

Subtitulado en Tiempo Real. Sistemas y Tecnología

Virginia Fuentes Bueno (vfuentes@inf.uc3m.es), Israel González Carrasco (igcarras@inf.uc3m.es),
Belén Ruiz Mezcua (bruiz@inf.uc3m.es)

Departamento de Informática, Universidad Carlos III de Madrid.

Resumen. Las nuevas tecnologías de subtitulado para personas con discapacidad auditiva deben promover las oportunidades para la integración, el aprendizaje y el empleo y no constituir un conjunto de nuevas barreras. Se debe garantizar la accesibilidad de todos a la televisión ya que hoy en día constituye un medio de ocio en la sociedad, además de fomentar la participación de la vida política, social y cultural. Estas nuevas tecnologías se deben caracterizar por una mayor sencillez de uso y manejabilidad, independencia y mayor expansión de mercado. Hoy en día destacan como las más utilizadas la estenotipia computerizada, los sistemas de reconocimiento de voz, o el “live subtitling”. Sin embargo, aún queda mucho por hacer, por lo que se están desarrollando nuevos proyectos que se sustentan en estas tecnologías con el objetivo de ofrecer a los discapacitados auditivos la oportunidad de tener libre acceso a la información, a la formación y al entretenimiento y en general a cualquier situación comunicativa.

1. Introducción

Uno de los derechos fundamentales de los ciudadanos es el libre acceso a la información, formación y entretenimiento que ofrecen los medios de comunicación, y en general cualquier situación comunicativa (teatro, congresos, conferencias...), y es un elemento clave para analizar el nivel de integración de los individuos en el entorno social y cultural de un país. Por ello, desde las instituciones públicas se organizan esfuerzos dirigidos a favorecer la incorporación de los sectores menos favorecidos a las posibilidades que ofrece la denominada Sociedad de la Información.

Según los últimos datos oficiales del Instituto Nacional de Estadística, en España existen más de tres millones y medio de personas con alguna discapacidad. En términos relativos, esta cifra supone el 9% de la población total. Dentro de ese porcentaje, casi un millón de ciudadanos mayores de seis años padecen distintos grados de sordera o algún tipo de limitación auditiva.

Sea cual fuere la naturaleza de la limitación auditiva, quienes la sufren se encuentran con importantes barreras a la hora de establecer sus relaciones comunicativas, tanto desde el punto de vista interpersonal o grupal, como en lo referente al acceso a la información, entretenimiento y/o formación que ofertan los medios de comunicación audiovisuales (televisión, vídeo/DVD y cine) y, en general, cualquier situación audiovisual (conferencias, congresos, foros, teatros, sesiones formativas, información en medios de transporte...). Los actuales progresos tecnológicos abren un amplio abanico de posibilidades, a la vez que se constata la existencia de una voluntad de subrayar valores como la “accesibilidad universal” o el “diseño para todos”

El reconocimiento, profesional y social de la importancia de la subtitulación ha originado que en las cadenas de televisión, sobre todo en los países que han articulado legislaciones al respecto (Estados Unidos, Canadá, Reino Unido...), los porcentajes de programación subtitulada se hayan incrementado en los últimos años, con el objeto de lograr el 100% de la misma a corto o medio plazo. Sin embargo, esta situación no ha alcanzado los parámetros deseados en el contexto audiovisual español.

Los expertos admiten que queda mucho por hacer (establecer directrices, incrementar las ofertas de subtitulación...), y las organizaciones que defienden los intereses de las personas sordas vienen exponiendo desde hace años sus demandas. Así, se reclama la génesis de una legislación específica urgente que delimite el contacto de los discapacitados auditivos con los medios (y situaciones) de comunicación audiovisual, y que establezca una normativa concreta en materia de subtitulación. Por ejemplo: existencia de una ley de mínimos (en cuanto al porcentaje de subtitulación) y de planes de incremento progresivo de la oferta subtitulada, hasta llegar al 100% de la programación. [CESYA, 2006]

2. Subtitulado. Descripción general

Los subtítulos representan un método especialmente indicado para facilitar que los discapacitados auditivos puedan realizar el seguimiento de los programas de televisión, de las películas cinematográficas o de las reproducciones de vídeo o DVD y, en general, de cualquier experiencia audiovisual. De hecho, según algunos expertos, el subtitulado resultaría imprescindible para fomentar la accesibilidad del 90% de la población sorda que se comunica en lengua oral.

La subtitulación para las personas sordas es una simple transcripción, más o menos literal. En la articulación de estos recursos se debe pensar en ofrecer a los espectadores todas las ayudas posibles con el fin de identificar a los personajes, percibir los matices de la voz, comprender los mensajes que se envían sólo por la vía sonora, etc. Esto se traduce en otorgar colores específicos a los textos de los personajes principales, situar los rótulos justo debajo de los personajes que hablan, indicar cuándo el discurso está en off, señalar los sonidos esenciales para el seguimiento del argumento (tipos de música, disparos, ladridos, alarmas acústicas...). [CESYA, 2006]

La subtitulación es un servicio de apoyo a la comunicación que muestra en pantalla, mediante texto y gráficos, los discursos orales, la información suprasegmental y los efectos sonoros que se producen en cualquier obra audiovisual.

Consiste en la proyección sobre la pantalla de textos escritos que reproducen los mensajes hablados y gráficos asociados a los diversos sonidos de la obra audiovisual, permitiendo a las personas con limitaciones auditivas comprender la información suministrada [Forotécnico, 2005].

3. Tecnologías

Los avances técnicos y las nuevas aplicaciones y servicios de las tecnologías de la información y las comunicaciones deben permitir la existencia de oportunidades para la integración, el aprendizaje y el empleo, y no constituir un conjunto de nuevas barreras que incrementen la exclusión y la discriminación.

Al igual que ha sucedido en el mundo de los ordenadores y de los servicios on-line, se debe garantizar la accesibilidad de todos a la TV ya que ocupa una gran cantidad del tiempo de ocio en la Sociedad de la Información. La TV es un medio que resulta de fundamental importancia para que los ciudadanos puedan participar de la vida política, social y cultural.

Los problemas derivados del uso de las tecnologías son complejos cuando se pretende que estas tecnologías de la comunicación resulten accesibles para los usuarios con alguna discapacidad ya que las propias características del medio audiovisual y las posibilidades que ofrece el poder utilizar la imagen y el sonido para ayudar a interactuar con los telespectadores con discapacidades auditivas o visuales, por un lado, abren grandes facilidades para su acceso y, por otro, obligan a desarrollar protocolos que garanticen la usabilidad en el manejo de la televisión [Forotécnico, 2005].

El diseño de estas tecnologías debe caracterizarse por:

- Mayor sencillez de uso y manejabilidad
- Independencia
- Mayor expansión de mercado

3.1 Estenotipia Computerizada (Tv Digital)

Técnica para la subtitulación de espacios y retransmisiones en directo. Consiste en un sistema de recogida automatizada, de naturaleza fonético-silábica, de un discurso oral [Forotécnico, 2005].

La estenotipia es el sistema que permite a una persona entrenada en el método, escribir a la misma velocidad del discurso. Sintéticamente es un teclado manual estenotípico con inserción de papel testigo donde quedará impreso en código lo que se ha expresado verbalmente. Con ésta máquina se pueden ejecutar pulsaciones con una o más teclas que se presionan simultáneamente. La escritura posee letras alfabéticas en caracteres de imprenta. [StenotypeEspaña, 2006]

Los métodos o sistemas de estenotipia están diseñados sobre la base de un teclado de reducido número de teclas (21 la máquina Grandjean) con los que se logra obtener pulsaciones de sílabas o palabras completas. El resultado es poder coger en tiempo real íntegramente todo lo que se está diciendo en una reunión, consejo, congreso, etc. [EstenotipiaAsociación, 2006]



Figura 1. Máquina de Estenotipia

El sistema exige por parte del operador una alta concentración, excelente digitación y la capacidad de retención y memoria. El estenotipista recoge en un teclado los sonidos que recibe y un ordenador los convierte en texto escrito.



Figura 2. Máquina de estenotipia conectada a un ordenador

Existen diccionarios para traducir cómo se escribe cada palabra con los códigos de la estenotipia, para evitar los posibles errores ortográficos que puede provocar la existencia de palabras fonéticamente iguales.

Hoy por hoy es el único método que posibilita la escritura en tiempo real de la palabra hablada, muy útil para realizar subtítulos en Televisión para personas sordas, problema que no ha podido resolverlo ningún otro sistema de transcripción en directo por hoy es el único método que posibilita la escritura en tiempo real de la palabra hablada.

Sin embargo, actualmente están en desarrollo en diferentes televisiones sistemas de reconocimiento de voz para facilitar la subtitulación de los espacios en directo. Las iniciativas en este campo que se están llevando a cabo en España están funcionando perfectamente pero se echa en falta un mayor número de especialistas en nuestro país. La formación es fundamental ya que el éxito de esta técnica depende de la pericia de los estenotipistas.[Forotécnico, 2005]

3.2 Reconocimiento de Voz

3.2.1 *Historia de los sistemas ASR*

Los primeros intentos de desarrollo de sistemas de ASR datan de los años 50. Estos primeros trabajos abordaban el reconocimiento de un vocabulario reducido, del orden de 10 palabras, emitidas por un único locutor.

La década de los 60 marca el inicio de tres proyectos que han tenido gran repercusión en el área. Estos proyectos fueron desarrollados por:

- Martin (RCA Labs.), en el campo de la normalización de la voz,
- Vintsyuck (URSS), en métodos de programación dinámica,
- Reddy (CMU), que realizó los primeros trabajos en reconocimiento de voz continua.

En la década de los 70 se hicieron viables los sistemas de reconocimiento de palabras aisladas. En ésta década, los sistemas de reconocimiento estaban basados en los métodos de programación dinámica.

Sin embargo, en los 80 se produce un desplazamiento de estos métodos en favor de los modelos ocultos de Markov (HMM: Hidden Markov Models), ampliamente utilizados en la actualidad. También se comienzan a utilizar algunas aproximaciones basadas en redes neuronales.

En la actualidad existen varios sistemas de ASR, algunos de los cuales están ya siendo comercializados. Se pueden destacar los denominados SPHINX (CMU), BYBLOS (BBN), Dragon Dictate (Dragon Systems) y Naturally Speaking (IBM). La mayoría de estos sistemas están basados en HMM o su versión híbrida con redes neuronales y llegan a un reconocimiento del orden del 95%, en discurso continuo, para un único hablante, con un micrófono de buena calidad y en un ambiente de bajo ruido.

Los Modelos Ocultos de Markov (HMM), las Redes Neuronales (NN), o la combinación de Redes Neuronales y Modelos Ocultos de Markov (HMM + NN), permiten reconocer habla con la mayor tasa posible a nivel acústico, es decir, descodificar la señal acústica y generar la secuencia de palabras que más probablemente habría producido la secuencia de símbolos acústicos de entrada.

Los sistemas de reconocimiento automático de voz (ASR) se están incrementando al ser un sistema práctico de acceso al ordenador que permite la introducción de datos sin utilizar el teclado convencional, empleando un micrófono para introducir los datos mediante la voz [Rufiner, 2004]

Algunas de las diferentes posibilidades que ofrece el reconocimiento de voz son:

- Control de teléfonos, televisiones y otros aparatos del hogar.
- Mejora de la comunicación para personas con dificultades en el habla.
- Mejora de las telecomunicaciones para personas sordas.

3.2.2 *Etapas de un sistema de ASR*

La estructura general de los sistemas de ASR tiene esencialmente tres módulos o etapas [Rufiner, 2004]:

2.2.2.1 *Procesamiento o análisis del habla (front-end):*

En esta etapa se realiza algún tipo de análisis de la señal de voz en términos de la evolución temporal de parámetros espectrales (previa conversión analógica/digital de la señal). Esto tiene posibilidad de limpiar y reducir la dimensión de los patrones para facilitar su clasificación.

Los métodos de análisis espectral son generalmente considerados el núcleo de la etapa de procesamiento de señales en un sistema de ASR, junto con enfoques basados en otro tipo de representaciones más recientes.

Como ejemplo de técnicas de procesamiento del habla se pueden nombrar, entre otros [Rufiner, 2004]:

- Fourier (STFT, Short Time Fourier Transform) dentro del Análisis Tradicional

- Filtrado óptimo probabilístico (POF, Probabilistic Optimum Filtering): que constituye una técnica de limpieza de ruido en el dominio de las características extraídas de la señal de habla continua.

2.2.2.2 Clasificación de unidades fonéticas o modelo acústico

Esta etapa clasifica o identifica los segmentos de voz ya procesados con símbolos fonéticos (fonemas, dífonos o sílabas). El objetivo de esta etapa consiste, por tanto, en obtener una representación de la señal de voz como una cadena de símbolos asociados con los eventos acústico-fonéticos.

Para ello existen varias alternativas, como por ejemplo [Rufiner, 2004]:

- Modelos ocultos de Markov (HMM, Hidden Markov Models) y Árboles de decisión (DT, Decision Trees) dentro de las técnicas tradicionales
- Redes neuronales artificiales

2.2.2.3 Análisis en función de reglas del lenguaje o modelo del lenguaje:

En esta etapa se pueden aprovechar las reglas utilizadas en la codificación del mensaje contenido en la señal para mejorar el desempeño del sistema y producir una transcripción adecuada. Aquí se utilizan otras fuentes de conocimiento como la ortográfica, la sintáctica, la prosódica, la semántica o la pragmática.

Entre las técnicas de análisis en función de las reglas del lenguaje se encuentran:

- Incorporación información prosódica a un sistema de ASR basado en HMM: se trata de analizar y representar formalmente aquellos elementos suprasegmentales de la expresión oral, tales como el acento, los tonos y la entonación, y se asocia a las variaciones de la frecuencia fundamental, de la duración y de la intensidad que constituyen los parámetros prosódicos físicos. La información de acentuación es incorporada por medio del modelo de lenguaje en un reconocedor estándar basado en HMM [Rufiner, 2004].

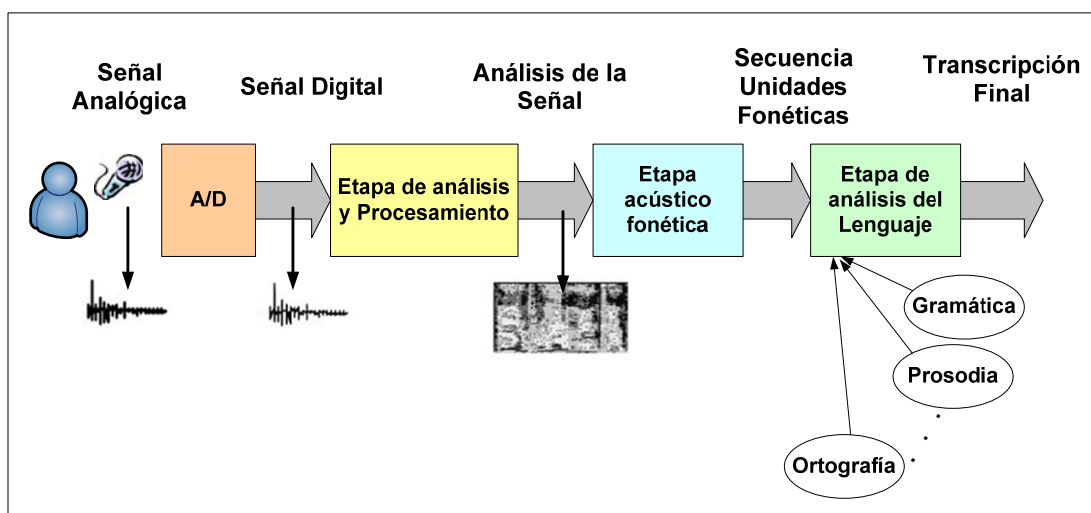


Figura 3. Componentes de un sistema ASR típico

3.3 Aspectos importantes en el reconocimiento de voz

El aprendizaje y el entrenamiento son dos factores clave a la hora de optimizar el resultado de los sistemas de reconocimiento de voz: es necesario aprender a interactuar con el sistema utilizando los comandos necesarios (que se deben aprender), el tono de voz, entonación, aprender a corregir los errores que comete el sistema, etc....

Además de los costes de aprendizaje, están estos costes cognitivos que son los que suponen la corrección de los errores producidos por el sistema. Los usuarios experimentados en el uso del ASR pueden llegar a conseguir una productividad similar a la utilización del teclado y el ratón, pero los errores cometidos por el sistema de reconocimiento de voz tienen consecuencias diferentes a los cometidos con el teclado.

La corrección posterior al dictado permite que el usuario se centre en la ejecución de la tarea principal (composición del texto) y posteriormente realiza la corrección. Este sistema parece ser que es más utilizado por personas habituadas al ASR.

Aunque los sistemas de reconocimiento de voz han mejorado considerablemente durante la última década aún existen algunos problemas técnicos o barreras como:

- La necesidad de un ordenador con más de potencia de lo habitual y esto supone un mayor coste.
- La compatibilidad del ASR con otros programas informáticos utilizados normalmente por los usuarios ya que aún resulta incompatible con determinados programas.
- La utilización del sistema de voz tiende a “saturar” el sistema

El problema del ASR presenta una naturaleza interdisciplinaria, y para solucionarlo es necesario aplicar técnicas y conocimientos procedentes de otras áreas como: procesamiento de señales, física (acústica), reconocimiento de patrones, teoría de la información y comunicaciones, lingüística, fisiología, informática y psicología.

A pesar de los avances logrados, aún quedan muchos problemas por resolver para la aplicación masiva de los sistemas ASR [Forotécnico, 2005]

3.4 Live Subtitling (Tv Digital)

Se usa para la transmisión de subtítulos en programas dónde estos no se pueden preparar. Estos subtítulos se usan en programas que son transmitidos en texto real, como son programas de noticias, deportes y shows en directo.

Este subtitulado en tiempo real, se compone combinación de subtítulos preparados y subtitulados en vivo. Empresas como FAB o SOFTEL, trabajan en esta técnica, que soporta las siguientes funcionalidades [FAB, 2006]:

- Interfaz para sistemas de noticias (ejemplo: iNews y AvStar)
- Teclado especial para inserción de texto mejorada
- Interfaz para software de reconocimiento de voz como el Via Voice de IBM.
- Editor de texto especial para la creación y el posterior formato de subtítulos.
- Editor de texto especial dónde el texto tecleado puede ser transmitido instantáneamente con autoduración.

El proceso se compone de dos fases:

1. Preparación del entorno de subtitulado antes del comienzo programa, en estaciones de trabajo.
2. Transmisión de subtítulos durante el programa

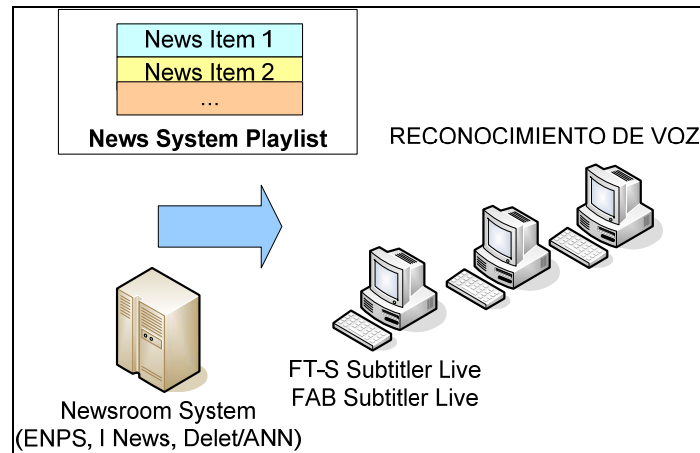


Figura 4. FAB live subtitling

4. Plataformas y Herramientas

4.1 MUSA: Multilingua Subtitling of Multimedia Content

MUSA apunta a la creación de un sistema multimodal multilingüe que convierte corrientes de audio en transcripciones de texto, genera subtítulos de estas transcripciones y luego traduce los subtítulos en otras lenguas.

MUSA operará en inglés, el francés y el griego. El estado del arte del Reconocimiento del discurso será incrementado y mejorado para encontrar los ajustes de proyecto. Un escenario innovador de la Máquina de Traducción será diseñado de modo que combine un motor de Traducción Automática con una Memoria de Traducción y un Módulo de Substitución de Términos.

La generación de subtítulos será realizada por un análisis automático de la estructura lingüística de la oración. MUSA combinará el discurso principal y tecnologías lingüísticas en una aplicación a la vida real dirigida a una audiencia audiovisual que depende de subtítulos para vencer las barreras lingüísticas.

Los retos del subtitulado que MUSA considera de mayor importancia son:

- El reto de la generación automática implica que debe haber un acuerdo entre los subtítulos, la fuente del lenguaje hablado y la imagen correspondiente.
- Generar subtítulos implica un conjunto de restricciones impuestas por el contexto visual del texto y los factores espacio-temporales
- El subtitulado no es un texto normal escrito, sino que se tiene que ajustar al texto oral

El objetivo principal del proyecto de MUSA es el desarrollo de un sistema que combina el análisis de texto avanzado, el reconocimiento vocal, la traducción automática y otras técnicas para ayudar en la preparación de subtítulos: un sistema que convierte corrientes de audio en transcripciones de texto, produce traducciones en borrador en al menos dos de lenguas y finalmente reformula el contenido para cumplir los requerimientos específicos espacio-temporales del proceso de subtitulado.

MUSA combina tres tecnologías [MUSA, 2006]:

- Reconocimiento Automático del Habla (ASR – Automatic Speech Recognition)
- Máquina de Traducción (MT – Machine Translation)
- Procesamiento de Lenguaje Natural (NLP - Natural Language Processing).

La arquitectura de la línea de producción multimedia de MUSA incluye los siguientes bloques funcionales [Piperidis, 2004]

- Una colección de material Audio Visual (AV) en los dominios de noticias de difusión y documentales, en tres lenguas: inglés, francés y griego
- Un subsistema ASR para la transcripción del audio a texto (en inglés)

- Un subsistema de Condensación de Subtítulos que produce subtítulos desde las transcripciones de audio, guiado por el objetivo de proporcionar la máxima comprensión mientras cumple con las restricciones espacio-temporales y parámetros lingüísticos.
- Un subsistema de Traducción Multilingüe que integra la Máquina de Traducción, la Memoria de Traducción y la Sustitución de Términos.

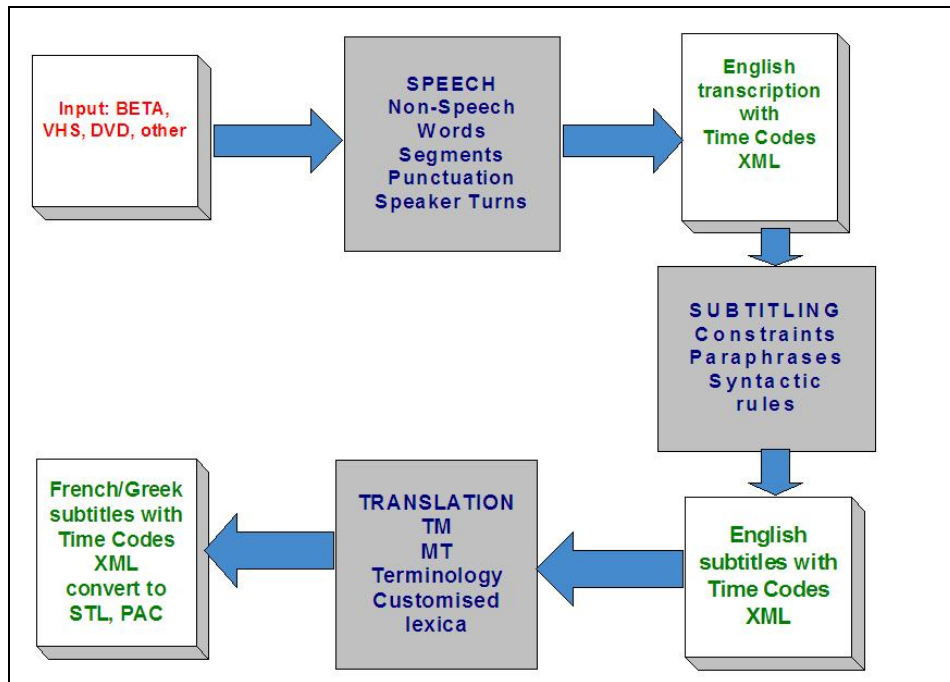


Figura 5. Arquitectura de MUSA

El módulo de reconocimiento automático del habla, desarrollado por K.U.Leuven/ESAT, incluye un estado de preproceso, el modelo acústico (AM), el modelo de lenguaje (LM), el léxico y el motor de búsqueda. La entrada al reconocedor de habla es un fichero de audio y la salida es un texto etiquetado por tiempo que es la transcripción palabra por palabra de la entrada de audio, con segmentos de transcripción correspondientes a frases. En caso de que la transcripción del programa ya esté disponible, el módulo ASR alinea el audio con la transcripción y proporciona los “timecodes” o códigos de tiempo.

El sistema de subtítulado comprende la formulación de restricciones y el módulo de cálculo, el módulo de condensación de texto de CNTS y el módulo de edición de subtítulado. (CNTS es un centro de investigación del Departamento de Lingüística de la Universidad de Antwerp en Bélgica) [CNTS, 2006]. La entrada al sistema de subtítulado es una transcripción en inglés, con códigos de tiempo, palabras, segmentos, puntuación interna y turnos de interlocutores, y la salida son subtítulos en inglés.

El subsistema de traducción comprende la Memoria de Traducción TM (TrAID) y el Motor de Traducción MT de Systran (Systran White papers, 2003). La entrada de los subsistemas de traducción son subtítulos en inglés y la salida son subtítulos en francés o griego con códigos de tiempo.

Todos los datos intercambiados entre los componentes del sistema son desarrollados en ficheros XML obedeciendo a unos DTDs predefinidos. Los subtítulos en griego y francés son lingüísticamente procesados y convertidos en formato STL (formato de salida estándar para programas de CAD o Diseño Asistido por Ordenador). Los subtítulos formateados son vistos y editados con un editor de subtítulos [Piperidis, 2005]

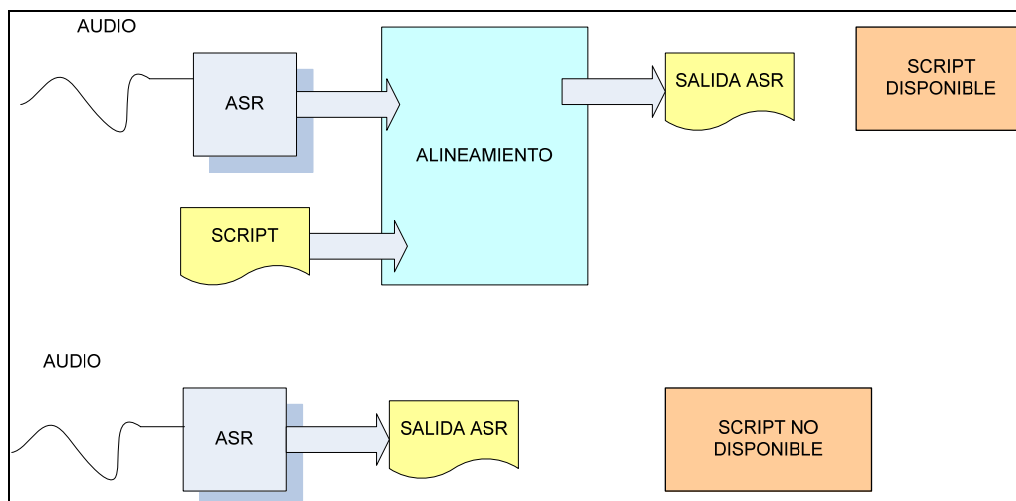


Figura 6. Funcionamiento ASR (Automatic Speech Recognition)

4.2 Iniciativas Técnicas Internacionales

4.2.1 Estados Unidos

En los Estados Unidos, NCAM (National Center for Accesible Media), que es la sección de Media Access Group de WGBH que se dedica a la investigación y al desarrollo, en coordinación con las organizaciones que establecen los estándares, con la industria y con los consumidores, se encarga de crear e implementar estándares y desarrollar técnicas para proveer accesibilidad a la plataforma multimedia, la televisión digital, la educación a distancia, el cine digital y los medios convergentes. Los proyectos de NCAM están dirigidos a facilitar la implementación, fomentar el crecimiento y establecer bases para el acceso igualitario a las nuevas tecnologías. Son ya muchos los avances en accesibilidad con los que cuenta este país, y no sólo en el ámbito de la TV digital [ForoTécnico, 2005]. Entre estos avances destacan:

4.2.1.1. MOPIX (Motion Picture Access Project)

Gracias al desarrollo del proyecto MOPIX, los teatros y salas de cine cuentan actualmente con sistemas para subtítulado y audiodescripción que los hacen completamente accesibles para el colectivo de discapacitados. Se han desarrollado para ello, sistemas patentados como [MoPix, 2006]:

- **Rear Window**

El sistema patentado de subtitulación Rear Window permite que las personas que van al cine y que tienen discapacidades auditivas, puedan ver las películas con subtítulos optativos al mismo tiempo que el resto de los espectadores, durante la proyección habitual de una película de este tipo.

El sistema Rear Window despliega una serie de subtítulos al revés en una pantalla de texto con LEDs (light-emitting diode) que es montada en la parte de atrás de la sala. En los asientos hay unos paneles transparentes que reflejan esos subtítulos por lo que se pueden ver los subtítulos superpuestos sobre la pantalla o debajo de ella. Los paneles son ajustables y portables, por lo que posibilita el subtítulado en cualquier sitio de la sala.



Figura 7. Dibujo representativo del sistema Rear Window

4.2.2 Proyectos de ámbito europeo

4.2.2.1 Proyecto VOICE

Proyecto desarrollado por el Instituto de sistemas, informática y seguridad (ISIS) del “Joint Research Centre (JRC)”. El proyecto VOICE investiga el uso del reconocimiento de voz en sistemas de conversación, conferencias, difusión de televisión y conversaciones telefónicas [ForoTécnico, 2005]

El equipo de investigadores que conforma el proyecto Voice trabaja sobre la creación de programas informáticos para el reconocimiento de la voz y en interfaces de usuario de fácil utilización que permitan convertir la voz en mensajes de texto en el ordenador y en subtítulos televisivos. Estos programas de reconocimiento de voz permiten la creación de documentos sin utilizar el teclado por lo que su utilidad se hace extensible a personas con discapacidades visuales y físicas. En uno y otro caso cabe destacar su fácil uso y aplicación a actividades cotidianas: conversación, clases, conferencias, llamadas telefónicas, televisión, etc [NoticiaVoice, 2003]

Las áreas de investigación se centran en el desarrollo de sistemas como [Voice, 2006]

- Un sistema de subtítulo de Voz a Texto
- Un sistema telefónico de Voz a Texto
- Un sistema para crear imágenes visuales: que corresponden a voces y sonidos en casa y en la televisión.

Por el momento, los técnicos de Voice ya han desarrollado un prototipo que convierte los programas de reconocimiento de voz en sistemas de subtitulación a través de la integración de hardware y software común de amplia difusión en el mercado lo que garantiza aplicaciones flexibles, un bajo coste y un fácil uso. [NoticiaVoice, 2003].

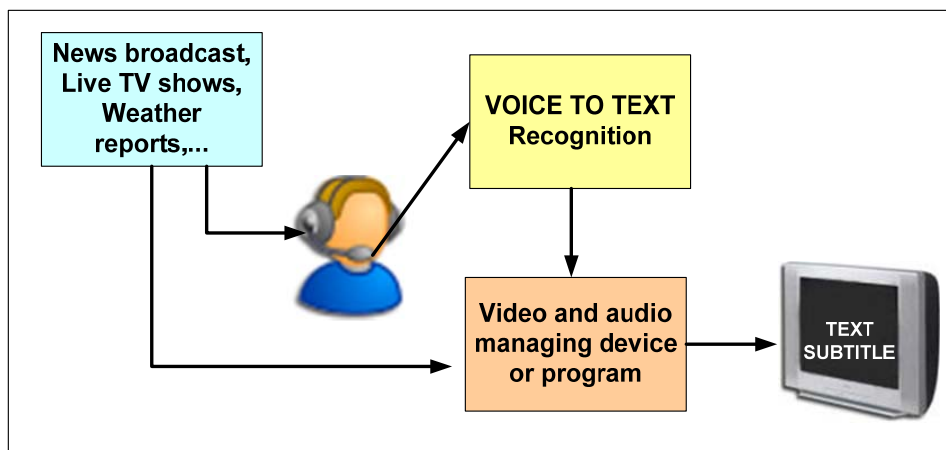


Figura 8. Funcionamiento del proyecto Voice para subtitulado

4.2.2.2 Propuesta para el desarrollo y armonización del subtitulado en la televisión europea (Dentro del proyecto VOICE):

Propuesta que se centra en el desarrollo de los servicios de accesibilidad a los contenidos y que estudia posibles mejoras en el subtitulado de difusiones de contenidos de televisión, cintas de video, DVD, aprendizaje a distancia (e-learning) y teletrabajo [Forotécnico, 2005]

4.2.3 Desarrollos técnicos a nivel nacional

4.2.3.1 Verbio ASR (Automatic Speech Recognition – Síntesis de voz)

Parte de la labor investigadora corresponde a la empresa ATLAS (Applied Technologies on Language and Speech S.L., empresa especializada en las tecnologías del habla, fundada en 1999). Es una empresa que trabaja a nivel internacional, sobretodo a nivel hispano-americano, pero en especial, en el mercado español.

Es un sistema de reconocimiento del habla que convierte, de forma automática, una locución cualquiera en un texto que se corresponda con ésta, con la máxima fiabilidad y en cualquiera de los idiomas disponibles.

Verbio ASR es una herramienta imprescindible en aplicaciones interactivas entre usuarios y sistemas automáticos. Un factor de éxito importante es poder disponer de vocabulario o gramática acorde al uso real de la aplicación [ForoTécnico, 2005]:

- Un sistema capaz de entender a cualquier a cualquier persona con gran exactitud gracias un diseño creado a partir de bases de datos de voz de miles de personas por idioma seleccionadas acorde a estrictos criterios geográficos y demográficos.
- Varios idiomas: reconocimiento en español, catalán, euskera, gallego y otras lenguas internacionales como inglés, francés, y la mayoría de variantes hispanoamericanas del español.
- Nuevo modelado acústico con mejores prestaciones para voz telefónica y, en especial, para telefonía móvil GSM, prácticamente con más uso que la propia telefonía fija, donde destaca por su robustez.
- Verbio ASR ofrece también un sistema con interpretación semántica del resultado, con lo cual, las tasas de error son muy bajas y permiten parametrizar el desarrollo de las aplicaciones o diálogos de voz ampliando las posibilidades del desarrollo de diálogos naturales.
- Gramáticas específicas básicas más habituales de reconocimiento ya incorporadas: cadenas de dígitos, deletreos etc....

- Sistema multilingüe que permite el uso simultáneo en varios idiomas.
- Sistema con interpretación semántica del resultado Posibilidad de incorporar un sistema de verificación de locutor con Verbio Speaker ID.

4.2.3.2 *Proyectos de CETTICO*

El Centro de Transferencia Tecnológica en Informática y Comunicaciones (CETTICO), es un centro sin ánimo de lucro perteneciente a la Fundación General de la Universidad Politécnica de Madrid (UPM). Para alcanzar los objetivos reseñados, el centro cuenta con la participación institucional de la UPM, a través de la Fundación General de la UPM y de la Facultad de Informática, la Organización Nacional de Ciegos de España (ONCE), la Fundación ONCE y la Dirección General de Electrónica y Nuevas Tecnologías del Ministerio de Industria y Energía.

Entre otros, CETTICO está desarrollando el proyecto "Diseño de una máquina de Estenotipia" que pretende sustituir las máquinas de estenotipia que existen en la actualidad por un ordenador estándar.

Este ordenador tiene que estar provisto de un teclado adaptado al método de trabajo y de un software capaz de convertir al castellano las diferentes combinaciones de teclas pulsadas durante la sesión. Se podría trabajar con varios métodos de estenotipia sin más que cambiar el teclado, ya que el software sería el encargado de detectar el método de Estenotipia con el cual se está trabajando antes de realizar la conversión de código. Teniendo esto en cuenta se pueden distinguir dos partes: una de ellas encargada de realizar el Hardware y otra sería la encargada de realizar el Software del proyecto [CETTICO, 2006].

5. Conclusiones

En la actualidad, por lo general, cada una de las empresas que se dedican al subtítulo para personas sordas en nuestro país posee su propia guía de estilo.

Es preciso elaborar códigos de buenas prácticas, que recogerán los requisitos de los usuarios de cara a los servicios de accesibilidad. Así pues, el cumplimiento de dichos códigos de buenas prácticas asegurará que los servicios de accesibilidad se prestan con la suficiente calidad y satisfacen a los usuarios. Como paso previo al inicio de los procedimientos de estandarización, han de realizarse estudios prenormativos.

Es previsible que los desarrollos técnicos y normativos se concentren especialmente en las siguientes áreas: elaboración de codecs adecuados para la codificación y transmisión de los servicios de accesibilidad; generación de intérpretes virtuales; uso de la tecnología del habla; revisión de las plataformas de servicios interactivos y multimedia.

La actual norma de audiodescripción se refiere a la elaboración del guión de modo que, en principio, será de aplicación a la televisión digital sin modificaciones. En cambio, la norma UNE actual de subtítulo se refiere al teletexto, por lo que es preciso adaptarla a otros soportes, teniendo en cuenta la flexibilidad que ofrecen los sistemas digitales (como la TV digital o el DVD) en cuanto a colores, identificación de parlamento por personaje, gráficos asociados a los diferentes sonidos de la banda sonora de la obra audiovisual, colocación sobre pantalla, etc.

En cuanto a la lengua de signos, deben analizarse los requisitos de cara a la interpretación en televisión digital, además de estudiar la idoneidad de las actuales bases de datos para ese medio de comunicación.

La falta de profesionales subtítulo y estenotipistas constituye un problema, ya que hay que dedicar muchas horas al subtítulo. En el momento presente, la formación de los subtítulo para las emisiones dirigidas a las personas sordas se resuelve de forma autónoma e independiente en función de los criterios que establecen cada una de las empresas que prestan este tipo de servicios. No existe una formación reglada e institucionalizada, y tampoco ningún tipo de certificación o acreditación oficial sobre la materia. Esto constituye una de las principales carencias a resolver en aras de promover el subtítulo de los productos y coyunturas audiovisuales en nuestro país. Como solución a esto, se apuesta por el

aprendizaje de estos conocimientos mediante un proceso validado por un organismo público y con la oferta, si se considera procedente, de una titulación adecuada.

La carencia de software específico de subtítulo destinado a personas sordas constituye otro problema, de modo que las empresas se ven obligadas a adaptar el software disponible en el mercado a sus necesidades.

Según FIAPAS, “aún no ha llegado ese día en que el espectador sordo pulse automáticamente la página 888 de su teletexto sin cuestionarse si lo que desea ver estará subtítulo” ya que, aunque la situación mejore progresivamente, aún hoy el 100% de horas de programación accesible constituyen una utopía [FIAPAS, 2006].

Entendiendo las barreras de comunicación como todos aquellos obstáculos que impiden a las personas sordas acceder a la información, desde la falta de subtítulos y de programas signados en la televisión y en el cine, hasta la ausencia de recursos técnicos (ayudas visuales y auditivas) e intérpretes de lengua de signos en las distintas esferas de la sociedad, son muchas las barreras de comunicación a las que se enfrentan las personas sordas. Las barreras en la comunicación persisten, y los avances no son lo suficientemente decisivos como para satisfacer las demandas de las principales organizaciones que trabajan en defensa de los derechos de las personas sordas.

En conclusión, es cierto que se están realizando proyectos e investigaciones para mejorar las tecnologías existentes de subtítulo y se están observando buenos resultados, sin embargo, se puede afirmar que todavía queda mucho que hacer en el campo de la subtítulo en general.

6. Bibliografía

- [CESYA, 2006] CESYA (Centro Español de Subtitulado y Autodescripción): <http://www.rpd.es/cesya.html>. Fecha última consulta: Junio-2006
- [StenotypeEspaña, 2006] Stenotype España: <http://www.stenotype-es.com/>. Fecha última consulta: Junio-2006
- [EstenotipiaAsociación, 2006] <http://www.estenotipiaasociacion.com/>. Fecha última consulta: Junio-2006
- [Rufiner, 2004] H.L. Rufiner, D.H. Milone, "Sistema de reconocimiento automático del habla", Revista Ciencia, Docencia y Tecnología, Año XV, no. 28, pp. 149-178, EdUNER, Concepción del Uruguay, May. 2004.
- [CETTICO, 2006] Proyectos CETTICO, <http://www.cettico.fi.upm.es/cettico/html/cettico/default.htm>. Fecha última consulta: Junio-2006
- [MoPix, 2006] MoPix : <http://ncam.wgbh.org/mopix/>. Fecha última consulta: Junio-2006
- [NoticiaVoice,2003] Noticia sobre el proyecto Voice: <http://www.manosquehablan.com.ar/noticias/2003/10/15948.php>. Fecha última consulta: Junio-2006
- [Voice, 2006] Proyecto VOICE: <http://voice.jrc.it/>. Fecha última consulta: Junio-2006
- [Piperidis, 2004] Piperidis, S., I. Demiros, P. Prokopidis, P. Vanroose, A. Hoethker, W. Daelemans, E. Sklavounou, M. Konstantinou, and Y. Karavidas. "Multimodal Multilingual Resources in the Subtitling Process" Proceedings of the 4th International Language Resources and Evaluation Conference. 2004
- [Piperidis, 2005] Piperidis, S., I. Demiros, and P. Prokopidis. "Infrastructure for a multilingual subtitle generation system" 9th International Symposium on Social Communication, Santiago de Cuba, Cuba. 2005
- [MUSA, 2004] Proyecto MUSA <http://sifnos.ilsp.gr/musa/>. Fecha última consulta: Junio-2006
- [CNTS, 2006] CNTS: research center of the Department of Linguistics of the University of Antwerp (UA) in Antwerp, Belgium. <http://www.cnts.ua.ac.be/cnts/index.php>. Fecha última consulta: Junio-2006.
- [FAB, 2006] FAB Teletext and Subtitling Systems: <http://www.fab-online.com/>. Fecha última consulta: Junio-2006
- [FIAPAS,2006] FIAPAS: Federación Española de Asociaciones de Padres y Amigos de los Sordos <http://www.fiapas.es/>. Fecha última consulta: Junio-2006.
- [ForoTécnico,2005]Foro Técnico de la Televisión Digital. Grupo de Trabajo 5 sobre accesibilidad del Foro Técnica de la Televisión Digital. Coordinado por la Subdirección General de Infraestructura y Normativa Técnica. Ministerio de Industria, Turismo y Comercio. Octubre 2005.