Informe de Evaluación De los Sistemas de Traducción de Voz a LSE y de LSE a Voz para el Dominio de la DGT



Fundación CNSE Para la supresión de las barreras de comunicación

Grupo de Tecnología del Habla (GTH). Departamento de Ingeniería Electrónica. ETSI Telecomunicación. Universidad Politécnica de Madrid.

INFORME 2009

Versión: V1.0

Fecha: 5 de marzo de 2009.

Índice

Índice		2
1. In	troducción	3
2. Pr	rocedimiento de renovación del carné de conducir	3
3. Pr	roceso de evaluación en la oficina de la DGT de Toledo	4
4. Es	scenarios y ejemplos de diálogo para cada escenario	5
5. E	valuación del sistema de traducción de voz a LSE	7
5.1.	Anotación de eventos	7
5.2.	Cambios a realizar en el software	8
5.3.	Medidas objetivas	8
5.4.	Medidas subjetivas: preguntas en los cuestionarios	9
6. E	valuación del sistema de traducción de LSE a voz	9
6.1.	Anotación de eventos	9
6.2.	Cambios a realizar en el software	10
6.3.	Medidas objetivas	10
6.4.	Medidas subjetivas: preguntas en los cuestionarios	11
7. C	uestionarios finales	11
8. R	esultados finales	14
8.1.	Evaluación del sistema de traducción de voz a LSE	14
8.2.	Evaluación del sistema de traducción de LSE a voz	18

1. Introducción

En este informe se pretende recoger las principales características de la evaluación de los sistemas desarrollados para la traducción de voz a Lengua de Signos Española (LSE) y viceversa aplicados al dominio de la renovación del carné de conducir. En primer lugar se describe el procedimiento de renovación con los papeles que son necesarios. Posteriormente se describe el proceso de evaluación, los escenarios considerados con ejemplos de diálogos, las medidas objetivas y subjetivas resultado de la evaluación, y finalmente, un resumen de resultados.

2. Procedimiento de renovación del carné de conducir

Nos vamos a centrar en la renovación del carné de conducir tipo B2 (coches) por ser el más común. Este carné hay que renovar lo cada 10 años hasta los 45, de 45 a 70 cada 5 años, y de 70 años en adelante cada 2 años.

La renovación del carné tiene como motivo más frecuente su caducidad. En este caso la documentación que hay que presentar es la siguiente:

- Impreso de solicitud que nos dan en la oficina de la Dirección General de Tráfico (DGT).
- El carné antiquo.
- El certificado médico (informe de aptitudes psicofísicas)
- DNI (o permiso de residencia)
- Una fotografía.

En el caso de que la renovación se realice por cambio de domicilio será necesario presentar un certificado de empadronamiento en el nuevo domicilio.

En todos los casos hay que pagar una tasa que ronda los 10-15 euros

En la oficina de la DGT de Toledo el procedimiento es el siguiente:

- 1. En la ventanilla de información nos dan el impreso de solicitud que debemos rellenar.
- 2. Posteriormente debemos pasar por caja para pagar la oportuna tasa.
- 3. Finalmente, con el justificante de haber pagado la tasa, el informe y el resto de documentación se pasa a una tercera ventanilla donde se deposita toda la documentación. En esta ventanilla te dan un carné provisional (con validez de tres meses) hasta que te llegue por correo el nuevo carné definitivo.

Para las tres ventanillas (información, caja y conductores) es necesario coger un número y guardar cola. En nuestro caso vamos a suponer que todas las etapas se realizan en un mismo puesto: en una mesa de despacho donde el usuario es atendido por el mismo funcionario para todos los pasos. En la siguiente figura podemos observar la preparación de la mesa para la evaluación:



Figura 1: Preparación del puesto para la interacción entre el funcionario y el usuario

3. Proceso de evaluación en la oficina de la DGT de Toledo

El proceso de evaluación con usuarios reales en la DGT de Toledo necesitó de las siguientes fases:

- 1. En una primera fase, hay que formar al funcionario en el manejo del programa de traducción de voz a LSE. Además, con el fin de garantizar la mejor tasa de reconocimiento se adaptarán los modelos acústicos a dicho locutor. Para realizar el proceso de adaptación se grabarán 100 frases pronunciadas por el funcionario que va a manejar el sistema. Esta fase requerirá al menos **1 hora** de trabajo con el funcionario que utilizará el sistema.
- 2. En segundo lugar, hay que formar a los usuarios en el manejo de la aplicación que permite traducir una secuencia de glosas (LSE) en voz. Hay que enseñar las diferentes formas de construir la secuencia y practicar generando algunas frases. Esta fase requerirá al menos 1 hora de trabajo incluyendo explicación y práctica. Esta segunda fase se puede hacer en paralelo con la primera.
- 3. En una tercera fase hay que explicar tanto a los usuarios, como al funcionario, el proceso de evaluación. Esta explicación debe incluir:
 - Descripción de los diferentes escenarios que se van a probar. Se les dará información escrita de dichos escenarios con ejemplos de diálogos traducidos a LSE.
 - Descripción de los cuestionarios que van a tener que rellenar al final de la evaluación. Se les dará información escrita de dichos cuestionario traducidos a LSE.

Esta fase necesitará de, al menos, media hora.

4. Finalmente se realiza la fase de evaluación en la que un usuario va a realizar la renovación del carné de conducir en los diferentes escenarios planteados. Esta fase estará supervisada, al menos, por un técnico de los programas, un observador externo del proceso de evaluación y un intérprete en LSE. Se estima que se necesite alrededor de media hora para completar 6 escenarios por cada uno de los usuarios que prueben el sistema.

4. Escenarios y ejemplos de diálogo para cada escenario

Los diferentes escenarios que se plantean en un proceso de evaluación son situaciones reales que se deben simular con usuarios reales. En nuestro caso, se han planteado 6 escenarios, uno de ellos en los que el proceso de renovación del carné de conducir se realiza sin ninguna incidencia, y otros 5 escenarios en los que aparecen alguna incidencia o problema que impide realizar la solicitud de renovación.

ESCENARIO 1: Renovación del carné sin ninguna incidencia (CARNÉ RENOVAR PROBLEMA NO).

EJEMPLO DE DIÁLOGO:

	Castellano	LSE		
Funcionario	Hola, buenos días	HOLA BUENOS DÍAS		
Usuario	Hola	HOLA		
Usuario	Me gustaría renovar el carné	YO CARNET RENOVAR QUERER		
Funcionario	Rellene este impreso, por favor	POR-FAVOR PAPEL ESTE TU ESCRIBIR		
Funcionario	Déme la documentación	DOCUMENTACIÓN DAR-A_MI		
Funcionario	Tiene que abonar dieciocho	PRECIO DIECIOCHO COMA		
	euros con cincuenta	CINCUENTA EUROS		
Usuario	¿Tengo que hacer algo más?	ALGO MÁS?		
Funcionario	Tome este documento que será	PAPEL ESTE CARNÉ PROVISIONAL		
	su carné provisional			
Funcionario	Antes de tres meses le llegará el	CARNET TU CASA ENVIAR-A_TI MÁS-		
	nuevo carné por correo a casa	O-MENOS TRES-MESES		
Funcionario	Muy bien, ya está todo	TODO BIEN		
Usuario	Muchas gracias	MUCHAS-GRACIAS		

ESCENARIO 2: Falta la foto (FOTO FALTAR).

EJEMPLO DE DIÁLOGO:

	Castellano	LSE			
Funcionario	Hola, buenos días	HOLA BUENOS DÍAS			
Funcionario	¿Qué desea?	QUERER QUÉ?			
Usuario	Me gustaría renovar el carné	YO CARNET RENOVAR QUERER			
Funcionario	Rellene este impreso	PAPEL ESTE TU ESCRIBIR			
Funcionario	Déme la documentación	DOCUMENTACIÓN DAR-A_MI			
Funcionario	Necesito una fotografía actual	FOTO ACTUAL YO NECESITAR			
Usuario	Lo siento no la tengo	LO-SIENTO FOTO HABER-NO			
Usuario	¿Dónde hay un fotomatón?	YO FOTOS FLASH DÓNDE?			
Funcionario	Vaya allí al fotomatón para sacarse	TU FOTO FLASH IR FIN VOLVER			
	la foto rápidamente y vuelve	RÁPIDO			
Usuario	Muchas gracias	MUCHAS-GRACIAS			

ESCENARIO 3: Faltan datos en el impreso (DATOS PAPEL FALTAR). EJEMPLO DE DIÁLOGO:

	Castellano	LSE
Funcionario	Hola, buenos días	HOLA BUENOS DÍAS
Funcionario	¿Qué desea?	QUERER QUÉ?
Usuario	Me gustaría renovar el carné	YO CARNET RENOVAR QUERER
Funcionario	Rellene este impreso	PAPEL ESTE TU ESCRIBIR
Funcionario	Falta la fecha y la firma	FECHA FIRMA TU ESCRIBIR FALTA
Funcionario	Déme la documentación	DOCUMENTACIÓN DAR-A_MI
Funcionario	Tiene que abonar dieciocho euros	PRECIO DIECIOCHO COMA
	con cincuenta	CINCUENTA EUROS
Usuario	¿Tengo que hacer algo más?	ALGO MÁS?
Funcionario	No, todo está bien	NO TODO BIEN
Funcionario	Tome este documento que será su	PAPEL ESTE CARNÉ PROVISIONAL
	carné provisional	
Funcionario	Antes de tres meses le llegará el	CARNET TU CASA ENVIAR-A_TI MÁS-
	nuevo carné por correo a casa	O-MENOS TRES-MESES
Funcionario	Muy bien, ya está todo	TODO BIEN
Usuario	Muchas gracias	MUCHAS-GRACIAS
Usuario	Hasta luego	HASTA-LUEGO

ESCENARIO 4: Falta el certificado médico (PAPEL ESPECIAL MÉDICO FALTA).

EJEMPLO DE DIÁLOGO:

	Castellano	LSE
Funcionario	Hola, buenos días	HOLA BUENOS DÍAS
Funcionario	¿Qué desea?	QUERER QUÉ?
Usuario	Me gustaría renovar el carné	YO CARNET RENOVAR QUERER
Funcionario	Rellene este impreso	PAPEL ESTE TU ESCRIBIR
Funcionario	Déme la documentación	DOCUMENTACIÓN DAR-A_MI
Funcionario	Necesito el certificado médico	PAPEL ESPECIAL MÉDICO FALTA
Usuario	Lo siento no lo tengo	LO-SIENTO PAPEL HABER-NO
Usuario	¿Dónde puedo conseguirlo?	PAPEL ESPECIAL MÉDICO DÓNDE?
Funcionario	Hay un centro en la calle Velazquez número 7	CALLE VELAZQUEZ NÚMERO SIETE
Usuario	Me puede repetir	REPETIR
Funcionario	Hay un centro en la calle Velazquez número 7	CALLE VELAZQUEZ NÚMERO SIETE
Usuario	Muchas gracias	MUCHAS-GRACIAS

ESCENARIO 5: Falta el DNI (DNI FALTA).

EJEMPLO DE DIÁLOGO:

	Castellano	LSE
Funcionario	Hola, buenos días	HOLA BUENOS DÍAS
Funcionario	¿Qué desea?	QUERER QUÉ?
Usuario Me gustaría renovar el carné YO CARNET RENOVAR QUE		YO CARNET RENOVAR QUERER
Funcionario	Rellene este impreso	PAPEL ESTE TU ESCRIBIR

Funcionario Déme la documentación		DOCUMENTACIÓN DAR-A_MI
Funcionario	Necesito su deneí	TU DNI YO NECESITAR
Usuario	Me puede repetir	REPETIR
Funcionario	Necesito su deneí, es obligatorio	TU DNI NECESITAR OBLIGATORIO
Funcionario	Va a tener que volver usted otro día	LO-SIENTO TU OTRO-DÍA VENIR
Usuario	Muchas gracias	MUCHAS-GRACIAS

ESCENARIO 6: Pagar con tarjeta de crédito (TARJETA_CREDITO PAGAR). EJEMPLO DE DIÁLOGO:

	Castellano	LSE
Funcionario	Buenas	HOLA BUENOS DÍAS
Funcionario	Dígame	QUERER QUÉ?
Usuario	Me gustaría renovar el carné	YO CARNET RENOVAR QUERER
Funcionario	Rellene este impreso	PAPEL ESTE TU ESCRIBIR
Usuario	Gracias	GRACIAS
Funcionario	Déme la documentación	DOCUMENTACIÓN DAR-A_MI
Funcionario	Tiene que abonar dieciocho euros	PRECIO DIECIOCHO COMA
	con cincuenta	CINCUENTA EUROS
Usuario	¿Puedo pagar con tarjeta de crédito?	TARJETA_CRÉDITO PODER?
Funcionario	No, tiene que ser en efectivo	NO DINERO DEBER
Funcionario	Vaya allí al cajero para sacar dinero	TU CAJERO IR DINERO FIN VOLVER
	rápidamente y vuelva	RÁPIDO
Usuario	Muchas gracias	MUCHAS-GRACIAS

5. Evaluación del sistema de traducción de voz a LSE

En este epígrafe se comentará la anotación de eventos, los cambios a realizar en el software, las medidas objetivas y las medidas subjetivas a recoger tanto del funcionario como del usuario.

5.1. Anotación de eventos

Se anota cada evento en una línea. Al comienzo de cada línea se escribe la fecha y la hora, y a continuación el botón o el mensaje que se quiere apuntar. Toda la información debe ir entre corchetes. A continuación veremos un fragmento del fichero de depuración:

```
01/16/09 16:07:24: [ESCENARIO 01]
01/16/09 16:07:38: [BOTÓN RECONOCER]
01/16/09 16:07:39: [RECONOCIENDO...]
01/16/09 16:07:39: [TRADUCIENDO...]
01/16/09 16:07:39: [SINTETIZANDO...]
01/16/09 16:07:41: [SIGNANDO...]
01/16/09 16:07:41: [FICHERO:LSE_1_1.wav|hola buenos días (0.95) {934ms}|HOLA BUENOS DÍAS (0.95) {233ms}|EJEMPLOS|0.25|{1478ms}]
```

Al final de cada turno anotamos el fichero donde se ha grabado la voz, la salida del reconocedor (incluyendo la medida de confianza y el tiempo en milisegundos que ha tardado en reconocer), y la salida del sistema de traducción junto con la medida de confianza y el tiempo en milisegundos que

ha tardado en traducir. Seguidamente se incluye el tipo de traductor utilizado (EJEMPLOS, REGLAS, ESTADÍSTICO), la distancia o la tasa según el modo de traducción, y finalmente, el tiempo que ha tardado en signar (en milisegundos).

5.2. Cambios a realizar en el software

Los principales cambios realizados en el software son los siguientes:

- En el interfaz hemos incluido una lista para seleccionar el escenario que se quiera grabar y un botón que permita volcar esta información al fichero de depuración.
- Los eventos a apuntar son los siguientes:
 - Cuando damos en el botón de ESCENARIO.
 - Cuando damos a los botones de RECONOCER, TRADUCIR y ESTIMAR RUIDO.
 - Cuando se envía un mensaje a la pantalla para actualizar el estado del sistema (reconociendo..., traduciendo..., sintetizando..., signando...)
 - Al final imprimimos el fichero donde se graba la voz, la salida del reconocedor, la salida del traductor (con los tiempos en milisegundos y las medidas de confianza), el tipo de traducción utilizada (EJEMPLOS, REGLAS o ESTADÍSTICA), la distancia o tasa, y el tiempo de signado.

5.3. Medidas objetivas

Las medidas objetivas calculadas a partir de la información recogida en el fichero de depuración son las siguientes:

- Tasa de reconocimiento: porcentaje de palabras reconocidas correctamente. Esta medida requiere la trascripción de los ficheros de audio.
- Medida de confianza del sistema de reconocimiento de voz.
- Tasa de traducción: porcentaje de glosas generadas correctamente (requiere la traducción manual de las frases).
- Medida de confianza del sistema de traducción de voz.
- Tiempo en milisegundos desde que se comienza a reconocer hasta que se obtiene la traducción.
- Tiempo desde que se obtiene la traducción hasta que se termina de signar.
- Tipo de traducción (porcentaje de cada tipo).
- Número de veces que se pulsan los botones de RECONOCER, TRADUCIR,...

- Número de veces que se pulsa el botón TRADUCIR para repetir la última intervención.
- Número de turnos del funcionario para cada uno de los escenarios (número de veces que pulsa el botón RECONOCER o TRADUCIR)

5.4. Medidas subjetivas: preguntas en los cuestionarios

Las medidas sujetivas se obtuvieron de preguntas que se realizaron, tanto a los usuarios como al funcionario, en un cuestionario que rellenó al final de la prueba cada uno de ellos.

En relación con la traducción de voz a LSE, al funcionario se le han preguntado los siguientes aspectos:

- Rapidez del sistema.
- Tasa de reconocimiento percibida.
- Facilidad de uso.
- Aprendizaje de manejo del sistema.
- Uso futuro en situaciones de necesidad.
- Valoración global del sistema.

A los usuarios se les ha preguntado sobre:

- La inteligibilidad de las frases en LSE.
- La naturalidad en la representación de los signos.
- Tasa de traducción percibida.
- Uso futuro en situaciones de necesidad.
- Valoración global del sistema.

6. Evaluación del sistema de traducción de LSE a voz

En este epígrafe se comenta la anotación de eventos, los cambios a realizar en el software, las medidas objetivas y las medidas subjetivas a recoger tanto del funcionario como del usuario.

6.1. Anotación de eventos

Se anota cada evento en una línea. Al comienzo de cada línea se escribe la fecha y la hora, y a continuación el botón o el mensaje que se quiere apuntar. Toda la información debe ir entre corchetes. A continuación veremos un fragmento del fichero de depuración:

```
01/16/09 16:07:24: [BOTÓN BORRAR TODO]
01/16/09 16:07:38: [BOTÓN D]
01/16/09 16:07:42: [LISTA DNI]
01/16/09 16:07:42: [LISTA DCLIK DNI]
01/16/09 16:07:43: [SIGUIENTES RENOVAR]
```

```
01/16/09 16:07:44: [SIGUIENTES QUERER]
01/16/09 16:07:46: [BOTÓN HABLAR|DNI RENOVAR QUERER|me gustaría renovar el deneí {133ms}|EJEMPLOS|0.20|{956ms}]
```

Al final de cada turno anotamos la pulsación del botón HABLAR, la frase en LSE (secuencia de signos), la salida del sistema de traducción (incluyendo el tiempo en milisegundos que ha tardado en traducir). Posteriormente se incluye el tipo de traductor utilizado (EJEMPLOS, REGLAS o ESTADÍSTICA), la distancia o tasa, y al final, el tiempo que se tarda en sintetizar la voz.

6.2. Cambios a realizar en el software

Los principales cambios realizados en el software son los siguientes:

- En el interfaz se ha incluido una lista para seleccionar el escenario que se quiere grabar y un botón que permita volcar esta información al fichero de depuración.
- Los eventos a apuntar son los siguientes:
 - Cuando damos en el botón de ESCENARIO.
 - Cuando damos a los botones de HABLAR, LETRAS, NÚMEROS. FECHA (calendario y añadir), HORA (listas de horas y minutos, y añadir), SIGNAR, BORRAR, BORRAR TODO, LISTA (de glosas, diferenciando el clic del doble-clik), SIGUIENTES (los más probables), ÚLTIMOS (las últimas glosas utilizadas, diferenciando el clic del doble-clic), DELETREO (y FIN DELETREO), HAMNOSYS (añadir y borrar un carácter), LISTA HAMNOSYS (de hamnosys, diferenciando el clic del doble-clik), FRASES-FRECUENTES (diferenciando el clic del doble-clik).

6.3. Medidas objetivas

Las medidas objetivas que se calculan a partir de la información recogida en el fichero de depuración son las siguientes:

- Tasa de traducción: porcentaje de palabras generadas correctamente (requiere la traducción manual de las frases).
- Tiempo que se tarda en la conversión texto a voz: general y por unidad de palabra.
- Tipo de traducción (porcentaje de cada tipo).
- Tiempo en segundos y número de clicks de ratón que se tarda en especificar una secuencia de glosas: desde que se "borra todo" hasta que se "habla".
- Tiempo en segundos y número de clicks de ratón que se tarda en especificar una secuencia de glosas normalizado al número de glosas de la secuencia generada.

- Número de veces que se utilizan unos u otros métodos para especificar un signo: lista de glosas, lista de siguientes, lista de últimos, deletreo,...
- Número de turnos del usuario para cada uno de los escenarios (número de veces que pulsa el botón HABLAR)

6.4. Medidas subjetivas: preguntas en los cuestionarios

Las medidas sujetivas se obtienen de preguntas que se realizaron, tanto a los usuarios como al funcionario, en un cuestionario que deberá rellenar al final de la prueba cada uno de ellos.

En relación con la traducción de LSE a voz, al funcionario se le han preguntado los siguientes aspectos:

- La inteligibilidad de las frases habladas.
- La naturalidad del habla.
- Uso futuro en situaciones de necesidad.
- Valoración global del sistema.

A los usuarios se les ha preguntado sobre:

- Rapidez del sistema.
- Facilidad de uso.
- Cobertura: disponibilidad de signos para generar las secuencias
- Uso futuro en situaciones de necesidad.
- Valoración global del sistema.

7. Cuestionarios finales

A continuación se muestran los cuestionarios utilizados tanto para el funcionario como para los usuarios.

CUESTIONARIO PARA EL FUNCIONARIO

Datos:

Edad:____Sexo: Varón ___ Mujer___

 Frecuencia de uso del ordenador: 						
o Muy baja Baja Alta Muy Alta_						
Valore los siguientes aspectos del sistema de traducció	on d	e v	oz a	a LS	SE:	
Manager Age Age						
			A DE			
Normal	7					
To State Sta	AL	RI	EGU	JLA	ıR	BIE
Describe par - Children - Childre		4	2	2	4	E
PREGUNTA Desides deleisters	0	1	2	3	4	5
Rapidez del sistema						
Tasa de reconocimiento						
Facilidad de uso						
Rapidez de aprendizaje para el manejo del sistema						
Si no hubiera un intérprete de LSE, ¿usaría este sistema?						
V 1 1/ 61 65 11			1	1		
Valoración GLOBAL						
Comentarios adicionales:						
Valore los siguientes aspectos del sistema de traducció	ón d	e L	SE	a v	oz:	
		_	_			
Q W E R T Y U I O P Entroom						
A S D F G H J K L Ñ Z X C V B N M SPACO						
1 2 3 4 5 6 7 8 9 0			400			9
DAR ವಿವರ್ಷ-೧೯೩೮ ದರ್ಜನ್ನ DARA_MI DARA_TI ವಿಷಣ-೫೯೩೯ ದೇವ್ಯರಪ್ಪ						
SOUNCE SO			della	4		
PRACES PROCUNCTES	AL	RI	EGU	JLA	R	BIE
Desarrollado por Filanciado por Manaciado po					_	
PREGUNTA	0	1	2	3	4	5
Inteligibilidad de las frases habladas						
Naturalidad de la voz generada						
Si no hubiera un intérprete de LSE, ¿usaría este sistema?						
		•	•	•		
Valoración GLOBAL						
Comentarios adicionales:						

MUCHAS GRACIAS

CUESTIONARIO PARA EL USUARIO

Datos: ■ Edad: Conductor: Sí No ■ Sexo: Varón Mujer ■ Frecuencia de uso del ordenador: □ Nunca Poco A veces Mucho ■ Comprensión del castellano escrito: □ Baja Normal Alta ■ Comprensión de las glosas: □ Baja Normal Alta Valore los siguientes aspectos del sistema de traducció		9 V	07.5	a I «	SE.	
To The same of the						
Name of the state	S IAL	R	EG	UL	AR	BIEN
The part of the pa	0	1	2	3	1	5
Los signos son correctos	0			3	4	3
Los signos se entienden						
Los signos son naturales						
Si no hubiera un intérprete de LSE, ¿usaría este sistema?						
, ,			l	l		
NOTA FINAL						
Comentarios:						
/alore los siguientes aspectos del sistema de traducció	n d	e L	SE	a v	oz:	
Q W E R T Y U I I O P Serioss 3						
A S D F G H J J K L R FOND 2 X C V B N M 5890.0 000 000 000 000 000 000 000 000 000						
	AL	RI	EGU	JLA	R	BIEN
Filenclade per: Filenclade per: PREGUNTA	0	1	2	3	4	5
Rapidez del sistema		-	_		7	<u> </u>
Se necesitan más signos						
Rapidez de aprendizaje para el manejo del sistema						
Si no hubiera un intérprete de LSE, ¿usaría este sistema?						
· · · ·						
NOTA FINAL						

MUCHAS GRACIAS

Comentarios:

8. Resultados finales

La evaluación se llevó acabo en las instalaciones de la Dirección General de Tráfico (DGT) de Toledo. El sistema fue evaluado por 10 personas sordas que interaccionaron con 2 funcionarios de la DGT a través de nuestro sistema. Estas 10 personas probaron la mayoría de los 6 escenarios planteados con anterioridad, generando 48 diálogos entre un funcionario y una persona sorda. En relación con las principales características de los usuarios cabe comentar los siguientes aspectos:

- La edad de los usuarios cabe comentar que oscilaba entre 22 y 55 años con una media de 40,9 años.
- La mayoría son varones (6 de 10) y están en posesión del carné de conducir (8 de 10).
- En cuanto al uso del ordenador cabe comentar que lo utilizan con bastante frecuencia.
- La mayoría posee una buena comprensión del castellano escrito y de la utilización de glosas para la especificación de los signos.
- La mitad de los usuarios eran de Madrid y la otra mitad de Toledo.

En la siguiente figura podemos ver algunas de las fotos recogidas durante la evaluación:





Figura 2: Fotos de la evaluación en la Jefatura Provincial de Tráfico de Toledo

8.1. Evaluación del sistema de traducción de voz a LSE

AGENTE	MEDIDA	VALOR
	Tasa de reconocimiento	95,2%
	Tasa de traducción	91,1%
	Tiempo de reconocimiento	3,3 seg
	Tiempo de traducción	0,0013 seg
	Tiempo de signado	4,7 seg
	Traducción basada en ejemplos	94,9%
Sistema	Traducción basada en reglas	4,2%
	Traducción estadística	0,8%
	Porcentaje de turnos con reconocimiento de voz	92,4%
	Porcentaje de turnos con traducción desde texto	0%
	Porcentaje de turnos de repetición (sin reconocimiento de voz)	7,6%
	Número de turnos del funcionario por diálogo	8,4
	Número de diálogos	48

Tabla 1: Resultados de las medidas objetivas.

En relación con las medidas objetivas, los resultados se muestran en la Tabla 1. Como se puede observar se ha conseguido una tasa de reconocimiento de voz superior al 95% y una tasa de traducción del 91%. Por otro lado el funcionamiento del sistema es muy rápido con tiempos de unos pocos segundos para completar el proceso de traducción de voz a LSE: reconocimiento de la voz, la traducción de las palabras a glosas y la representación posterior de los signos con el agente animado. Estos números aseguran una calidad apropiada para poder llegar a realizar estas pruebas reales.

En cuanto a las estrategias de traducción se puede observar que en este caso la traducción basada en ejemplos ha sido la predominante lo que refleja el buen estudio lingüístico realizado en las primeras tareas del proyecto. Este estudio permitió recoger, con un alto grado de fiabilidad, las expresiones que con mayor frecuencia pronuncia el funcionario en este tipo de diálogos.

En cuanto a la tipología de turnos podemos observar que en el 7,6% de los turnos el funcionario no tuvo que volver a pronunciar la frase, fue suficiente con pulsar el botón de traducir el texto anterior y el sistema vuelve a generar los mismos signos. El porcentaje de turnos con traducción de texto es cero lo que indica que no fue necesario que el funcionario introdujera mediante el teclado la frase a traducir. En todos los casos el sistema reconoció la frase con el suficiente grado de fiabilidad como para ser traducida con pocos errores (se entendió).

En relación con las medidas subjetivas los resultados son:

AGENTE	MEDIDA	VALOR (0-5)
	Rapidez del sistema	4
	Tasa de reconocimiento	3,5
Funcionario	Facilidad de uso	3,5
Funcionano	Rapidez de aprendizaje para el manejo del sistema	3,5
	Si no hubiera un intérprete de LSE, ¿usaría este sistema?	3,5
	Valoración GLOBAL	3,5
	¿Los signos son correctos?	2,1
	Comprendo los signos	2,2
Usuario	El signado es natural	0,8
	Si no hubiera intérprete, ¿usaría el sistema?	2,0
	Valoración GLOBAL	2,2

Tabla 2: Resultados de medidas subjetivas.

La valoración del funcionario es bastante positiva ofreciendo en todos los aspectos una puntuación de 3,5. Quizás el comentario más importante es que a veces era incómodo tener la pantalla mirando hacia el usuario (ver Figura 3). Es cierto que el sistema va realimentando con voz la frase reconocida y emite un pitido cuando termina de signar, para que el funcionario continúe interviniendo, pero quizás en el futuro haya que plantear conectar una doble pantalla para permitir al funcionario observar con detalle lo que va ocurriendo.



Figura 3: Intervención de Marta (la funcionaria de la JPT de Toledo) con la pantalla del Tablet PC girada hacia el usuario.

En relación con la valoración de los usuarios podemos comentar que es bastante baja, rozando el aprobado en la valoración global. Quizás la crítica más importante recae sobre la naturalidad en la representación de los signos (con una puntuación de 0,8). En relación con este aspecto hay bastante unanimidad: el sistema requiere la inversión de un mayor esfuerzo en el diseño del agente animado para incrementar su naturalidad y expresividad.

En relación con la inteligibilidad de los signos se consigue una mayor nota (2,1 y 2,2) que pero aun así, sigue siendo baja. Los principales motivos de estas valoraciones son los siguientes:

- Sin duda la poca naturalidad / expresividad del agente animado hace que la comprensión de los signos sea menor. Para mejorar esta naturalidad es necesario seguir invirtiendo esfuerzo en mejorar la flexibilidad y movimientos del agente animado, sobre todo en la parte de la cara.
- Por otro lado, también es justo reconocer que con bastante frecuencia se produjeron discrepancias entre los propios usuarios sobre la correcta ejecución de algunos de los signos (por ejemplo, el signo FOTO, sobre la flexión del dedo índice de ambas manos o sólo de la mano derecha) o en la utilización de uno u otro signo (por ejemplo la utilización del signos FECHA en lugar de DÍA). Estas diferencias se resuelven en la práctica remarcando la expresividad o la pronunciación en el signo, características que nuestro agente animado tiene que mejorar sensiblemente. El signado realizado por el sistema se ha apoyado en el diccionario DILSE III generado por la Fundación CNSE. Estas discrepancias deben animar a seguir trabajando en la estandarización de la Lengua de Signos Española (LSE). Aunque no existen datos significativos, sí que se percibió un mayor grado de acuerdo entre las personas procedentes de Madrid.

- Otra fuente de discrepancias entre los usuarios fue la estructura de algunas frases. La LSE, al igual que cualquier otra lengua, ofrece una importante flexibilidad. Dicha flexibilidad a veces no es entendida de esta manera y algunas de las posibilidades se pueden considerar como erróneas en lugar de frases alternativas. Algunos ejemplos son:
 - En el caso de la pregunta "¿Qué desea?" que se puede traducir a LSE como QUERER QUÉ? o como TU QUERER?. El sistema utilizaba la primera de ellas aunque para algunos usuarios les parecía raro y preferían la segunda.
 - La utilización del signo CAJERO y la necesidad o no de ir acompañado por el signo BANCO o DINERO para que se entienda completamente.
 - La utilización de los signos FOTO FLASH para hacer referencia a un fotomatón en lugar de CABINA.
 - En la frase "DNI CARNET CONDUCIR LOS-DOS DAR-A_MI" también se produjo problemas al entender el signo LOS-DOS a qué hacía referencia. Se sugirió cambiarlo de posición o incorporar algún signo explicativo adicional.
- El agente animado presenta excesiva rigidez en la ejecución de algunos de los signos que hace que el ángulo de presentación del agente animado pueda influir en la percepción de algunos signos. Por ejemplo, en el caso del signo VENIR (en el que hay un movimiento que combina un desplazamiento hacia abajo y otro hacia el signante), si la orientación del avatar es perfectamente enfrentada con el usuario, los movimientos de profundidad no se perciben, observándose únicamente los movimientos verticales. Para resolver este problema se optó por girar ligeramente el agente animado para percibir todos los movimientos.
- Existe un conjunto de signos (signos deícticos: que hacen referencia a una persona, cosa o lugar situados en una posición concreta) cuya ejecución implica la señalización de un objeto en una dirección concreta, por ejemplo, cuando se hace referencia a "esta ventanilla" traducido por ESTE VENTANILLA. El signo ESTE deberá ejecutarse en la dirección donde esté situada la ventanilla. Esta situación es muy dependiente del entorno donde esté situado el sistema. Para resolver este problema se evitaron este tipo de signos ofreciendo una especificación mayor del objeto concreto: "ventanilla de conductores" en LSE "VENTANILLA ESPECÍFICO CONDUCTOR".

Todos estos motivos que dificultan la percepción de los signos suponen que el reconocimiento de voz y la traducción a glosas se han realizado sin errores. Pero, ¿cómo influye un error de traducción en la comprensión de las frases en LSE?. Pues la respuesta es que la influencia es decisiva. Si ya hay dificultades de comprensión por los motivos comentaros anteriormente, la introducción de un error en la secuencia de glosas hace que el usuario se despiste completamente y no entienda el significado de la frase:

- En el caso de que se produzca la inserción de una glosa errónea o la sustitución de una glosa por otra, la consecuencia es desastrosa. En la mayoría de los casos, el usuario rechaza completamente la frase y solicita la repetición de la misma.
- En el caso del borrado de alguna de las glosas, el resultado no es tan drástico y en algunos casos el usuario ha sido capaz de deducir el significado de la frase en LSE aun faltando algún signo.

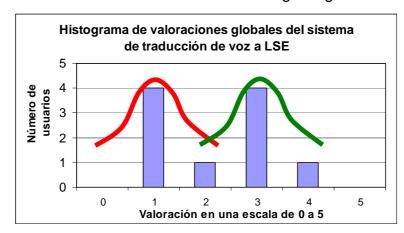


Figura 4: Distribución de número de usuarios según la valoración global del sistema.

Para terminar la discusión sobre los resultados de valoración del sistema de traducción de voz a LSE, en la Figura 4 se representa el número de usuarios en función de la valoración global ofrecida al sistema. Como se puede observar existen dos tipologías de usuarios muy diferenciadas: un grupo de usuarios que valoran positivamente (3,2) el sistema mientras que otro grupo de usuarios lo valoran muy negativamente (1,2).

En líneas generales se puede decir que aunque las valoraciones reflejan la necesidad de importantes mejoras, los comentarios recibidos sí que reflejan un interés importante por este tipo sistemas, lo que anima a seguir invirtiendo esfuerzo para abordar las mejoras necesarias.

Finalmente, una conclusión muy importante del proyecto es que se debe complementar los sistemas desarrollados (de traducción de voz a LSE) con sistemas que permitan traducir LSE a voz. Estos sistemas permitirían hacer llegar al funcionario las preguntas que desea formular el usuario sin necesidad de utilizar un intérprete.

8.2. Evaluación del sistema de traducción de LSE a voz

En relación con las medidas objetivas, los resultados han sido los siguientes:

AGENTE	MEDIDA	VALOR
Sistema	Tasa de traducción	98%
	Tiempo de traducción	0,001 seg
	Tiempo de conversión de texto a voz	1,7 seg
	Traducción basada en ejemplos	91,9%
	Traducción basada en reglas	8,1%
	Traducción estadística	0,0%

Tiempo en segundos que se tarda en especificar una secuencia de glosas.	18,5
Número de clicks de ratón que se tarda en especificar una secuencia de glosas.	8,6 clicks
Tiempo en segundos que se tarda en especificar una secuencia de glosas normalizado al número de glosas.	7,4 seg
Número de clicks de ratón que se tarda en especificar una secuencia de glosas normalizado al número de glosas.	3,3 clicks
Número de glosas por turno	2,6
Número de veces que se selecciona un signo según:	
- Lista de glosas	50,4%
- Lista de Hamnosys	0,0%
- Lista de siguientes	48,1%
- Lista de últimos	1,5%
- Fecha	0,0%
- Hora	0,0%
- Deletreo	0,0%
Número de turnos con frases frecuentes por diálogo	0,4
Número de turnos del usuario (botón HABLAR) por diálogo	4,0
Número de diálogos	48

Tabla 3. Medidas objetivas obtenidas de la traducción de LSE a voz

La muy buena tasa de traducción y los tiempos de traducción y conversión texto a voz (tiempos de unos pocos segundos) permiten la utilización del sistema en tiempo real con buenas garantías.

Al igual que ocurría en el caso del sistema de traducción de voz a LSE, se puede observar que la traducción basada en ejemplos ha sido la predominante, reflejando el buen estudio lingüístico realizado. Este estudio permitió recoger, con un alto grado de fiabilidad, las preguntas características que hacen los usuarios cuando van a renovar el carné de conducir.

En cuanto al manejo del programa, se tarda menos de 20 segundos en especificar una secuencia de glosas. Este valor es bastante bajo si consideramos que los usuarios utilizan esta herramienta por primera vez sin apenas tiempo para aprender su manejo (el uso de las frases frecuentes es muy reducido).

En relación con el uso de las diferentes opciones para especificar un signo, la gran mayoría utiliza la búsqueda basada en glosas organizadas en listas según la letra de comienzo (50,4%) y la herramienta que predice los siguientes signos (48,1%). Esta herramienta ha sido muy importante a la hora de agilizar el proceso de especificación de las glosas. Las fechas y horas apenas se han utilizado porque no han sido necesarias en los escenarios planteados. Por otro lado, tampoco ha sido necesario especificar nombres propios con lo que la opción del deletreo no se ha utilizado.

En relación con las medidas subjetivas los resultados son:

AGENTE	MEDIDA	VALOR (0-5)
Funcionario	Inteligibilidad de las frases habladas	4
	Naturalidad de la voz generada	3
	Si no hubiera un intérprete de LSE, ¿usaría este sistema?	4
	Valoración GLOBAL	4
Usuario	Rapidez del sistema	3,2

Facilidad de aprendizaje del programa	2,4
Necesita más signos para expresarse	2,6
Si no hubiera intérprete, ¿usaría el sistema?	2,8
Valoración GLOBAL	3,00

Tabla 4. Medidas subjetivas obtenidas de la traducción de LSE a voz

El funcionario ha valorado positivamente tanto la inteligibilidad como la naturalidad de la voz utilizada. Se ha utilizado el sistema de conversión de texto a voz de Loquendo en su voz masculina "Jorge".

En relación con esta herramienta los usuarios la evalúan más positivamente consiguiendo un aprobado en la valoración global. Sin embargo, la puntuación no ha sido muy elevada debido principalmente a varios factores como se desprende de los comentarios que nos han hecho llegar:

- En primer lugar apenas se le dejó tiempo para practicar con la herramienta con lo que nunca han llegado a conocer todas las posibilidades u opciones disponibles.
- La notación en glosas no está estandarizada ni extendida lo suficientemente como para ser utilizado por las personas sordas de forma inequívoca: un mismo signo puede estar representado por varias glosas (por ejemplo el signo YO y MI). Para resolver este problema se puede plantear el diseño de un interfaz que incorpore la posibilidad de seleccionar los signos en base a dibujos o animaciones que resulten más intuitivas. A sugerencia de algunos de los usuarios se debería añadir dos nuevas utilidades a la herramienta. Además estas utilidades son muy útiles y fáciles de incorporar:
 - Que todas las frases necesarias en la aplicación estén predefinidas en una lista con la frase en castellano y en LSE.
 Para usuarios que no se sienten cómodos con las glosas.
 - O Por otro lado, la posibilidad de escribir la frase directamente para aquellos usuarios que sepan escribir en castellano. Según comentan los propios usuarios, aunque no puedan escribir una frase en castellano perfecta, lo pueden hacer lo suficientemente bien como para que se entiendan dentro de un contexto concreto.
- Por último, perciben la herramienta poco natural: según ellos lo natural sería reconocer los signos cuando se están signando (mediante procesado de vídeo) en lugar de tener que especificar los signos manualmente. Si al funcionario se le reconoce la voz automáticamente, ¿por qué a ellos no se les reconoce el signado?. En relación con este último comentario tienen toda la razón del mundo pero el problema es que la tecnología de reconocimiento de vídeo todavía no está lo suficientemente madura para ser aplicada al reconocimiento de signos. Sin embargo, entre los usuarios más jóvenes se detectó un parecido del programa con la escritura de SMS en el teléfono móvil y su aceptación fue mayor.

En la Figura 5 podemos ver la foto de un ordenador con la aplicación de edición de glosas y su traducción a voz en castellano.



Figura 5: Ordenador con la aplicación de edición de glosas y traducción a voz en castellano.

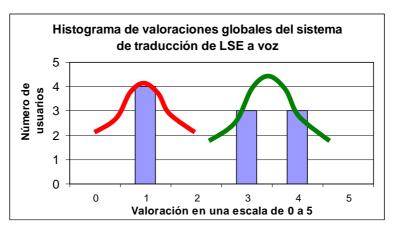


Figura 6: Distribución de número de usuarios según la valoración global del sistema.

Para terminar la discusión sobre los resultados de valoración del sistema de traducción de LSE a voz, en la Figura 6 se representa el número de usuarios en función de la valoración global ofrecida al sistema. Como se puede observar, en este caso también existen dos tipologías de usuarios muy diferenciadas: un grupo de usuarios que valoran positivamente (3,5) el sistema mientras que otro grupo de usuarios lo valoran negativamente (1).

Finalmente, una sugerencia muy interesante de los usuarios es la de incorporar un mecanismos de semáforos (en la pantalla de los ordenadores) que permita gestionar los turnos del diálogo. De forma que el funcionario no pueda hablar hasta que el usuario no haya terminado de generar sus frases y ceda su turno al funcionario, y viceversa. En la siguiente figura podemos observar la organización de los ordenadores.



Figura 7: Organización de los ordenadores durante la evaluación.