean particulares del canal que se está visualizando. El control se haría mediante botones de selección y desplazamiento, y mediante gestos. En la zona central se ubican un teclado numérico virtual y la barra de volumen. El teclado numérico permite introducir directamente un número de canal. Aunque se propone su ubicación en la zona izquierda de la pantalla, ésta y sus dimensiones podrían variar en función del televisor. La barra de control de volumen permite al usuario ajustar el volumen o silenciar el televisor. Finalmente, la zona de programa está diseñada para mostrar la información del programa que se está visualizando: nombre, espacio de tiempo, descripción, etc. Igualmente, se permitirían realizar acciones sobre el programa como activar el teletexto, cambiar de idioma o activar/seleccionar los subtítulos.

Con independencia de estos aspectos, en el modo de visionado estándar se plantea la posibilidad de realizar algunos gestos sobre el TVPad para las siguientes acciones:

- Ampliar o reducir la imagen, con un movimiento de pinza en diagonal.
- Subir o bajar el volumen, con un movimiento de desplazamiento vertical con dos dedos.
- Cambiar de modo de funcionamiento, con un movimiento de desplazamiento horizontal con dos dedos.

VI. CONCLUSIONES

En este trabajo se presenta un proyecto liderado por el grupo de investigación de Sistemas de Distribución

Multimedia de la Universidad de Oviedo. En este proyecto se ha evaluado el rendimiento obtenido con distintos métodos de escritura de texto susceptibles de ser utilizados en servicios de televisión digital interactiva.

En una primera etapa de la evaluación, se ha tenido en cuenta que el dispositivo mayoritariamente disponible en estos servicios es el mando a distancia, por lo que los métodos se han seleccionado acorde a esta situación. Así, hemos visto cómo los métodos más rápidos provenían del mundo de los dispositivos móviles (Multitap y T9), pero a costa de unas tasas de error altísimas. La mayoría de los errores se producen porque los usuarios no se dan cuenta de que han cometido un error y continúan escribiendo. Esta tasa de error se incrementa notablemente cuando los textos incorporan símbolos que podríamos denominar complejos (signos de puntuación, tildes, etc.) Dado que estos símbolos son necesarios en un contexto real, los teclados virtuales son la opción más recomendada. Con las diferentes disposiciones de símbolos que hemos probado no hemos obtenido diferencias significativas, pero la retroalimentación de los usuarios inclina la balanza hacia las disposiciones QWERTY y Genética. Finalmente, tampoco hemos encontrado diferencias significativas cuando se altera la forma del mando a distancia o la disposición de sus teclas.

En una fase posterior del proyecto se han realizado experimentos con otros dispositivos, terminando con una propuesta de lo que en nuestra opinión podría ser el dispositivo ideal. En términos objetivos de rendimiento, los mejores resultados se obtienen con un teclado convencional, a costa de unas elevadas tasas de error cuando se introducen símbolos complejos. No obstante, los usuarios manifiestan claramente que este periférico no es cómodo para ser utilizado en un entorno de visionado de televisión. Hemos visto que con la experiencia los usuarios mejoran su rendimiento con otros periféricos, acercando su rendimiento al del teclado convencional. Por ello, objetiva y subjetivamente el mejor dispositivo es el touchpad. Tomando este periférico como referencia, el proyecto concluye con un diseño de un touchpad adaptado a la televisión que hemos denominado TVPad. Al diseño del periférico incorporamos una propuesta de interface gráfica de usuario ajustada al modo de visionado de televisión.

En las pruebas realizadas en el marco del proyecto, se ha constatado la importancia de elegir un método de escritura adecuado, dado que esta elección impacta notablemente sobre la velocidad de escritura y el número de errores que cometen los usuarios. Pese a que la elección de este método depende en gran medida de los fabricantes de televisores o set-top-box, no es menos cierto que los resultados de este proyecto son directamente aplicables por los desarrolladores de aplicaciones, que sin duda tendrán un peso específico en el ya presente sector de los servicios de idTV.

AGRADECIMIENTOS

Este trabajo se realiza con el apoyo parcial del proyecto UsabiliTV (ID 1103 521 28462), financiado por el Ministerio de Educación Nacional de Colombia a través de COLCIENCIAS y de la alianza de la RedAUTI (CYTED 512RT0461).

REFERENCIAS

- [1] Brown, M., Coughlan, T., Lawson, G., Goulden, M., Houghton, R.J., and Mortier, R. 2013, Exploring Interpretations of Data from the Internet of Things in the Home. In *Interacting with Computers*, 25:3, pp. 204-217.
- [2] Jakob Nielsen. Remote Control Anarchy. Disponible en la dirección <http://www.useit.com/alertbox/20040607.html>, 2012
- [3] Iatrino, A. y Modeo, S. Text Editing in Digital Terrestrial Television: a Comparison of Three Interfaces. *Proceedings of EuroITV'06*, Atenas
- [4] Bellman, T., and MacKenzie, I. S. 1998, A probabilistic character layout strategy for mobile text entry. In *Proceedings of Graphics Interface '98*. pp. 168-176, Mississauga, Canada.
- [5] Zhai, S., Hunter, M. and Smith, B.A., 2000, The Metropolis Keyboard – An Exploration of Quantitative Techniques for Virtual Keyboard Design. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 119-128, San Diego, CA, USA.
- [6] Brewbaker, C.R. (2008). *Optimizing Stylus Keyboard Layouts With a Genetic Algorithm: Customization and Internationalization*. Department of Computer Science, Iowa State University
- [7] Silfverberg, M., MacKenzie, I.S. and Korhonen, P. (2000). Predicting Text Entry Speed on Mobile Phones. En *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 9-16. NY
- [8] Butts, L. & Cockburn, A. (2002). An evaluation of mobile phone text input methods. *Journal of Australian Computer Science Communications*. 24:4, 55-59.
- [9] MacKenzie, I.S., Kober, H., Smith, D., Jones T. and Skepner, E. (2001). LetterWise: Prefix-based Disambiguation for Mobile Text Input. En *Proceedings del ACM Symposium on User Interface Software and Technology*, pp. 111-120. New York
- [10] MacKenzie, I. S., & Tanaka-Ishii, K. (2007a). Evaluation of text entry techniques. In I. S. MacKenzie & K. Tanaka-Ishii (Eds.), *Text entry systems: Mobility, accessibility, universality* (pp. 75–101). San Francisco, CA: Morgan Kaufmann.
- [11] Wigdor, D., & Balakrishnan, R. (2004). A comparison of consecutive and concurrent input text entry techniques for mobile phones. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 81–88.
- [12] Oniszczak, A., & MacKenzie, I. S. (2004). A comparison of two input methods for keypads on mobile devices. *Proceedings of the Third Nordic Conference on Human-Computer Interaction*, 101–104.
- [13] MacKenzie I. S., Lopez M. H. and Castellucci S., 2009, Text Entry with the Apple iPhone and the Nintendo Wii. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, USA.
- [14] Siek K. A., Rogers Y. and Connelly K. H., 2005, Fat Finger Worries: How Older and Younger Users Physically Interact with PDAs. In *Proceedings of the 2005 IFIP TC13 international conference on Human-Computer Interaction*, pp. 267-280, Rome, Italy.
- [15] Castellucci, S. J., and MacKenzie, I. S. (2008). Unigest: text entry using three degrees of motion. *CHI'08 Extended Abstracts on Human Factors in Computing Systems* (pp. 3549–3554).
- [16] Jones, E., Alexander, J., Andreou, A., Irani, P., and Subramanian, S. (2010). GesText: accelerometer-based gestural text-entry systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2173–2182).
- [17] Költringer T., Isokoski P. and Grechenig T.: TwoStick: Writing with a Game Controller. En *Proceedings of Graphics Interface 2007 (GI 2007)*, 103-110
- [18] Wobbrock, J.O., Myers, B.A. & Aung, H.H. (2004). Writing with a Joystick: A Comparison of Date Stamp, Selection Keyboard, and EdgeWrite. *Proceedings of Graphics Interface '04*, pp. 1-8, Ontario, Canada.
- [19] Ingmarsson, M., Dinka, D., Zhai, S. (2004). TNT – A Numeric Keypad Based Text Input Method. In *Proceedings del SIGCHI Conference*, pp. 639-646, Vienna, Austria
- [20] Geleijnse G., Aliakseyeu D, Sarroukh E. (2009). Comparing Text Entry Methods for Interactive Television Applications. En *Proceedings del EuroITV'09*, Leuven, Belgium.
- [21] Orbist, M., Bernhaupt, R. and Tscheligi, M. (2008). Interactive TV for the Home: An Ethnographic Study on Users' Requirements and Experiences. En *International Journal of Human-Computer Interaction*, 24:2, 174-196
- [22] Gargi U., Gossweiler R.: QuickSuggest: Character Prediction on Web Appliances. En *WWW 2010*, April 26-30, 2010, Raleigh, North Carolina, USA
- [23] Sporka, A.J., Polacek, O. and Slavik, P., 2012, Comparison of two text entry methods on interactive TV. In *Proceedings of the 10th European conference on Interactive TV and Video*, pp. 49-52, Berlin, Germany.
- [24] MacKenzie, I. S. (1991). *Fitts' law as a performance model in human-computer interaction*. Doctoral dissertation. University of Toronto: Toronto, Ontario, Canada
- [25] Barrero, A., Melendi, D., Pañeda, X.G., García, R., Cabrero, S. (2014) An empirical investigation into text input methods for interactive digital television applications. En *International Journal of Human-Computer Interaction*, 30:4, 321-341
- [26] Barrero, A., Melendi, D., Pañeda, X.G., García, R., Cabrero, S. (2015) An empirical investigation into typing errors in interactive digital television applications. En *International Journal of Human-Computer Interaction*, 31:3, 210-225
- [27] Barrero, A., Melendi, D., Pañeda, X.G., García, R., Cabrero, S. (2014) TVpad: Un novedoso sistema de interacción con la televisión. En *II Jornadas Iberoamericanas de Difusión y Capacitación sobre Aplicaciones y Usabilidad de la Televisión Digital Interactiva*, 10 a 12 de Septiembre, 2013, España.



Aurora Barrero is a Computer Science Engineer from the University of Oviedo and PhD student. Her current research interest is in the area of multimedia systems and services, content distribution networks and interactive digital TV services. Currently, she is a Research Engineer in the Department of Computer Science of the University of Oviedo.



David Melendi is a Computer Science Engineer with a PhD from the University of Oviedo. His current research interest is in multimedia systems and services, content distribution networks, idTV services, mobile ad-hoc networks and efficient driving. He is an Associate Professor at the University of Oviedo and member of the W3C.



Xabiel G. Pañeda is a Computer Science Engineer with a PhD from the University of Oviedo. His current research interest is in multimedia systems and services, content distribution networks, idTV services, mobile ad-hoc networks and efficient driving. He is an Associate Professor at the University of Oviedo and member of the W3C.



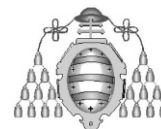
Roberto García is a Telecommunications Engineer from The Technical University of Madrid with a PhD from the University of Oviedo. His current research interest is in telecommunication networks and services, performance analysis, modelling and simulation of systems and services and efficient driving. He is an Associate Professor at the University of Oviedo.



Laura Pozueco is a Telecommunications Engineer from University of Oviedo with a PhD from the Spanish University for Distance Education (UNED). She is a contracted researcher in the group of Distributed MultiMedia Systems and her current research interests are in the area of telecommunication networks and services and efficient driving.




Jose L. Arciniegas is Telecommunication Engineer and PhD from the University Politechnique of Madrid. He is full Titular Professor with the Department of Telematics of the University of Cauca (Colombia). His current research interests are in the area of services and application using digital interactive television, software architecture and quality of software. He is also taking part in several research projects on Colombia, Iberoamerica and European levels.



6.4 Evaluation of Text Entry Methods for Interactive Digital Television Applications with Devices Alternative to Conventional Remote Controls

A. Barrero, D. Melendi, X. G. Pañeda, R. García, y L. Pozueco, «Evaluation of Text Entry Methods for Interactive Digital Television Applications with Devices Alternative to Conventional Remote Controls», *International Journal of Human-Computer Interaction*, vol. 32, n.o 10, pp. 765–776, 2016 DOI: 10.1080/10447318.2016.1195586

Evaluation of Text Entry Methods for Interactive Digital Television Applications with Devices Alternative to Conventional Remote Controls

Aurora Barrero, David Melendi , Xabiel G. Pañeda, Roberto García, and Laura Pozueco

Department of Informatics, University of Oviedo, Xixón, Asturias, Spain

ABSTRACT

The popularity of Interactive Digital Television (IDTV) applications has grown in recent years, playing a significant role in today's society. This new type of television allows users to access interactive applications in order to look for information, for communication or educational purposes, or just for fun. Most of these applications require text entry and many users expect a user experience similar to that of computers. Although the most common device to interact with television is the conventional remote control, other devices available on the market may be used in this context. Thus, in this article we aim to evaluate and compare different text input methods for IDTV applications using devices alternative to conventional remote controls. We have carried out an empirical study with 52 participants. We analyze entry speeds, error rates, and subjective impressions for six different entry methods, taking into account the particular characteristics of the users. As devices we have used a full-sized keyboard, a palm-sized keyboard, a gyroscopic remote point-select, and a modified touchpad. The fastest method is the standard keyboard, but in an IDTV context many users report discomfort and problems under low lighting conditions. Also, error rates are considerably higher with both keyboards when modifier keys need to be used. The results obtained with the gyroscopic remote and the touchpad are similar. Nevertheless, while users complain about fatigue problems with the former, their feedback about the latter is very positive. We have also observed that age is a major factor affecting the performance of the users. We expect our results to contribute to the design of new text entry methods for IDTV.

1. Introduction

There is a huge emerging market in the interactive digital television (IDTV) realm thanks to the evolution of communications networks and the proliferation of Internet connections in the home (Brandtzæg, Heim, & Karahasanović, 2011). This has motivated several content providers to move toward a convergent service, drawing the web and television closer together. Moreover, the number of IDTV users is growing considerably. This is creating a highly competitive market in which different providers offer interactive multimedia services by means of a connected TV or set-top box (STB).

Nowadays, televisions can be used for web browsing, social networking, and to share data, in an attempt to provide viewers with the ability to create, edit, share, and control contents (Cesar & Chorianopoulos, 2009). However, the main interaction device available is the traditional remote control which may be suitable for the IDTV context to carry out simple interactions. For instance, users may move through menus using a reduced set of keys of the remote control. Nevertheless, there are other types of interaction that may require further development to improve the experience of the users, one of which is writing text. Users may need to write their username and password to authenticate themselves on a pay-per-view platform. They may need to

perform searches on a video streaming service or interact with other users in a social TV environment. The problem is that conventional remote controls were not designed to write text, making it a relatively difficult task for some users (Iatrino & Modeo, 2007).

In order to find effective typing methods for IDTV applications, we have carried out an empirical research on several text entry methods suitable for this context. The main goal of this research was to analyze text entry methods that may be used with commonly available technology that most people will probably already have in their homes. We started with several experiments aimed at finding effective mechanisms to type text with traditional remote controls. The results were published in Barrero, Melendi, Pañeda, García, and Cabrero (2014). But given the fact that other devices may be used in an IDTV environment to allow users to write text efficiently, we have further extended our research to perform experiments with devices different from conventional remote controls. This study involved 52 participants between the ages of 20 and 69 with heterogeneous characteristics. They had to write several sentences in different sessions, using a full-sized QWERTY keyboard, a mini QWERTY keyboard, a modified touchpad, and a gyroscopic remote control. Both the touchpad and the gyroscopic remote are used with a QWERTY and a Genetic (Brewbaker, 2008)

virtual keyboard layout. We studied effectiveness (speed and error rates) and users' preferences.

We chose a full-sized keyboard because it is the most common computer input device nowadays. For the same reason, we chose a keyboard with a QWERTY layout, despite previous works identifying certain limitations (Swanson, Galinsky, Cole, Pan, & Sauter, 1997) and several pathologies caused by its use (Rose, 1991). Mini keyboards (a.k.a. palm-sized keyboards) are a miniature version of the traditional full-sized desktop keyboard. We have included one of these devices because they are starting to become very popular in home entertainment systems. Its use is usually performed with two fingers. Most users hold the device with both hands and use both thumbs to press the keys. Although such thumb typing may seem more difficult than the full 10-finger typing, Clarkson, Clawson, Lyons, and Starner (2005) showed that significant speeds can be acquired after very little practice. A touchpad is a pointing device with a flat surface that detects the position of the finger. We have included this device because although it is usually found in laptops, it is also available as an independent device. For older users, touchpads may be problematic depending on their size (Taveira & Choi, 2009). Also, these users may encounter problems when tapping touchpads (Wood, Willoughby, Rushing, Bechtel, & Gilbert, 2005). Furthermore, they may lead to localized fatigue if used during long periods of time (Taveira & Choi, 2009). Gyroscopic remote controls are pointing devices that detect hand movements and translate those movements into variations in the position of the cursor. They allow users to control the cursor of the mouse as if they were pointing remotely with a finger. We have included them due to the popularity of the Wii system and its Wiimote (Aoki, Maeda, Watanabe, Kobayashi, & Abe, 2010). However, these pointing devices may potentially cause problems of fatigue (Nacke, 2010). Also, precision may be affected by hand tremor and our own pointing precision limitations (König, Gerken, Dierdorf, & Reiterer, 2009).

The remaining part of this article is structured as follows: some previous work is covered in Section 2. Section 3 presents the evaluation method of the study and its characteristics. The analysis of the results is detailed in Section 4. Finally, conclusions and future work are presented in Section 5.

2. Related Work

The study of different text input methods is a wide field where a lot of research has been carried out, although most works have been focused on mobile devices (MacKenzie & Soukoreff, 2002a). However, in the recent Human-Computer Interaction (HCI) literature one can find a variety of studies with different devices and text entry methods of interaction in the context of IDTV. Some of these devices are different to traditional remote controls, with keyboards, touch screens, microphones (speech analyzers), or remote pointing devices to name a few examples.

There are commonly two methods to write text in an IDTV application with a conventional remote control: using cursors and OK keys on a virtual keyboard or using a reduced set of keys. In the case of virtual keyboards, the QWERTY layout (Bellman & MacKenzie, 1998) is the most common configuration. On the other hand, Multitap is the most common

technique when a reduced set of keys is used. The former was very common on mobile phones and young and adult users are familiar with it. As an example of evaluations of this method, Silfverberg, MacKenzie, and Korhonen (2000) compared three text input techniques using a reduced set of keys: the Multipress (a.k.a. Multitap) method, in which three or four symbols are assigned to each key (to enter a character users have to repeatedly press the same key), a method based on two keystrokes called 2-key (first users select a group of symbols and afterwards the desired symbol), and the predictive method T9 (based on the theory that each combination of keystrokes produces a single word or a reduced set of words). The evaluation was carried out with 12 volunteers aged between 23 and 47. The authors highlighted that the best results were obtained with T9 and the worst with 2-key.

One of the first works based on evaluating text input methods in the IDTV context was that of Ingmarsson, Dinka, and Zhai (2004). They designed The Numpad Typer (TNT) method. Each character was accessible with two keystrokes on a 9-button keypad on the remote control with a similar approach to 2-key. The results show writing speeds comparable or superior to Personal Digital Assistant (PDA) handwriting or the Multitap method. Users found the method very easy to use and to understand.

Iatrino and Modeo (2007) evaluated the aforementioned ways of typing text with a conventional remote control. They compared three methods: Multipress (a.k.a. Multitap), Multipress with visual feedback, and a QWERTY virtual keyboard. Thirty-six subjects participated in an experiment in which users had to write an e-mail address and a short sentence in Italian. Their results show that Multipress was better than the virtual keyboard in terms of speed and user satisfaction.

More recently, Barrero et al. (2014) conducted a comparative experiment in which 96 participants had to use different typing methods. These methods were several virtual keyboard layouts (QWERTY, Square Alphabetic and Genetic) and three methods based on a reduced set of keys (Multitap, T9, and 2-key). They also performed experiments with three remote control devices of different sizes, shapes, and keysets. The results show that, when writing simple texts, Multitap proved the best method. However, when complex texts had to be written (such as in a real context), virtual keyboards presented the same or even better speeds than Multitap but with significantly lower error rates. They also found that the design of the remote control did not have a significant impact on performance.

Although the traditional remote control is still the main interaction device available for televisions, it was not designed to type text (Iatrino & Modeo, 2007) or to perform complex interactions. Thus, new devices are appearing on the television and STB market, creating new interaction techniques in this context. Nevertheless, some authors are skeptical. For instance, Obrist, Bernhaupt, and Tscheligi (2008) stated that it is not feasible to rely on the popularity of external peripherals for televisions different to a conventional remote control. They carried out an ethnographic study. The aim of the research was to gain a deeper understanding of household types and their members, daily habits, social background, TV viewing behavior and experiences, and expectations toward IDTV. They concluded

recommending voice recognition as an alternative to conventional remote controls.

A different approach to the problem of typing text would be to use a conventional physical keyboard. In fact, Marshall, Foster, and Jack (2001) claimed that this was the best solution for IDTV text entry. On the other hand, other authors such as Lee and Schmidt (2005) argue that the desktop metaphor should be avoided because a conventional keyboard is difficult to handle in many postures. There are, however, some previous works in which physical keyboards have been used. For instance, Geleijnse, Aliakseyeu, and Sarroukh (2009) compared Multitap, T9, and a QWERTY virtual keyboard, with the speed of a conventional keyboard. The goal was to search for author-track pairs on YouTube. Twenty-two people aged between 21 and 32 participated in the experiments. The authors concluded that the conventional keyboard was faster than remote control methods. Another conclusion was that they found no indications about users not accepting keyboards in a living room-like setting. Furthermore, there are also studies in which mini keyboards are analyzed such as that of MacKenzie and Soukoreff (2002b) who proposed a model for two-thumb text entry on mobile keyboards. Their model provides a behavioral description of the interaction as well as a predicted text entry rate in words per minute (wpm). The prediction obtained was 60.74 wpm. Also, Clarkson et al. (2005) conducted a study using two different mini-QWERTY keyboards. They recruited 21 subjects who had not used a mini-QWERTY keyboard more than once before. Both keyboards were oval-shaped. The Dell keyboard had a single space key in the middle of the bottom row of keys, while the Targus had two triangular space keys set below the rest of the keys. The results show an evolution of participants from 31.72 wpm in the first session to 60.03 wpm in the 20th session.

Apart from physical keyboards, point-and-click devices such as gyroscopic or touchpad devices may be used to type text in IDTV applications. However, there are not many studies in which these devices are used for text entry. The majority of these studies use the Wiimote as a method to recognize gestures. For instance, Castellucci and MacKenzie (2008) carried out a study with 18 participants of ages ranging from 18 to 42. They used a technique called UniGest, based on a gesture alphabet and a Wiimote. Gesture durations ranged from 296 to 417 ms. Combined with the proposed gesture alphabet, an upper-bound text entry rate of 27.9 wpm is possible. Also, Jones, Alexander, Andreou, Irani, and Subramanian (2010) evaluated the performance of a Wiimote device with two accelerometer-based input techniques: Tri-center and Matrix. Fifteen users aged 19 to 28 with experience using accelerometer-based controllers participated in the study. The matrix interface text entry rate was of 3.7 wpm and the Tri-center interface was 3.3 wpm. On the other hand, MacKenzie, Lopez, and Castelluci (2009) conducted an empirical study with 16 participants using an Apple iPhone and a Wiimote. Based on a conventional virtual keyboard, the authors compared text entry speeds and error rates. The participants entered text at 18.5 wpm using the iPhone and at 9.2 wpm using the Wiimote. The results suggested that using a touch screen to enter text was far faster than using a remote pointer such as a Wii remote. In the case of touchpad devices, there have been few studies applied in the context of

IDTV applications. Enns and MacKenzie (1998) presented a new style of remote control device combining a touchpad with gesture-based interactions. Unistrokes were used to issue commands and select menu items presented on a television screen. More recently, Choi, Han, Lee, Lee, and Lee (2011) carried out a study comparing a “clickable” touchpad with a “clickable” joystick (an Xbox 360 controller). Twelve university students participated in the first part of the study. They had an average age of 25.4 and all had some experience with touchpads. In the second part of the study they improved the design of the touchpad thanks to the feedback obtained in the first part. In the second part, the participants had to perform three tasks (numeric gestures, program selection, and text entry) using the touchpad device under two different conditions (Hover-Tracking and No-Hover-Tracking). Only 11 of the initial 12 students participated in this part of the study. Some results were that there was no difference in text-entry speeds between conditions and that the best speeds that they could achieve were around 8.5 wpm.

There are also some studies that have used game controllers different from the Wiimote to enter text. For example, Költringer, Isokoski, and Grechenig (2007) compared a technique named TwoStick with a QWERTY layout in the Microsoft Xbox 360. In the study, users increased their initial speed of 4.3 wpm to 14.9 wpm after 5 hr of practice. Also, Wobbrock, Myers, and Aung (2004) designed a new joystick text entry method called EdgeWrite. They used a game controller to enter text based on gestures over a unistroke alphabet. They compared this method and the datestamp and selection keyboards. Results showed that EdgeWrite was at least 1.5 times faster than a selection keyboard and 2.4 times faster than with datestamp.

Finally, there are also papers aiming to solve the problem of writing text by using other techniques. Most of them rely on voice recognition, interactive surfaces, or freehand gestures. This is the case of Vega-Oliveros, Pedrosa, Pimentel, and De Mattos Fortes (2010) who explored a multimodal text input on IDTV combining speech recognition, a cell keypad, and a virtual keyboard. Also, Rick (2010) studied the efficiency of entering text by using interactive surfaces, such as a touch-based tabletop, in which users move through the letters of a word on a virtual keyboard without lifting their finger. Similarly, Varcholik, LaViola, and Hughes (2012) presented two studies in which a virtual QWERTY keyboard was used on a multi-touch platform. More recently, Ren and O’Neill (2013) investigated text entry using freehand gestures captured with a sensor system using three selection techniques and two layouts: QWERTY and Dual-Circle.

As can be seen, writing text in an IDTV environment is a problem that has received some attention in recent years. Some authors have designed ad hoc methods or devices to solve this problem, but these are not available in the market or require specific training. Most authors perform tests with people, but the number of participants is generally low and they are usually homogeneous. The texts used in the experiments are normally short sentences in English, so no complex symbols are needed.

The main contributions of our article are as follows: the results have been obtained from an experiment that was

conducted with a heterogeneous set of users: non-expert and expert, young and old, male and female. These participants come from different cultural backgrounds than users in other experiments in previous works, so the results may be compared allowing us to know whether they can be extrapolated globally or not. Moreover, four devices different to a standard remote control that most people will more than likely already have in their homes were tested. Furthermore, we use simple texts as in previous works, but also texts including special characters as in a real service. Finally, we report objective results about the performance of the users and their error rates, with in addition, subjective information about the feelings of the users and the way they use each of the devices.

3. Evaluation Method

3.1. Participants

Fifty-two users took part in the experiments: 22 were female and 30 were male. They all had different technological skills, level of studies, and habits. Six participants were left-handed. Their ages ranged from 20 to 69 with an average of 38.90 (SD = 12.70) and they did not receive any compensation for participating in the experiments.

The users have been classified into three groups depending on their age. There were 21 “young” users who belonged to the mobile phone generation (aged between 18 and 30). Fifteen were “adult” and belonged to the “computer generation” (aged between 31 and 45). Finally, 16 users were “older” (aged over 45). We have considered age in our analyses due to the fact that changes in perceptual and motor skill abilities that often accompany aging bring implications to the design of human–computer interfaces (Taveira & Choi, 2009). Also, we have considered gender and laterality because these are, together with age, common types of demographic groups used in usability experiments.

Initially, we expected differences between users depending on their habits, skills, and prior experience toward technology. According to the information gathered in an initial questionnaire, 88.46% of the participants had a certain degree of previous experience with traditional keyboards; 67.31% of them had some experience with mini keyboards. Finally, 53.85% of the users had some experience with touch devices and only 28.85% of them had used gyroscopic devices.

Table 1 shows the groups of users according to their age and experience with each of the devices in the experiment. We can see that most of the users state that they have considerable experience with the traditional keyboard. Only six users report no experience with this device (most of them in the older group). On the other hand, most of the users report no

experience with the gyroscopic remote control. Only 15 users report certain experience with this device. Moreover, we can see that the majority of young users report some experience with all the devices, whereas most older users report no experience with any device. The group of Adult users is more balanced. This situation clearly reflects the current social environment.

3.2. Apparatus

The experiments were carried out in a lab with conditions similar to those in the sitting room of a house. Participants were left alone, sat in a comfortable armchair positioned 2 m from a 32” LCD television, following the recommendations of the manufacturer. The television was an LG 32LG5000 with dimensions 814 × 599 mm and a resolution of 1280 × 720 pixels. It was connected to a PC that ran the application designed to carry out the tests. The communications between the PC and the devices were carried out using their own adapters.

The application shown in Figure 1 was created to perform the experiments. It was developed in C++ and used the QT framework. This application was run in full-screen mode and the fonts used were Verdana 65 points for the sentences and Arial 60 points for virtual keyboard layouts. As shown in Figure 1, the topmost area displayed the sentence to be written and an input text field was used to display the text written by the users. When virtual keyboards were necessary, layouts were placed in the bottom part of the screen as shown in the figure.

The application showed a 5-s timer at the beginning of each sentence, allowing participants to read what had to be written. Afterwards users were allowed to write. Symbols could be deleted in order to perform corrections. Nevertheless, no feedback was provided to users if they made a mistake as in a real IDTV environment. When the text written coincided with the proposed sentence, the application moved to the next sentence, and this process was repeated until the end of the test. Once all the sentences had been written, the application showed a final questionnaire designed to gather subjective information. This questionnaire had several questions that could be answered using a Likert scale (Likert, 1932).

The actions of the users and the answers collected in the final questionnaire were stored in an XML file that was later used to perform the analyses.

Table 1. Classification of users according to their experience and age.

	Number of experienced users			Number of inexperienced users		
	Young	Adult	Older	Young	Adult	Older
Keyboard	21	14	11	0	1	5
Mini keyboard	19	11	5	2	4	11
Touchpad	16	10	2	5	5	14
Gyroscopic	9	4	2	12	11	14



Figure 1. Application used in the experiments.

3.3. Text Input Techniques

The application described in the previous section was used in combination with different text input methods and devices. For the devices, we used a full-sized keyboard, a mini-sized keyboard, a remote point-select, and a modified touchpad. With the devices requiring a virtual keyboard layout, we used the QWERTY and Genetic layouts adapted to the Spanish language. In this section, we describe each of these devices and layouts.

Traditional Full-Sized Keyboard

In the experiments we have used a Logitech Wireless Touch Keyboard K40. It is an ultrathin and compact keyboard with low-profile keys. Its dimensions are $13.716 \times 35.303 \times 2.286$ cm and it has a weight of 357.20 g.

When the participants use this technique in the evaluation test, they only have to read the sentences and type the appropriate keys. The user holds down the Alt Gr key to enable certain symbols and the SHIFT key to enable the uppercase chars.

Mini Keyboard

In the experiments we used a Logitech diNovo Mini. Its dimensions are $15.2 \times 9 \times 2.75$ cm and it weighs 317.51 g. It has a backlight mechanism that makes text entry effortless, even in low light. When the participants use this technique in the evaluation test, they simply have to read the sentences and type the appropriate keys. Users have to hold down the FN key to enable certain symbols and the SHIFT key to enable uppercase chars.

TouchPad

We used an Apple Magic Trackpad of dimensions $19 \times 13 \times 4$ cm and a weight of 200 g. As these devices usually need to be used on a surface for clicking purposes, we slightly modified the device as shown in Figure 2a. An attached surface allows users to use the device without a table. Given the fact that we aimed to experiment with commonly available technology, the TouchPad worked with its out-of-the-box c-d ratio and with a relative mapping between it and the application. Taking into account the resolution of the screen, users had to use clutching to interact with the system.

Gyroscopic Remote Controls

In our experiments we used a Fujitsu Air CommAnd Plus. It has a 3-axis accelerator and a 2-axis gyroscope sensor. Its dimensions are $21.0 \times 3.2 \times 3.5$ cm and its weight is 160 g. When used in the experiments, users need to point at the desired symbol on the screen and then press a button with the index finger. Figure 2b demonstrates the two possible positions of the device in the user's hand, with the index finger on the activation trigger. We aimed to experiment with commonly available technology, thus, this control worked with its out-of-the-box behavior. Whenever the cursor went beyond the borders of the screen, users had to press a button to freeze the cursor and then refocus. Otherwise the cursor started working unaligned with the control.

Layouts for Pointing Devices

Although physical keyboards come with an out-of-the-box layout, when pointing devices are used it is necessary to present to the user an on-screen virtual keyboard (or soft keyboard). Users point and select the symbols they would like to write. Virtual keyboards can be as easily customized and improved as any other software solution. Previous work has broadly studied virtual keyboards (Barrero et al., 2014; MacKenzie, Zhang, & Soukoreff, 1999; Zhai, Hunter, & Smith, 2000). They require constant visual attention by the users.

In our research, we have used two virtual keyboard layouts: QWERTY and the Genetic layout (Brewbaker, 2008). Both layouts have been slightly modified to include vowels needed in Spanish, capital letters, and special symbols. As shown in Figure 2c, QWERTY has an additional row for vowels with tildes. In the case of the Genetic layout, we have used the algorithm of Brewbaker (2008) as in Barrero et al. (2014). Using a genetic algorithm and Don Quixote as a reference text, the Genetic layout positions the most common symbols in the center of the keyboard. The result is shown in Figure 2d.

In both layouts, four additional keys were used for different purposes. One allows the user to switch to the same layout but with capital letters. Another allows the user to switch to another layout with special symbols, as shown in Figure 2e. Finally, two more buttons allow the users to write blank spaces and delete.

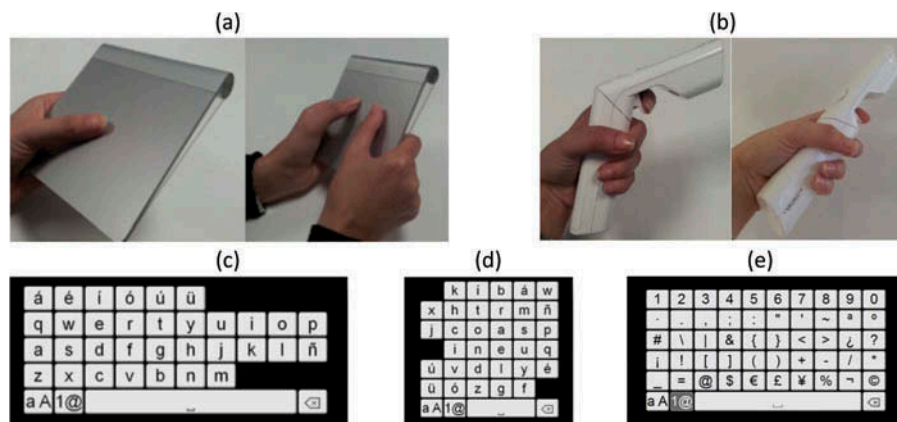


Figure 2. Pointing devices and virtual keyboards.

3.4. Procedure

The experiment had a 3×6 within-subjects design, following MacKenzie and Soukoreff (2002a). It was divided into three sessions with a break of two days between sessions. In each session, users had to write three sentences with each of the devices and methods described previously. Participants were divided into six groups to counterbalance the impact of the order of the methods.

There were two independent variables: input methods and sessions. Also, the dependent variables were entry speed and error rate, following the criteria in Soukoreff and MacKenzie (2003). Entry speed was measured in words per minute considering an average of five characters per word. On the other hand, error rate was calculated using the MSD error rate, by comparing the number of errors with the total amount of symbols written.

In the first session, users received basic instructions about how to interact with the application and how to use each input method. Also, the type and number of sentences to be written was described, as well as the method for correcting mistakes. These instructions were verbal and they were reinforced with a computer presentation and a practical demo. Occasionally, it was necessary to perform an initial training for the results to be reliable. This happened with some of the oldest users and users without any technological experience. This was carried out through a special session in which they had to write a set of sample sentences.

In all of the sessions, participants were left alone in the room to perform the test with no time limit. They had to write “as quickly and accurately as possible”. The sentences were representative of the Spanish language. In the first and second sessions the sentences were gathered from Internet newspapers considering only lower-case letters (particular Spanish symbols were removed). The sentences ranged from 113 to 121 characters with an average length of 115.6 characters. While the main aim of these sessions was to evaluate how users learn, the third session was meant to check their performance in a real IDTV environment. Thus, the sentences in the third session included numbers, Spanish tildes, upper-case letters, and special symbols. In this case, users had to write an e-mail address, a regular address, and an Internet URL. In all 2808 sentences were written during the experiments (52 participants \times 6 devices/methods \times 3 sessions \times 3 sentences).

In order to collect details of the users and subjective information, we designed several questionnaires. During the first session of the experiments, users had to fill out a web questionnaire with approximately 30 questions. This allowed us to know certain details of the users such as their experience with computers or with some of the devices used in the experiments. Also, at the end of each session, the users had to answer three 5-point Likert questions to rank the methods according to ease of use, speed, and overall satisfaction. This questionnaire was integrated into the test application. Finally, after the experiment the users had to complete another WEB questionnaire in which they had to describe each of the methods in their own words.

4. Results and Discussion

To perform our analyses, we have used the most common statistical tests for this type of analysis (Crawley, 2012). Also, for the sake of simplicity, we have used the following abbreviations to refer to each device and method: KT for Keyboard-Traditional, KM for Keyboard-Mini, TQ for Touchpad + QWERTY, TG for Touchpad + Genetic, GQ for Gyroscopic + QWERTY, and GG for Gyroscopic + Genetic. Finally, regarding handedness, we have only considered two groups of users, one for strictly right-handed users and another for the rest of the participants.

4.1. Writing Speed

Figure 3 shows the average writing speeds for each method. As shown in the figure, physical keyboards are much better than the rest of the devices. In fact, the best results are obtained with the traditional keyboard, while the worst are obtained with the Gyroscopic + Genetic method. This is consistent in all the sessions. Using *t*-tests it is possible to analyze the evolution of writing speeds for each method from Session 1 to Session 2. Table 2 shows that there are no significant changes for physical keyboards, but there is a significant improvement of entry speeds for the touchpad and the gyroscopic device ($p < 0.05$). The decrease in speed in Session 3 is caused by the increased level of complexity of the sentences used in this session.

In order to check whether the differences between methods have any statistical significance, we have performed pairwise

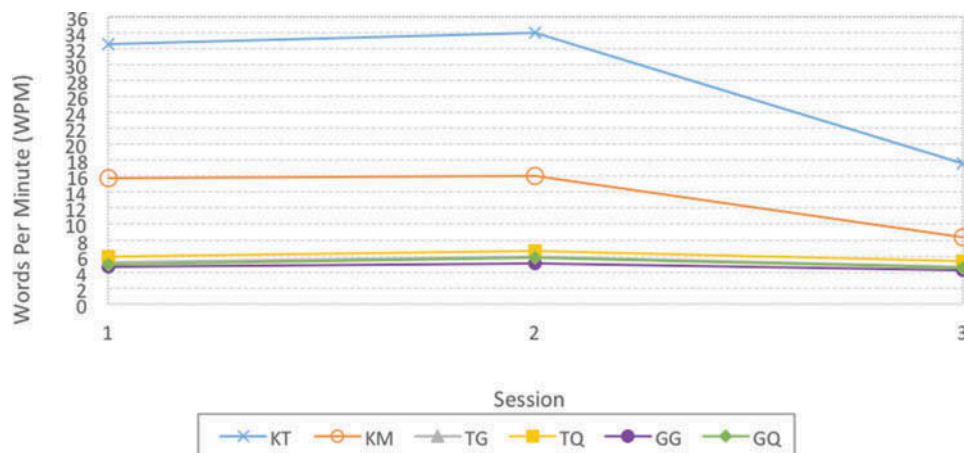


Figure 3. Average writing speeds per session and method in words per minute.

Table 2. Results of *t*-tests for entry speeds in sessions 1 and 2.

KT S1—KT S2	$t(52) = -1.3688$, not significant	TQ S1—TQ S2	$t(52) = -5.0945$ ($p < 0.001$)
KM S1—KMS2	$t(52) = -0.4424$, not significant	GG S1—GG S2	$t(52) = -2.2209$ ($p < 0.05$)
TG S1—TG S2	$t(52) = -3.6996$ ($p < 0.001$)	GQ S1—GQ S2	$t(52) = -6.0262$ ($p < 0.001$)

comparisons with Tukey tests with a 95% family-wise confidence level. Figure 4 shows two Tukey 95% confidence interval plots with the mean differences of writing speeds for Sessions 1 and 3. If an interval does not contain zero, the corresponding means are significantly different. As shown in Figure 4, this happens several times. In general, there are statistically significant differences between the keyboards and remote pointing devices. First, the traditional keyboard is significantly faster than the touchpad and gyroscopic methods ($p < 0.001$ in both sessions). Second, the differences between the mini keyboard and the touchpad and gyroscopic methods change slightly from Session 1 to Session 3. In the first session there are no significant differences between touchpad methods and the mini keyboard, but this changes in the third session ($p < 0.01$). A similar situation happens with the gyroscopic methods; although there are significant differences in the first session ($p < 0.05$), they increase in the third session ($p < 0.001$). This is caused by the increased level of complexity of the sentences used in the third session. If we compare both keyboards, we can also see that there are significant differences between them ($p < 0.01$). The traditional keyboard is much faster than the mini keyboard. Finally, we have not found any significant difference between the touchpad and the gyroscopic methods.

Previous work shows that older users are usually slower, especially in selecting and tapping keys (Bosman, 1993; Taveira & Choi, 2009). In order to check the influence of age on writing speed, we have performed several analyses. Figure 5 shows the empirical writing speeds per age obtained in Sessions 1 (left) and 3 (right). The correlation between age and writing speed in Session 1 is relatively high ($-0.7199 < r < -0.6171$), negative, and significant ($p < 0.05$). This means that the older the user is, the slower he

or she is. This is also the case in Session 3 ($-0.7314 < r < -0.6017$; $p < 0.05$). Thus, we can conclude that, in general, age is a major factor affecting entry speeds and should be taken into account in theoretical studies.

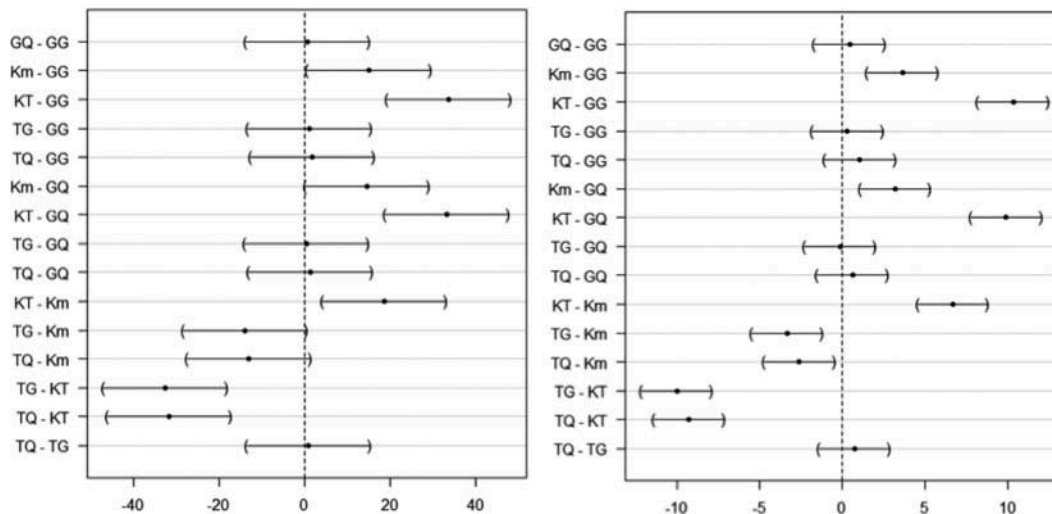
We have seen that, in general, older users are slower than younger users. If we analyze age groups separately, we can observe that this sentence still holds true. Table 3 shows average and standard deviations of writing speed per method and age group for Sessions 1 and 3. As we can see, younger users are faster than adult users and adult users are also faster than older users. Pairwise comparisons of average speeds show that young users and adult users are faster than older users with statistically significant differences ($p < 0.001$) in all sessions. Also, young users are faster than adult users, and this difference is statistically significant ($p < 0.05$) but for TG and GQ in Sessions 1 and 3 and except for KT and KM in Session 3.

Moreover, in the results we have seen that there are statistically significant differences between male and female users, men being faster than women in certain techniques. As shown in Table 4, in Session 1 men are faster than women with all the methods ($p < 0.05$) except the mini keyboard. However, in Session 3 the differences are obtained in all the methods apart from those using the touchpad.

Regarding handedness, we have not found any difference between strictly right-handed users and the rest of the participants.

4.2. Error Rate

The data gathered during the sessions of the experiment have allowed us to analyze error rates for each technique. Average error rates for each method and session are shown in Figure 6. As shown in the figure, error rates are similar for all the methods when simple texts are used, but when we require users to write complex texts, the number of errors obtained with physical keyboards increases considerably. With *t*-tests it is possible to analyze the evolution of error rates for each method from Session 1 to Session 2. Table 5 shows that there are no significant

**Figure 4.** Pairwise comparisons of average text entry speeds in Sessions 1 (left) and 3 (right) (95% family-wise confidence level).

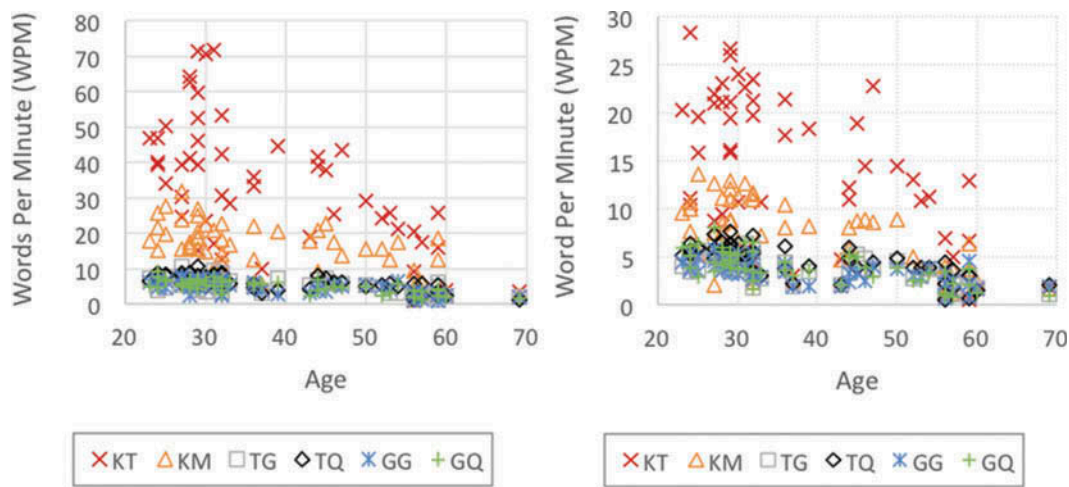


Figure 5. Overlay scatter plot of writing speeds per age and method during Sessions 1 (left) and 3 (right).

Table 3. Mean and standard deviations of writing speeds per method and age group.

	Session 1			Session 3		
	Young	Adult	Older	Young	Adult	Older
KT	$M = 45.25, SD = 15.26$	$M = 31.43, SD = 16.61$	$M = 16.98, SD = 12.12$	$M = 18.57, SD = 5.87$	$M = 13.23, SD = 7.67$	$M = 7.93, SD = 6.51$
KM	$M = 20.61, SD = 4.78$	$M = 15.23, SD = 5.65$	$M = 9.81, SD = 6.23$	$M = 9.23, SD = 2.94$	$M = 7.25, SD = 3.30$	$M = 4.08, SD = 2.75$
TG	$M = 6.38, SD = 1.75$	$M = 5.66, SD = 1.62$	$M = 3.41, SD = 1.71$	$M = 4.68, SD = 0.76$	$M = 3.74, SD = 1.4$	$M = 2.51, SD = 1.26$
TQ	$M = 7.37, SD = 1.45$	$M = 5.79, SD = 1.76$	$M = 4.07, SD = 1.84$	$M = 5.81, SD = 0.81$	$M = 4.3, SD = 1.45$	$M = 2.92, SD = 1.32$
GG	$M = 5.84, SD = 1.31$	$M = 4.58, SD = 1.52$	$M = 3.31, SD = 1.73$	$M = 4.31, SD = 0.74$	$M = 3.13, SD = 0.96$	$M = 2.53, SD = 1.15$
GQ	$M = 6.05, SD = 1.14$	$M = 5.08, SD = 1.49$	$M = 3.2, SD = 1.42$	$M = 4.99, SD = 1.22$	$M = 3.86, SD = 1.35$	$M = 2.41, SD = 1.14$

Table 4. Entry speeds and results of t-tests for gender groups in sessions 1 and 3.

	Session 1			Session 3		
	p-Value	Women	Men	p-Value	Women	Men
KT	0.0392*	$M = 26.33, SD = 18.13$	$M = 37.13, SD = 18.20$	0.0111*	$M = 10.56, SD = 8.20$	$M = 16.10, SD = 6.89$
KM	0.0924	$M = 13.80, SD = 8.56$	$M = 17.14, SD = 5.47$	0.0424*	$M = 5.88, SD = 3.82$	$M = 7.95, SD = 3.32$
TG	0.0243*	$M = 4.49, SD = 2.17$	$M = 5.81, SD = 1.89$	0.0693	$M = 3.32, SD = 1.72$	$M = 4.05, SD = 1.12$
TQ	0.0369*	$M = 5.18, SD = 2.48$	$M = 6.43, SD = 1.72$	0.184	$M = 4.12, SD = 2.13$	$M = 4.75, SD = 1.24$
GG	0.0217*	$M = 4.03, SD = 1.96$	$M = 5.19, SD = 1.57$	0.0142*	$M = 2.95, SD = 1.25$	$M = 3.77, SD = 1.07$
GQ	0.0457*	$M = 4.32, SD = 2.14$	$M = 5.32, SD = 1.36$	0.033*	$M = 3.31, SD = 1.75$	$M = 4.28, SD = 1.43$

changes. The increase in error rates in Session 3 is caused by the increased level of complexity of the sentences.

In order to check whether the differences between each method have any statistical significance, we have performed pairwise comparisons with Tukey tests with a 95% family-wise confidence level. The results of these comparisons show that there are several statistically significant differences between the methods. In Session 1, both gyroscopic methods are significantly worse than the traditional keyboard ($p < 0.039$). Also, the GQ method is significantly worse than both touchpad methods ($p < 0.029$). This changes in Session 3, in which error rates are significantly higher for physical keyboards than for the rest of the methods ($p < 0.045$).

Furthermore, we have observed that age is a major factor affecting average error rates in all the sessions. In general, the older the user is, the higher the number of errors he or she makes. Nevertheless, this is only significant for some of the methods. For instance, Figure 7 shows pairwise comparisons of error rates per age group and method in Session 1. Older users make more errors than young users with statistically significant differences ($p < 0.05$) with the TG, GG, and GQ methods

in Sessions 1–3, and with the TQ in Session 3. Moreover, in Session 3 pairwise comparisons of error rates show that adult users also make more errors than young users with statistically significant differences ($p < 0.01$) with the GQ method.

In addition, in the results we have seen that there are no statistically significant differences between male and female users. On the other hand, Table 6 shows that there are differences between right-handed and left-handed participants with the KT, KM, and TG methods in Session 1 and with the KT, KM, and GQ methods in Session 3.

4.3. Users' Point of View

As we have commented previously, users had to fill in a final web questionnaire after the last session in which they had to give, in their own words, their opinion about each technique. Their answers gave us some feedback and allowed us to provide some interesting remarks, as follows:

- Regarding the traditional keyboard, 95% of the participants used the keyboard on their lap. Also, 50%

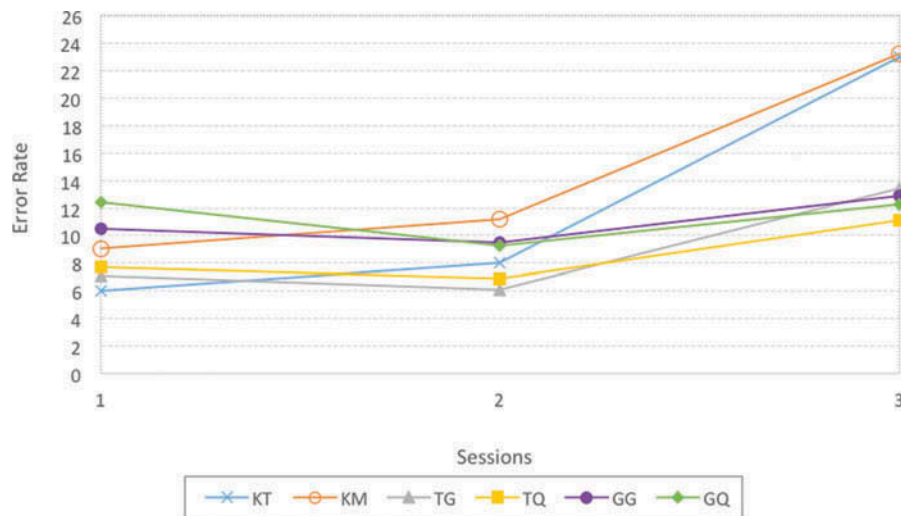


Figure 6. Average error rates per session and method.

Table 5. Results of *t*-tests for error rates in sessions 1 and 2.

KT S1—KT S2	$t(52) = -1.1241$, not significant	TQ S1—TQ S2	$t(52) = 0.4604$, not significant
KM S1—KM S2	$t(52) = -0.2681$, not significant	GG S1—GG S2	$t(52) = 0.4006$, not significant
TG S1—TG S2	$t(52) = 0.4092$, not significant	GQ S1—GQ S2	$t(52) = 1.2921$, not significant

used all the fingers to press the keys, 14% used the index finger and thumb, and 24% used only the index finger. Moreover, although the majority of the users considered the traditional keyboard as the fastest technique, 35% of them explicitly indicated that it was

very uncomfortable to use in a TV context. This is consistent with the conclusions of Lee and Schmidt (2005). Other users reported that lack of light made it difficult to see the keys on the keyboard. Finally, some participants, who do not use computers frequently, reported problems when writing symbols that require pressing more than one key at a time (17% of the users).

- With regard to the mini keyboard, 55% of the users held the keyboard with both hands and pressed the keys with both thumbs. However, 19% of them held the keyboard

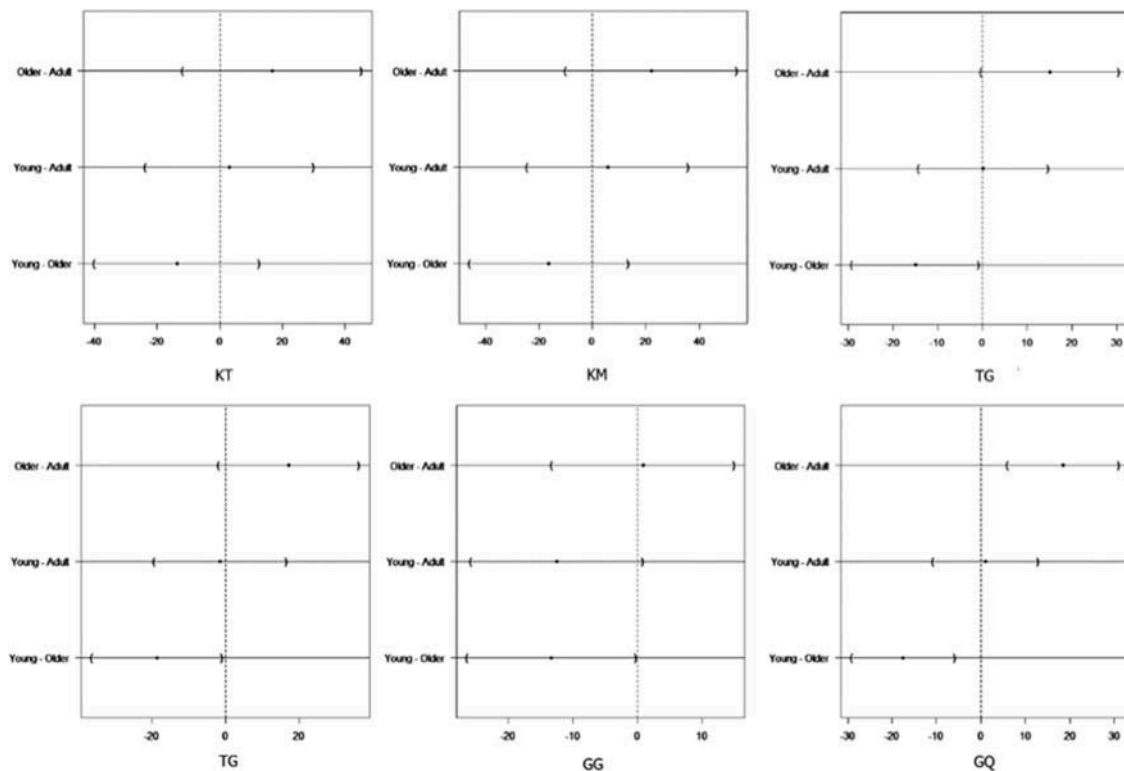


Figure 7. Pairwise comparisons of error rates per age group and method in Session 1 (95% family-wise confidence level).

Table 6. Error rates and results of *t*-tests for handedness groups in sessions 1 and 3.

	Session 1			Session 3		
	<i>p</i> -Value	Left-handed	Right-handed	<i>p</i> -Value	Left-handed	Right-handed
KT	0.01351	$M = 2.51, SD = 1.62$	$M = 6.44, SD = 9.37$	6.485e-05	$M = 2.89, SD = 3.34$	$M = 25.66, SD = 34.15$
KM	0.000208	$M = 0.88, SD = 1.77$	$M = 10.14, SD = 14.9$	0.001124	$M = 4.68, SD = 4.78$	$M = 25.69, SD = 39.03$
TG	0.004429	$M = 1.47, SD = 1.99$	$M = 7.81, SD = 13.3$	0.9033	$M = 12.67, SD = 14.79$	$M = 13.51, SD = 19.25$
TQ	0.5753	$M = 10.47, SD = 12.01$	$M = 7.41, SD = 11.49$	0.9628	$M = 10.88, SD = 9.95$	$M = 11.14, SD = 24.38$
GG	0.9863	$M = 10.62, SD = 14.74$	$M = 10.50, SD = 16.46$	0.3778	$M = 18.73, SD = 21.77$	$M = 12.09, SD = 16.61$
GQ	0.2792	$M = 21.24, SD = 19.86$	$M = 11.27, SD = 12.10$	0.002623	$M = 4.41, SD = 2.89$	$M = 13.29, SD = 17.19$

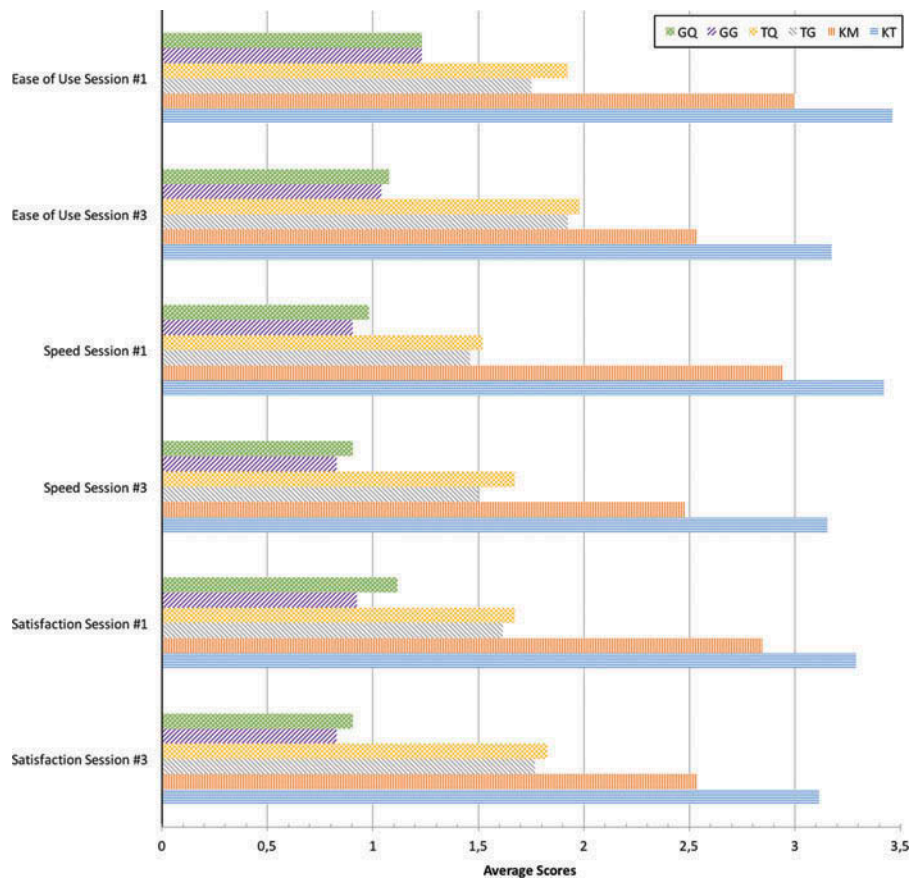
with the left hand and pressed the keys with a finger of the right hand. Finally, 12% of the users used the keyboard on their lap and pressed the keys with their index fingers. Also, the mini keyboard presented additional problems to the traditional keyboard because of its size: 36% of the users complained about the size of the keyboard, the position of certain keys, and/or the size of each individual key. Finally, some of the users reported problems when writing symbols that require pressing more than one key at a time (26%) as happened with the traditional keyboard.

- Fatigue problems with the gyroscopic device during the test sessions were reported by 48% of the participants. Despite the fact that 90% of the users chose the *Gun* mode instead of the *Remote* mode in order to avoid wrist and hand fatigue, 45% of the users still reported these problems during the experiment for both virtual layouts. This is consistent with previous work (Nacke,

2010). Moreover, 57% of the users complained about problems when pointing (focus lost) or clicking, as in König et al. (2009).

- With regard to the TouchPad device, 64% of the users reported that they held it with both hands, using the right thumb to select the key and the left thumb to confirm the selection. Also, 55% of the users stated explicitly that this was a simple and easy device to use. Some of them even stated that “this is the best method for old people or people with no typing experience “or” this is the best method as remote control.”

Apart from the final questionnaire, users had to rate three aspects of each of the methods: ease of use, speed of use, and overall satisfaction. Figure 8 shows the mean scores for these aspects for each of the methods used in the experiment, in Sessions 1 and 3. As we can see, users rank physical keyboards higher than the rest of the methods. In second place we find

**Figure 8.** Mean scores of subjective impressions per method in Sessions 1 and 3.

the touchpad with both layouts and the gyroscopic device obtains the lowest scores. In Figure 8, we can also see that scores decrease, in general, from Session 1 to Session 3. This coincides with the worse results obtained in this session, due to the increase in the complexity of the texts. Nevertheless, there is an exception with the touchpad, which increases its scores from Session 1 to Session 3 with both layouts. This indicates that the impression of the users with this device is not affected by the complexity of the texts (or of the layout supporting the increased number of symbols). This is in line with the aforementioned statements by some users about this device.

5. Conclusions and Future Work

The increasing impact of IDTV on our daily lives, together with the design of traditional remote controls, reveals the need for further effort in finding effective text entry methods for this context. It is increasingly evident that the popularity of devices to interact with televisions different to conventional remote controls will increase in the near future (Anido, Valladares, Fernandez-Iglesias, Rivas, & Gomez, 2013). Thus, after our research into text input methods based on traditional remote controls (Barrero et al., 2014), we have evaluated the performance and user acceptance of methods designed for other types of devices. With the same philosophy as in our previous research, these text entry methods may be used with commonly available technology that most people will more than likely already have in their homes.

In general, our experiments show that the best results, in terms of writing speed, are obtained with the conventional keyboard. Nevertheless, the number of errors under real conditions is very high, because users have problems when pressing more than one key at a time while holding the keyboard. Also, a high number of users reported discomfort and problems under semidarkness conditions. Considering entry speeds, the second method in the ranking is the mini keyboard, but it also registers high error rates when special symbols need to be written. Furthermore, a considerable amount of users complained about the size of the keyboard and the size of the keys. The entry speed is slightly better for the touchpad than for the gyroscopic remote and error rates are similar for both devices even when complex symbols are used. Nevertheless, a large number of users reported fatigue problems when using the latter. They also complained about problems when focus was lost. Despite the low speed of the touchpad, we obtained fairly positive feedback for this device. More than half of the users explicitly stated that it was simple and easy to use. Another conclusion that may be drawn is that age is a major factor affecting entry speeds and error rates. This is consistent with previous works of our research group (Barrero et al., 2014).

Future work will be based mainly on experiments with new or modified devices. We would like to check the performance obtained with devices that have been designed for other purposes, but may be used to interact with televisions. This can be the case of smartphones or tablet devices. The main reason is that these devices are already available in most

homes, so their usage in this context is totally feasible. Also, given the fact that we achieved good entry speeds with physical keyboards, in this new iteration we would like to further experiment in order to see whether we can reduce error rates or not. For instance, we could use sticky instead of quasi-mode modifiers, or experiment with a full-sized keyboard with backlight. Also, and after the results reported in this article, we would like to design, develop, and test a device specifically to interact with IDTV applications, based on a touchpad. Finally, we would also like to measure other factors that may influence the users' experience. By incorporating further equipment, we would like to measure stress levels, brain activity, the effect on eyesight, and similar parameters, which may allow us to draw even more interesting conclusions.

Funding

This work was partially supported by the University of Oviedo and the Principality of Asturias through the project SV-PA-13-ECOEMP-75.

ORCID

David Melendi  <http://orcid.org/0000-0001-8251-5646>

References

- Anido, L. E., Valladares, S. M., Fernandez-Iglesias, M. J., Rivas, C., & Gomez, M. (2013). Adapted interfaces and interactive electronic devices for the smart home. In *Proceedings of 8th International Conference on Computer Science & Education (ICCSE)*. Colombo, Sri Lanka.
- Aoki, R., Maeda, A., Watanabe, T., Kobayashi, M., & Abe, M. (2010). Twist tap: Text entry for TV remotes using easy-to-learn wrist motion and key operation. *IEEE Transactions on Consumer Electronics*, 56(1), 161–168.
- Barrero, A., Melendi, D., Pañeda, X. G., García, R., & Cabrero, S. (2014). An empirical investigation into text input methods for interactive digital television applications. *International Journal of Human-Computer Interaction*, 30(4), 321–341.
- Bellman, T., & MacKenzie, I. S. (1998). A probabilistic character layout strategy for mobile text entry. In *Proceedings of Graphics Interface '98* (pp. 168–176). Toronto, Canada.
- Bosman, E. A. (1993). Age-related differences in the motoric aspects of transcription typing skill. *Psychology and Aging*, 8(1), 87–102.
- Brandtzæg, P. B., Heim, J., & Karahasanović, A. (2011). Understanding the new digital divide—A typology of Internet users in Europe. *International Journal of Human-Computer Studies*, 69(3), 123–138.
- Brewbaker, C. R. (2008). *Optimizing stylus keyboard layouts with a genetic algorithm: Customization and internationalization*. Department of Computer Science, Iowa State University, Ames, IA.
- Castellucci, S. J., & MacKenzie, I. S. (2008). Unigest: Text entry using three degrees of motion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3549–3554). Florence, Italy.
- Cesar, P., & Chorianopoulos, K. (2009). The evolution of TV systems, content, and users toward interactivity. *Foundations and Trends in Human-Computer Interaction*, 2(4), 373–395.
- Choi, S., Han, J., Lee, G., Lee, N., & Lee, W. (2011). RemoteTouch: Touch-screen-like interaction in the TV viewing environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 393–402). Vancouver, Canada.
- Clarkson, E., Clawson, J., Lyons, K., & Starner, T. (2005). An empirical study of typing rates on mini-QWERTY keyboards. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1288–1291). Portland, OR.

- Crawley, M. J. (2012). *The R book*. Hoboken, NJ: John Wiley and Sons.
- Enns, N., & MacKenzie, I. S. (1998). Touchpad-based remote control devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 229–230). Los Angeles, CA.
- Geleijnse, G., Aliakseyeu, D., & Sarroukh, E. (2009). Comparing text entry methods for interactive television applications. In *Proceedings of the 7th European Conference on European Interactive Television Conference* (pp. 145–148). Leuven, Belgium.
- Iatrino, A., & Modeo, S. (2007). Text editing in digital terrestrial television: A comparison of three interfaces. In G. Lekakos, K. Chorianopoulos, & G. Doukidis (Eds.), *Interactive digital television: Technologies and applications* (pp. 224–241). Hershey, PA: Idea Group Publishing.
- Ingmarsson, M., Dinka, D., & Zhai, S. (2004). TNT: A numeric keypad based text input method. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 639–646). Vienna, Austria.
- Jones, E., Alexander, J., Andreou, A., Irani, P., & Subramanian, S. (2010). GesText: Accelerometer-based gestural text-entry systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2173–2182). Atlanta, GA.
- Költringer, T., Isokoski, P., & Grechenig, T. (2007). TwoStick: Writing with a game controller. In *Proceedings of Graphics Interface 2007* (pp. 103–110). New York, NY.
- König, W. A., Gerken, J., Dierdorf, S., & Reiterer, H. (2009). Adaptive pointing–design and evaluation of a precision enhancing technique for absolute pointing devices. In T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, & M. Winckler (Eds.), *Human–computer interaction–INTERACT 2009* (pp. 658–671). Germany: Springer.
- Lee, E., & Schmidt, M. (2005). Revolutionizing TV with user-centered design and research. *User Experience*, 2(4), 10–13.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 140.
- MacKenzie, I. S., Lopez, M. H., & Castelluci, S. (2009). Text entry with the apple iphone and the Nintendo Wii. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Boston, MA.
- MacKenzie, I. S., & Soukoreff, R. W. (2002a). Text entry for mobile computing: Models and methods, theory and practice. *Human–Computer Interaction*, 17(2–3), 147–198.
- MacKenzie, I. S., & Soukoreff, R. W. (2002b). A model of two-thumb text entry. In *Proceedings of Graphics Interface 2002* (pp. 117–124). Toronto, Canada: Canadian Information Processing Society.
- MacKenzie, I. S., Zhang, S. X., & Soukoreff, R. W. (1999). Text entry using soft keyboards. *Behavior and Information Technology*, 18(4), 235–244.
- Marshall, D., Foster, J. C., & Jack, M. A. (2001). User performance and attitude towards schemes for alphanumeric data entry using restricted input devices. *Behavior and Information Technology*, 20(3), 167–188.
- Nacke, L. E. (2010). Wiimote vs. controller: Electroencephalographic measurement of affective gameplay interaction. In *Proceedings of the International Academic Conference on the Future of Game Design and Technology* (pp. 159–166).
- Obrist, M., Bernhaupt, R., & Tscheligi, M. (2008). Interactive TV for the home: An ethnographic study on users' requirements and experiences. *International Journal of Human–Computer Interaction*, 24(2), 174–196.
- Ren, G., & O'Neill, E. (2013). Freehand gestural text entry for interactive TV. In *Proceedings of the 11th European Conference on Interactive TV and Video* (pp. 121–130).
- Rick, J. (2010). Performance optimizations of virtual keyboards for stroke-based text entry on a touch-based tabletop. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology, UIST '10* (pp. 77–86). New York: ACM.
- Ritter, F. E., & Schooler, L. J. (2002). *The learning curve. international encyclopedia of the social and behavioral sciences*, (Oxford) (pp. 8602–8605).
- Rose, M. J. (1991). Keyboard operating posture and actuation force: Implications for muscle over-use. *Applied Ergonomics*, 22(3), 198–203.
- Silfverberg, M., MacKenzie, I. S., & Korhonen, P. (2000). Predicting text entry speed on mobile phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 9–16). New York, NY.
- Soukoreff, R. W., & MacKenzie, I. S. (2003). Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03* (pp. 113–120). New York, NY.
- Swanson, N. G., Galinsky, T. L., Cole, L. L., Pan, C. S., & Sauter, S. L. (1997). The impact of keyboard design on comfort and productivity in a text-entry task. *Applied Ergonomics*, 28(1), 9–16.
- Taveira, A. D., & Choi, S. D. (2009). Review study of computer input devices and older users. *International Journal of Human–Computer Interaction*, 25(5), 455–474.
- Varcholik, P. D., LaViola, Jr., J. J., & Hughes, C. E. (2012). Establishing a baseline for text entry for a multi-touch virtual keyboard. *International Journal of Human–Computer Studies*, 70(10), 657–672.
- Vega-Oliveros, D. A., Pedrosa, D. C., Pimentel, M. G. C., & De Mattos Fortes, R. P. (2010). An approach based on multiple text input modes for interactive digital TV applications. In *Proceedings of the 28th ACM International Conference on Design of Communication* (pp. 191–198).
- Wobbrock, J. O., Myers, B. A., & Aung, H. H. (2004). Writing with a joystick: A comparison of date stamp, selection keyboard, and edge-write. In *Proceedings of Graphics Interface 2004, GI '04* (pp. 1–8). Ontario, Canada.
- Wood, E., Willoughby, T., Rushing, A., Bechtel, L., & Gilbert, J. (2005). Use of computer input devices by older adults. *Journal of Applied Gerontology*, 24(5), 419–438.
- Zhai, S., Hunter, M., & Smith, B. A. (2000). The metropolis keyboard—an exploration of quantitative techniques for virtual keyboard design. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology* (pp. 119–128). San Diego, CA.

About the Authors

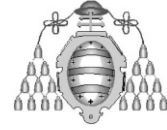
Aurora Barrero is a computer science engineer from the University of Oviedo and a PhD student, with an interest in the area of multimedia systems and services, content distribution networks, and interactive digital TV services. She is a research engineer in the Department of Computer Science at the University of Oviedo.

David Melendi is a computer science engineer with a PhD from the University of Oviedo, with an interest in multimedia systems and services, content distribution networks, IDTV services, and mobile ad hoc networks. He is an associate professor at the University of Oviedo and a member of the W3C.

Xabiel G. Pañeda is a computer science engineer with a PhD from the University of Oviedo, with an interest in multimedia systems and services, content distribution networks, IDTV services, and mobile ad hoc networks. He is an associate professor at the University of Oviedo and a member of the W3C.

Roberto García is a telecommunications engineer from The Technical University of Madrid with a PhD from the University of Oviedo, with an interest in telecommunication networks and services, applied to performance analysis, modeling and simulation of systems and services. He is an associate professor at the University of Oviedo.

Laura Pozuelo is a telecommunications engineer from the University of Oviedo with a PhD from the Spanish University for Distance Education (UNED). She is a contracted researcher in the group of Distributed Multimedia Systems and her current research interests are in the area of telecommunication networks and services.



7 Informe sobre la calidad de las publicaciones

En este capítulo se incluye información sobre la calidad y el factor de impacto de las publicaciones presentadas en el capítulo anterior.

Artículo 1

A. Barrero, D. Melendi, X. G. Pañeda, R. García, y L. Pozueco, «Evaluation of Text Entry Methods for Interactive Digital Television Applications with Devices Alternative to Conventional Remote Controls», *International Journal of Human-Computer Interaction*, vol. 32, n.o 10, pp. 765–776, 2016

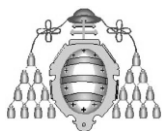
- **DOI:** 10.1080/10447318.2016.1195586
- **Estado:** Publicado
- **Factor de impacto** (JCR* 2015): 1.26
- **Factor de impacto** (5 años*): 1.459
- **Categorías:**
COMPUTER SCIENCE, CYBERNETICS: 11/22(Q2) *
ERGONOMICS: 8/16(Q2) *
- **Link:** <http://dx.doi.org/10.1080/10447318.2016.1195586>

* Los datos de Factor de Impacto de 2016 no ha sido publicado en esta fecha por lo que se ha utilizado el último disponible.

Artículo 2

A. Barrero, D. Melendi, X. G. Pañeda, R. García, y S. Cabrero, «An Empirical Investigation Into Typing Errors in Interactive Digital Television Applications», *International Journal of Human-Computer Interaction*, vol. 31, n.o 3, pp. 210–225, 2015.

- **DOI:** 10.1080/10447318.2014.994195
- **Estado:** Publicado
- **Factor de impacto** (JCR 2015): 1.26
- **Factor de impacto** (5 años): 1.459



- **Categorías:**

COMPUTER SCIENCE, CYBERNETICS: 11/22(Q2)

ERGONOMICS: 8/16(Q2)

- **Link:** <http://www.tandfonline.com/doi/full/10.1080/10447318.2014.994195>

Artículo 3

A. Barrero, D. Melendi, X. G. Pañeda, R. García, y S. Cabrero, «An empirical investigation into text input methods for interactive digital television applications», *International Journal of Human-Computer Interaction*, vol. 30, n.o 4, pp. 321–341, 2014.

- **DOI:** 10.1080/10447318.2013.858461

- **Estado:** Publicado.

- **Factor de impacto** (JCR 2014): 0.850

- **Factor de impacto** (5 años): 1.291

- **Categorías:**

COMPUTER SCIENCE, CYBERNETICS: 16/24 (Q3)

- **Link:** <http://dx.doi.org/10.1080/10447318.2013.858461>

Artículo 4

A. Barrero, D. Melendi, X. G. Paneda, R. Garcia, L. Pozueco, y J. L. Arciniegas, «A research on typing methods for interactive Digital Television Applications», *IEEE Latin America Transactions*, vol. 13, n.º 11, pp. 3612–3620, 2015.

- **DOI:** 10.1080/10447318.2013.858461

- **Estado:** Publicado.

- **Factor de impacto** (JCR 2014): 0.436

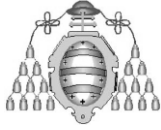
- **Factor de impacto** (5 años): 1.291

- **Categorías**

COMPUTER SCIENCE, CYBERNETICS: 133/144 (Q4)

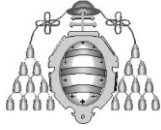
ENGINEERING, ELECTRICAL & ELECTRONI: 219/257(Q4)

- **Link:** <http://www.tandfonline.com/doi/full/10.1080/10447318.2016.1195586>



8 Otras Publicaciones de la Tesis

En este capítulo contiene las publicaciones de congresos que también han contribuido a la tesis doctoral.



8.1 Metodología

A. Barrero, D. Melendi, X. G. Pañeda, R. García, y S. Cabrero, «Metodología de Test de Usuario y Pruebas Subjetivas para Métodos de Inserción de Texto en Aplicaciones IDTV», presentado en Jornadas de Ingeniería Telemática: JITEL 2013, Granada (Spain), 2013.

Metodología de test de usuario y pruebas subjetivas para métodos de inserción de texto en aplicaciones iDTV

Aurora Barrero, David Melendi, Xabiel G. Pañeda, Roberto García, Sergio Cabrero

Departamento de Informática

Universidad de Oviedo

Campus de Viesques, SN 33204, Gijón, Asturias, España

{barrerourora, melendi, xabiel, garciaroberto, cabrerosegio}@uniovi.es

Resumen- En la actualidad las aplicaciones interactivas en televisión demandan una introducción de texto de manera similar a como se hace en el ordenador. En este trabajo proponemos una metodología ágil que nos permita establecer un proceso para realizar test de usuario y pruebas subjetivas en el ámbito de la inserción de texto en televisión digital interactiva. Esta metodología abarcará todas sus etapas: desde la definición de métricas, pasando por el desarrollo de prototipos y realización de las pruebas hasta concluir con el análisis de resultados. Su aplicación a un caso de estudio muestra la importancia de elegir tanto un mecanismo de escritura como los dispositivos de interacción adecuados, puesto que su elección influye sobre la velocidad de escritura y el número de errores que se comenten.

Palabras Clave- Metodología, test usuario, pruebas subjetivas, televisión, iDTV, escritura de texto.

I. INTRODUCCIÓN

El mercado de la televisión digital interactiva está en auge debido a que muchos proveedores de contenidos han apostado por una convergencia entre la WEB y la televisión. Esta situación, ligada al hecho de que el número de usuarios está en continuo crecimiento, ha provocado un clima de competitividad elevado entre los diferentes proveedores que ofrecen sus servicios multimedia interactivos mediante una televisión conectada o un set-top-box (STB). Las aplicaciones para ofrecer estos servicios son mucho más avanzadas (navegar por internet, leer el correo electrónico, ver vídeos de YouTube, escribir comentarios en nuestro Facebook, etc...) debido a que los telespectadores han pasado de tener un rol pasivo a tener un rol cada vez más activo. Con ello se ha conseguido dar un giro al tipo de aplicaciones interactivas en las que los usuarios demandan una experiencia de usuario (UX) similar a la de los ordenadores, especialmente en el tema de la introducción de texto.

Aunque en los últimos años se ha producido una importante proliferación de test de usuario y evaluaciones subjetivas, no existen muchos trabajos en los que se trate el tema específico de la escritura de texto en aplicaciones de televisión digital interactiva (iDTV). Esto es debido a la problemática que se genera a la hora de abordar, de manera eficiente, el elevado número de pruebas que es necesario hacer por cada tipo de experimentación, puesto que el tiempo y el coste de algunos recursos suele ser elevado. Por tanto, surge la necesidad de encontrar un proceso que facilite la ejecución de las tareas necesarias a la hora de realizar los test,

definir el entorno de pruebas, diseñar la interfaz de aplicación y analizar los resultados obtenidos. De la realización de este tipo de test de usuario y pruebas subjetivas, con usuarios reales, se pueden extraer conclusiones interesantes para el campo a estudiar: ¿Cuál es el mecanismo de inserción de texto óptimo?, ¿Cuál es el mejor dispositivo de interacción con iDTV?, ¿Cómo influye la edad de las personas?, ¿Cómo influye el nivel tecnológico de la persona?, ¿Está el usuario satisfecho con el método? etc.

En este artículo se presenta una metodología ágil e intuitiva, cuya aplicación será medir tanto la usabilidad, como la satisfacción de los usuarios en métodos de inserción de texto en iDTV. Se tendrán en cuenta cuestiones importantes como el diseño de la interfaz, las características del medio de interacción a utilizar y el entorno de pruebas.

El resto del artículo está organizado de la siguiente forma. En la Sección II se analizarán los trabajos relacionados. La sección III describe la metodología desarrollada. La sección IV relata cómo hemos aplicado la metodología en un caso práctico. Por último, en la sección V se presentan las conclusiones y trabajos futuros.

II. TRABAJOS RELACIONADOS

En sus orígenes la experiencia del usuario (UX) y la calidad de la experiencia (QoE) fueron promovidas por investigadores de Human-Computer Interaction (HCI). Desde entonces la usabilidad en iDTV ha sido un campo de investigación en continuo crecimiento en el que los investigadores hacen hincapié en la preocupación por los resultados de las personas que experimentan con tecnología.

Uno de los primeros trabajos en inserción de texto en iDTV fue presentado por Ingmarsson et al. [1] quienes diseñaron una nueva técnica llamada TNT. Se evaluó a cinco personas con edades comprendidas entre los 27 y 32 años. La prueba consistía en escribir una novela en sueco durante 10 sesiones de 45 minutos cada una. Los participantes recibieron una compensación económica.

Iatrino y Modeo [2] compararon 3 interfaces de introducción de texto para TDT (Digital Terrestrial Television). Participaron 36 personas que se dividieron en 2 grupos en función del nivel de experiencia en escribir SMS. La prueba consistió en realizar 6 tareas (2 con cada método) donde el orden de los métodos y de las tareas fue aleatorio para minimizar el efecto aprendizaje.

Por otra parte, Geleijnse et al [3] compararon tres métodos (Multitap, T9 y un teclado virtual) usando un mando a distancia con la escritura realizada con un teclado QWERTY convencional. El objetivo del experimento era buscar vídeos en YouTube escribiendo el par “artista-pista” utilizando cada uno de esos métodos. Participaron 22 usuarios (15 hombres y 7 mujeres) de edades comprendidas entre los 21 y 32 años. El experimento fue llevado a cabo en un laboratorio similar a una sala de estar.

Poco después, Aoki et al. [4] diseñaron un método de entrada de texto para mandos a distancia llamado Twist&Tap. Se evaluó a un conjunto de 14 mujeres diestras que fueron seleccionadas de una agencia de trabajo temporal. Su rango de edad estaba comprendido entre 24 y 33 años.

Gargi y Gossweiler [5] presentaron un nuevo sistema predictivo diseñado para mejorar la velocidad de escritura en teclados virtuales: QuickSuggest. Realizaron un estudio teórico del rendimiento y un experimento con 10 usuarios reales.

Más recientemente, Sporka et al.[6] realizaron un estudio comparativo de dos métodos: el ya conocido TNT [1] y TwiceTap que ellos desarrollaron. Participaron 18 usuarios (10 hombres y 8 mujeres) cuya media de edad fue 22.7 años. Los usuarios no eran expertos en introducir texto en iDTV pero estaban familiarizados con la escritura en teléfono móvil. El experimento fue organizado en 10 sesiones y cada una de ellas de unos 20 minutos de duración. El orden de los métodos fue contrabalanceado para minimizar el efecto aprendizaje.

Existen estudios con dispositivos diferentes al mando a distancia tradicional. Költringer et al. [7] compararon una técnica llamada Twostick con teclado QWERTY con el mando de una consola Xbox 360 Microsoft. En el experimento participaron 8 usuarios diestros (4 hombres y 4 mujeres), de entre 22 y 29 años. No tenían experiencia con el método Twostick pero si en el uso del teclado QWERTY.

Mackenzie et al. [8], compararon los teclados virtuales de un iPhone de Apple y de una Nintendo Wii. El estudio se realizó en un salón con ambiente relajado. Participaron 16 personas (9 hombres y 7 mujeres) cuyo rango de edad estaba comprendido entre los 18 y los 30 años. Fueron reclutadas del campus universitario local y se les dio una compensación económica por su colaboración.

Como se ha comprobado existen trabajos en los que se han realizado evaluaciones que tratan el tema específico de inserción de texto en aplicaciones iDTV, pero en ninguno de ellos se ha utilizado una metodología para su realización. No se han guiado por un proceso metodológico sino de forma intuitiva o por su propia experiencia. La mayoría realizan evaluaciones con usuarios reales pero seleccionan un grupo muy reducido de personas y/o con rangos de edad equivalentes y/o con un nivel tecnológico parecido. En definitiva con características muy similares. Tampoco suelen utilizar un entorno de pruebas equivalente al que un usuario tendría en su propia casa.

Hasta donde alcanza el conocimiento de los autores, no se tiene constancia de la existencia ninguna metodología específica para la realización de test de usuario y pruebas subjetivas para métodos de inserción de texto en aplicaciones iDTV. Este trabajo propone una metodología que indique a los experimentadores cómo tienen que proceder y qué tareas

deben realizar para alcanzar los objetivos que se planteen en un estudio.

III. METODOLOGÍA

La metodología que se propone establece un método que abarca todas las etapas de una experimentación, desde la definición de métricas y objetivos, pasando por el diseño de la misma, el desarrollo de prototipos y la realización de las pruebas de usuario hasta finalizar con la recogida y validación de datos y el análisis de resultados.

Esta metodología está basada en una serie de fases, que a su vez, se van a subdividir en una serie de tareas. Para definir todas las fases se han tenido en cuenta las recomendaciones y reglas de la iDTV y de la experiencia del usuario encontradas en la literatura.

El diagrama de la Fig. 1 muestra un flujo de trabajo formal de las fases y tareas propuestas en notación Business Process Model and Notacion (BPMN).

A. Fase 1. Definición de Métricas

En esta fase se definirán los objetivos y se especificarán las métricas que los satisfagan.

Una métrica aportará información sobre la interacción entre el usuario y el sistema de inserción de texto en iDTV. Se tendrán en cuenta aspectos como la efectividad (capacidad para completar una tarea), la eficiencia (la cantidad de esfuerzo requerido para completar una tarea) y la satisfacción (el grado con el que el usuario es feliz con su propia experiencia de realización de la tarea).

1) Definir Objetivos

Los objetivos marcarán lo que se quiere conseguir, y para tener una ayuda en su definición se responderá a preguntas del siguiente tipo: *¿Qué se va a estudiar?*; será necesario especificar un objetivo genérico. *¿Cuál es la motivación y el propósito?* es decir, precisar para qué se hará, (para evaluar, para mejorar...) y qué se pretende con ello (analizar los mecanismos de inserción de texto, evaluar los dispositivos, analizar un sector de la población...). *¿Desde qué punto de vista se definen los objetivos?*, por ejemplo, desde un punto de vista de usuario o desde un punto de vista investigador.

2) Estudiar Métricas

En esta tarea se tendrá que realizar un estado del arte para conocer las métricas y estándares a utilizar en la evaluación de mecanismos de inserción de texto aplicables a iDTV.

En el campo de la inserción de texto algunas métricas típicas son: velocidad de escritura, tasa de errores cometidos, satisfacción del usuario y curva de aprendizaje.

3) Especificar Métricas

Para poder evaluar el rendimiento de un usuario los objetivos y las métricas tienen que estar alineados. Así que, lo que se hará, en primer lugar, será transformar los objetivos en preguntas, y en segundo lugar, se buscará respuestas a esas preguntas, y a partir de ahí, se podrán especificar las métricas.

Es necesario establecer las unidades de medida de las métricas seleccionadas. Por ejemplo, la velocidad de escritura de un método de inserción de texto, puede medirse en WPM (palabras por minuto), CPS (caracteres por segundo) o KSPS (pulsaciones de tecla por segundo).

B. Fase 2. Diseño de la Prueba

En esta fase se diseñará la experimentación a utilizar para realizar los test de usuario y evaluaciones subjetivas.

1) *Seleccionar Tipos Dispositivos de Interacción*

En esta tarea se seleccionarán los diferentes tipos de dispositivos con los que el usuario interactuará durante la realización de la prueba (mandos a distancia, teclados, apuntadores giroscópicos, touchpads, etc.). Se detallarán las características específicas de las que deberán disponer (disposición de teclas, tamaño, ergonomía, conexión, etc.).

2) *Seleccionar Métodos de Inserción de Texto*

En la literatura existen una gran variedad de mecanismos a utilizar para la inserción de texto. Por un lado, existen teclados virtuales: QWERTY, Square Alphabetic, Metropolis, FOLC, ATOMIK, Genetic, Wide Alphabetic, etc. Por otro lado, existen otros mecanismos como los de tipo teléfono móvil: MULTITAP, T9, 2-KEY, TNT, TwoStick, etc.

En esta tarea será necesario realizar una selección de mecanismos que deben ser probados en base a los objetivos del estudio, los dispositivos seleccionados, los resultados publicados y la conveniencia del método para servicios de televisión digital interactivos.

3) *Seleccionar la Población*

Cuando diseñamos un estudio de usabilidad es necesario prestar especial atención a las características de la población. Por eso, se hará necesario responder a cuestiones como: ¿Qué tipo de participantes se necesitarán?, ¿Cuántos participantes?, ¿Se compararán los datos con un único grupo o con varios grupos de participantes?, ¿Se necesitará contrabalancear el orden de las tareas?

En cuanto a la selección de participantes, la primera cuestión a tener en cuenta es que estos deberán ser seleccionados cuidadosamente debido a que deben ser representativos de la población objetivo del estudio. Se deberá establecer un criterio de selección de la población por el cual se determinará si una persona concreta es elegible para participar en el estudio. El criterio debe ser tan específico como sea posible para reducir la posibilidad de reclutar a un participante que no dé el perfil.

La segunda cuestión será saber si vamos a dividir los datos por diferentes tipos de participantes. Si es así, se tendrán que separar a los participantes en distintos grupos y entonces habrá que pensar qué grupos son y cuantos participantes irán en cada grupo. Los grupos más comunes de segmentos [9] son:

- Nivel de experiencia en un dominio: (novato, intermedio, experto).
- Frecuencia de uso.
- Demográficos (genero, edad, lugar de procedencia).
- Actividades (uso de funcionalidades o características particulares).

La tercera cuestión será la estrategia a seguir para seleccionar la muestra. El objetivo de un estudio de usabilidad es generalizar los “hallazgos” a una población mayor. Es importante conocer cómo de bien la muestra refleja a la población general y será necesario estar al tanto de cualquier posible sesgo que se produzca en los datos. Para localizar a los participantes se pueden emplear métodos como poner un anuncio o usar una lista de gente que ha participado en otras pruebas.

Respecto al tamaño de la muestra, no existe una regla que diga el número mínimo de participantes de un estudio. Pero el tamaño de la muestra se deberá escoger en base a dos factores: los objetivos del estudio y la tolerancia al margen de error. Se recomienda tener de 50 a 100 usuarios representativos [9], ya que si escogemos un número menor, la varianza en los datos será más alta, haciendo difícil generalizar los resultados a la población objeto del estudio. En el caso de que la población sea agrupada será de ayuda tener al menos 4 participantes de cada grupo.

A veces, el orden en el que los usuarios realizan las tareas del estudio, tiene un impacto significativo en los resultados debido a que estos aprenden con la experiencia de uso. Si se sospecha que puede haber un efecto en el orden de realización de las tareas que influya en el aprendizaje de otra, se recomienda realizar un contrabalanceado.

Contrabalancear involucra cambiar simplemente el orden en el que diferentes tareas son realizadas. Se puede asignar aleatoriamente el orden de las tareas para cada participante, o crear previamente grupos con el orden de las tareas diferente y asignar a cada usuario a un grupo. Sin embargo, hay algunas situaciones en las que no es necesario o incluso pueden ser perjudicial para los objetivos del estudio. Un ejemplo sería, si las tareas no tienen ninguna relación entre sí, es poco probable que se dé el aprendizaje entre ellas, es decir, la realización de una tarea con éxito no va a ayudar a ninguna otra tarea. Otro ejemplo, que se puede dar, es cuando el orden de las tareas es inamovible porque, en ese caso, la sesión de prueba no tendría sentido. En esta situación habría que conocer el orden de los efectos como parte general del proceso de aprendizaje y no pensar que es un síntoma de un estudio de usabilidad mal diseñado.

4) *Diseñar el Proceso de Evaluación*

En esta tarea se establecerá el proceso de evaluación y las experimentaciones que se realizarán. También será necesario establecer el número de sesiones de cada experimentación en las que participará un usuario y el tiempo mínimo y máximo entre sesiones. Además, se concretará el número de frases y los textos que deberán escribir los usuarios, así como las encuestas y escalas de votación a las que se someterá cada uno de ellos. Consecuentemente, se establecerán los datos necesarios que deberán ser recabados para calcular las métricas que se han definido.

El número de sesiones de una experimentación dependerá de los objetivos y de las métricas especificadas. Puede suceder que las sesiones tengan la misma complejidad, se evalúen los mismos métodos y/o mismos dispositivos y se quiera medir la curva de aprendizaje [10] de los participantes. O puede suceder, que las sesiones sean independientes y en cada una de ellas se mida el rendimiento de un método o dispositivo diferente. También será necesario establecer el rango de duración de una cada sesión y el tiempo mínimo y máximo entre cada una de ellas.

Normalmente habrá que diseñar el corpus de frases que se utilizarán en el estudio y que serán distribuidas entre las distintas sesiones. La complejidad de las frases dependerá de los objetivos del estudio. Podrán contener minúsculas, y/o mayúsculas, y/o signos de puntuación, y/o números, y/o caracteres especiales, etc. Dependiendo del caso, se podrán seleccionar de algún medio (periódico, artículos de internet [11], etc.) o ser diseñadas a medida para el estudio [12]. El

orden en que las frases aparecerán durante una sesión tendrá que ser establecido. Por ejemplo, podrá ser aleatorio o tener un orden predeterminado.

En la mayoría de los casos será conveniente realizar una encuesta previa para conocer a la población. Las preguntas suelen ser sobre cuestiones demográficas, tecnológicas, de habilidades, de necesidades y hábitos. La encuesta se puede diseñar en formato WEB o papel. Es recomendable hacerla antes de la primera sesión de pruebas.

También es fundamental hacer encuestas de satisfacción, de preferencias y/o deseos del participante, ya que es muy importante conocer la información de la percepción de los usuarios del sistema y su interacción con él. Para recoger estos datos subjetivos, se suelen realizar encuestas mediante escalas de votación y/o cuestionarios donde los participantes describen con sus palabras como se han sentido y qué problemas se han encontrado.

Para las escalas de votación se recomienda la escala de Likert [13]. Esta propone una serie de preguntas y en las repuestas se especifica el nivel de acuerdo o desacuerdo con la pregunta. El formato típico de 5 niveles de respuesta sería: Totalmente en desacuerdo, En desacuerdo, Ni de acuerdo ni en desacuerdo, De acuerdo y Totalmente de acuerdo.

Los cuestionarios con escala de votación suelen hacerse al final de cada sesión de pruebas y suelen ir incorporadas en la aplicación de pruebas. Los cuestionarios en los que los usuarios describen sus impresiones suelen realizarse en la última sesión del estudio y su formato suele ser WEB o papel.

5) *Diseñar el Prototipo*

En esta tarea se diseñará la aplicación que permitirá a los participantes realizar la prueba y recogerá los datos necesarios que permitirán calcular las métricas seleccionadas.

Para la creación de una interfaz adaptada al entorno de la televisión será necesario utilizar reglas y recomendaciones para iDTV, ya que no debe de hacerse al azar. La norma ISO 14915-3:2002 [14] proporciona recomendaciones y orientaciones para el diseño, selección y combinación de interfaces de usuario interactivas que integran y sincronizan diferentes medios. Algunos autores como Perrinet al. [15] señalan algunas recomendaciones respecto a colores, tamaño del texto, elementos de navegación o número de palabras por pantalla. Y Graham [16] mantiene que algunas de las Leyes de Gestalt, como las leyes de figura/fondo, proximidad, similaridad y simetría se pueden aplicar al diseño de la interfaz para aplicaciones iDTV.

Del mismo modo, será necesario seleccionar tanto la metodología de desarrollo como el lenguaje de programación, así como el formato de recogida de datos (por ejemplo, si se utilizará una base de datos o un fichero XML) para satisfacer los objetivos marcados y las métricas seleccionadas.

Respecto a la funcionalidad del prototipo se diseñarán ciertas cuestiones tales como disposición y/o combinación de teclas y qué métodos de alternancia utilizar (por ejemplo, mayúsculas y minúsculas) para cada uno de los dispositivos seleccionados. También serán importantes cuestiones relativas a cómo capturar los eventos de pulsación de teclas, cómo mostrar listados de desambiguación y si se implementarán o no métodos de recomendación.

Asimismo, se tendrán en cuenta las escalas de votación que irán incorporadas en la aplicación y el método de

propuesta de frases diseñado en la tarea de “Diseñar el Proceso de Evaluación”.

6) *Diseñar el Entorno de Experimentación*

Normalmente, los espectadores suelen interactuar con la televisión en un ambiente relajado y la distancia a esta suele ser de unos pocos metros. Esta separación espacial provoca que los usuarios tengan que alternar el foco de atención entre el dispositivo de interacción y la televisión.

Para que el resultado de la experimentación sea lo más realista posible, será necesario realizarlo en unas condiciones controladas y lo más parecidas posible a un entorno real. Así pues, en esta tarea se diseñará lo referente al entorno de experimentación. Será necesario establecer cuestiones tales como las relativas a qué características tiene que tener la sala donde se realizarán las experimentaciones: iluminación, mobiliario y recursos hardware (por ejemplo se necesitará una Televisión, qué características tiene que tener, cuantas pulgadas, qué resolución será necesaria, etc). En definitiva todo lo relacionado con los recursos del entorno.

En relación a los recursos hardware se necesitará al menos, una televisión y un ordenador. Por un lado, se seleccionarán las características de la televisión (pulgadas, resolución, distancia entre el espectador y la TV, etc.). Por otro lado, los requisitos hardware y software del ordenador tendrán que ir en consonancia con las necesidades del software a desarrollar.

Para la selección de los parámetros relacionados con iluminación de la sala, distancia de observación, formato de pantalla, resolución, pulgadas etc. es aconsejable utilizar las recomendaciones de la ITU-R BT.500-1 [17] que cubren todos estos aspectos.

El resultado de esta tarea generará un listado con todos los recursos necesarios para establecer un entorno apropiado para la realización de los test.

C. *Fase 3 Adquisición de Equipamiento*

1) *Adquirir Equipamiento Específico de la Prueba*

Se realizará un estudio de los distintos productos disponibles en el mercado para los tipos de dispositivos seleccionados en la fase anterior. Una vez concluido el estudio se procederá a la adquisición de los mismos.

2) *Adquirir Equipamiento del Entorno de Pruebas*

La tarea de Diseño del Entorno de experimentación nos proveerá el listado de equipamiento necesario. Se realizará un estudio de mercado para cada punto del listado, y se procederá a la adquisición del equipamiento.

D. *Fase 4. Desarrollo del Prototipo*

1) *Desarrollar el Prototipo*

En esta tarea se realizará la implementación de la aplicación de pruebas. Como entradas de la tarea se utilizarán el diseño del prototipo, las recomendaciones iDTV y la metodología de desarrollo de software elegida.

La aplicación será el medio para realizar las pruebas y medirá tanto la usabilidad, como la satisfacción de los usuarios de métodos de inserción de texto en iDTV.

2) *Adaptar el Prototipo*

Esta tarea solo será necesaria en caso de que de la fase de mejora continua se haya identificado una mejora y sus acciones correctivas impliquen la modificación de alguna funcionalidad del prototipo desarrollado.

E. Fase 5. Recogida de Datos

1) Realizar Adiestramiento

Para obtener resultados fiables, en algunos casos será necesario enseñar a los participantes mediante una sesión especial de adiestramiento. Esta se realizará con antelación a las sesiones de prueba.

Las frases utilizadas en la sesión de adiestramiento deberán ser diferentes de las utilizadas en la prueba. Sería recomendable la utilización de todos los dispositivos y/o métodos que se usarán en la prueba.

Un caso de utilización de adiestramiento sería, por ejemplo, en participantes de edades avanzadas y/o de perfil no tecnológico donde el uso de los métodos de inserción de texto y/o de los dispositivos puede ser una labor compleja.

2) Realizar la Prueba

Para llevar a cabo la evaluación, se concertarán citas con los usuarios que participarán en el experimento. Cuando el usuario acuda una sesión, se le darán indicaciones sobre lo que tendrá que hacer para interactuar con la aplicación: qué métodos de inserción de texto utilizará, con qué dispositivos interactuará, cuantas frases tendrán que escribir, cómo podrán corregir errores y cómo se tiene que rellenar la escala de votación. Estas indicaciones serán verbales y podrán ser reforzadas con una presentación de ordenador. También se recomienda hacer una demostración práctica.

Una vez que el participante realice la prueba, sus datos quedarán registrados para su posterior evaluación.

F. Fase 6. Análisis de Datos

1) Validar Datos

Una vez que se hayan recogidos los datos, se comprobará su validez, mediante el análisis de la información, verificando que sean correctos, estén completos y sean consistentes. Normalmente se realizará un pre-procesado de los mismos, que generará el conjunto de datos final, que se introducirá en el programa estadístico escogido para su análisis.

En el pre-procesado del conjunto de datos obtenido, será necesario realizar tareas tales como *Eliminar Datos Perdidos* (datos que por alguna razón no son conocidos), *Discretizar Datos* (convertir variables reales o enteras en nominales), *Filtrar Atributos y/o Eliminar Ruido*. Una vez realizadas estas tareas tendremos el conjunto de datos preparado para su análisis estadístico.

2) Analizar Resultados

Tras concluir las pruebas y una vez que se hayan valido los datos registrados, se realizará el análisis de los mismos y se elaborará la documentación necesaria.

El tipo de análisis estadístico, dependerá de la experimentación e irá en consonancia con los objetivos y métricas del estudio. Un ejemplo de operativa de análisis de datos podría ser la que sigue. En primer lugar, se comprobará la normalidad de los datos con test de Saphiro-Wilk y su homocedasticidad con test de Barlett. Dependiendo de estas características, se utilizarán Tests de ANOVA o de Kruskal-Wallis y, en caso de existir diferencias entre los datos que se estén analizando se utilizarán los test de Tukey con un coeficiente de confianza del 95% para realizar comparaciones dos a dos.

3) Extraer Conclusiones

Es esta tarea, una vez analizados los datos estadísticamente, se interpretarán sus resultados. Se

recomienda generar un informe en el que se incluirán todas las observaciones, interpretaciones y conclusiones derivadas del análisis estadístico y de los comentarios subjetivos de los participantes.

G. Fase 7. Mejora Continua

En esta fase será necesario evaluar el estudio, dejar constancia de las lecciones aprendidas y seleccionar las acciones a realizar para que las distintas fases de la metodología se apliquen de forma eficiente. Así se obtendrá un mayor beneficio y los resultados obtenidos estarán alineados con los objetivos que se persiguen.

1) Analizar el Estudio

Después de efectuar una experimentación se realizará una evaluación de todo el proceso.

En primer lugar, es necesario aprender del proceso en sí, tanto de sus éxitos como de sus fracasos, y es necesario que quede reflejado para que sea una lección aprendida a tener en cuenta la siguiente vez que se realice un estudio.

En segundo lugar, hay que comprobar si con los resultados obtenidos se han alcanzado los objetivos esperados. Y si es el caso, habrá que encontrar cuál ha sido el motivo por el cual los objetivos y los resultados no están alineados.

En tercer lugar habría que identificar qué mejoras podremos aplicar al proceso. Por ejemplo, se puede mejorar el rendimiento o eficiencia de una fase o tarea. También puede suceder que a raíz del estudio surjan ampliaciones del estudio o estudios alternativos que complementen el estudio actual.

Para ayudarnos a identificar las mejoras se evaluarán los resultados obtenidos a partir de las conclusiones de la fase del análisis de datos y de los comentarios de los participantes que fueron recogidos en los cuestionarios.

2) Diseñar Acciones a Aplicar

En esta tarea se diseñarán las acciones a aplicar por cada una de las mejoras identificadas en la tarea anterior. Será necesario priorizarlas basándose en los objetivos del estudio, y valorar si son aplicables o no, puesto que en algún caso tendrán un coste (de recursos, de tiempo, etc.) demasiado elevado.

Un ejemplo de mejora sería ampliar la población objeto del estudio porque se ha detectado que la población está sesgada. Esto implicaría como acciones correctivas conseguir más participantes y realizar las pruebas con ellos para ampliar los resultados del estudio.

Otro ejemplo sería que, después de unos meses realizando las pruebas al hacer una revisión de los trabajos relacionados se encuentren nuevos dispositivos de interacción. Esto podría implicar como acciones a realizar añadir otra sesión para probar ese dispositivo y/o rediseñar el prototipo.

IV. CASO DE ESTUDIO

En esta sección se propone un caso de estudio resumido que ilustra la metodología propuesta.

A. Fase 1. Definición de Métricas

El principal objetivo de la experimentación es determinar la eficacia y la eficiencia de los diferentes métodos de entrada de texto, teniendo en cuenta diferentes perfiles de usuario y distintos contextos de uso. Con el fin de alcanzar este

objetivo, se eligieron varias métricas teniendo en cuenta, tanto los aspectos objetivos como subjetivos:

- Velocidad entrada de texto: palabras por minuto (WPM)
- Tasa de error: porcentaje de caracteres erróneos escritos
- Impresión subjetiva de la facilidad de uso, velocidad de escritura y satisfacción general: medidos con escalas de Likert de 0 a 4.

B. Fase 2. Diseño de la Prueba

1) Seleccionar Tipos Dispositivos

Se eligieron como dispositivos a utilizar 3 mandos a distancia que incorporan las teclas más comunes de los mandos a distancia convencionales: números, tecla OK y cursores. Estos serán diferentes entre sí, es decir, tendrán distinta ergonomía, tamaño y/o disposición de teclas.

2) Seleccionar Métodos de inserción de texto

Se eligieron para su evaluación dos métodos de inserción de texto virtuales (QWERTY y Genético) y un método tipo teléfono móvil Multitap.

3) Población

Participaron 42 usuarios, que se escogieron de forma heterogénea; con diferentes edades (comprendidas entre los 20 y los 70 años), de distinto género; así como, con diferentes habilidades tecnológicas y de diferente nivel de estudios.

Los usuarios se agruparon de acuerdo a su edad. Por lo tanto, varios usuarios pertenecen a la llamada “generación móvil”, con edades comprendidas entre los 18 y 30 años; otros pertenecen a la llamada “generación ordenador”, con edades comprendidas entre los 31 y 45 años. Y finalmente, los usuarios “pre-ordenador” que son mayores de 45 años.

4) Diseñar el Proceso de Evaluación

Se decidió que los participantes deberían completar tres sesiones, una por cada uno de los mandos a distancia escogidos. Durante una sesión (de unos 35 minutos) se le pedirá al usuario que introduzca 6 frases (que contendrán caracteres en minúsculas, mayúsculas, números, signos de puntuación y caracteres especiales), 5 serán datos personales (nombre y apellidos, DNI, fecha de nacimiento, dirección postal, correo electrónico) que ellos mismos habrán insertado en un formulario previo a la sesión, y otra que será una URI común para todos. El usuario tendrá que introducir las 6 frases (que se mostrarán en orden aleatorio) con cada uno de los 3 métodos de inserción de texto. Para evitar un “efecto aprendizaje”, se realizará un contrabalanceado, asignando a cada usuario un grupo distinto, lo que cambiará el orden de uso de cada uno de los métodos y se neutralizará ese efecto.

Con el fin de recopilar datos de los usuarios y la información subjetiva, se diseñaron varios cuestionarios. Antes de la primera sesión de la experimentación, los usuarios rellenarán un cuestionario que permitirá saber el sexo, edad, profesión, nivel y tipo de estudios, hábitos de uso de la televisión, del teléfono móvil y del ordenador, y si son zurdos o diestros. Además, en la última sesión y después de realizar la prueba, los usuarios completarán un cuestionario final en el que los participantes describirán con sus palabras cada método, cómo se sienten, si tienen la impresión de estar mejorando, etc.

Los participantes tendrán que votar mediante escalas de Likert los siguientes parámetros: facilidad de uso, velocidad de escritura, y satisfacción global. El sistema de votación se implementará en el prototipo y se realizará al final de cada una de las sesiones de prueba.

5) Diseñar el Prototipo

Se diseñó una aplicación que más tarde fue implementada con Adobe® AIR®. Las fuentes utilizadas serán Verdana 65 puntos para las frases mostradas y los campos de texto, y Arial 60 puntos para la distribución de los teclados.

La aplicación se visualizará a pantalla completa sobre fondo de color negro y mostrará los textos que se tienen que escribir con cada uno de los métodos de inserción de texto.

Cuando se inicie una sesión en la aplicación, se dejarán 5 segundos antes de permitir al usuario escribir. Una vez que el usuario escriba una frase, la aplicación comprobará si la frase escrita coincide exactamente con la frase propuesta. Si coinciden, la aplicación mostrará una nueva frase o cambiará a un método de entrada de texto diferente.

La aplicación se comportará de manera diferente dependiendo del tipo de método usado, teclado virtual o tipo teléfono móvil (SMS). Cuando se utilicen teclados virtuales, los usuarios utilizarán los cursores del mando a distancia para moverse a través de una distribución del teclado en pantalla y el botón "OK" para confirmar una acción. Es importante tener en cuenta que cada vez que un usuario quiera escribir algo, necesitará encontrar la tecla deseada, moverse desde la posición actual a esa tecla y pulsar el botón OK.

Para hacer frente a errores, la aplicación deberá proporcionar mecanismos para borrar los caracteres mediante una tecla del mando a distancia.

La aplicación procesará los eventos que se producen durante la ejecución en un fichero XML. Se grabarán tanto los textos propuestos, como lo que escriben los usuarios. Quedará constancia de todas las pulsaciones que se realicen. El propósito será el de disponer de información para su posterior análisis estadístico.

6) Diseñar el Entorno de Experimentación

Se decidió que durante una sesión del experimento, el participante se quedaría solo en una habitación para evitar distracciones. Para crear la situación más realista posible, cada usuario se acomodará como si estuvieran en su propia sala de estar. Para ello será necesario un sillón donde se sentará el participante, un ordenador para ejecutar la aplicación y una televisión donde esta se mostrará.

El sillón se colocará frente a una televisión de 32” con 1024 x 768 de resolución; y ésta se situará encima de una mesa de alrededor de 1 metro de altura. El sillón se ubicará a una distancia de entre 2 o 3 metros.

El ordenador será necesario para ejecutar la aplicación y llevar a cabo las pruebas. Este necesitará como sistema operativo Windows XP y como característica Hardware 1 Gb de Memoria RAM. Además tendrá que disponer de un puerto USB donde se conectarán los receptores de radio frecuencia e infrarrojos de los diferentes mandos a distancia.

Fase 3. Adquisición de Equipamiento

1) Adquirir Equipamiento Específico de la Prueba

Del tipo de mandos seleccionados se compraron los siguientes (Fig.2): Un primer mando a distancia modelo SnapStream Firefly con el teclado numérico en la zona superior y las flechas y la tecla OK en el centro. Un segundo mando a distancia que se eligió porque los números y las flechas están situadas a un lado, es decir, tiene los números en la parte superior izquierda y las flechas y la tecla OK en la parte centro-derecha. Se trata de un mando a distancia Golden Interstar. Y un tercer mando que se eligió para comprobar si

el tamaño es importante o no, ya que es notablemente más pequeño que los otros dos. Además, este mando permitirá comprobar el impacto de invertir el orden de sus elementos, ya que tiene los números en la parte inferior y las flechas y la tecla OK en la parte superior. Se trata de un mando a distancia Aver Media.



Fig. 2. Mandos a distancia usados en la experimentación.

2) Adquirir Equipamiento del Entorno de Pruebas

El hardware adquirido fue: una televisión LG LCD de 32" con dimensiones de 814x599mm y un ordenador Dell Pentium 4, 3 GHz con 1 Gb de RAM y sistema operativo Windows XP.

Como mobiliario fue necesario adquirir un sillón modelo TULLSTA de color gris cuyas medidas son 80x70x77 cm y una mesa modelo BESTA de color blanco con las siguientes dimensiones 121x72x8 cm.

C. Fase 4. Desarrollo del Prototipo

El prototipo se desarrolló mediante Adobe® AIR®.

D. Fase 5. Recogida de Datos

Se tardaron dos meses en la realización de la evaluación. Para llevarla a cabo, se concertaron las citas con los usuarios que tomaron parte en el experimento. Se citó a cada participante con una diferencia de 1 semana cada sesión.

Cuando el usuario acudía a una sesión, se le dieron indicaciones sobre qué métodos de inserción de texto utilizaría, que mandos a distancia usaría, se le explicaba como corregir errores y cómo rellenar la escala de votación. Estas indicaciones se hicieron mediante una breve exposición mediante una presentación de ordenador. Después se hizo una breve demostración práctica de utilización de los mandos a distancia. Una vez que el participante entendió lo que tenía que hacer, se ejecutó la aplicación desarrollada y dio comienzo la prueba. Cuando finalizó la prueba el registro de sus datos quedaron almacenados en un fichero XML para su posterior evaluación.

E. Fase 6. Análisis de Datos

Una vez recogidos los datos de los usuarios en el XML que genera la aplicación, se procesaron y se validaron para poder analizar sus resultados. El análisis estadístico de los datos se realizó con el paquete estadístico RCommander.

Como resultado destacable se obtuvo que, en general, la forma y la localización de las teclas no producen diferencias significativas ni en la velocidad de escritura ni en la tasa de error de los métodos utilizados.

V. CONCLUSIONES Y TRABAJOS FUTUROS

A través de este trabajo, se ha diseñado una metodología, para avanzar en la interacción humana en el ámbito de los métodos de inserción de texto para iDTV. Su aplicación será medir tanto la usabilidad, como la satisfacción de los usuarios. Abarcará todas las etapas necesarias para la realización de los test de usuario y pruebas subjetivas; desde

la definición de métricas, pasando por el desarrollo de prototipos y realización de las pruebas hasta concluir con el análisis de resultados. Esta metodología se validó de forma práctica con un caso de estudio, el cuál otorgó una importante retroalimentación que permitió aplicar ciertas mejoras.

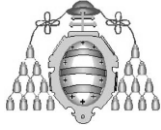
La metodología desarrollada permite organizar estudios con usuarios reales de forma ordenada pero en el ámbito de la inserción de texto. Sería interesante realizar mejoras para que fuese más genérica y su ámbito de aplicación sea cualquier tipo de aplicaciones en iDTV con o sin inserción de texto.

AGRADECIMIENTOS

El trabajo presentado en este artículo ha sido financiado por la Universidad de Oviedo y por el Principado de Asturias a través del Proyecto con referencia SV-PA-13-ECOEMP-75.

REFERENCIAS

- [1] M. Ingmarsson, D. Dinka, y S. Zhai, «TNT: a numeric keypad based text input method», en *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004, pp. 639–646.
- [2] A. Iatrino y S. Modeo, «Text editing in digital terrestrial television: a comparison of three interfaces», *Euro Itv*, 2006.
- [3] G. Geleijnse, D. Aliakseyeu, y E. Sarroukh, «Comparing text entry methods for interactive television applications», en *Proceedings of the seventh european conference on European interactive television conference*, 2009, pp. 145–148.
- [4] R. Aoki, A. Maeda, T. Watanabe, M. Kobayashi, y M. Abe, «Twist tap: text entry for TV remotes using easy-to-learn wrist motion and key operation», *Ieee Trans. Consum. Electron.*, vol. 56, n.º 1, pp. 161–168, feb. 2010.
- [5] U. Gargi y R. Gossweiler, «QuickSuggest: character prediction on web appliances», en *Proceedings of the 19th international conference on World wide web*, North Carolina, USA, 2010, pp. 1249–1252.
- [6] A. J. Sporka, O. Polacek, y P. Slavik, «Comparison of two text entry methods on interactive TV», en *Proceedings of the 10th European conference on Interactive tv and video*, 2012, pp. 49–52.
- [7] T. Költringer, P. Isokoski, y T. Grechenig, «TwoStick: writing with a game controller», en *Proceedings of Graphics interface 2007*, 2007, pp. 103–110.
- [8] I. S. MacKenzie, M. H. Lopez, y S. Castelluci, «Text Entry with the Apple iPhone and the Nintendo Wii», presentado en CHI2009, Boston, MA, USA, 2009.
- [9] T. Tullis y W. Albert, *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Morgan Kaufmann, 2008.
- [10] F. E. Ritter y L. J. Schooler, «The learning curve», *Int. Encycl. Soc. Behav. Sci. W Kintch N Smelser P Baltus Eds Oxf.*, pp. 8602–8605. Amsterdam: Pergamon, 2001.
- [11] T. Bellman y S. Mackenzie, «A Probabilistic Character Layout Strategy for Mobile Text Entry», en *Graphics Interface*, 1998, pp. 168–176.
- [12] I. S. MacKenzie, H. Kober, D. Smith, T. Jones, y E. Skepner, «LetterWise: prefix-based disambiguation for mobile text input», en *Proceedings of the 14th annual ACM symposium on User interface software and technology*, 2001, pp. 111–120.
- [13] R. Likert, «A technique for the measurement of attitudes.», *Arch. Psychol.*, 1932.
- [14] ISO, «14915-3:2002: Software ergonomics for multimedia user interfaces -- Part 3: Media selection and combination», *Int. Organ. Stand.*
- [15] J. Perrinet, Xabiel G. Pañeda, Claudia Acebedo, Jose Luis Arciniegas, Sergio Cabrero, David Melendi, y Roberto García, «Adaptación de una aplicación de E-Learning A T-LEARNING», presentado en V Congreso Iberoamericano de Telemática. CITA 2009.
- [16] L. Graham, «Gestalt theory in interactive media design», *J. Humanit. Soc. Sci.*, vol. 2, n.º 1, 2008.
- [17] «Recommendation ITU-R BT.500-12: Methodology for the subjective assessment of the quality of television pictures», International Telecommunication Union, vol.12, 2009.



8.2 Sistemas de introducción de texto en aplicaciones de TV interactiva.

A. Barrero, D. Melendi, X. G. Pañeda, R. García, y S. Cabrero, «Sistemas de introducción de texto en aplicaciones de TV interactiva», en *jAUTI2012 - I Jornadas Iberoamericanas de Difusión y Capacitación sobre Televisión Digital Interactiva*, La Plata (Argentina), 2013.

Sistemas de introducción de texto en aplicaciones de TV interactiva

AURORA BARRERO, DAVID MELENDI, XABIEL G. PAÑEDA, ROBERTO GARCÍA,
SERGIO CABRERO

Departamento de Informática, Universidad de Oviedo
Campus de Xixón/Gijón, s/n, Xixón/Gijón, 33203, Asturias, España.
{barreroaurora.uo, melendi, xabiel, garciaroberto, cabrerosergio}@uniovi.es

Resumen. En los últimos años se han llevado a cabo cambios regulatorios en muchos países para la implantación de servicios de televisión digital interactivos. La aparición de estos servicios conduce a un clima de mayor competencia, en el que los proveedores de contenidos ofrecen servicios cada vez más innovadores. Muchos de estos servicios requieren capacidades de interacción y, más concretamente, de inserción de texto. No obstante, surge un nuevo problema que emana del hecho de que el principal dispositivo de interacción en estos servicios es el mando a distancia (control remoto), elemento que no ha sido pensado para la escritura de texto. En este trabajo se presenta un proyecto que persigue encontrar métodos efectivos de escritura de texto para entornos de Televisión Digital Interactiva mediante un mando a distancia convencional.

Palabras clave: HCI, idTV, escritura de texto

147

1 Introducción

Hace no tantos años los contenidos digitales eran principalmente proporcionados por compañías privadas de satélite o cable. Atraían nuevos clientes ofreciendo mejores calidades y contenidos, además de una amplia variedad de canales, pero no servicios adicionales. Como no había muchos proveedores, podemos decir que tampoco había un clima de competencia que favoreciese la aparición de servicios innovadores. Recientemente esta situación ha cambiado drásticamente con la promoción de la televisión digital terrestre y de nuevos estándares para televisión digital. Por ejemplo, la Unión Europea estableció el año 2012 como límite para completar la

migración a televisión digital [1]. Estos cambios regulatorios han creado un nuevo escenario: un mercado gigantesco en el que navega una amplia oferta de proveedores. Con ello, se ha creado igualmente un clima de competencia que anima a los proveedores de contenidos a innovar y ofrecer servicios avanzados para ser competitivos.

No obstante, muchos de estos servicios avanzados requieren capacidades de interacción. Esta interacción presentaba muchas dificultades en el pasado, pero los avances tecnológicos y los índices de penetración de Internet en el hogar han cambiado este escenario. La televisión e Internet se han mezclado dando lugar a nuevos servicios como los de IPTV e Internet TV. Adicionalmente, nuevos dispositivos set-top-box y televisiones con capacidad computacional están disponibles en el mercado y permiten a los usuarios recibir señales de televisión y conectarse a Internet al mismo tiempo, haciendo posible la Televisión Digital Interactiva y, con ello, permitiendo a los proveedores de contenidos ofrecer los servicios avanzados que se habían comentado con anterioridad.

Muchos de estos servicios han surgido como métodos complementarios de acceso a información proporcionada por operadores de televisión. En ellos, el usuario puede interactuar con facilidad utilizando el mando a distancia de su televisión o set-top-box, dado que lo único que hace es desplazarse por menús de aplicaciones. Raramente se pide al usuario que introduzca texto, tarea que puede ser obligatoria en otros servicios como pueden ser los juegos interactivos, banca electrónica, navegación por Internet o sistemas de aprendizaje basados en la televisión. Por ello, aunque las tecnologías detrás de los servicios de televisión digital interactivos han evolucionado considerablemente, si se desean desarrollar aplicaciones complejas también pueden llegar a ser necesarios métodos efectivos para interactuar con los usuarios. En la actualidad, estas interacciones no pueden ser muy complejas dado que el elemento de interacción principal es el mando a distancia (control remoto) que viene con el equipo correspondiente. Esto supone un problema cuando se pretende escribir texto, ya que los mandos a distancia no han sido diseñados para esta tarea [2]. Por otro lado, la actual falta de estandarización tampoco facilita esta tarea [3]. Es cierto que hay otros tipos de

dispositivos como teclados externos o elementos apuntadores. No obstante, no son muy populares y su diseño no los hace confortables para su utilización desde el sofá. Así, si necesitamos usuarios que introduzcan texto de aplicaciones de televisión digital tenemos que responder a preguntas como ¿qué métodos hay disponibles para la introducción de texto utilizando un mando a distancia? ¿cómo están de familiarizados los usuarios con estos métodos? ¿cómo afectan las experiencias previas o la edad a la efectividad de estos métodos? [4] y, en definitiva ¿cuál es el mejor método para escribir texto utilizando un mando a distancia?

Dadas las dificultades expuestas, el Grupo de Investigación de Sistemas de Distribución Multimedia de la Universidad de Oviedo se planteó un proyecto de usabilidad en el ámbito de las aplicaciones de televisión digital interactiva. El problema concreto es el de encontrar métodos efectivos de inserción de texto considerando las limitaciones actuales en cuanto a los dispositivos de interacción disponibles en el mercado.

Los resultados del proyecto pueden suponer un avance en el estado del arte en el terreno de la usabilidad en el campo de los servicios de televisión digital interactivos. Pueden ser de gran interés no sólo para la comunidad investigadora, sino también para diseñadores de aplicaciones o fabricantes de dispositivos, en cuanto al planteamiento de los mecanismos de interacción a utilizar en sus servicios o equipos. Este trabajo pretende presentar el proyecto de investigación citado, así como algunos de los resultados obtenidos hasta el momento.

El resto del trabajo se ha organizado como sigue. En la sección 2 se presentan algunos trabajos anteriores relacionados con lo aquí publicado. En la sección 3 se muestra el plan de trabajo seguido en el proyecto. En la sección 4 se describe el entorno de experimentación utilizado durante las pruebas realizadas a los usuarios reales. En la sección 5 se presentan algunos resultados obtenidos en las pruebas, para concluir con las conclusiones y trabajos futuros en la sección 6.

2 Trabajos previos

Aunque el estudio de métodos de inserción de texto es un campo que ha generado muchos resultados de investigación, la mayoría de los avances más recientes corresponde al mundo de los dispositivos móviles como las tabletas, las pantallas táctiles o los teléfonos móviles [5]. Antes del año 2006 casi no se han encontrado estudios en los que se tenía en cuenta el escenario de la televisión digital interactiva tal y como mencionan Iatrino y Modeo [2]. Todo indica que esta situación se mantiene incluso en la actualidad.

Los mandos a distancia pueden utilizarse de dos formas posibles: utilizando los cursores y la tecla OK o usando un conjunto reducido de teclas como en los teléfonos móviles. Iatrino y Modeo [2] evalúan ambas posibilidades en un experimento grupal. Se hacen pruebas con tres métodos: multipress (el método SMS), el multipress con una retroalimentación visual y un teclado virtual QWERTY. 36 personas participaron en el experimento teniendo que llevar a cabo dos tareas con cada método. La primera tarea consistía en escribir una dirección de correo electrónico, mientras que la segunda se trataba de escribir una frase corta en italiano. Los autores concluyen que el mejor método es el multipress y destacan múltiples problemas relacionados con la internacionalización.

Ingmarsson et al. [6] presentan una nueva técnica llamada TNT similar al ya comentado TwoStick [7]. El sistema se basa en una rejilla de 3x3 en la que cada celda se subdivide en otras 9 celdas menores. Cada una de las celdas menores tiene un carácter, haciendo un total de 81 posibilidades. Para escribir un carácter el usuario debe seleccionar una de las celdas principales utilizando el teclado numérico del mando a distancia y luego seleccionar el carácter que desea escribir con una nueva pulsación. Como se puede observar, cada carácter es accesible con sólo dos pulsaciones. Cinco personas pagadas con edades entre los 27 y los 32 años probaron el sistema durante 10 sesiones de 45 minutos para escribir una novela corta en Sueco. Los resultados muestran velocidades comparables o superiores a la escritura manual en una PDA o un método multipress. Los usuarios destacaban la sencillez del método.

Por otra parte, Geleijnse et al. [8] han comparado tres técnicas mediante un mando a distancia (multitap, T9 y un teclado virtual) con escritura realizada con un teclado convencional QWERTY (con y sin autocompletado) El objetivo del experimento era buscar vídeos en Youtube escribiendo el par “artista-pista” utilizando cada uno de estos métodos. La evaluación se basó en las respuestas de 22 participantes con edades comprendidas entre los 21 y los 32 años a un cuestionario y medidas de tiempos requeridos para completar la tarea. A pesar de que los autores no especifican la experiencia de los usuarios con nuevas tecnologías, no es de sorprender que los resultados muestren una diferencia significativa entre las técnicas del mando a distancia y las del teclado convencional. Lo que es sorprendente, es que los autores no encontraron ninguna diferencia significativa entre los distintos métodos utilizados con el mando a distancia. Los autores también concluyeron que, bajo las condiciones del experimento, “las encuestas no han demostrado ningún indicio de que los usuarios no acepten un teclado en un salón”. Esta conclusión contrasta con la de Orbist et al. [9] que abogan por sistemas basados en voz. Los resultados de su estudio etnográfico indican que es poco realista centrarse en estudios dependientes de la popularidad de periféricos externos a la televisión distintos de los convencionales.

Más recientemente, Gargi y Gossweiler [10] presentan un nuevo sistema predictivo diseñado para mejorar la velocidad de escritura en teclados virtuales: QuickSuggest. Esta técnica muestra un anillo circundante al carácter actual, mostrando los cuatro caracteres mas frecuentes que siguen al seleccionado. Cuando se escribe un carácter el anillo aparece y si el usuario iba a escribir uno de los caracteres que se sugieren solamente tiene que seleccionarlo y presionar OK para utilizarlo. El cursor se mueve inmediatamente a la tecla correspondiente. Si el carácter deseado no es ninguno de los sugeridos, el usuario debe moverse a la posición del teclado en el que se encuentra el carácter deseado. Este método requiere un mínimo de dos pulsaciones por carácter. En cuanto a la evaluación, los autores la han realizado en dos pasos. En primer lugar realizan un estudio teórico del método, mientras que en el segundo se realiza un estudio con 10 participantes. Los autores emplean la métrica de pulsaciones por carácter (KSPC) en ambos estudios

pero también calculan el tiempo requerido por los usuarios en la segunda fase del estudio.

Como se puede apreciar, no hay muchos trabajos en los que se trate el tema de la escritura de texto en aplicaciones de televisión. Muchos trabajos presentan conclusiones basadas en los modelos de predicción de Fitt [11]. Otros confían en técnicas no muy extendidas como teclados externos o reconocimiento de voz. En otros trabajos se ejecutan evaluaciones con usuarios reales pero se selecciona un grupo o muy reducido de personas o con características muy similares: comúnmente jóvenes y con experiencia en nuevas tecnologías.

3 Plan de trabajo

El trabajo que se presenta tiene como objetivo evaluar distintos métodos de inserción de texto en aplicaciones de televisión digital interactiva. Para ello se ha seguido un plan de trabajo que está estructurado en cinco tareas:

- Tarea #1: Experimentación con métodos populares. Se han seleccionado una serie de métodos de escritura populares y se realizan unas pruebas con unos textos sencillos. En concreto se ha elegido utilizar teclados virtuales QWERTY, Alfabético y Genético y el método Multitap. QWERTY utiliza la misma disposición de letras que el teclado de un ordenador. El Alfabético utiliza el orden alfabético para ubicar las letras en pantalla. Por otro lado, el Genético consiste en agrupar los caracteres más populares en el centro del teclado, de acuerdo con el algoritmo publicado en [12]. Finalmente, el método Multitap, utilizado masivamente en el pasado para la escritura de mensajes SMS, asigna grupos de caracteres a los números del mando a distancia.
- Tarea #2: Experimentación con métodos optimizados. En base a las pruebas realizadas en la tarea anterior, se plantean una serie de optimizaciones para los métodos previamente utilizados. Entre estas optimizaciones, está la utilización de algunos botones del mando a distancia para

el borrado de caracteres y la escritura de espacios en blanco para los teclados virtuales. Igualmente, se eligen algunas optimizaciones seleccionadas de los trabajos previos, en concreto los métodos T9 y 2-Key diseñados para la escritura en teléfonos móviles [13].

- Tarea #3: Experimentación en entornos específicos. En tareas anteriores se utilizan textos sencillos, pero ahora se pretende utilizar textos más complejos, en los que aparecen vocales con tildes y caracteres especiales. Esta situación se aproxima a un entorno real en la que el usuario tiene que escribir una URI, una dirección de correo o sus datos personales. Para ello, se modifican los teclados virtuales para incluir las vocales con tildes (como se muestra en la parte superior izquierda de la Fig. 1), además de presentar al usuario juegos de caracteres complementarios para la escritura de símbolos (como se muestra en la parte inferior de la Fig. 1). Por otro lado, el método Multitap se altera para incluir nuevas tablas de símbolos a las que se accede pulsando botones del teclado numérico del mando a distancia, tal y como se muestra en la parte derecha de la Fig. 1.

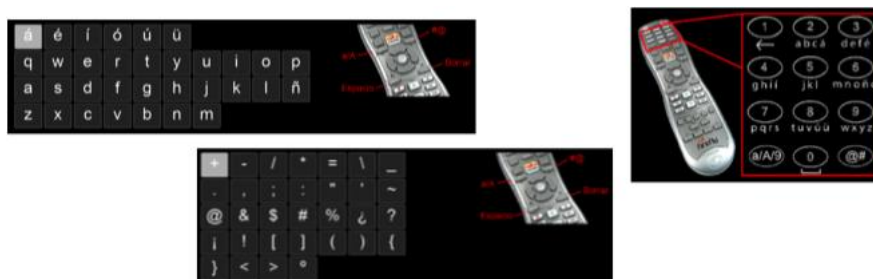


Fig. 1. Disposiciones de caracteres especiales. En la parte superior izquierda teclado virtual QWERTY que incluye vocales con tildes. En la parte inferior central, juego complementario de caracteres especiales para teclados virtuales. En la parte derecha, juego de símbolos modificado para el método Multitap y botones de acceso para la escritura de otros caracteres especiales.

- Tarea #4: Experimentación con mandos heterogéneos. Hasta el momento en todas las pruebas realizadas se utiliza un único mando. Para evaluar el impacto que tiene

la forma del mando a distancia o la disposición de teclas que utiliza, se realiza una nueva tarea en la que se comparan los resultados obtenidos con mandos a distancia de características heterogéneas. Para ello, se repiten las pruebas de la Tarea #3 con dos nuevos mandos a distancia, que tienen una disposición de teclas y una forma distintas de las del mando que se venía utilizando hasta el momento. Los resultados de esta tarea se comparan con los de la anterior.

- Tarea #5: Experimentación con dispositivos competidores. Como es indudable que en el futuro las televisiones incorporarán otros elementos de interacción distintos del mando a distancia, se plantea esta tarea para analizar elementos de interacción diferentes. Se plantea la utilización en las pruebas de teclados de distintas dimensiones, elementos táctiles y periféricos giroscópicos, entre otros. El objetivo es tanto comparar estos periféricos entre sí, como comparar los resultados obtenidos en esta tarea con los de tareas anteriores. Esta tarea está todavía en ejecución y no se dispone de resultados todavía.

Para cada una de estas tareas, se convoca a un conjunto de personas para participar en una o varias sesiones de pruebas. Durante estas sesiones, los participantes utilizan una aplicación desarrollada a tal efecto, que nos permite capturar datos para determinar la siguiente información:

- Velocidad de escritura: En caracteres por minuto, se calcula tomando como referencia el tiempo que un usuario ha necesitado para escribir una determinada sentencia. Lógicamente, los errores cometidos durante la escritura penalizan en el tiempo utilizado para escribir una frase.
- Porcentaje de error: Se calcula comparando el número total de caracteres escritos por el usuario con el número de caracteres que en realidad debía escribir.

- Curva de aprendizaje: Estimada en aquellas tareas en las que los usuarios participan en varias sesiones, se calcula según describen Ritter y Schooler [14]
- En base a cuestionarios finales con escalas Likert (de 0 a 4), la impresión subjetiva del usuario respecto a la facilidad de uso, la velocidad de escritura y la satisfacción general.

4 Entorno de experimentación

Para realizar cada experimento se dispone una sala en la que cada participante puede realizar las pruebas sin distracciones. Para crear una situación lo más realista posible, se le deja en un sofá enfrente de una televisión de 32 pulgadas situada a una distancia de 2 metros, siguiendo las recomendaciones del fabricante. La televisión está conectada a un PC con Windows XP. A ese PC se conectan los dispositivos que se vayan a utilizar en cada prueba. En la mayoría de los casos se trata de un mando a distancia SnapStream Firefly, que dispone de un receptor USB de radiofrecuencia. En la Fig. 2 se observa el mando a distancia utilizado.



Fig. 2. Mando a distancia SnapStream FireFly y captura de la pantalla de la aplicación utilizada durante las pruebas.

Por otro lado, en el PC se ejecuta una aplicación Adobe® AIR® diseñada para realizar los experimentos. La aplicación se ejecuta a pantalla completa sobre fondo negro, y muestra los controles necesarios para la captura de texto, así como propuestas sobre los

textos que deben ser escritos en cada experimento. En la Fig. 2 se muestra esta aplicación. Cuando se inicia una sesión en la aplicación, se dejan unos instantes antes de permitir que el usuario comience a escribir. A continuación, se le presenta el texto que tiene que escribir y no se le da un tiempo límite. Si en la prueba se proponen un texto, la captura finaliza cuando el texto escrito coincide exactamente con el texto propuesto. Si no se propone un texto, es el usuario el que indica que ha terminado. Una vez finalizada la captura se muestra un nuevo texto o se cambia el método de escritura, informando al usuario y dándole unos instantes para que se adapte al cambio.

En este momento la aplicación permite escribir textos mediante dos mecanismos: teclados virtuales y métodos heredados de los teléfonos móviles. Para utilizar los teclados virtuales, la aplicación presenta en pantalla las letras acorde a la disposición con la que se desee experimentar. El usuario se desplaza por esa proyección utilizando los cursores del mando a distancia. Cuando desea escribir una letra, presiona el botón de confirmación. En función de la prueba, el teclado virtual dispone de un elemento especial para el borrado de caracteres, con lo que a efectos prácticos borrar un carácter supone desplazarse por el teclado virtual igualmente. En cuanto a los métodos heredados de los dispositivos móviles, todos están pensados para escribir texto con los botones numéricos del mando a distancia. Por ello, lo que se presenta en pantalla es solamente información de apoyo. Por ejemplo, la relación de caracteres que corresponde a una tecla determinada.

Para proponer los textos, se ha creado un corpus de frases cortas extraídas de medios de comunicación populares en España. En función del diseño de cada experimento, se extrae un número determinado de frases del corpus.

Por otro lado, la aplicación dispone de algunos mecanismos que pretenden hacer la escritura más fluida:

- Se dispone de un mecanismo de sugerencias inspirado en LetterWise [15]. En función de lo que escribe el usuario, el sistema cambia el color de las seis letras que con mayor probabilidad siguen a la última que se ha escrito. Para ello, se utiliza un diccionario de Castellano [16].

- Se permite a los usuarios que mantengan los cursores del mando a distancia apretados para moverse por los teclados virtuales más rápidamente.
- Los bordes de los teclados virtuales están conectados como si se tratase de una esfera, lo que permite a los usuarios pasar de la fila superior a la inferior o de la columna izquierda a la derecha sin atravesar los elementos intermedios.

La aplicación captura los eventos que se producen durante su ejecución en un fichero XML. Se anotan tanto los textos propuestos, como lo que escriben los usuarios. Queda constancia de todas las pulsaciones que realizan. El propósito no es otro que el de disponer de información para su posterior análisis estadístico. Este análisis se produce una vez que las pruebas de cada tarea han concluido. El proceso seguido es siempre el mismo. En primer lugar se comprueba la normalidad de los datos con tests de Saphiro-Wilk y su homocedasticidad con tests de Bartlett. Dependiendo de estas características, utilizamos tests de ANOVA o de Kurskal-Wallis y, en caso de existir diferencias entre los datos, utilizamos tests de Tukey con un coeficiente de confianza del 95% para realizar comparaciones dos a dos.

5 Resultados obtenidos

En esta sección se presenta un resumen de los resultados obtenidos en las tareas #1 a #4. No se dispone de resultados sobre la tarea #5, que se encuentra en curso en la actualidad. Por otro lado, algunos de estos resultados se han publicado de forma detallada en [17].

En las pruebas participaron un total de 96 personas con características heterogéneas, tal y como se desprende de la tabla 1.

Colectivo	Número total	Mujeres	Profesión o Estudios TI	Estudios Primaria	Educación Secundaria	Educación Universitaria
Jóvenes	57	25	39	0	7	50
Mediana edad	27	10	14	0	5	22
Sénior	12	4	1	5	1	6

Tabla 1. Personas participantes en los experimentos

Una primera conclusión de las pruebas realizadas es que el método impacta claramente sobre la velocidad de escritura. Por ejemplo, la Fig. 3 muestra la velocidad de escritura registrada en la tarea #1. Se pueden apreciar diferencias notables entre la velocidad de los distintos métodos, existiendo, en algunos casos, diferencias estadísticamente significativas.

Por otro lado, en la Fig. 3 también se puede apreciar que con la repetición se mejoran los resultados. Aunque es algo lógico hasta cierto punto, la mejoría se produce con todos los métodos. Esto incluye mejorar en métodos que a priori son conocidos por los usuarios como el caso de los teclados virtuales QWERTY o Alfabético, para los que nos se esperaba una mejora paulatina de rendimiento.

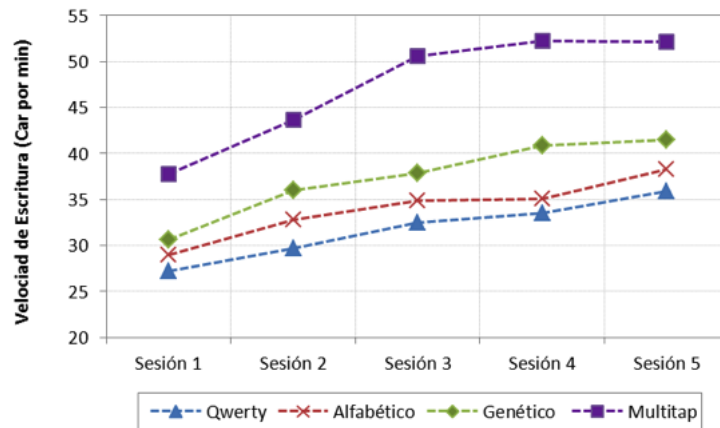


Fig. 3. Evolución de la velocidad media de escritura durante las cinco sesiones de la tarea #1 del proyecto.

Otra conclusión de las pruebas es que un método que registra una velocidad de escritura muy alta no necesariamente es el método con el que se produce un menor número de errores. En la Fig. 4 se muestra la tasa de errores obtenida en la tarea #1. En ella se puede apreciar cómo el método Multitap, que era el más rápido según la Fig. 3, es el que mayor número de errores registra.

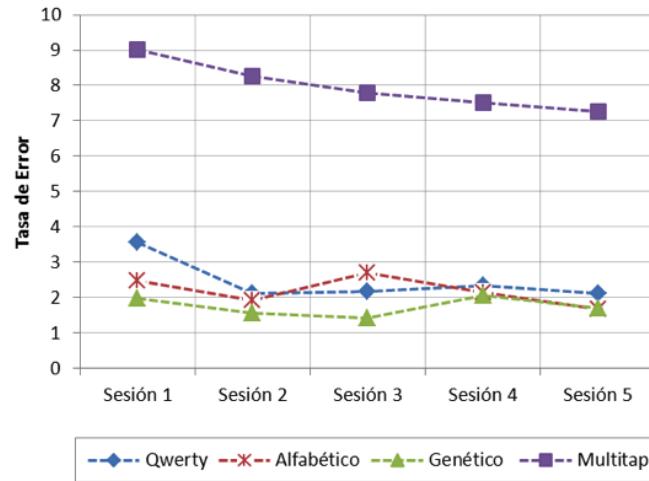


Fig. 4. Evolución de la tasa de error durante las cinco sesiones de la tarea #1 del proyecto.

En el estudio también se ha constatado que las impresiones subjetivas no coinciden con los datos empíricos que se registraron durante las pruebas. En muchos casos, la apreciación de velocidad de los usuarios se aleja de la velocidad real registrada para un método concreto. También se ha constatado que métodos que en principio eran desconocidos para los usuarios, como los teclados virtuales genéticos, son poco valorados durante las primeras sesiones, pero ganan popularidad a medida que los usuarios se familiarizan con ellos.

En la tarea #2 se ha constatado que, en algunos casos, las optimizaciones sobre el papel no producen los efectos deseados. Por ejemplo, el resultado de utilizar botones específicos en el mando a distancia para el borrado de caracteres y la inserción de espacios en blanco no produce ninguna mejora sobre los resultados obtenidos con los teclados virtuales. Del mismo modo, la utilización del método 2-Key empeora la velocidad obtenida con el método Multitap en un 38,51%. Solamente se mejora la velocidad con el método T9, comparando éste con el método Multitap tradicional.

Durante la tarea #3 también se constata que, con la inserción de disposiciones alternativas para la escritura de caracteres especiales se empeora en mucho el rendimiento. El hecho de tener que

cambiar de juego de caracteres tanto en los teclados virtuales como en el método Multitap provoca un empeoramiento significativo en todos los casos. En concreto, para el método QWERTY modificado se obtienen unos resultados que empeoran la velocidad de escritura en un 20,15%, para el Genético el empeoramiento es del 32,30% y en el Multitap el empeoramiento llega al 48,09%.

En la tarea #4 también se ha observado que el cambio en la disposición de teclas en el mando a distancia o en su forma, no afecta a la velocidad de escritura ni a la tasa de error significativamente.

En todas las pruebas se ha apreciado una influencia de la edad en el rendimiento de los usuarios, tanto en la velocidad de escritura como en la tasa de error. En general, esta diferencia no es estadísticamente significativa entre los usuarios jóvenes y los de mediana edad, pero sí entre estos usuarios y los de mayor edad. Finalmente, se ha observado que los hábitos de utilización de servicios de televisión y TI no tienen, en general, una relación con el rendimiento de los usuarios.

6 Conclusiones y trabajos futuros

En este trabajo se presenta un proyecto liderado por el grupo de investigación de Sistemas de Distribución Multimedia de la Universidad de Oviedo. En este proyecto se ha evaluado el rendimiento obtenido con distintos métodos de escritura de texto diseñados para servicios de televisión digital interactiva. En la evaluación se ha tenido en cuenta que el dispositivo mayoritariamente disponible en estos servicios es el mando a distancia, por lo que los métodos se han seleccionado acorde a esta situación.

En las pruebas se ha constatado la importancia de elegir un método de escritura adecuado, dado que esta elección impacta notablemente sobre la velocidad de escritura y el número de errores que cometen los usuarios. Otra conclusión interesante es que la utilización de textos complejos en las aplicaciones produce un incremento en la complejidad de los métodos de escritura y un gran empeoramiento en el rendimiento de los usuarios. Esta situación puede cambiar mediante la inclusión de algunos símbolos

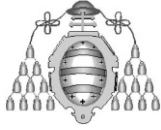
comúnmente utilizados durante el proceso de escritura en primera línea. Así, los usuarios podrían escribir sencillamente símbolos como las comas, los puntos o el @. De esta forma, se podrían relegar los símbolos menos utilizados a los juegos complementarios de caracteres, accesibles mediante un botón del mando.

En la actualidad el proyecto está en su última fase, experimentando con periféricos distintos del mando a distancia convencional. Estas pruebas se justifican ante la indudable presencia masiva de estos dispositivos en el futuro. Por otro lado, también se espera aumentar la masa crítica de usuarios participantes mediante la colaboración con otros grupos de investigación, pudiendo hacer comparaciones en función del idioma o de la procedencia de los usuarios.

Referencias

1. Unión Europea, *Communication from the Commission to the Council, the European Parliament, the European Economic and Social committee and the Committee of the Regions on accelerating the transition from analogue to digital broadcasting* {SEC(2005)661}, 2005.
2. Iatrino, A. y Modeo, S. "Text Editing in Digital Terrestrial Television: a Comparison of Three Interfaces". En *Proceedings of EuroITV'06*, Atenas, Grecia. 2006.
3. Jakob Nielsen. "Remote Control Anarchy". Disponible en la dirección <http://www.useit.com/alertbox/20040607.html>, 2012.
4. Taveira, Alvaro D. and Choi, Sang D. (2009). "Review Study of Computer Input Devices and Older Users". En *International Journal of Human-Computer Interaction*, 25: 5, 455-474.
5. MacKenzie, I. S. y Soukoreff, R. W. "Text Entry for Mobile Computing: Models and Methods, Theory and Practice". En *Human-Computer Interaction*, Vol. 17, pp 147-198. 2002.
6. Ingmarsson, M., Dinka, D., Zhai, S. (2004). "TNT – A Numeric Keypad Based Text Input Method". En *Proceedings del SIGCHI Conference*, pp. 639-646, Vienna, Austria.
7. Költringer T., Isokoski P. and Grechenig T.: "TwoStick: Writing with a Game Controller". En *Proceedings of Graphics Interface 2007 (GI 2007)*, 103-110.
8. Geleijnse G., Aliakseyeu D, Sarroukh E. (2009). "Comparing Text Entry Methods for Interactive Television Applications". En *Proceedings del EuroITV'09*, Leuven, Belgium.

9. Orbist, M., Bernhaupt, R. and Tscheligi, M. (2008). "Interactive TV for the Home: An Ethnographic Study on Users' Requirements and Experiences". En *International Journal of Human-Computer Interaction*, 24:2, 174-196.
10. Gargi U., Gossweiler R.: "QuickSuggest: Character Prediction on Web Appliances". En *WWW 2010*, April 26-30, 2010, Raleigh, North Carolina, USA.
11. MacKenzie, I. S. (1991). *Fitts' law as a performance model in human-computer interaction*. Doctoral dissertation. University of Toronto: Toronto, Ontario, Canada.
12. Brewbaker, C. R. (2008). *Optimizing Stylus Keyboard Layouts With a Genetic Algorithm: Customization and Internationalization*. Department of Computer Science, Iowa State University.
13. Silfverberg, M., MacKenzie, I.S. and Korhonen, P. (2000). "Predicting Text Entry Speed on Mobile Phones". En *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 9-16. New York.
14. Ritter, F. E., & Schooler, L. J., 2002, "The learning curve. In *International encyclopedia of the social and behavioral sciences*". 8602-8605. Amsterdam: Pergamon.
15. MacKenzie, I. S., Kober, H., Smith, D., Jones T. and Skepner, E. (2001). "LetterWise: Prefix-based Disambiguation for Mobile Text Input". En *Proceedings del ACM Symposium on User Interface Software and Technology*, pp. 111-120. New York.
16. Real Academia de la Lengua Española (2001), *Diccionario de la Lengua Española*. 22 edición (disponible en <http://www.rae.es/rae.html>).
17. Perrinet, J., Pañeda, X.G., Cabrero, S., Melendi, D., García, R. & García, V., 2011, "Evaluation of Virtual Keyboards for Interactive Digital Television Applications". En *International Journal of Human-Computer Interaction*, vol. 27, n. 8, pp. 703-728.



8.3 TVpad: Un novedoso sistema de interacción con la televisión.

A. Barrero, D. Melendi, y X. G. Xabiel, «TVPAD: Un Novedoso Sistema de Interacción con la Televisión», en *II Jornadas Iberoamericanas de Difusión y Capacitación sobre Aplicaciones y Usabilidad de la Televisión Digital Interactiva*, 2014.

TVpad: Un novedoso sistema de interacción con la televisión

Aurora Barrero, David Melendi, Xabiel G. Pañeda, Roberto García, Sergio Cabrero

Departamento de Informática, Universidad de Oviedo
Campus de Xixón/Gijón, s/n, Xixón/Gijón, 33203, Asturias, España.
{barreroaurora, melendi, xabiel, garciaroberto, cabrerosergio}@uniovi.es

Resumen. En los últimos años ha habido una gran evolución en el mercado de los sistemas de televisión digital interactivos. Esta evolución hacia lo que hoy conocemos como *smartTV*, aproxima la experiencia de los telespectadores a la de los usuarios de un ordenador conectado a Internet. Lógicamente, muchos de los servicios que se ofrecen en este entorno requieren cierta capacidad de interacción. No obstante, surge un nuevo problema que emana del hecho de que el principal dispositivo de interacción en estos servicios es el mando a distancia, elemento que no ha sido pensado para proporcionar la capacidad de interacción que requiere este nuevo entorno. Por ello, en este trabajo se presenta un novedoso sistema de interacción que persigue mejorar la experiencia de los usuarios en entornos de Televisión Digital Interactiva. Este sistema se ha pensado para alcanzar las capacidades de interacción requeridas por este tipo de aplicaciones, al mismo tiempo que se conserva la filosofía tradicional de uso de la televisión.

Palabras clave: HCI, idTV, control remoto, dispositivo de interacción, touchpad

1 Introducción

Recientes innovaciones en el campo de los servicios de televisión IP han creado un nuevo escenario de uso en los servicios de televisión digital interactiva [1]. En el pasado, las capacidades de interacción en este entorno eran muy limitadas por condicionantes técnicos. No obstante, recientes avances tecnológicos y la mejora en los índices de penetración de Internet en el hogar han cambiado este escenario. Nuevos dispositivos set-top-box y televisiones con capacidad computacional están disponibles en el mercado y permiten a los usuarios recibir señales de televisión y conectarse a Internet al mismo tiempo. La televisión e Internet se han mezclado dando lugar a nuevos servicios y a un escenario de real interacción en este contexto.

Estos dispositivos set-top-box y televisores incorporan como elemento principal de interacción un mando a distancia convencional. En muchos de los servicios que se ofrecen en estos entornos, los usuarios pueden interactuar con facilidad con este mando a distancia. Permiten al usuario interactuar con facilidad dado que lo único que hace es desplazarse por menús de aplicaciones. No obstante, el mando a distancia convencional puede presentar algunos problemas cuando se necesitan capacidades de

interacción avanzadas. Estas capacidades pueden incluir la inserción de texto, la ejecución de acciones de ampliación y reducción, la selección de elementos en pantalla y/o de fragmentos de texto, el uso del portapapeles o el accionamiento de un control habilitado en pantalla.

Dadas las dificultades expuestas, en este trabajo se presenta un novedoso sistema de interacción con la televisión que pretende ofrecer una alta capacidad de interacción al mismo tiempo que mantiene la filosofía de confort que se supone a un entorno de uso de la televisión.

El resto del trabajo se ha organizado como sigue. En la sección 2 se presentan otros dispositivos susceptibles de ser utilizados en un entorno de televisión digital interactiva. En la sección 3 se presenta el sistema diseñado, para concluir con las conclusiones y trabajos futuros en la sección 4.

2 Otros dispositivos de interacción

2.1 Mandos a distancia convencionales

Hasta la actualidad, los dispositivos de televisión se comercializan incorporando como único elemento de interacción un mando a distancia. Con independencia de las limitaciones funcionales que ya se han expuesto con anterioridad, los mandos a distancia también adolecen de otros problemas. Por un lado, no existe un estándar que se encargue de especificar cómo tienen que ser estos mandos a distancia, solo una serie de recomendaciones de alto nivel [2]. Esto hace que se produzcan numerosos problemas de accesibilidad dado que cada fabricante aplica sus propios criterios a la hora de diseñar estos elementos [3]. Por otro lado, cuando se tienen que realizar tareas complejas el mando a distancia puede ser limitante porque las teclas y la pantalla se encuentran en dispositivos diferentes, con lo que los usuarios no pueden mirar al mismo tiempo a la pantalla y al elemento de interacción [4].

2.2 Teclados

Pensando en la aproximación del mundo de la televisión al mundo del ordenador, podríamos plantear la utilización de un teclado como método de interacción con la televisión. Los teclados son los dispositivos de interacción con el ordenador más antiguos. La configuración más extendida de teclas es la QWERTY, que mantiene prácticamente su diseño de 1868 con la incorporación de algunas teclas adicionales. Algunas de las limitaciones de los teclados QWERTY [5], así como las patologías que provoca su uso [6], ya fueron identificadas en el pasado. No obstante, aunque para la realización de algunas tareas su utilidad en un servicio de televisión digital interactiva es innegable [8], su utilización en un entorno de confort alejado de una mesa puede no ser adecuado. Esto con independencia de otros trabajos que descartan de forma rotunda su utilización en un contexto de televisión [9].

2.3 Dispositivos apuntadores

Con el diseño de un interface de usuario adecuado, se podrían utilizar dispositivos apuntadores para controlar el televisor. Podríamos encuadrar en esta categoría a los ratones, *trackballs*, *touchpads* y dispositivos de tipo *wiimote*.

Los ratones son el periférico de interacción con el ordenador más utilizado después del teclado [10]. Con independencia de algunos problemas que genera su utilización [11], este periférico no es adecuado para el televisor ya que necesita una superficie sobre la que desplazarse. Esto es inadecuado en el entorno de confort que se presupone al uso de la televisión.

Por otro lado, los *trackballs* pueden ser una alternativa viable en un entorno de televisión digital interactiva. No requiere una superficie determinada, ya que puede sujetarse con una mano, puede integrarse en un mando a distancia convencional, ampliando sus posibilidades y parece tener buenos resultados en personas con discapacidad o de edad avanzada [12]. No obstante, los *trackballs* pueden causar fatiga y problemas de precisión, acompañados de una caída en el rendimiento cuando se utiliza al mismo tiempo que se pulsa un botón [13].

La utilización de *touchpads* en un entorno de televisión también puede ser factible. Los *touchpads* modernos proporcionan una gran riqueza de interacción en forma de gestos. No obstante, pueden presentar algunos problemas. En la actualidad muchos de estos dispositivos están diseñados para operar sobre una superficie, lo que permite controlar los eventos de pulsación. Esto puede suponer un problema en este contexto. Por otro lado, los usuarios pueden experimentar problemas en función de su tamaño [14] pese a algunos intentos en esta línea como el de [15].

Finalmente, otra posibilidad es la de utilizar algún tipo de dispositivo similar al *wiimote* que, a modo de pistola, permita apuntar sobre alguna zona de la pantalla, como en [16]. No obstante, este método tiene potenciales problemas de fatiga [17]. Además, este tipo de dispositivo puede presentar problemas de precisión a la hora de realizar ciertas tareas, debido al temblor natural de las manos y a limitaciones inherentes en las personas a la hora de apuntar [18].

2.5 Pantallas táctiles

Es también posible disponer de un televisor con una pantalla táctil. No obstante, este mecanismo no se considera adecuado ya que los usuarios tendrían que estar sentados muy cerca del televisor. Normalmente, los fabricantes recomiendan al telespectador sentarse a una distancia de uno a varios metros en función del tamaño de pantalla. Por ello, este mecanismo no se considera viable de forma generalizada a pesar de ser una posibilidad desde el punto de vista tecnológico.

2.6 Smartphones y tabletas

Algunos fabricantes de soluciones que operan en el mercado de la televisión y en el mercado de los dispositivos móviles ofrecen a los usuarios la posibilidad de disponer

de una experiencia de usuario multimodal de experiencia. Esta experiencia permite controlar el televisor desde un Smartphone o una tableta. Este es el caso de fabricantes como Samsung, LG o Apple.

Si bien esta experiencia puede ser todo lo rica que se desee, puesto que estos dispositivos son muy versátiles, la duda viene del hecho de si es viable que uno de estos dispositivos se convierta en el dispositivo de interacción principal del televisor. Es decir, que llegue a sustituir al mando a distancia y se distribuya junto con el televisor. Podríamos afirmar que no, dado su elevado coste. Este método de interacción podría darse en aquellos casos en los que, circunstancialmente, el usuario tenga varios dispositivos del mismo fabricante.

La utilización de estos dispositivos de forma masiva para el control de las televisiones probablemente requiera el desarrollo de algún tipo de estándar que, hasta donde sabemos, no existe en la actualidad.

2.7 Comandos de voz

Otra alternativa para interactuar con la televisión es la voz. Este mecanismo, ya cuenta con varios productos en el mercado. No obstante, este método presenta una serie de inconvenientes. En primer lugar, el usuario debe recordar multitud de comandos para poder realizar la tarea que desea realizar. Por otro lado, pueden presentarse problemas en algunas situaciones. En entornos ruidosos, en entornos en los que la voz puede resultar molesta para otras personas, cuando la boca del usuario está ocupada o cuando existe algún tipo de dificultad en expresar algo [19]. Estos problemas suelen tener que solucionarse con algún tipo de interfaz multimodal que combina las instrucciones de voz con algún tipo de dispositivo complementario.

2.8 Reconocimiento de gestos

Una última posibilidad que ya tiene algunas implementaciones en el mercado, es la de interactuar con la televisión a través de gestos. Estos gestos pueden ser capturados mediante una cámara e interpretados por el televisor o set-top-box para ejecutar instrucciones concretas. No obstante, este método presenta una serie de problemas. En primer lugar, este sistema es poco intuitivo y requiere de un periodo de adaptación determinado. Además, la definición de un espacio individual de interacción que evite la interpretación errónea de movimientos enfrente de la cámara, es un tema sin resolver. Finalmente, trabajos anteriores han registrado una elevada carga mental y física en los usuarios en comparación con otros métodos de interacción [20], además de la preferencia de algunos colectivos por otros métodos de interacción [21].

3 Diseño de TVpad

El diseño de TVpad se sustenta sobre la idea de que, en general, el mecanismo más versátil y sencillo de utilizar para interactuar con la televisión es mediante algún tipo de dispositivo apuntador. No obstante, la utilización de un dispositivo apuntador para controlar la televisión necesita un cambio en el interface que se muestra a los usuarios en pantalla. Por ello, en esta sección se hace una propuesta tanto del interface a mostrar en pantalla, como del dispositivo de interacción en sí mismo.

3.1 Diseño del dispositivo de interacción

De los dispositivos de interacción comentados en este trabajo, hemos seleccionado el *touchpad* como base para el diseño de TVpad.

En los últimos años, hemos estado trabajando en un estudio de evaluación de diferentes métodos de inserción de texto en aplicaciones de televisión digital interactiva. En la última fase de este estudio, hemos comparado la escritura de texto con teclados convencionales de distintos tamaños, con un *touchpad* y con un mando de tipo *wiimote*. Los resultados de este estudio vienen a completar los trabajos publicados en [22] y [23]. Pese que estos resultados están pendientes de publicación, hemos considerado relevantes las valoraciones subjetivas de los usuarios del experimento. Aquí trasladamos algunos comentarios:

- De forma generalizada los usuarios consideran que el método más rápido de escritura de texto es el teclado convencional. No obstante, un 36% de los usuarios indicaron explícitamente que era incómodo de utilizar en el contexto de visionado de la televisión. Otros usuarios proporcionaron comentarios sobre la falta de iluminación, que dificulta ver las teclas del teclado. Finalmente, algunos usuarios que no utilizan el ordenador frecuentemente indicaron problemas a la hora de escribir símbolos que necesitan la pulsación simultánea de varias teclas (un 18% de los usuarios).
- Los teclados de reducidas dimensiones, presentaron problemas adicionales a los teclados convencionales debido a su tamaño. Un 45% de los usuarios proporcionaron comentarios relativos a las dimensiones del teclado, la reorganización de las teclas y/o la dimensión de cada tecla.
- Un 46% de los usuarios indicaron que habían sufrido problemas de fatiga durante la prueba. Todos ellos coinciden en el cansancio que produce la utilización del dispositivo de tipo *wiimote*.
- Un 56% de los usuarios indica explícitamente que el *touchpad* es un dispositivo muy sencillo y cómodo de utilizar, llegando a afirmar que “es el mejor método para personas que no saben mecanografía” o que “es el mejor método de todos como mando”

Basándonos en las pruebas que hemos realizado, creemos que un *touchpad* puede ser un dispositivo totalmente adecuado para interactuar con una televisión. Obviamente, no es el mejor dispositivo para escribir texto, pero consideramos que es adecuado para reemplazar al mando a distancia convencional. No obstante, el

dispositivo debería tener un tamaño adecuado [14], además de tener la posibilidad de prescindir de una superficie de apoyo. En nuestro estudio, utilizamos un *Magic Trackpad* de Apple, que dispone de unas dimensiones que hemos considerado adecuadas. Para poder ser utilizado con libertad en el experimento, incorporamos una pieza de metacrilato en la parte inferior tal y como se muestra en la figura 1. Esto permite, por ejemplo, sujetar el dispositivo con una mano e interactuar con la otra, sujetar e interactuar con ambas manos a la vez.



Fig. 1. *Magic Trackpad* de Apple utilizado en los experimentos.

Se propone que *TVPad* tenga unas dimensiones similares a las del *Magic Trackpad* de Apple, además de desarrollarlo de forma que pueda ser utilizado sin apoyos, tal y como hicimos en nuestro experimento. El dispositivo podría ser utilizado para controlar la televisión, con el desarrollo previamente comentado de un interface de usuario adecuado. Adicionalmente, para permitir el acceso a los comandos más utilizados, se propone incorporar unos botones básicos en el dispositivo tal y como se muestra en el prototipo de la figura 2.

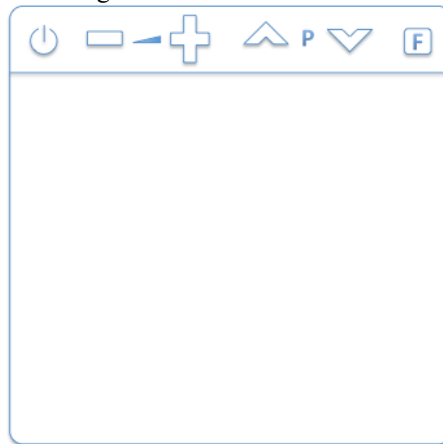


Fig. 2. Prototipo del dispositivo de interacción propuesto.

Como se aprecia en la figura 2, la mayoría de la superficie se destina a la interacción táctil, mientras que en la parte superior se dispone de una serie de botones. Estos botones permitirían realizar las siguientes acciones:

- Encender y apagar el televisor.
- Controlar el volumen del televisor.
- Cambiar de programa en el televisor o desplazarse por una lista en otros modos de funcionamiento.
- Cambiar el modo de funcionamiento del televisor, permitiendo pasar de la Televisión al modo interactivo o seleccionar una fuente de entrada alternativa.

Adicionalmente, se propone ubicar en el frontal del dispositivo un interruptor para apagar el dispositivo. Se propone utilizar Bluetooth o una tecnología similar para evitar la necesidad de apuntar al televisor durante la interacción. No obstante, si se desea utilizar una transmisión por infrarrojos puede ubicarse un transmisor en el frontal del dispositivo.

3.2 Diseño del interface de usuario

Partiendo del diseño del dispositivo apuntador presentado con anterioridad, se ha realizado un diseño conceptual de cómo sería la interacción con el televisor.

Con toda probabilidad, el televisor tendrá varios modos de funcionamiento. Con independencia de algunos botones que se pueden ubicar en el propio televisor, el cambio de modo de funcionamiento se puede controlar desde el dispositivo de interacción. En el modo “smartTV” el dispositivo podría utilizarse tal y como se hace en un ordenador convencional. Por ello, aquí se presenta solamente un diseño conceptual de cómo sería el interfaz de usuario en el modo de visionado de televisión.

En el modo de visionado de televisión, el usuario podría utilizar los botones disponibles en el *TVPad* para las acciones más comunes. Adicionalmente, cuando el usuario toca la superficie táctil del dispositivo, aparecería el cursor en pantalla además de tres recuadros tal y como se muestra en la figura 3. El recuadro superior se utilizaría como zona de control. El recuadro del lado derecho se utilizaría para controlar el volumen. Finalmente, el recuadro inferior se utilizaría como zona habilitada para el programa que se está viendo en ese momento.

En la zona de control, se propone ubicar una serie de botones para las siguientes acciones:

- Acceder a la guía electrónica de programación. La guía aparecería en pantalla y el usuario podría desplazarse por ella mediante botones y mediante gestos para las acciones de *scroll* horizontal y vertical. De igual modo, para visualizar o programar un canal, se pueden utilizar botones en pantalla.
- Acceder a la configuración del televisor. Las pantallas de configuración se podrían manipular mediante botones de selección y desplazamiento y mediante gestos.

- Otros controles que, en función de las capacidades del televisor, puedan ser necesarios en este modo de funcionamiento y que no sean particulares del canal que se está visualizando en ese momento.



Fig. 3. Diseño conceptual del interface de usuario.

Por otro lado, la barra de control de volumen permite al usuario ajustar el volumen o silenciar el televisor.

Finalmente, la zona de programa está diseñada para:

- Mostrar la información del programa que se está reproduciendo en este momento: nombre, espacio de tiempo, descripción, etc. En caso de que los datos ocupen más espacio del disponible en el recuadro, se pueden utilizar botones de desplazamiento en pantalla, además de gestos para controlar un *scroll* vertical.
- Realizar acciones sobre el programa. Mediante unos botones se pueden ofrecer al usuario acciones sobre el programa que está viendo en ese momento como las de activación de teletexto, cambio de idioma o activación/selección de subtítulos.

Si tras mostrar el cursor y los recuadros en pantalla el usuario no ha realizado ninguna acción, o han transcurrido algunos segundos sin actividad, el televisor recuperaría el modo de visionado estándar.

Con independencia de estos aspectos, en el modo de visionado estándar se plantea la posibilidad de realizar algunos gestos sobre el *TVPad* para llevar a cabo las siguientes acciones:

- Ampliar o reducir la imagen. Con un movimiento de pinza en diagonal se realizaría una operación de zoom sobre el área en la que aparece el cursor.
- Subir o bajar el volumen. Con un movimiento de desplazamiento vertical con dos dedos.
- Cambiar de modo de funcionamiento. Con un movimiento de desplazamiento horizontal con dos dedos.

6 Conclusiones y trabajos futuros

En este trabajo se presenta un novedoso sistema de interacción que persigue mejorar la experiencia de los usuarios en entornos de Televisión Digital Interactiva. Este sistema se ha pensado para alcanzar las capacidades de interacción requeridas por este tipo de aplicaciones, al mismo tiempo que se conserva la filosofía tradicional de uso de la televisión. Para ello, se presenta el diseño de un dispositivo hardware de interacción, así como un diseño conceptual de lo que sería el interface de usuario en el modo de visionado de televisión.

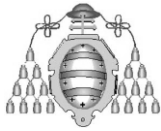
El dispositivo hardware que se presenta está basado en un *touchpad*. No obstante, se han incorporado los botones más frecuentes en un entorno de visionado de televisión, para mayor comodidad de los usuarios. Esos botones permiten acceder a las funciones básicas del televisor. Para interacciones más complejas, se utiliza la superficie táctil con un cursor en pantalla. El cursor se desplazaría sobre la pantalla permitiendo la utilización de determinados controles que se proponen en un diseño conceptual de interface de usuario. El sistema se ha diseñado tomando como referencia las valoraciones subjetivas de los usuarios de un servicio de televisión digital interactivo. Creemos que el sistema es muy fácil de utilizar por cualquier usuario. Tanto por lo comentado con anterioridad como por su coste de producción, creemos que puede ser utilizado como alternativa al mando a distancia convencional. Nos atreveríamos a decir que sería una posible solución a los problemas causados por la complejidad e inconsistencia de los interfaces de usuario actuales [24].

Como principal trabajo futuro está el de realizar una implementación real del interface de usuario y del dispositivo hardware. Esta implementación real nos permitirá realizar pruebas de usabilidad con usuarios reales, dado que las pruebas que hemos hecho hasta el momento se han limitado a tareas de escritura de texto. Estas pruebas de usuario permitirán probar algunas hipótesis planteadas en este trabajo.

Referencias

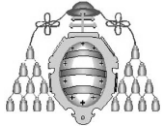
1. Spira, J.B. (2011). Internet TV: almost ready for prime time [Tools & Toys], En *IEEE Spectrum*, 48:7, 24-26.
2. European Broadcasting Union (2006). Digital Video Broadcasting (DVB); Multimedia Home Platform (MHP) Specification 1.0.3. ETSI specification ETSI ES 201 812 V1.1.2.
3. Jakob Nielsen. Remote Control Anarchy. Disponible en la dirección <http://www.useit.com/alertbox/20040607.html>, 2013
4. Iatrino, A. y Modeo, S. Text Editing in Digital Terrestrial Television: a Comparison of Three Interfaces. En Proceedings of EuroITV'06, Atenas, Grecia. 2006
5. Swanson, N.G., Galinsky, T.L., Cole, L.L., Pan, C.S. y Sauter, S.L. (1997). The impact of keyboard design on confort and productivity in a text-entry task. En *Applied Ergonomics*, 28(1), pp 9-16.
6. Rose, M.J. (1991). Keyboard operating posture and actuation forcé: implications for muscle over-use. En *Applied Ergonomics*, 22, pp 198-203
8. Geleijnse G., Aliakseyeu D, Sarroukh E. (2009). Comparing Text Entry Methods for Interactive Television Applications. En Proceedings del EuroITV'09, Leuven, Belgium.

9. Orbist, M., Bernhaupt, R. and Tscheligi, M. (2008). Interactive TV for the Home: An Ethnographic Study on Users' Requirements and Experiences. En *International Journal of Human-Computer Interaction*, 24:2, 174-196
10. Atkinson, S., Woods, V., Haslam, R.A. y Buckle, P. (2004). Using non-keyboard input devices: Interviews with users in the workplace. En *International Journal of Industrial Ergonomics*, 38, pp. 763-776
11. Keir, P.J., Bach, J.M. y Rempel, D. (1999). Effects of computer mouse design and task on carpal tunnel pressure. En *Ergonomics*, 42, pp. 1350-1360
12. Wobbrock, J.O. y Myers, B.A. (2006). Trackball text entry for people with motor impairments. En *Proceedings de CHI 2006*, pp. 479-488
13. Chaparro, A., Bohan, M., Fernández, J., Kattel, B. y Choi, S.D. (1999). Is the trackball a better input device for the older computer user?. En *Journal of Occupational Rehabilitation*, 9, 1, pp. 33-43
14. Taveira, A.D. y Choi, S.D. (2009). Review study of computer input devices and older users. En *International Journal of Human-Computer Interaction*, 25, 5, pp. 455-474
15. Choi, S., Han, J., Lee, G., Lee, N. y Lee, W. (2011). RemoteTouch: Touch-Screen-like Interaction in the TV Viewing Environment. En *Proceedings de CHI 2011*, pp. 393-402
16. Aoki, R., Maeda, A., Watanabe, T., Kobayashi, M. y Abe, M. (2010). Twist tap: text entry for TV remotes using easy-to-learn wrist motion and key operation. En *IEEE Transactions on Consumer Electronics*, 56, pp. 161–168.
17. Lennart E.N. (2010). Wiimote vs. controller: electroencephalographic measurement of affective gameplay interaction. En *Proceedings del International Academic Conference on the Future of Game Design and Technology (Futureplay '10)*
18. König, W.A., Gerken, J., Dierdorf, S. y Reiterer, H. (2009) Adaptive Pointing — Design and Evaluation of a Precision Enhancing Technique for Absolute Pointing Devices, En *Proceedings del 12th IFIP TC 13 International Conference on Human-Computer Interaction*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 658–671.
19. Berglund, A., Johansson, P. (2004). Using speech and dialogue for interactive TV navigation. En *Universal Access in the Information Society*. Vol. 3, núm. 3-4, pp 224-238
20. Haiyue, Y., Calic, J., Fernando, A. y Kondo, A. (2013). Investigation and evaluation of pointing modalities for interactive stereoscopic 3D TV, En *Proceedings del IEEE International Conference on Multimedia and Expo*, pp. 15-19
21. Bobeth, J., Schmehl, S., Kruijff, E., Deutsch, S. Y Tscheligi, M. (2012) Evaluating Performance and Acceptance of Older Adults Using Freehand Gestures for TV Menu Control. En *Proceedings del 10th European conference on Interactive tv and video*, EuroITV'12, pp. 35-44
22. Perrinet, J., Pañeda, X.G., Cabrero, S., Melendi, D., García, R. & García, V. (2011). Evaluation of Virtual Keyboards for Interactive Digital Television Applications. En *International Journal of Human-Computer Interaction*, 27:8, 703-728.
23. Barrero, A., Melendi, D., Pañeda, X.G., García, R., Cabrero, S. (2013) An empirical investigation into text input methods for interactive digital television applications. En *International Journal of Human-Computer Interaction*, DOI: 10.1080/10447318.2013.858461
24. Nielsen, J. (2013). Remote Control Anarchy. Disponible en la dirección <http://www.useit.com/alertbox/20040607.html>

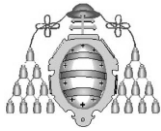


9 Bibliografía

- [1] J. W. Jeong y D. H. Lee, «Inferring search intents from remote control movement patterns: a new content search method for smart TV», *IEEE Trans. Consum. Electron.*, vol. 60, n.º 1, pp. 92-98, feb. 2014.
- [2] A. Boztas, A. R. J. Riethoven, y M. Roeloffs, «Smart TV forensics: Digital traces on televisions», *Digit. Investig.*, vol. 12, Supplement 1, pp. S72-S80, mar. 2015.
- [3] P. Cesar y K. Chorianopoulos, «The evolution of TV systems, content, and users toward interactivity», *Found. Trends Hum.-Comput. Interact.*, vol. 2, n.º 4, pp. 373-95, 2009.
- [4] «Remote Control Anarchy». [En línea]. Disponible en: <https://www.nngroup.com/articles/remote-control-anarchy/>. [Accedido: 20-mar-2017].
- [5] A. Iatrino y S. Modeo, «Text editing in digital terrestrial television: a comparison of three interfaces», *Euro ITV*, 2007.
- [6] A. Barrero, D. Melendi, X. G. Pañeda, R. García, y S. Cabrero, «Metodología de Test de Usuario y Pruebas Subjetivas para Métodos de Inserción de Texto en Aplicaciones iDTV», presentado en Jornadas de Ingeniería Telemática: JITEL 2013, Granada (Spain), 2013.
- [7] A. Barrero, D. Melendi, y X. G. Xabiel, «TVPAD: Un Novedoso Sistema de Interacción con la Televisión», en *II Jornadas Iberoamericanas de Difusión y Capacitación sobre Aplicaciones y Usabilidad de la Televisión Digital Interactiva*, 2014.
- [8] I. S. MacKenzie y R. W. Soukoreff, «Text Entry for Mobile Computing: Models and Methods, Theory and Practice», *Human-Computer Interact.*, vol. 17, n.º 2-3, pp. 147-198, 2002.
- [9] T. Bellman y S. Mackenzie, «A Probabilistic Character Layout Strategy for Mobile Text Entry», en *Graphics Interface*, 1998, pp. 168-176.
- [10] S. Zhai, M. Hunter, y B. A. Smith, «The metropolis keyboard - an exploration of quantitative techniques for virtual keyboard design», 2000, pp. 119-128.
- [11] C. R. Brewbaker, «Optimizing stylus keyboard layouts with a genetic algorithm: customization and internationalization», *Dept Comput. Sci. Iowa State Univ.*, 2008.
- [12] M. Silfverberg, I. S. MacKenzie, y P. Korhonen, «Predicting text entry speed on mobile phones», en *Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, 2000, pp. 9-16.
- [13] L. Butts y A. Cockburn, «An evaluation of mobile phone text input methods», en *Australian Computer Science Communications*, 2002, vol. 24, pp. 55-59.
- [14] I. S. MacKenzie, H. Kober, D. Smith, T. Jones, y E. Skepner, «LetterWise: prefix-based disambiguation for mobile text input», en *Proceedings of the 14th annual ACM symposium on User interface software and technology*, 2001, pp. 111-120.
- [15] I. S. MacKenzie y K. Tanaka-Ishii, «Evaluation of Text Entry Techniques», en *Text entry systems: mobility, accessibility, universality*, Morgan Kaufmann, 2007, pp. 75-101.
- [16] D. Wigdor y R. Balakrishnan, «A comparison of consecutive and concurrent input text entry techniques for mobile phones», en *Proceedings of the SIGCHI conference on Human factors in computing systems*, Viena, Austria, 2004, pp. 81-88.



- [17] A. Oniszczak y I. S. MacKenzie, «A Comparison of Two Input Methods for Keypads on Mobile Devices», en *Proceedings of the Third Nordic Conference on Human-computer Interaction*, New York, NY, USA, 2004, pp. 101–104.
- [18] I. S. MacKenzie, M. H. Lopez, y S. Castelluci, «Text Entry with the Apple iPhone and the Nintendo Wii», presentado en CHI2009, Boston, MA, USA, 2009.
- [19] K. A. Siek, Y. Rogers, y K. H. Connelly, «Fat finger worries: how older and younger users physically interact with PDAs», *Hum.-Comput. Interact.-INTERACT 2005*, pp. 267–280, 2005.
- [20] E. Jones, J. Alexander, A. Andreou, P. Irani, y S. Subramanian, «GesText: accelerometer-based gestural text-entry systems», en *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 2173–2182.
- [21] S. J. Castellucci y I. S. MacKenzie, «Unigest: text entry using three degrees of motion», en *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, 2008, pp. 3549–3554.
- [22] M. H. Lopez, S. Castelluci, y I. S. MacKenzie, «Text Entry with the Apple iPhone and the Nintendo Wii», presentado en CHI2009, 2009.
- [23] S. Choi, J. Han, G. Lee, N. Lee, y W. Lee, «RemoteTouch: touch-screen-like interaction in the tv viewing environment», *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 393–402, 2011.
- [24] Y. M. Choi y J. Li, «Usability evaluation of a new text input method for smart TVs», *J. Usability Stud.*, vol. 11, n.º 3, pp. 110–123, ene. 2016.
- [25] T. Költringer, P. Isokoski, y T. Grechenig, «TwoStick: writing with a game controller», en *Proceedings of Graphics interface 2007*, 2007, pp. 103–110.
- [26] J. O. Wobbrock, B. A. Myers, y H. H. Aung, «Writing with a Joystick: A Comparison of Date Stamp, Selection Keyboard, and EdgeWrite», en *Proceedings of Graphics Interface 2004*, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2004, pp. 1–8.
- [27] M. Ingmarsson, D. Dinka, y S. Zhai, «TNT: a numeric keypad based text input method», en *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004, pp. 639–646.
- [28] G. Geleijnse, D. Aliakseyeu, y E. Sarroukh, «Comparing text entry methods for interactive television applications», en *Proceedings of the seventh european conference on European interactive television conference*, 2009, pp. 145–148.
- [29] M. Obrist, R. Bernhaupt, y M. Tscheligi, «Interactive TV for the Home: An Ethnographic Study on Users' Requirements and Experiences», *Intl J. Human-Computer Interact.*, feb. 2008.
- [30] D. Marshall, J. C. Foster, y M. A. Jack, «User performance and attitude towards schemes for alphanumeric data entry using restricted input devices», *Behav. Inf. Technol.*, vol. 20, n.º 3, pp. 167–188, ene. 2001.
- [31] E. Lee y M. Schmidt, «Revolutionizing TV with User-Centered Design and Research», *User Exp.*, vol. 2, n.º 4, pp. 10–13, 2005.
- [32] R. Aoki, A. Maeda, T. Watanabe, M. Kobayashi, y M. Abe, «Twist tap: text entry for TV remotes using easy-to-learn wrist motion and key operation», *IEEE Trans. Consum. Electron.*, vol. 56, n.º 1, pp. 161–168, feb. 2010.
- [33] U. Gargi y R. Gossweiler, «QuickSuggest: character prediction on web appliances», en *Proceedings of the 19th international conference on World wide web*, North Carolina, USA, 2010, pp. 1249–1252.



- [34] A. J. Sporka, O. Polacek, y P. Slavik, «Comparison of two text entry methods on interactive TV», en *Proceedings of the 10th European conference on Interactive tv and video*, 2012, pp. 49–52.
- [35] T. Chen, Y. Yesilada, y S. Harper, «What input errors do you experience? Typing and pointing errors of mobile Web users», *Int. J. Hum.-Comput. Stud.*, vol. 68, n.º 3, pp. 138-157, mar. 2010.
- [36] S. Trewin y H. Pain, «Keyboard and mouse errors due to motor disabilities», *Int. J. Hum. Comput. Stud.*, vol. 50, n.º 2, pp. 109–144, 1999.
- [37] N. Henze, E. Rukzio, y S. Boll, «Observational and experimental investigation of typing behaviour using virtual keyboards for mobile devices», 2012, p. 2659.
- [38] D. A. Vega-Oliveros, D. C. Pedrosa, M. G. C. Pimentel, y R. P. de Mattos Fortes, «An approach based on multiple text input modes for interactive digital TV applications», en *Proceedings of the 28th ACM International Conference on Design of Communication*, 2010, pp. 191–198.
- [39] J. Rick, «Performance optimizations of virtual keyboards for stroke-based text entry on a touch-based tabletop», en *Proceedings of the 23rd annual ACM symposium on User interface software and technology*, New York, NY, USA, 2010, pp. 77–86.
- [40] P. D. Varcholik, J. J. LaViola Jr., y C. E. Hughes, «Establishing a baseline for text entry for a multi-touch virtual keyboard», *Int. J. Hum.-Comput. Stud.*, vol. 70, n.º 10, pp. 657-672, oct. 2012.
- [41] G. Ren y E. O’Neill, «Freehand gestural text entry for interactive TV», 2013, p. 121.
- [42] G. Ren, W. Li, y E. O’Neill, «Towards the design of effective freehand gestural interaction for interactive TV», *J. Intell. Fuzzy Syst.*, vol. 31, n.º 5, pp. 2659-2674, oct. 2016.
- [43] I. S. Mackenzie, «Fitts’ Law As a Performance Model in Human-computer Interaction», University of Toronto, Toronto, Ont., Canada, Canada, 1991.
- [44] R. Likert, «A technique for the measurement of attitudes.», *Arch. Psychol.*, 1932.
- [45] A. Barrero, D. Melendi, X. G. Pañeda, R. García, y S. Cabrero, «An empirical investigation into text input methods for interactive digital television applications», *Int. J. Hum.-Comput. Interact.*, vol. 30, n.º 4, pp. 321–341, 2014.
- [46] A. Barrero, D. Melendi, X. G. Pañeda, R. García, y S. Cabrero, «An Empirical Investigation Into Typing Errors in Interactive Digital Television Applications», *Int. J. Hum.-Comput. Interact.*, vol. 31, n.º 3, pp. 210–225, 2015.
- [47] C. J. Roe, W. H. Muto, y T. Blake, «Feedback and Key Discrimination on Membrane Keypads», *Proc. Hum. Factors Soc. Annu. Meet.*, vol. 28, n.º 3, pp. 277-281, oct. 1984.
- [48] J. R. Lewis, K. M. Potosnak, y R. L. Magyar, «Keys and Keyboards», en *Handbook of Human-Computer Interaction*, Elsevier, 1997, pp. 1285-1315.
- [49] A. Barrero, D. Melendi, X. G. Paneda, R. Garcia, L. Pozueco, y J. L. Arciniegas, «A research on typing methods for interactive Digital Television Applications», *IEEE Lat. Am. Trans.*, vol. 13, n.º 11, pp. 3612–3620, 2015.
- [50] A. Barrero, D. Melendi, X. G. Pañeda, R. García, y L. Pozueco, «Evaluation of Text Entry Methods for Interactive Digital Television Applications with Devices Alternative to Conventional Remote Controls», *Int. J. Human-Computer Interact.*, vol. 32, n.º 10, pp. 765–776, 2016.