

Apéndice G. Evaluación de campo del sistema de información y reserva ferroviaria según el tipo de usuario.

En el apartado 6.4.6. se ha descrito la evaluación del sistema de información ferroviaria al nivel general, considerando las consultas de los 105 usuarios y ofreciendo los resultados globales obtenidos. En esta evaluación han participado dos tipos diferentes de usuario: en un primer proceso de evaluación, se solicitó la colaboración de estudiantes universitarios de la Escuela de Telecomunicaciones y de otras facultades como la Facultad de Medicina. En este proceso participaron 55 personas que completaron 220 consultas (4 consultas/escenarios por persona). Posteriormente se solicitó la colaboración de empleados de la empresa RENFE consiguiendo la participación de otros 50 usuarios que completaron 135 consultas (en este caso no todos los usuarios completaron los 4 escenarios). Los resultados y las conclusiones obtenidas en uno y otro caso son bastante diferentes. Por esta razón, hemos creído conveniente incorporar este apéndice en el que se discutiera cada resultado por separado.

G.1 Evaluación en el entorno universitario.

En las tablas G-1 y G-2 se muestran los resultados obtenidos de las medidas del sistema y medidas de los cuestionarios respectivamente, para los 55 usuarios provenientes del entorno universitario.

Medida	Valor
Duración media de la consulta (segundos)	205
Nº medio de preguntas realizadas por el sistema	19,21
Nivel medio de destreza del usuario	1,91
Varianza del nivel de destreza del usuario	0,34
Nº medio de veces que se vuelve a empezar en la mitad de una consulta	0,21
Nº de veces que el usuario solicita la corrección de un dato	0,47
Duración de la negociación de la opción del tren (segundos)	57

Tabla G-1: Medidas obtenidas de las anotaciones realizadas por el sistema para los 55 usuarios provenientes del entorno universitario.

La duración media por consulta es de 205 (3 min. 25 seg.) con un número medio de 19,21 preguntas. En las 220 consultas sólo en un 6,3% el usuario acabó colgando. Esta tasa tan reducida no es representativa del funcionamiento del sistema porque estos usuarios eran conscientes de que estaban evaluando el sistema, e intentaban completar los escenarios invirtiendo el tiempo que fuera necesario.

Medida subjetiva	Valor
El sistema comprende lo que le dices.	3,8
Las respuestas del sistema son claras y concisas.	4,3
Entiendo lo que el sistema me dice.	4,5
Se accede a la información de trenes rápidamente.	3,6
El sistema es fácil de usar.	4,2
Es fácil de aprender su funcionamiento.	4,3
El sistema me ayuda durante la conversación. Tengo información de lo que está haciendo el sistema.	3,9
En caso de error la corrección fue fácil.	3,5
El sistema me hace las preguntas en un orden lógico.	4,3
En general, es un buen sistema.	3,9

Tabla G-2: Medidas subjetivas recogidas de los cuestionarios para los 55 usuarios procedentes del entorno universitario (valoración de 1 a 5).

Como se puede observar la valoración subjetiva que realizan los usuarios es francamente muy positiva superando de manera importante los resultados globales mostrados en la tabla 6-11. Los usuarios son bastante benévolo en su valoración, ofreciendo para todos los aspectos más de un 3 (en una escala de 1 a 5). Como valoración global, se ha obtenido una puntuación de 3,9.

Se puede decir que en este caso, los usuarios observaban al sistema como una alternativa válida y útil para la adquisición de información ferroviaria. Con esta puntuación los usuarios querían reconocer y a la vez apoyar el trabajo realizado. Como resumen cabe comentar la reflexión de una de las personas que evaluaron el sistema: "Funciona bastante bien. A veces tienes que tener paciencia porque se equivoca al reconocer y hay que repetir, pero en general, sí que puedes conseguir la información de forma automática. Parece útil". Los usuarios valoran positivamente que se consiga la información siendo más condescendientes con los errores que comete el sistema.

G.2 Evaluación de los empleados de RENFE.

Veamos a continuación los resultados obtenidos para la evaluación realizada por los 50 empleados de RENFE. De las 135 consultas realizadas, en 57 el usuario abandonó la llamada. Este número supone un porcentaje del 42,2%, muy superior al presentado en el apartado anterior. En este caso, únicamente se resuelven satisfactoriamente algo más de la mitad de las consultas (57,8%). En la tabla G-3 se muestran las principales medidas obtenidas de las anotaciones del sistema. Como se puede observar, estas medidas son bastante buenas, mejores incluso que las obtenidas en el caso anterior. El tiempo medio de consulta es de 182 segundos (3 min. 2 seg.) y el número medio de preguntas de 18,28.

Medida	Valor
Duración media de la consulta (segundos)	182
Nº medio de preguntas realizadas por el sistema	18,28
Nivel medio de destreza del usuario	1,86
Varianza del nivel de destreza del usuario	0,30
Nº medio de veces que se vuelve a empezar en la mitad de una consulta	0,29
Nº de veces que el usuario solicita la corrección de un dato	0,23
Duración de la negociación de la opción del tren (segundos)	60

Tabla G-3: Medidas obtenidas de las anotaciones realizadas por el sistema para los 50 empleados de la empresa RENFE.

A tenor de estos resultados se revela una mayor exigencia por parte de los usuarios. Ante cualquier error del sistema, el usuario cuelga y vuelve a llamar. De esta manera, tenemos menos llamadas completas, pero de mejor calidad (menor duración y menor número de preguntas). Otro efecto importante es la gran reducción en el número de veces que el usuario solicita la corrección de un dato (reducción a la mitad). Generalmente a los empleados de RENFE les resulta más cómodo abandonar la llamada que intentar la corrección de un dato erróneo.

En cuanto a la valoración subjetiva, los resultados se muestran en la tabla G-4.

Medida subjetiva	Valor
El sistema comprende lo que le dices.	2,3
Las respuestas del sistema son claras y concisas.	2,5
Entiendo lo que el sistema me dice.	2,5
Se accede a la información de trenes rápidamente.	2,2
El sistema es fácil de usar.	2,8
Es fácil de aprender su funcionamiento.	3,1
El sistema me ayuda durante la conversación. Tengo información de lo que está haciendo el sistema.	2,2
En caso de error la corrección fue fácil.	2,2
El sistema me hace las preguntas en un orden lógico.	2,8
En general, es un buen sistema.	2,1

Tabla G-4: Medidas subjetivas recogidas de los cuestionarios para los 50 empleados de RENFE (valoración de 1 a 5).

Los resultados obtenidos vuelven a sorprender pero en este caso negativamente. Casi ninguno de los aspectos ha conseguido más de un 3 y en la valoración global se ha obtenido un 2,1 (1,8 puntos menos que los resultados vistos en el apartado G.1).

La gran exigencia mostrada por los usuarios y la pobre valoración obtenida se debe fundamentalmente a dos causas:

1. En primer lugar, a la hora de evaluar el sistema, los empleados han tomado como punto de referencia el servicio que actualmente ofrece la empresa RENFE a través de operadores humanos. En este servicio el tiempo medio por llamada es menor de 150 segundos (2 min. 30 seg.), se ofrece una mayor variedad de servicios: reclamaciones, información sobre descuentos, venta de billetes, y además, no se cometen errores de reconocimiento, ofreciendo una mayor flexibilidad. Cuando se compara directamente un sistema automático con la atención personalizada que ofrecen los operadores, es muy difícil competir, mas aún cuando se utiliza voz sintética que identifica claramente al sistema como automático.
2. Por otro lado, la mayoría de los empleados de RENFE han ejercido como operadores de atención al público ofreciendo información y venta de billetes, bien a través de ventanillas o bien a través del teléfono. La aparición de un servicio automático capaz de ofrecer un servicio parecido, aunque muy limitado en algunos aspectos, no deja de ser cierta competencia al trabajo que desempeñan.

Con el análisis realizado en este apéndice hemos pretendido mostrar una vez más la dificultad que supone evaluar un Servidor Vocal Interactivo. Se puede observar cómo la percepción que recibe el usuario depende de su interacción con el sistema, pero también, de su relación e implicación con el servicio que está evaluando. En nuestro caso hemos considerado dos grupos de usuarios muy dispares: en primer lugar se ha analizado la opinión de un grupo de usuarios provenientes del entorno universitario donde los retos y avances tecnológicos son acogidos y apoyados en mayor medida. En segundo lugar, se ha obtenido la opinión de los empleados de RENFE que actualmente ofrecen un servicio similar con operadores humanos, con el que nos han comparado y ha sido muy difícil competir.