UNIVERSIDAD CARLOS III DE MADRID
Escuela Politécnica Superior

Universidad
Carlos III de Madrid

PROYECTO FIN DE CARRERA

---

# HERRAMIENTA DE TRANSCRIPCIÓN AUTOMÁTICA DE LOS COMUNICADOS AL PASAJE AÉREO

---

**INGENIERÍA DE TELECOMUNICACIÓN**

Departamento de Teoría de la Señal y Comunicaciones

*Autora:*
Isabel VÁZQUEZ RUFINO
100060773

*Tutora:*
Carmen PELÁEZ MORENO
*Directora:*
Ascensión GALLARDO ANTOLÍN

Leganés, 21 de marzo de 2014

*Cátedra EADS - Fundación ADECCO* con la colaboración de CESyA




*A mi padre y a mi madre.*
*A mi hermano.*

# Datos

| | |
|---|---|
| **TÍTULO**: | HERRAMIENTA DE TRANSCRIPCIÓN AUTOMÁTICA DE LOS COMUNICADOS AL PASAJE AÉREO |
| **TITULACIÓN**: | INGENIERÍA DE TELECOMUNICACIÓN |
| **AUTORA**: | ISABEL VÁZQUEZ RUFINO |
| **TUTORA**: | CARMEN PELÁEZ MORENO |
| **DIRECTORA**: | ASCENSIÓN GALLARDO ANTOLÍN |
| **DEPARTAMENTO**: | TEORÍA DE LA SEÑAL Y COMUNICACIONES |
| **CURSO**: | 2013 - 2014 |
| **UNIVERSIDAD**: | UNIVERSIDAD CARLOS III DE MADRID |

# El tribunal

| | |
|---|---|
| **PRESIDENTE**: | FERNANDO DÍAZ DE MARIA |
| **SECRETARIO**: | FERNANDO FERNÁNDEZ MARTINEZ |
| **VOCAL**: | LUIS A. PUENTE RODRÍGUEZ |

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día 21 de Marzo de 2014 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid.

# Declaración

- Este trabajo ha sido realizado para el Proyecto Fin de Carrera (PFC) de la titulación Ingeniería de Telecomunicación de esta universidad.
- Cualquier parte de este proyecto que haya sido utilizada previamente por la universidad o cualquier otra institución, ha sido claramente constatado.
- Todas las obras pertenecientes a otros autores consultadas han sido siempre atribuidas a sus autores.
- Todas las citas han sido correctamente referenciadas al autor correspondiente. A excepción de aquellas que pertenecen claramente a este trabajo.
- Todas las fuentes de ayuda han sido reconocidas.
- Cuando el proyecto se basa en el trabajo realizado por el autor junto con otros, se deja claro exactamente el trabajo de los demás y la contribución del autor.

Leganés, 21 de marzo de 2014.

# Agradecimientos

El proyecto se ha realizado en la Universidad Carlos III de Madrid (UC3M) dentro de la "*Cátedra EADS - Fundación ADECCO para la integración laboral de personas con discapacidad en entornos aeronáuticos*" con la colaboración del Centro Español de Subtitulado y Audiodescripción [1] (CESyA)[8].

[1] CESyA es un centro dependiente del Real Patronato sobre Discapacidad - Ministerio de Sanidad, Servicios Sociales e Igualdad, cuyo proyecto multidisciplinar es favorecer la accesibilidad en el entorno de los medios audiovisuales, a través de los servicios de subtitulado y audiodescripción.

# Resumen

Este proyecto consiste en conseguir la integración y autonomía de las personas con discapacidad auditiva en la industria aeronáutica mediante la realización de una herramienta de subtitulado de comunicados de la tripulación al pasaje aéreo. Se asume la ausencia de aplicaciones que faciliten esta comunicación, y por ello se desarrolla una nueva herramienta. El objetivo es transcribir y mostrar los mensajes orales a las personas con discapacidad sensorial, en particular, auditiva.

Para el diseño y desarrollo se tienen en cuenta las limitaciones del escenario. En primer lugar, el ruido procedente de la aeronave, que contamina el audio. En segundo lugar, el hecho de que la locución sea habla continua en varios idiomas, en concreto inglés y español. Por último, pero no menos importante, la transcripción se muestra como un subtítulo, semejante al utilizado por la industria televisiva y cinematográfica. El punto de partida es un motor de reconocimiento comercial (Dragon NaturallySpeaking) y se ha realizado, por una parte, su integración en la herramienta y por otra, la adaptación de sus modelos a las limitaciones anteriormente mencionadas.

Se han realizado un conjunto de pruebas para simular la locución en el interior de la cabina. Para ello se ha transcrito un audio-libro en inglés, mezclado a distintos niveles de ruido, y se han obtenido las tasas de acierto de palabra. Los resultados sirven para mostrar la capacidad de la herramienta a la hora de transcribir habla continua cuando es adaptada a las necesidades de la aplicación.

**Palabras clave:** accesibilidad, discapacidad auditiva, Reconocimiento Automático del Habla (RAH), subtitulado, aeronave, ruido, página web, robustez, adaptación.

# Abstract

The aim of this project is to achieve the integration and autonomy of people with hearing disabilities in the aviation industry by implementing a system which provides captions of the crew's announcement to airplane passengers. Assuming the absence of applications that help with this communication, a new tool is developed. The aim is to transcribe and show the speech to people with sensory disabilities, in particular, with hearing problems.

For the design and development of the tool the limitations of the environment are taken into account. First of all, the noise from the aircraft, which degrades the audio quality. Secondly, the fact that it deals with continous speech in two different languages, English and Spanish. And last but not least, the captions must be shown as a subtitle, similar to the ones used by the film and television industry. We depart from a comercial speech recognition engine (Dragon NaturallySpeaking) we have integrated into our application, on the one hand and adapted to the previously mentioned constraints, on the other.

A set of objective tests was run to simulate the speech inside of an airplane. For this purpose, an English audiobook, mixed with different noise levels, was transcribed and the word accuracy rate obtained. These results profile the performance of the tool when transcribing continuous speech with the particular needs of the application.

**Keywords:** accessibility, hearing disability, Automatic Speech Recognition (ASR), subtitle, airplane, noise, web page, robustness, adaptation.

# Índice general

# Índice de figuras

# Índice de tablas

# CAPÍTULO 1

# Introducción

Este proyecto se genera por petición de la empresa EADS con el fin de desarrollar soluciones técnicas para dotar de accesibilidad a personas con discapacidad sensorial en ámbitos y entornos relacionados con las actividades y trabajos implicados en la industria aeronáutica.

Esta investigación es realizada a través de la "*Cátedra EADS - Fundación ADECCO para la integración laboral de personas con discapacidad en entornos aeronáuticos*" y la Universidad Carlos III de Madrid en colaboración con CESyA.

## 1.1. Introducción y motivación

La creación de medidas de accesibilidad ha sido durante años únicamente asociada a entornos físicos, tales como barreras arquitectónicas. Sin embargo, en la actualidad este concepto está evolucionando, teniéndose en cuenta más tipos de discapacidades, y abarcando un amplio conjunto de medidas que han de ser tenidas en cuenta en las estrategias sociales y políticas, con el objetivo de lograr una accesibilidad universal.

La motivación de este proyecto es conseguir la integración y autonomía de las personas con discapacidad auditiva en la industria aeronáutica.

Esta motivación proviene de la concienciación social de crear entornos plenamente accesibles. De este modo, una persona con discapacidad no ve interrumpida o dificultada la realización de sus actividades, porque el espacio o entorno no sea accesible, y no le permita avanzar de forma autónoma.

El escenario en el que se presenta es el interior de una aeronave, donde la información de los tripulantes no es accesible para personas con discapacidad auditiva. Este acceso a la comunicación de los mensajes orales debe ser posible para todos los pasajeros de forma autónoma.

Las alternativas existentes que pueden encontrase son la comunicación en lengua de signos, ya sea por parte de la tripulación o por un asistente para la persona con discapacidad. O la utlización de imágenes impresas en folletos, con algún o ningún texto explicativo para el caso de la información básica (señalización de salidas de emergencia, cierre del cinturón, etc.). Pero en el primer caso exige tener una

persona formada para ello, y en el segundo no existe comunicación entre tripulante y pasajero para el resto de los mensajes orales notificados en un avión.

## 1.2. Objetivo

El objetivo es diseñar e implementar una herramienta capaz de generar una transcripción en directo del mensaje oral transmitido por la tripulación y presentarlo en formato subtítulo al pasajero. Se utilizará un motor de reconocimiento automático del habla para obtener la transcripción.

Debido a la imposibilidad de probar la herramienta en el interior de la aeronave, el escenario ha sido reproducido de manera virtual para la investigación. En el caso de obtener los resultados esperados, se podría llevar a cabo un estudio en un escenario real. Por ello se ha utilizado una grabación de ruido obtenida del interior de una cabina y se ha mezclado a distintos niveles de ruido con archivos de audio procedentes de la lectura de un libro.

Por último, se pretende realizar un ejercicio de concienciación social y hacer entender que el problema expuesto existe. En lo que respecta a la accesibilidad, se realizan continuamente grandes esfuerzos y siempre se buscan nuevas soluciones que sean de utilidad y por lo tanto obtener una mejoría de la calidad de vida en la sociedad.

## 1.3. Estructura del documento

El presente documento se compone de seis capítulos así como de anexos que ayudan a complementar el mismo.

1. *Introducción*: contiene el motivo y objetivo por el que se realiza el proyecto. Una breve introducción de del marco en el que se presenta y las alternativas existentes.

2. *Estado del arte*: contexto tecnológico y social. Se introducirá la discapacidad sensorial y la accesibilidad. Y se presentarán las tecnologías de subtitulado y de reconocedores del habla.

3. *Desarrollo del proyecto*: describe en detalle la solución técnica. Incluye una breve discusión sobre la decisión de las herramientas utilizadas.

4. *Pruebas y resultados*: conjunto de pruebas realizadas. Discusión de resultados y consideraciones correspondientes.

5. *Gestión del proyecto*: planificación y estimación de recursos y costes del proyecto.

6. *Conclusiones*: conclusiones finales del desarrollo llevado acabo. Contiene las líneas futuras de trabajo.

# CAPÍTULO 2

# Estado del arte

En este capítulo se procede a describir el contexto social, tecnológico y el entorno de operación en el que se encuentra delimitado el actual proyecto.

## 2.1. Discapacidad y accesibilidad

Las personas con discapacidad sensorial, son aquellas personas afectadas por ciertas deficiencias del oído, la vista o ambas; que encuentran limitaciones en la actividad y restricciones en la participación debido a factores socio-ambientales. Las necesidades básicas son el acceso a la información y la comunicación.

La aplicación de medidas de accesibilidad que contribuyan a la integración de las personas con discapacidad está alcanzando un mayor número de escenarios en la sociedad actual. Se ha convertido en una idea de profundo calado cultural que demuestra la eficacia que al respecto han tenido las campañas de concienciación, marcando el camino por el que deben continuar los desarrollos y despliegues de herramientas enfocadas a la mejora de la calidad de los servicios.

### 2.1.1. En la sociedad internacional

Según los datos de la Organización Mundial de la Salud (OMS), más del 5 % de la población del mundo (aproximadamente 360 millones de personas) tienen pérdida de audición. Aproximadamente 328 millones de adultos y 32 millones de niños (Figura 2.1). Aproximadamente un tercio de las personas mayores de 65 años de edad se ven afectados por este problema y más de la mitad de las personas con deficiencias auditivas están en edad laboral.

Una persona con discapacidad auditiva es aquella que no es capaz de oír igual que una persona con capacidad normal (umbrales de 25dB o mejores en ambos oídos). Esta discapacidad puede ser leve, moderada, severa o profunda y puede afectar a un oído o ambos, dificultando oír el habla coloquial o sonidos fuertes. Las causas de la pérdida de audición y la sordera pueden ser causas congénitas y causas adquiridas [37].

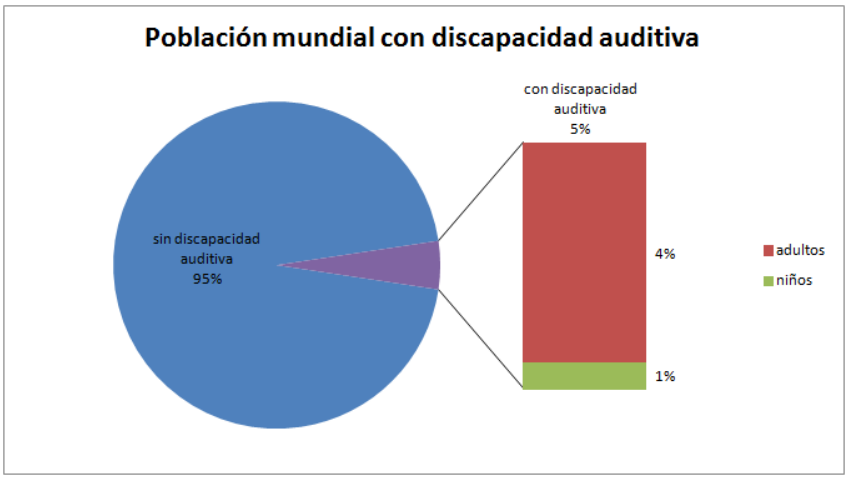



Figura 2.1: Población mundial con discapacidad auditiva

El mayor impacto que tiene la discapacidad auditiva es la reducción de la capacidad para comunicarse con los demás, pudiendo provocar aislamiento.

Este colectivo se ve cada vez más apoyado por la sociedad. Algunas de las existentes soluciones de accesibilidad se pueden agrupar en dos categorías:

1. **Apoyo y accesibilidad individual**

   Soluciones personalizadas según las características de la discapacidad.

   - Implantes cocleares y audífonos:

     Actualmente parte de las personas con discapacidad auditiva utilizan implantes cocleares o audífonos, si no tienen sordera. Fabricar audífonos asequibles y correctamente adaptados y procurar que los servicios de seguimientos sean accesibles resulta beneficioso para las personas con pérdida de audición. Sin embargo, la fabricación de audífonos satisface menos de un 10 % de las necesidades mundiales.

     En los países en vías de desarrollo, el audífono es utilizado por una relación menor de 1 por cada 40 personas que lo necesitan [37].

   - Sistemas de inducción magnética y sistemas de FM:

     Estos sistemas mitigan el ruido ambiente y salvan las dificultades que impone la distancia con el interlocutor o el soporte emisor, evitando interferencias y solventando situaciones auditivas desfavorables. Se trata de equipos de pequeño tamaño que constan de un transmisor que usa el interlocutor y un receptor del que dispone el usuario en conexión con su prótesis [13].

2. **Apoyo y accesibilidad colectiva**

   Soluciones generalizadas para todo el conjunto de la sociedad con discapacidad auditiva.

   - Sistemas de subtitulado:

     Sistemas basados en la muestra de un texto por pantalla, ya sea la locución de una conversación o la descripción del ruido ambiente. Permite al usuario acceder a la información, con el valor añadido de tener la locución representada visualmente. Se detalla en el apartado 2.2.

   - Sistemas de reconocimiento automático del habla:

     Son sistemas que permiten transcribir la locución a texto. Estos sistemas suelen estar integrados en aplicaciones accesibles al usuario. Se detalla en el apartado 2.3.

### Normativa internacional

Existen normas internacionales, no vinculantes, que representan un compromiso moral y político de los gobiernos respecto a la adopción de medidas encaminadas a lograr la igualdad de oportunidades para las personas con discapacidad. A continuación se mencionan algunas de las más representativas:

- **48/96. Normas Uniformes sobre la igualdad de oportunidades para las personas con discapacidad** [38]:

  > *"La finalidad de estas Normas es garantizar que niñas y niños, mujeres y hombres con discapacidad, en su calidad de miembros de sus respectivas sociedades, puedan tener los mismos derechos y obligaciones que los demás."*

- **La Carta de los Derechos Fundamentales de la Unión Europea** [18]:

  > *"La Carta reúne en un único documento los derechos que hasta ahora se repartían en distintos instrumentos legislativos, como las legislaciones nacionales y comunitarias, así como los Convenios internacionales del Consejo de Europa, de las Naciones Unidas (ONU) y de la Organización Internacional del Trabajo (OIT). Al dar mayor visibilidad y claridad a los derechos fundamentales, establece una seguridad jurídica dentro de la UE."*

## 2.1.2. En la sociedad española

La población con discapacidad auditiva en España, mayores de 6 años, se cifra en torno al millón de personas (Figura 2.2), según la encuesta del Instituto Nacional de Estadística (INE) [12]. De éstas, más de cien mil padecen sordera profunda. Sobre el millón de afectados, más del 90 % se comunica en lengua oral y entre el 6-8 % lo hace en lengua de signos. Esta misma proporción de usuarios de una y otra lengua se reproduce en el entorno europeo y en otros países con similar avance en sanidad y educación [13].

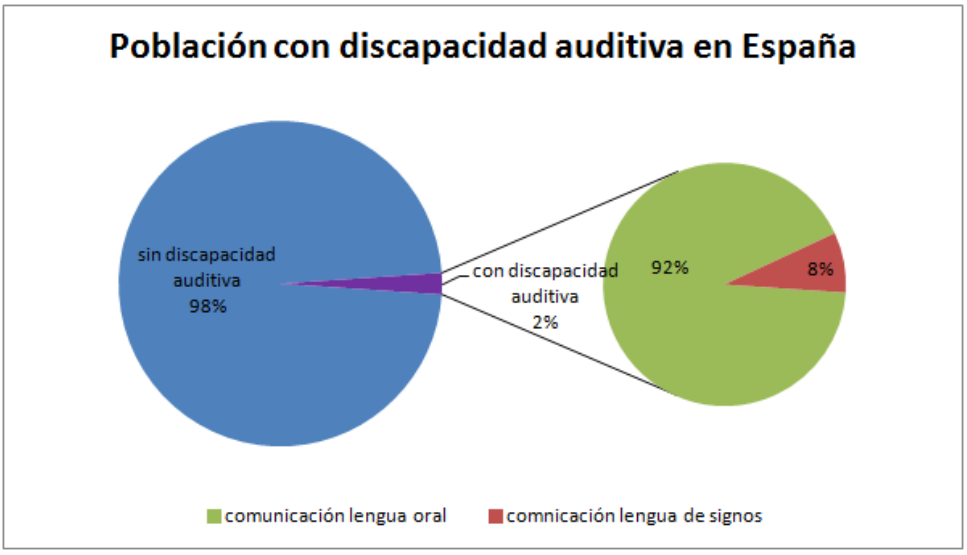



Figura 2.2: Población española con discapacidad auditiva

### Leyes y normativas españolas

Los servicios de accesibilidad, principalmente el subtitulado, la audiodescripción, la lengua de signos y la audionavegación, son elementos clave de acceso a los medios que, de otro modo, no sería posible para más de un millón de personas en España debido a su discapacidad auditiva o visual.

Para ello se cuenta con leyes y normas, donde se regulan las condiciones básicas de accesibilidad en cada uno de sus ámbitos de intervención: telecomunicaciones y sociedad de la información, espacios públicos urbanizados, infraestructuras y edificación, transportes, bienes y servicios a disposición del público y relaciones con las Administraciones Públicas.

A continuación se mencionan algunas de las medidas más importantes que existen:

- **Constitución Española** [14]:

  *"Art.9.2 Libertad e igualdad:*

  *Corresponde a los poderes públicos promover las condiciones para que la libertad y la igualdad del individuo y de los grupos en que se integra sean reales y efectivas; remover los obstáculos que impidan o dificulten su plenitud y facilitar la participación de todos los ciudadanos en la vida política, económica, cultural y social."*

  *"Art.10 Derechos de la persona:*

  1. *La dignidad de la persona, los derechos inviolables que le son inherentes, el libre desarrollo de la personalidad, el respeto a la ley y a los derechos de los demás son fundamento del orden político y de la paz social.*
  2. *Las normas relativas a los derechos fundamentales y a las libertades que la Constitución reconoce se interpretarán de conformidad con la Declaración Universal de Derechos Humanos y los tratados y acuerdos internacionales sobre las mismas materias ratificados por España."*

  *"Art.14 Igualdad ante la ley:*

  *Los españoles son iguales ante la ley, sin que pueda prevalecer discriminación alguna por razón de nacimiento, raza, sexo, religión, opinión o cualquier otra condición o circunstancia personal o social."*

  *"Art.49 Atención a los disminuidos físicos:*

  *Los poderes públicos realizarán una política de previsión, tratamiento, rehabilitación e integración de los disminuidos físicos, sensoriales y psíquicos, a los que prestarán la atención especializada que requieran y los ampararán especialmente para el disfrute de los derechos que este Título otorga a todos los ciudadanos."*

- **Ley 51/2003, de 2 de diciembre, de igualdad de oportunidades, no discriminación y accesibilidad universal de las personas con discapacidad** [15]:

  *"Esta ley tiene por objeto establecer medidas para garantizar y hacer efectivo el derecho a la igualdad de oportunidades de las personas con discapacidad, conforme a los artículos 9.2, 10, 14 y 49 de la Constitución.*

  *A estos efectos, se entiende por igualdad de oportunidades la ausencia de discriminación, directa o indirecta, que tenga su causa en una discapacidad, así como la adopción de medidas de acción positiva orientadas a evitar o compensar las desventajas de una persona con discapacidad para participar plenamente en la vida política, económica, cultural y social."*

- **Ley 7/2010, de 31 de marzo, General de Comunicación Audiovisual (LGCA)** [16]:

  > *"Esta Ley regula la comunicación audiovisual de cobertura estatal y establece las normas básicas en materia audiovisual sin perjuicio de las competencias reservadas a las Comunidades Autónomas y a los Entes Locales en sus respectivos ámbitos."*

- **Norma UNE 153010:2012** [1]:

  Estándares de calidad sobre subtitulación.

  > *"Esta norma especifica los requisitos y recomendaciones sobre la presentación de subtitulado para personas sordas o con discapacidad auditiva como medio de apoyo a la comunicación para facilitar la accesibilidad de los contenidos audiovisuales de la Sociedad de la información."*

## 2.2. Tecnología del subtítulo

El subtítulo es un elemento de comunicación procedente de la conversión en formato visual, generalmente texto, de los elementos de audio. El formato de este texto ofrece una lectura cómoda que facilita el flujo comunicativo de conversaciones entre una o más personas, así como la descripción de elementos auditivos relevantes, tales como la entonación, el ruido del entorno, el énfasis...

Mediante el subtitulado se facilita este proceso de comunicación tanto a personas con discapacidad auditiva como a las personas que desconocen el idioma original en el que se realiza la comunicación. Por supuesto es un medio accesible en múltiples espacios públicos ya que facilita la difusión de información a todo el mundo en aquellos entornos donde existe un alto nivel de ruido y donde la difusión acústica es deficiente.

### 2.2.1. Clasificación y generación del subtitulado

Es necesario diferenciar entre subtitulado simple y subtitulado adaptado. El primero muestra el texto procedente de una conversación en pantalla. El segundo no sólo muestra el texto de la conversación, sino también información relevante para su comprensión, por ejemplo si hay música en el ambiente, ruido de maquinaria, etcétera.

Se pretende con el subtítulo representar qué se dice, así como cualquier otro sonido relevante en la comunicación.

A continuación, se presenta una clasificación de los subtítulos según:

- **La audiencia que tiene acceso**:
  - Abierto: los subtítulos se muestran a un público colectivo.
  - Cerrado: los subtítulos se muestran a un público individual.
- **La información que transmiten**:
  - Narrativos: aquel subtítulo que contiene transcripción de la acción que está sucediendo en ese momento.
  - Forzados: aquel subtítulo que busca transcribir diálogos o audio del entorno en un idioma distinto al materno. Pueden aparecer traducidos o no según la necesidad.
  - Contenidos: aquel subtítulo con información adicional sin estar asociados a un audio, tales como introducciones al contexto de la obra.

    - Informativos: aquel subtítulo que añade información adicional a la obra pero no forman parte de la misma, como por ejemplo definiciones de palabras utilizadas.
    - Descriptivos: aquel subtítulo que muestra información relevante del contexto, por ejemplo el ruido de fondo.

- **Literalidad del texto**:
    - Literal: el subtítulo se ajusta a la transcripción real.
    - Adaptado: el subtítulo ha sido adaptado / modificado para facilitar su lectura manteniendo la coherencia y el contenido de la transcripción real.

- **Por su distribución**:
    - Incrustado: el subtítulo insertado en la imagen a la que acompaña y no puede ser separado.
    - Aislado: el subtítulo independiente de la imagen, que contiene las marcas de tiempo y el texto compatibles para su representación.
    - Flotante: el subtítulo en formato de imagen pero separado del flujo de vídeo y sin ofrecer cambios en la configuración.

### 2.2.2. Métodos de subtitulado

Con el avance de las tecnologías han ido apareciendo nuevas fórmulas de subtítulos en función de lo que se desee subtitular. Alguno de los métodos más utilizados son:

- **Transcripción manual**: transcripción manual de audio a texto. Método clásico, sencillo pero laborioso, que garantiza alta calidad de transcripción. Esta técnica es bastante utilizada para los subtítulos en diferido.

- **Estenotipia**: utilización de teclados especialmente diseñados para lograr una escritura rápida mediante abreviaturas. Aumenta la velocidad de transcripción con una baja probabilidad de error. Debido a la falta de personal especializado, no es un método extendido, excepto en transmisiones en directo donde se requiera precisión y velocidad.

- **Rehablado**: personas que repiten la locución y posteriormente la transcribe. Actualmente este método se ve apoyado por los sistemas de reconocimiento del habla que permiten realizar la transcripción en directo de la locución rehablada, obteniendo así una alta tasa de acierto en la transcripción.

- **Reconocedores automáticos**: sistemas automáticos de reconocimiento del habla que obtienen la transcripción directamente del audio. Pretenden sustituir al rehablador reduciendo en costes y recursos la obtención de subtítulos en directo [53].

### 2.2.3. Escenarios donde se utilizan

Existen en la actualidad múltiples contextos donde se utiliza el subtitulado como apoyo visual para la accesibilidad y comunicación. Entornos como eventos públicos, medios de transporte, congresos, escuelas...

Teniendo en cuenta los distintos escenarios donde esta tecnología puede ser utilizada, se pueden clasifican como:

- **Diferido**: cuando la generación de subtítulos no sucede en el mismo instante en el que se genera el audio [35], permitiendo así transcribir el contenido con algún elemento del entorno según la norma UNE 153010:2012 [1].

- **Directo**: cuando es necesario generar de forma inmediata la transcripción del contenido. En medios audiovisuales es necesario que la sincronización entre el audio, la imagen y el subtítulo sea sin retardo [29].

## 2.3. Tecnología de RAH

Actualmente hay una nueva ola de desarrollo de aplicaciones centradas en el consumidor que requieren del reconocimiento automático del habla, tales como la interacción con dispositivos móviles mediante la voz [41], sistemas de comunicación entre piloto y controlador para sistemas de control de tráfico aéreo [25] o sistemas de reconocimiento del habla en automóviles [21].

### 2.3.1. Funcionamiento de los reconocedores

El objetivo del reconocimiento automático del habla es detectar qué se ha dicho mediante un reconocimiento de patrones [39]. Dada una secuencia de observaciones acústicas se ha de encontrar una secuencias de palabras más probables asociadas a dicha observación. Su funcionamiento se muestra en la Figura 2.3, la cual muestra los principales bloques:

- **Modelo acústico**: define cómo se generan las observaciones a partir las palabras. Se utilizan los modelos ocultos de Markov (HMM) [40] para generarlo. Contiene el conocimiento léxico, fonológico y fonético de las palabras.

- **Modelo de lenguaje**: define cómo se concatenan las palabras para formar oraciones. Requiere de un conocimiento sintáctico y semántico del lenguaje y su gramática. Para generarlo se utilizan modelos de n-gramas.

- **Vocabulario**: define la lista de palabras y su subdivisión según se haya especificado en el uso del reconocedor.

- **Extracción de características**: consiste en obtener la información esencial para identificar sonidos y palabras. Debe ser capaz de comprimir la información y de eliminar todo aquello que sea irrelevante. Existen diversas formas de realizar la extracción de características [2]:

  - Basados en el modelo de predicción: Predicción lineal (LPC).
  - Basados en el modelo de percepción: Coeficientes ceptrales escalados MEL (MFCC) y predicción lineal perceptual (PLP).

- **Decisor**: decide el patrón que más se aproxima a la palabra.

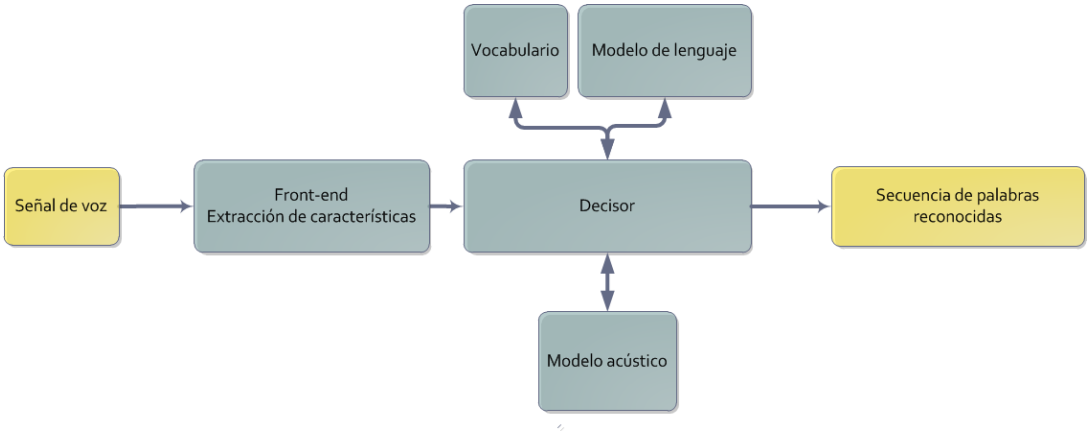



Figura 2.3: Diagrama de bloques de RAH

### 2.3.2. Utilización y dimensionamiento del RAH

Con el objetivo anteriormente mencionado se puede dimensionar la utilización del RAH adaptándolo a las necesidades de la aplicación:

1. **Dependiente / independiente del locutor**:

   Existen dos modos de funcionamiento de los RAH, dependiente o independiente del locutor [51], que vendrán fijados por la necesidad de la aplicación.

   Debido a que se compararán patrones con observaciones, aquellos reconocedores que permitan adaptación obtendrán una alta tasa de acierto para el locutor especificado.

2. **Uso de palabras aisladas o habla continua**:

   Existen mecanismos de detección de principio y final de palabra en una señal de audio, pero esta segmentación es compleja e imprecisa. Por ese motivo la utilización de los reconocedores con palabras aisladas [47] (por ejemplo comandos o dígitos), donde el usuario introduce pausas entre palabras, obtiene una mayor tasa de acierto que con habla continua [34]. En este último caso la coarticulación ha de ser modelada, aumentando la complejidad del reconocedor.

3. **Uso de un vocabulario corto, medio o largo**:

   Si el vocabulario a utilizar es muy reducido, se podrán construir modelos acústicos de palabra. La utilización de un vocabulario largo [49] requiere de subdivisiones en sílabas, fonemas, difonemas, trifonemas... para la construcción del modelo acústico.

   La ambigüedad de las palabras y la confusión acústica no son modelables y la decisión de uno u otro patrón dependerá del contexto en el que opere el reconocedor.

4. **Dependientes del entorno de operación**:

   Cuanto mayor sea la diferencia entre el entorno de entrenamiento y la adaptación del reconocedor frente al de operación, peor será la tasa de acierto, perjudicando al reconocedor.

   El entorno introduce distintas distorsiones, tales como: degradación de transmisión, variación de locutores (por edad, dialecto, acento, estado anímico...) o ruido.

5. **Uso integrado, distribuido o en red**:

   El despliegue de reconocedores en dispositivos como teléfonos móviles o tablets, que limitan en capacidad de procesado, duración de batería y memoria, requiere un diseño de la aplicación adecuado a los distintos requisitos de los aparatos.

   Existen distintos planteamientos para solventar este problema tales como: el uso de un sistema integrado, donde el servidor realiza toda la operación de reconocimiento (incluyendo captura de audio y *front-end*). El uso de un sistema distribuido, donde la captura de audio y el *front-end* se realiza en el cliente y el resto de las operaciones en el servidor. O en red, donde el cliente envía el audio al servidor y éste realiza el resto de operaciones [46].

### 2.3.3. Reconocimiento robusto en entornos ruidosos

La tecnología de los RAH ha avanzado rápidamente permitiendo su aplicación en más entornos. Para dicha gran escala de aplicaciones en el mundo real, la robustez frente al ruido se está convirtiendo en un factor muy importante en la tecnología de los RAH, ya que estos necesitan trabajar en entornos acústicos más adversos que los requeridos en tiempos pasados [28].

El problema del RAH robusto [20] (Figura 2.4) surge en el entorno acústico de entrenamiento, donde normalmente se realiza con habla limpia y por lo tanto la extracción de características devuelve parámetros acústicos limpios. En el momento en el que ese modelo entrenado se utiliza en un entorno acústico real, donde el habla contiene ruido, se produce un desajuste entre los nuevos parámetros obtenidos y los procedentes del entrenamiento.

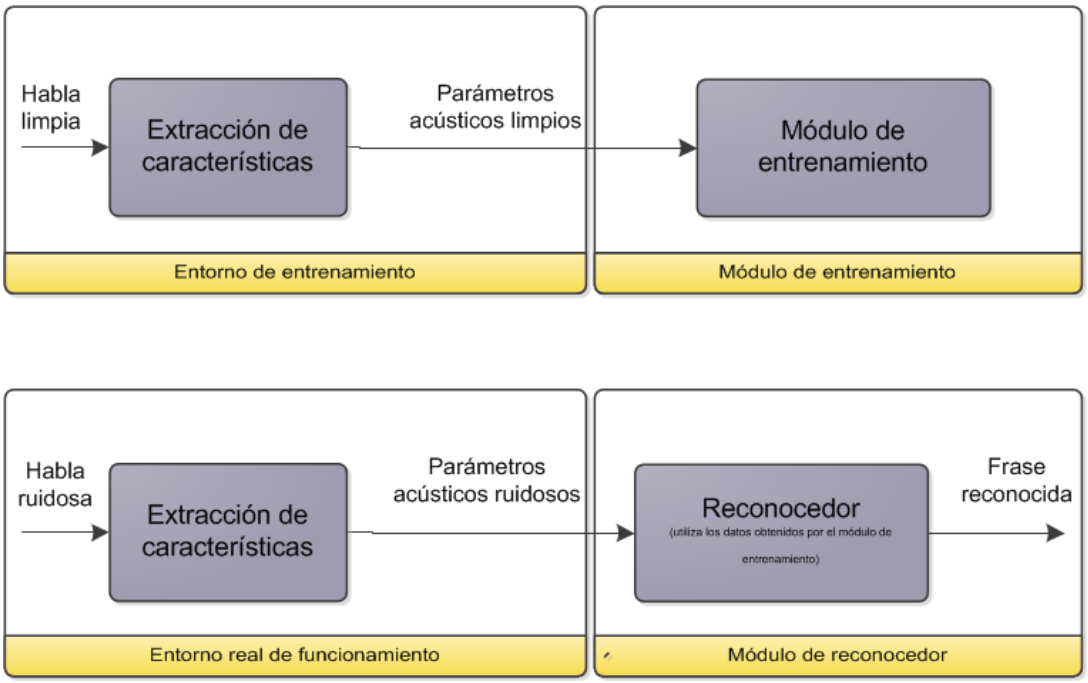



Figura 2.4: Problemática del RAH robusto

En la actualidad existen numerosos métodos propuestos para hacer robusto un reconocedor, algunos de ellos han tenido gran impacto académico o de uso comercial. A continuación se describen los más comunes y se clasifican:

### Técnicas de normalización y transformación del espacio

Estas técnicas normalizan y/o transforman el espacio donde se obtiene la parametrización del audio. A continuación se describen las más comunes:

- **Basadas en transformación de modelos**:

  Estas técnicas se basan en la modificación del modelo existente con uno nuevo (normalmente ruido) para obtener como resultado un modelo de habla ruidoso parametrizado.

  a) Combinación paralela de modelos (PMC): esta técnica utiliza por un lado el modelo del habla limpia y por otro el modelo del ruido los cuales son combinados mediante la función de disparidad[1] para posteriormente obtener los parámetros [19]. Utiliza una aproximación log-normal que convierte los estadísticos cepstrales en estadísticos espectrales lineales procedentes del habla limpia y del ruido.

  b) Vector de series de Taylor (VTS): esta técnica realiza una compensación de parámetros, tanto estáticos como dinámicos, y puede ser aplicada en entrenamiento adaptativo de ruido. Utiliza las series de Taylor para aproximar la relación entre habla ruidosa y habla limpia [45].

- **Basadas en la mejora de la señal de voz**:

  Esta técnica tiene como finalidad reducir los efectos del ruido aditivo.

  a) Substracción espectral (SS): elimina el ruido estacionario del habla mediante una substracción espectral del ruido previamente obtenido durante la grabación sin habla.

  Esta técnica suele ser utilizada en combinación con el resto de técnicas mencionadas en este apartado y en sistemas de autentificación de locutor mediante reconocedores [4].

- **Basadas en normalización de parámetros**:

  La finalidad de esta técnica es que el rango dinámico de los MFCC sea el mismo independientemente de las condiciones de ruido. El ruido produce diversos efectos sobre las distribuciones de los vectores de parámetros, tales como desplazamiento de media o cambio de varianza.

  a) Normalización en la media (CMN): consiste en eliminar la media temporal de cada componente del vector de parámetros, eliminando el efecto del canal. Esta técnica se utiliza en combinación con otros como la normalización en ganancia [52] (CGN) para mejorar los resultados.

  b) Normalización en la varianza (CVN): normaliza los vectores de parámetros para que su fdp tenga media cero y varianza unidad.

  c) Ecualización de histogramas (HEQ): es una técnica de mejorar la robustez reduciendo la diferencia entre los parámetros limpios y los ruidosos. Normaliza los parámetros acústicos para que su fdp sea gaussiana de media cero y varianza unidad. Existen mejoras de la técnica como por ejemplo adaptar HEQ para maximizar la probabilidad de características procesadas por el HEQ en el modelo acústico, con restricción de parámetros [50].

  Las técnicas basadas en normalización de parámetros ofrecen buenas prestaciones tanto para ruidos aditivos como convolucionales, siendo mejor HEQ frente a CMN y CVM.

[1] mismatch function

### Técnicas de adaptación del modelo acústico

Otro de los métodos para conseguir resultados robustos es adaptar el modelo. Estas técnicas se basan en la adaptación del modelo entrenado a las características particulares del locutor. Utilizan modelos básicos entrenados y mejoran su precisión para otro locutor utilizando sólo los datos de adaptación [11].

Un modelo adaptado ha de cumplir los siguientes requisitos:

1. Deberá mejorar la precisión del reconocedor incluso con una poca cantidad de datos para adaptar.

2. A medida que los datos para la adaptación aumenten, deberá hacer que la precisión del reconocedor se aproxime a la del modelo coincidente.

A continuación se describen los más característicos [43]:

- **Maximum a posteriori (MAP)**:

  Esta técnica permite incorporar información estimada a priori al proceso de entrenamiento. MAP es un método para estimar los parámetros [42].

- **Regresión lineal (MLLR)**:

  Esta técnica proporciona una adaptación de modelos mediante transformaciones de éstos. En la aproximación MLLR, los parámetros originales del sistema basado en HMM son mapeados a los valores adaptados a través de un conjunto de transformaciones afines que se estiman a partir de una pequeña cantidad de datos de adaptación. MLLR fue propuesto por primera vez para la adaptación del locutor con el fin de mejorar el rendimiento de los sistemas de reconocimiento de voz, y más tarde una variedad de extensiones se han desarrollado con aplicaciones a otras áreas, tales como la síntesis de voz [24].

Utilizando la estimación MAP con MLLR ofrece mejoras frente al uso de ellos de forma independiente.

### Técnicas de modelado del lenguaje

En el reconocimiento del habla continua con gran vocabulario y varios locutores posibles se vuelve extremadamente difícil de interpretar la señal acústica en la secuencia de palabras correctas, y malas interpretaciones son comunes. Algún conocimiento sobre las posibles estructuras dentro de un idioma puede ayudar en gran medida a esta interpretación. Tal conocimiento puede ser proporcionado como estimaciones de la probabilidad de cada secuencia de palabras que se considera probable dada la señal acústica.

Los modelos de lenguaje son entrenados para cada conjunto de textos en todos sus distintos contextos. Estas colecciones de modelos obtenidos son eficientemente utilizados en habla continua cuando el reconocedor se centra en el modelo correspondiente más cercano al tema actual de discusión [27].

### Utilización de equipos de alta calidad

En la utilización de cualquier reconocedor del habla es importante la calidad de la captura del audio. Tanto la colocación del micrófono, próximo a la persona, como la calidad del dispositivo. Proporcionando ganancia en la captura, mejoran los resultados de transcripción en el reconocedor sin tener que preprocesar la señal.

Es importante elegir un buen micrófono adaptado al entorno de utilización [30].

### 2.3.4. Algunos de los RAH actuales

Con mayor velocidad van apareciendo sistemas de RAH tanto de código abierto como privados. Se enumeran algunos de los más conocidos y utilizados:

1. **CMUSphinx** [10]:

   CMUSphinx es una herramienta de reconocimiento del habla que puede utilizarse para aplicaciones de pequeño, mediano y gran vocabulario. Utiliza la plataforma Java para su desarrollo.

   Es una herramienta de código abierto (open-source) que puede obtenerse gratuitamente en la web.

2. **HTK** [33]:

   Hidden Markov Model Toolkit (HTK) es un conjunto de herramientas, disponibles en código C, para la construcción y manipulación de modelos ocultos de Markov (HMM). HTK se utiliza principalmente para la investigación de reconocimiento de voz aunque se ha utilizado en numerosas otras aplicaciones, incluyendo la investigación de síntesis de voz y reconocimiento de caracteres.

   Esta herramienta es open-source que puede obtenerse gratuitamente previo registro en la web. Una vez instalada en la aplicación, no requiere de internet para su utilización.

3. **Julius** [23]:

   Julius es un decodificador software de amplio vocabulario de alto rendimiento de habla continua (Large Vocabulary Continuous Speech Recognition, LVCSR) para investigadores y desarrolladores. Permite crear diversos tipos de sistemas de reconocimiento de voz adecuados al uso que se le quiera dar. También adopta formatos estándar para hacer frente a otro conjunto de herramientas tales como HTK, CMU-Cam SLM kit de herramientas, etc.

   Esta herramienta es open-source que puede obtenerse gratuitamente en la web. Una vez instalada en la aplicación, no requiere de internet para su utilización.

4. **DNS** [36]:

   Dragon NaturallySpeaking (DNS) es un software de reconocimiento de voz que permite a los usuarios interactuar con el sistema con la voz, Dragon reconoce lo que dice y cómo lo dice con una precisión del 99 %[2]. Ésta herramienta ha sido creada por Nuance Communications Inc. dentro de la gama de herramientas Dragon Naturally Speaking Solutions.

   Este software es privado y no se tiene acceso al código fuente. Se puede utilizar la herramienta en otras aplicaciones sin la necesidad de estar conectado a internet.

5. **iSpeech** [22]:

   iSpeech es un software que realiza sintetización del habla a texto (Text -To-Speech, TTS) y reconocimiento automático del habla.

   Este software pertenece a una compañía privada, no se tiene acceso al código fuente, que ofrece una SDK para múltiples lenguajes de programación, permitiendo poder utilizar la herramienta en otras aplicaciones. Por cada palabra transcrita se realiza un pago. Requiere de internet para su utilización.

6. **MS Speech Platform** [32]:

   Microsoft Speech Platform SDK es un conjunto de herramientas de desarrollo para aplicaciones de voz. Provee un reconocedor automático del habla y un sintetizador de habla (TTS) para la interacción del usuario con su aplicación.

[2] Según la compañía

Este software es privado y no se tiene acceso al código fuente, pero sí a la SDK de forma gratuita. Una vez instalado, no requiere de internet para su utilización.

7. **Google ASR** [3]:

   Google ASR es un software de reconocimiento perteneciente a Google que posee una gramática abierta y en varios idiomas. Ofrece distintas opciones de uso para servicios web y aplicaciones móviles.

   Este software es privado y no se tiene acceso al código fuente pero existe una Web Speech API[44] para su utilización gratuita. Requiere de internet para su utilización.

## 2.4. Accesibilidad en medios de transporte

La mejora de la accesibilidad de los medios de transporte es un campo que ha cobrado importancia en los últimos tiempos. Se apuesta más por la accesibilidad universal, con el fin de facilitar el uso de los sistemas de transporte público en superficie a cualquier persona de forma segura y autónoma, con independencia de su condición física, psíquica o sensorial [17].

Los sistemas más comunes son el uso de grabaciones de voz para anunciar las paradas, llegadas y/ o salidas acompañadas de un rótulo que muestra por escrito dicho mensaje. Estos sistemas sólo ofrecen la información guardada, en ningún caso pueden cambiar su información durante la utilización del transporte, ni informar de cambios o locuciones del operador. Perdiendo de esta forma información relevante.

### 2.4.1. Accesibilidad en aviones comerciales

En la actualidad, las únicas ayudas, no asistidas por personas, que reciben las personas con discapacidad auditiva que deseen viajar en avión se encuentran en las terminales de los aeropuertos, y en puntos muy específicos. Se utilizan equipos de inducción magnética para orientar a las personas sordas en los aeropuertos [7]. Una vez que las personas suben abordo, estas ayudas desaparecen y con ellas toda la comunicación.

#### Ruido en el interior de cabinas de aviones comerciales

El transporte aéreo es uno de los más utilizados. El ruido en el interior de la aeronave es importante, especialmente en vuelos de larga duración, afectando a la salud, comodidad y comunicación de los pasajeros con la tripulación. Los niveles de ruido varían según las diferentes acciones del vuelo, se pueden agrupar por ruido durante el despegue y aterrizaje y ruido a nivel de crucero.

En la actualidad se realizan medidas genéricas de la cabina obteniendo entre 60 y 65 dBA antes del despegue, 80-85 dBA durante el vuelo y 75 a 80 dBA en el aterrizaje. Con niveles discontinuos de ruido en la cabina que han llegado a 81-88 dBA [26].

# CAPÍTULO 3

# Desarrollo del proyecto

En este capítulo se procede definir la solución y a desarrollarla. Se especificarán los requisitos y la funcionalidad de la herramienta así como el diseño técnico correspondiente a la solución. Finalmente se describirán las herramientas utilizadas para el desarrollo del sistema.

## 3.1. Solución propuesta

Como se ha indicado antes el proyecto busca mantener informado a los pasajeros de un vuelo comercial de las instrucciones que la tripulación les comunica. El escenario base es aquel en el que el tripulante, ya sea comandante o auxiliar de vuelo, genera un comunicado a través de la megafonía de la aeronave. Estos mensajes no son recibidos por pasajeros con discapacidad auditiva.

La solución que se plantea es la presentación de estos mensajes en formato de texto, a través de pantallas o bien del navegador web del dispositivo personal, mediante una transcripción en directo del comunicado. La Figura 3.1 muestra la descripción del servicio que se proporcionará con esta herramienta.

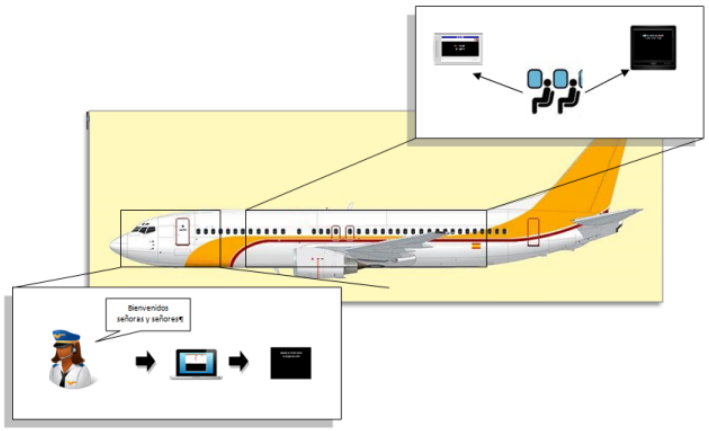



Figura 3.1: Descripción del servicio

El proyecto se centra en la captura de audio, su transcripción y posteriormente su presentación a través de una pantalla o del navegador web.

### 3.1.1. Especificación de necesidades

Este proyecto especifica varias necesidades que determinarán la decisión de desarrollo del mismo.

En primer lugar, el proyecto se centra en la obtención de la locución en texto del comunicado del tripulante al pasajero. Dada la relevancia de la información, el reconocedor deberá garantizar una alta tasa de acierto de palabra, cumpliendo las siguientes condiciones:

1. Utilización de un vocabulario genérico y largo dado que se pretende transcribir habla continua.
2. Preferentemente ha de ser dependiente del locutor para garantizar una alta tasa de acierto de palabra.
3. Ha de poder utilizarse en inglés y español, por lo que ha de tener modelos entrenados en ambos idiomas.

En segundo lugar, el escenario es un entorno ruidoso pero hay que tener en cuenta que para una correcta utilización del reconocedor, el micrófono ha de estar lo más cerca del tripulante (no se pueden utilizar micrófonos en medio de la cabina), por lo que el ruido será menor al mencionado en el apartado *Ruido en el interior de cabinas de aviones comerciales* 2.4.1.

En tercer lugar, debido a que el texto obtenido es en directo, el subtítulo será abierto, literal e incrustado. Las características de presentación (número de líneas, número de caracteres por línea y persistencia) serán configuradas por el administrador de la herramienta.

Finalmente se tiene en cuenta que el sistema va dirigido a un público no técnico y por lo tanto se ha de facilitar de la manera más sencilla el acceso a la herramienta, por ello se utilizará como sistema de presentación las pantallas de la aeronave o, de no tener, una página web a la que tendrán acceso desde sus dispositivos móviles, tablets o pcs (mediante una intranet).

### 3.1.2. Herramientas utilizadas para el desarrollo

En esta sección se describen brevemente las principales herramientas utilizadas para el desarrollo del sistema.



#### .NET de Microsoft

.NET Framework [1] es una popular plataforma de desarrollo para la creación de aplicaciones para Windows, Windows Phone, Windows Server y Windows Azure. La plataforma .NET Framework incluye los lenguajes de programación C# y Visual Basic.

Para el desarrollo de este proyecto se ha elegido esta plataforma y el lenguaje de programación C# debido a la gran variedad de funcionalidades y bibliotecas ya implementadas que ofrece y la facilidad con la que se puede generar código rápidamente. Esta herramienta permitirá generar los subtítulos con la configuración que se especifique así como interactuar con los dispositivos del avión, como pueden ser pantallas conectadas al sistema principal o bien crear un servidor web que de acceso a una página para mostrar el subtítulo.

[1] ©2013 Microsoft

### Dragon Naturally Speaking

Actualmente existen múltiples motores de RAH pero éste ha sido elegido porque cumple con la especificación de necesidades del proyecto.

En primer lugar, es la necesidad de un motor robusto de RAH, en este caso DNS ofrece una alta tasa de acierto de palabra con un modelo de locutor adaptado. La herramienta contiene unos modelos entrenados, tanto en inglés como en español con un vocabulario largo y genérico. Permite la adaptación de los modelos acústicos mediante grabaciones de audio, por lo que no requiere que el usuario final a utilizarlo esté presente en el momento de la adaptación. Sólo será necesaria una grabación de audio de 15 minutos para obtener su modelo y exportarlo a la herramienta.

En segundo lugar, se necesita un motor que no precise de conexión a internet para operar, DNS lo cumple.

Finalmente, debido a su sencillo manejo utilización y un coste razonable (los motores con conexión a internet cobran por palabra transcrita).

## 3.2. Requisitos y especificación funcional

En esta sección se muestra la información y el proceso de generación de la especificación del sistema. Presenta los requisitos de usuario y los requisitos técnicos obtenidos.

### 3.2.1. Requisitos de usuario

Los requisitos de usuario recogen la descripción detallada de las peticiones y restricciones a las que el sistema debe someterse. Corresponde a la conversión en especificaciones de la descripción del funcionamiento del servicio, previamente mencionada, a la que se añaden las nuevas necesidades y limitaciones que han de soportar y gestionar dicha funcionalidad.

| **Requisitos de usuario** | | | | |
|---|---|---|---|---|
| **Id** | **Nombre** | **Descripción** | **Categoría** | **Antecesor** |
| RU01 | General | La herramienta deberá presentar la transcripción literal del comunicado | Capacidad | |
| RU02 | Captura | La herramienta deberá captura el audio del comunicado | Capacidad | RU01 |
| RU03 | Transcripción | La herramienta deberá transcribir el audio capturado | Capacidad | RU01 |
| RU04 | RAH | Para transcribir se utiliza un motor de Reconocimiento Automático del Habla | Capacidad | RU03 |
| RU05 | Formato | La herramienta deberá formatear el texto obtenido a la salida del motor de RAH a modo de subtítulo de acuerdo a la configuración especificada | Capacidad | RU03 |
| RU06 | Presentación | La herramienta deberá mostrar el subtítulo en las pantallas y la web correspondiente | Capacidad | RU01 |
| RU07 | Persistencia y líneas | La persistencia en la pantalla y el número de líneas de los subtítulos será configurable | Restricción | RU05 |

*Continúa en la siguiente pág.*

Tabla 3.1 – *Proviene de la pág. anterior*

| Id | Nombre | Descripción | Categoría | Antecesor |
|---|---|---|---|---|
| RU08 | Configuración | Existirá una interfaz de usuario que permitirá la configuración de parámetros | Capacidad | RU01 |
| RU09 | Perfiles de locutor | La herramienta permitirá la selección y utilización de perfiles de locutor existentes en la herramienta | Capacidad | RU04 |
| RU10 | Arranque | La herramienta deberá iniciar automáticamente la transcripción con la última configuración seleccionada | Arranque | RU01 |
| RU11 | Interfaz | La herramienta deberá presentar un interfaz de manejo muy sencillo para los tripulantes y otro con la funcionalidad completa para el personal de administración | Interfaz | RU01 |
| RU12 | Idioma | La herramienta deberá transcribir en español o en inglés según el modelo de locutor seleccionado | Capacidad | RU01 |

Tabla 3.1: Requisitos de usuario

La Tabla 3.1 muestra los requisitos de usuario de este proyecto. Se identifican por una etiqueta y un nombre, se describe la necesidad del usuario, se clasifica según la categoría en la que se agrupa y finalmente si tiene o no un antecesor.

### 3.2.2. Especificación funcional

Como previamente se ha descrito el proyecto surge de la necesidad de solucionar el acceso a los comunicados de la tripulación a personas con discapacidad auditiva, tal y como muestra la Figura 3.1. A continuación se describen las funcionalidades principales del proyecto y posteriormente se detallan los casos de uso de la herramienta.

#### Funcionalidades

1. Transcripción en directo:

   Consiste en obtener en directo el audio del sistema de microfonía de la aeronave y mediante un motor de RAH transcribirlo con el fin de poder generar subtítulos.

   Una de las limitaciones de esta funcionalidad será un breve retraso entre la muestra del subtítulo y su locución debido a que el motor de RAH sólo devuelve texto transcrito cuando encuentra una pausa.

2. Muestra del subtítulo:

   Consiste en presentar el texto procedente de la transcripción en formato subtítulo según una configuración previamente especificada por el usuario. El fin de esta funcionalidad es facilitar la lectura de las transcripciones y poder presentarlo en diversos dispositivos.

   Al igual que la funcionalidad anterior, esta poseerá un breve retraso ya que se requiere un mínimo de 4 segundos para que una persona pueda leer cómodamente. Este tiempo es configurable por el administrador, el cual decidirá la persistencia del subtítulo.

### Casos de uso

Los siguientes casos definen las acciones que un actor podrá realizar sobre la herramienta. Se distinguen los siguientes actores:

- Administrador:

  Persona autorizada que gestionará la herramienta, los cuales serán miembros del equipo de mantenimiento y técnicos de desarrollo durante el proceso de desarrollo de la herramienta. Se le asignan los casos de uso: *Seleccionar perfil*, *Administrar perfil* y *Administrar modelo*.

- Tripulante:

  Persona perteneciente a la tripulación de la aeronave y habilitado para un manejo básico de la herramienta, el cual se limita a la selección del perfil. Y es por ello que su único caso de uso es *Seleccionar perfil*.

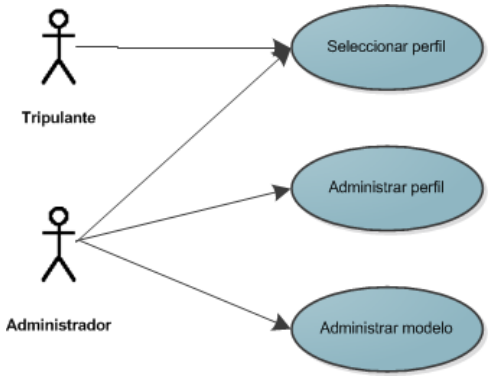



Figura 3.2: Casos de uso

La Figura 3.2 muestra los casos de uso que se detallan a continuación.

1. CU01 Seleccionar perfil:

   El objetivo es definir cómo el actor selecciona un perfil concreto. La Tabla 3.2 describe el escenario de este caso.

| **ESC01 Seleccionar perfil** | | | |
|---|---|---|---|
| **Objetivo** | | | |
| Definir cómo se selecciona un perfil concreto de los disponibles en el sistema. | | | |
| **Precondición** | | | |
| Haber iniciado la herramienta como tripulante o administrador. | | | |
| **Actividad** | | | |
| **Acción actor** | | **Respuesta herramienta** | |
| Solicitar el listado | [*Listar (perfiles)]* | Lista los perfiles | [*Presentar (perfiles)]* |
| Seleccionar el más apropiado | [*Seleccionar (perfil)]* | Perfil seleccionado | [*Perfil seleccionado]* |
| **Post-condición** | | | |
| Haber iniciado la herramienta como tripulante o administrador | | | |

Tabla 3.2: Seleccionar perfil

2. CU02 Administrar perfil:

   El objetivo es describir las operaciones del administrador sobre la gestión de los perfiles. Las Tablas 3.3,3.4, 3.5 y 3.6 describen los escenarios para este caso.

| **ESC02 Crear perfil** | | | |
|---|---|---|---|
| **Objetivo** | | | |
| Definir cómo se crea un perfil en la herramienta. | | | |
| **Precondición** | | | |
| Haber iniciado la herramienta como administrador. | | | |
| **Actividad** | | | |
| **Acción actor** | | **Respuesta herramienta** | |
| Solicitar el listado | [*Listar (perfiles)]* | Lista los perfiles | [*Presentar (perfiles)]* |
| Crear nuevo perfil | [*Crear (perfil)]* | Limpia datos | |
| Asignar nombre único | [*Nombrar]* | | |
| Solicita el listado de modelos | [*Listar (modelos)]* | Lista los modelos | [*Presentar (modelos)]* |
| Seleccionar el modelo | [*Seleccionar (modelo)]* | Modelo seleccionado | |
| Solicitar el listado de idioma | [*Listar (idiomas)]* | Lista idiomas | [*Presentar (idiomas)]* |
| Seleccionar idioma | [*Seleccionar (idioma)]* | Idioma seleccionado | |
| Guardar perfil | [*Guardar]* | Confirma | [*Guardar]* |
| **Post-condición** | | | |
| La herramienta tiene un nuevo perfil con el nombre, modelo e idioma asignados. | | | |

Tabla 3.3: Crear perfil

| **ESC03 Eliminar perfil** | | | |
|---|---|---|---|
| **Objetivo** | | | |
| Definir cómo elimina un perfil en la herramienta. | | | |
| **Precondición** | | | |
| Haber iniciado la herramienta como administrador. | | | |
| **Actividad** | | | |
| **Acción actor** | | **Respuesta herramienta** | |
| Solicitar el listado | [*Listar (perfiles)]* | Lista los perfiles | [*Presentar (perfiles)]* |
| Seleccionar el perfil a eliminar | [*Seleccionar (perfil)]* | Confirma selección | [*Confirma]* |
| Eliminar | [*Eliminar]* | Confirma selección | [*Confirma]* |
| **Post-condición** | | | |
| La herramienta ya no tiene el perfil eliminado. | | | |

Tabla 3.4: Eliminar perfil

| **ESC04 Importar perfil** | | | |
|---|---|---|---|
| **Objetivo** | | | |
| Definir cómo se importan los datos de un perfil de la herramienta. | | | |
| **Precondición** | | | |
| Haber iniciado la herramienta como administrador. | | | |
| **Actividad** | | | |
| **Acción actor** | | **Respuesta herramienta** | |
| Seleccionar importar | [*Importar (perfil)]* | | |
| Seleccionar ruta origen | [*Seleccionar ruta]* | | |
| Aceptar | [*Aceptar]* | Confirma selección | [*Confirma]* |
| **Post-condición** | | | |
| La herramienta tiene un nuevo perfil importado. | | | |

Tabla 3.5: Importar perfil

| **ESC05 Exportar perfil** | | | |
|---|---|---|---|
| **Objetivo** | | | |
| Definir cómo se exportan los datos de un perfil de la herramienta. | | | |
| **Precondición** | | | |
| Haber iniciado la herramienta como administrador. | | | |
| **Actividad** | | | |
| **Acción actor** | | **Respuesta herramienta** | |
| Solicitar el listado | [*Listar (perfiles)]* | Lista los perfiles | [*Presentar (perfiles)]* |
| Seleccionar el más apropiado | [*Seleccionar (perfil)]* | Perfil seleccionado | |
| Seleccionar exportar | [*Exportar (perfil)]* | | |
| Seleccionar ruta destino | [*Seleccionar ruta]* | | |
| Aceptar | [*Aceptar]* | Confirma selección | [*Confirma]* |
| **Post-condición** | | | |
| La herramienta habrá copiado el perfil en la ruta seleccionada. | | | |

Tabla 3.6: Exportar perfil

3. CU03 Administrar modelo:

   El objetivo es definir cómo el actor gestiona los modelos de locutor. Las Tablas 3.7, 3.8 y 3.9 describen los escenarios de este caso.

| **ESC06 Eliminar modelo** | | | |
|---|---|---|---|
| **Objetivo** | | | |
| Definir cómo elimina un modelo de locutor en la herramienta. | | | |
| **Precondición** | | | |
| Haber iniciado la herramienta como administrador. | | | |
| **Actividad** | | | |
| **Acción actor** | | **Respuesta herramienta** | |
| Solicitar el listado | [*Listar (modelos)]* | Lista los modelos | [*Presentar (modelos)]* |
| Seleccionar el modelo a eliminar | [*Seleccionar (modelo)]* | Selecciona modelo | |
| Eliminar | [*Eliminar]* | Confirma selección | [*Confirma]* |
| **Post-condición** | | | |
| La herramienta ya no tiene el modelo eliminado. | | | |

Tabla 3.7: Eliminar modelo de locutor

| **ESC07 Importar modelo** | | | |
|---|---|---|---|
| **Objetivo** | | | |
| Definir cómo se importa el modelo de locutor a la herramienta. | | | |
| **Precondición** | | | |
| Haber iniciado la herramienta como administrador. | | | |
| **Actividad** | | | |
| **Acción actor** | | **Respuesta herramienta** | |
| Seleccionar importar | [*Importar (modelo)]* | | |
| Seleccionar ruta origen | [*Seleccionar ruta]* | | |
| Aceptar | [*Aceptar]* | Confirma selección | [*Confirma]* |
| **Post-condición** | | | |
| La herramienta tiene un nuevo modelo importado. | | | |

Tabla 3.8: Importar modelo de locutor

| **ESC08 Exportar modelo** | | | |
|---|---|---|---|
| **Objetivo** | | | |
| Definir cómo se exporta el modelo de locutor de la herramienta. | | | |
| **Precondición** | | | |
| Haber iniciado la herramienta como administrador. | | | |
| **Actividad** | | | |
| **Acción actor** | | **Respuesta herramienta** | |
| Solicitar el listado | [*Listar (modelos)]* | Lista los modelos | [*Presentar (modelos)]* |
| Seleccionar el más apropiado | [*Seleccionar (modelo)]* | Modelo seleccionado | |
| Seleccionar exportar | [*Exportar (modelo)]* | | |
| Seleccionar ruta destino | [*Seleccionar ruta]* | | |
| Aceptar | [*Aceptar]* | Confirma selección | [*Confirma]* |
| **Post-condición** | | | |
| La herramienta habrá copiado el modelo de locutor en la ruta seleccionada. | | | |

Tabla 3.9: Exportar modelo de locutor

### 3.2.3. Requisitos técnicos

Las Tablas 3.10, 3.15, 3.11, 3.12 y 3.14 muestran los requisitos técnicos de este proyecto agrupados por funcionalidades. Se identifican mediante una etiqueta y un nombre, se describe su funcionalidad y cuál es la fuente (requisito de usuario) cuya necesidad resuelve. Finalmente si tiene o no un antecesor.

| **Requisitos técnicos generales** | | | | |
|---|---|---|---|---|
| **Id** | **Nombre** | **Descripción** | **Fuente** | **Antecesor** |
| RTG01 | Objetivo | La herramienta deberá procesar una señal de audio a la entrada y convertirla en transcripción a la salida | RU01 | |

Tabla 3.10: Requisitos técnicos generales

| **Requisitos técnicos de audio** | | | | |
|---|---|---|---|---|
| **Id** | **Nombre** | **Descripción** | **Fuente** | **Antecesor** |
| RTA01 | Conexión | La herramienta se conectará al sistema de audio de la aeronave. Si la señal es analógica será digitalizada | RU01, RU02 | RTG01 |

Tabla 3.11: Requisitos técnicos de audio

| Requisitos técnicos de transcripción | | | | |
|---|---|---|---|---|
| **Id** | **Nombre** | **Descripción** | **Fuente** | **Antecesor** |
| RTT01 | Motor RAH | La herramienta realizará la transcripción mediante un motor de RAH que se conectará a la entrada de flujo de audio. Realizará la transcripción siempre que haya presencia de voz en la señal | RU03, RU04 | RTG01 |
| RTT02 | Idioma | La herramienta transcribirá en español y en inglés | RU04, RU12 | RTG01 |
| RTT03 | Cambio de idioma | El cambio de idioma no será automático y se establece por el cambio a un perfil diseñado para el nuevo idioma | RU04, RU12 | RTT02 |

Tabla 3.12: Requisitos técnicos de transcripción

| Requisitos técnicos de presentación | | | | |
|---|---|---|---|---|
| **Id** | **Nombre** | **Descripción** | **Fuente** | **Antecesor** |
| RTP01 | Salida | La salida serán textos en formato subtítulo correspondientes a la transcripción de los comunicados | RU05, RU06 | RTG01 |
| RTP02 | Formato | Se podrá configurar el formato del subtítulo mediante: la persistencia, número de líneas y número de caracteres por línea | RU06, RU07, RU08 | RTP01 |

Tabla 3.13: Requisitos técnicos de presentación

| Requisitos técnicos de configuración | | | | |
|---|---|---|---|---|
| **Id** | **Nombre** | **Descripción** | **Fuente** | **Antecesor** |
| RTC01 | Repositorio de perfiles y modelos | La herramienta incorporará un repositorio de perfiles y modelos | RU09 | RTI01 |
| RTC02 | Configuración | La herramienta permitirá configurar los repositorios pudiendo crear (sólo perfiles), eliminar, importar y exportar | RU08 | RTC01 |

Tabla 3.14: Requisitos técnicos de configuración

| Requisitos técnicos de interfaz de usuario | | | | |
|---|---|---|---|---|
| **Id** | **Nombre** | **Descripción** | **Fuente** | **Antecesor** |
| RTI01 | Perfiles | La herramienta contendrá un número indefinido de perfiles y cualquiera podrá ser seleccionado como perfil de funcionamiento | ESC01, RU09 | RTG01 |
| RTI02 | Administrar perfiles | La herramienta permitirá gestionar perfiles: crear, eliminar, importar y exportar | ESC02, ESC03, ESC04, ESC05 | |
| RTI03 | Listar perfiles | La herramienta presentará una lista de los perfiles existentes en su repositorio | | |
| RTI04 | Seleccionar perfiles | La herramienta permitirá seleccionar un perfil de la lista presentada | | RTI03 |
| RTI05 | Administrar modelos | La herramienta permitirá gestionar modelos: eliminar, importar y exportar | ESC06, ESC07, ESC08 | |
| RTI06 | Listar modelos | La herramienta presentará una lista de los modelos existentes en su repositorio | | |
| RTI07 | Seleccionar modelos | La herramienta permitirá seleccionar un modelo de la lista presentada | | RTI06 |
| RTI08 | Confirmar | Para todas las operaciones que se soliciten y sean irreversibles la herramienta pedirá una confirmación | | |
| RTI09 | Arranque | La herramienta debe iniciar automáticamente la transcripción de audio una vez arrancada | RU10 | |
| RTI10 | Interfaz tripulante | La herramienta deberá iniciar por defecto con esta interfaz en la que sólo se podrá seleccionar el perfil que se utilizará durante el vuelo. Existirá desde esta interfaz un acceso al administrador. | RUI10 | |
| RTI11 | Interfaz administrador | La herramienta presentará una nueva interfaz si el usuario se registra como administrador en la interfaz tripulante. Esta interfaz permitirá realizar las acciones de gestión de la herramienta | RU11 | RTI10 |

Tabla 3.15: Requisitos técnicos de interfaz de usuario

## 3.3. Diseño técnico de la solución

En esta sección se presenta el diseño completo de la solución para satisfacer los requisitos técnicos del usuario y ajustarse a las funcionalidades descritas. Se describe la arquitectura del sistema y se detalla el proceso de generación de los subtítulos. Finalmente se realizan unos comentarios sobre las consideraciones operativas de la herramienta.

### 3.3.1. Arquitectura del sistema

Una vez establecidos los requisitos se determina la estructura de la arquitectura desde el punto de vista conceptual, funcional y de despliegue.

Analizando el problema a resolver, una solución es primero, obtener el texto procedente de la locución utilizando un motor de RAH y después dividir ese texto en palabras para poder formar los subtítulos con las características previamente especificadas por el administrador.

#### Arquitectura conceptual

La Figura 3.3 muestra la arquitectura conceptual. Esta arquitectura presenta la herramienta desde el punto de vista de la lógica del proceso y la divide en componentes que representan los elementos principales.

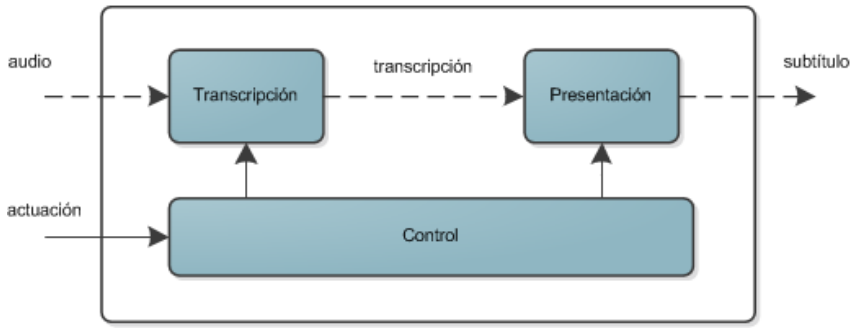



Figura 3.3: Arquitectura conceptual

Tal y como se ha descrito en la especificación funcional, se muestra un flujo de datos. El proceso comienza por la captura del **audio** que será transcrito (**Transcripción**) y una vez obtenida la **transcripción** se presentará (**Presentación**), lo que permitirá obtener el subtítulo en los dispositivos. Todo controlado en base a la configuración que se haya establecido mediante la **actuación** sobre la herramienta (**Control**).

#### Arquitectura funcional

La Figura 3.4 muestra la arquitectura funcional. Esta arquitectura presenta la herramienta desde el punto de vista de sus capacidades, agrupando éstas en los componentes principales según criterios de servicio y operación.

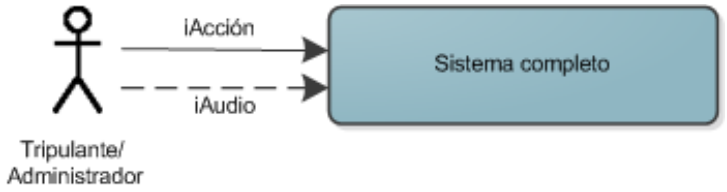



Figura 3.4: Arquitectura funcional

Se representan los interfaces del **sistema completo** que mantendrá con el mundo exterior. En este caso existe un único rol del **tripulante/ administrador**, el cual actuará sobre el sistema o bien enviará audio al sistema.

A continuación se especifican los interfaces:

- **iAcción**: interfaz gráfica de usuario que permite la gestión (si es administrador) o la utilización (si es tripulante) de la herramienta.

  Podrán realizar las siguientes acciones: seleccionar, listar perfiles y confirmar (tripulantes); confirmar, listar, seleccionar, crear, modificar, eliminar, importar y exportar perfiles, listar, seleccionar, eliminar, importar y exportar modelos (administrador).

- **iAudio**: interfaz que permite el envío de audio al sistema.

  Señal de audio de la microfonía de la aeronave. Dicha señal será capturada para ser transcrita.

## Arquitectura despliegue

La Figura 3.5 muestra la arquitectura de despliegue. Esta arquitectura presenta la herramienta desde el punto de vista de distribución física, identificando los artefactos independientes y las unidades sobre las que se implantarán.

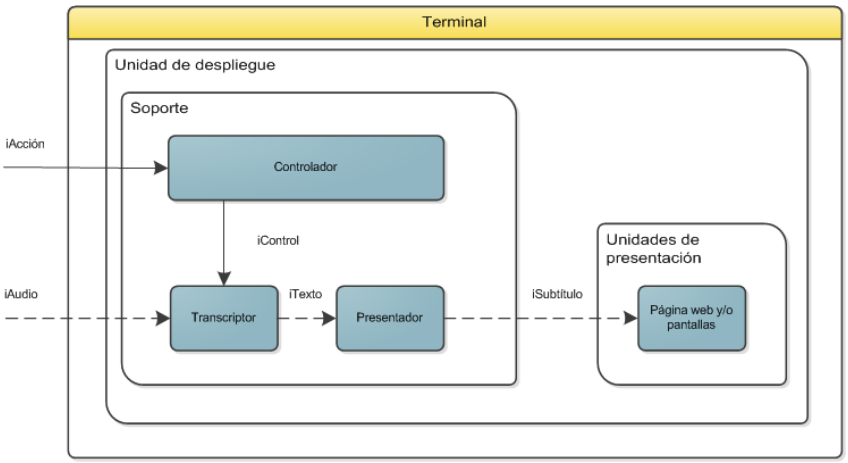



Figura 3.5: Arquitectura de despliegue

Se plantea una única unidad de despliegue en el terminal constituida por dos artefactos.

- El primero es el **Soporte** que realizará todas las operaciones necesarias para la funcionalidad de la herramienta.

  Compuesto por los siguientes módulos:

- **Controlador**: permitirá administrar y gestionar los datos de la herramienta y las interfaces gráficas de usuario (GUI), permitirá la configuración del **Presentador** y controlará la ejecución del **Transcriptor**.
- **Transcriptor**: captará la señal de audio y la transcribirá haciendo uso de la infraestructura software del reconocedor compuesto por la librería *dnstools.dll*.
- **Presentador**: formateará el texto procedente del **Transcriptor** en subtítulo.

- El segundo es **Unidades de presentación** que se compondrá de una página web, accesible para los pasajeros, donde se mostrará el subtítulo; y de las pantallas que se conecten al terminal.

A continuación se especifican los interfaces:
Los interfaces **iAcción** y **iAudio** son las mismas que en el aparatado 3.3.1 *Arquitectura funcional*.

- **iControl**: interfaz que permite el control, iniciar o detener, del módulo **Transcriptor**.
- **iTexto**: interfaz que permite que el texto sin formato procedente del módulo **Transcriptor** fluya al **Presentador**. Será el texto que contiene la transcripción del audio procedente del sistema de microfonía.
- **iSubtítulo**: interfaz que permite que el subtítulo fluya hacia las unidades de presentación, ventana de texto en las pantallas y página web para los navegadores. Será el texto formateado procedente de la transcripción de la locución.

## 3.3.2. Diseño detallado

A continuación se detallan en pseudocódigo[2] del **Controlador** y del **Transcriptor** y el algoritmo del **Presentador** para mostrar el subtítulo, en las Figuras 3.6, 3.7 y 3.8. Este diseño detallado permitirá una mayor comprensión del desarrollo del sistema y facilitará su implementación en código.

### Controlador

El **Controlador** controlará la acción del **Transcriptor** y permitirá gestionar la configuración del subtitulado, así como administrar y gestionar la herramienta. Se define el pseudocódigo de la siguiente forma:

```
Controlador
{
                Iniciar_sistema(perfil){  inicia el sistema y el Transcriptor automáticamente con el perfil que se
                                          indique}
                      Detener_sistema(){  detiene la herramienta deteniendo el Transcriptor}
                        Salir_sistema(){  cierra el sistema}
                       Cambiar_perfil(){  cambia el perfil actual por el del perfil seleccionado}
                Listar (perfil/ modelo){  lista todos los perfiles de usuarios o modelos existentes}
                 Crear (perfil/ modelo){  crea un nuevo perfil o modelo}
               Eliminar (perfil/modelo){  elimina el perfil o modelo seleccionado}
     Importar/Exportar (perfil/modelo){  importa/exporta perfiles de usuario o modelos}
   Configurar_subtitulo (configuracion){  guarda la nueva configuración del subtítulo (nº líneas, nº caracteres por
                                          línea y persistencia)}
}
```

Figura 3.6: Pseudocódigo Controlador

[2]El pseudocódigo se presenta como guía para implementar la herramienta.

### Transcriptor

El **Transcriptor** obtendrá el texto de la transcripción a partir de la señal de audio capturada. Se define el pseudocódigo de la siguiente forma:

```
Transcriptor
{
                    Iniciar_rah (modelo){  inicia el reconocedor con el modelo acústico indicado}
                          Detener_rah(){  detiene el reconocedor }
                      Obtener_palabras(){  obtiene el texto resultante de la transcripción realizada para que pueda
                                           ser formateado}
}
```

Figura 3.7: Pseudocódigo Transcriptor

### Presentador

El **Presentador** realizara la generación de los subtítulos. Para formatear el texto procedente de la transcripción se tendrán en cuenta las siguientes características especificadas por el administrador:

- Número de caracteres por línea = ncaract
- Número de líneas = nlineas
- Persistencia = persistencia

Y la lista con las palabras procedentes de la transcripción = lista_de_palabras.
Se define el pseudocódigo de la siguiente forma:

```
Presentador
{
   Generar_subtitulo (lista_de_palabras){  formateará la lista de palabras para generar un subtítulo con el nº de
                                           líneas y nº de caracteres previamente configurados por el administrador}
                     Mostrar_subtitulo(){  muestra el subtítulo generado en las unidades de presentación durante el
                                           tiempo de persistencia previamente configurado por el administrador}
}
```

Figura 3.8: Pseudocódigo Presentador

- **Generar subtítulo**:

  Inicialmente, cada vez que se termina de añadir palabras a la lista de palabras (**lista_de_palabras**), el generador detendrá el temporizador de 5 segundos[3] y comenzará extrayendo una a una contando la longitud de la frase que está formando y comprobando que no supera en conjunto el máximo de línea (**ncaract**).

  Si lo superase, guardará la línea sin incluir la última palabra (puesto que era la que desbordaba) y posteriormente comprobará que no ha superado el máximo de líneas (**nlineas**). En caso de ser mayor, mostraría el subtítulo, si es menor empezaría la nueva línea en una nueva frase. Si se ha llegado al final de la lista, se muestra el subtítulo con las palabras restantes.

  La Figura 3.9 muestra este proceso.

[3]Temporizador de 5 segundos: este temporizador se utiliza para que, en el caso de que se haya terminado de hablar, no quede ninguna palabra suelta en la lista de palabras sin mostrar.

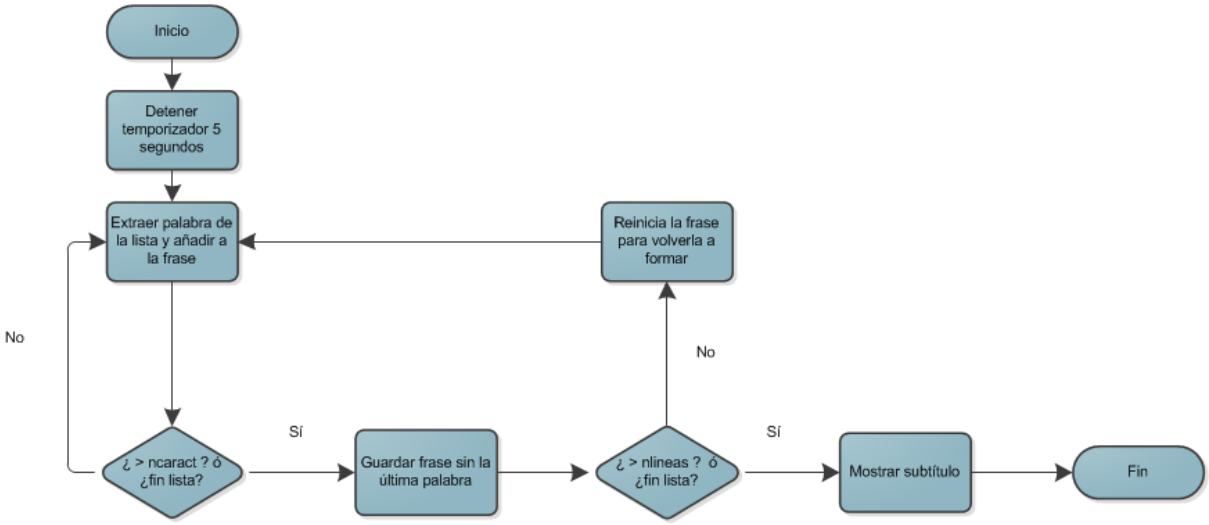



Figura 3.9: Generar subtítulos

- **Mostrar subtítulo**:

  Seguidamente, para mostrar los subtítulos se tiene en cuenta la **persistencia**. Para ello se activará otro temporizador con el tiempo indicado por la persistencia, siempre en segundos, y cuando termine mostrará el siguiente subtítulo.

  Por otra parte, cada vez que se termina de mostrar se activa un temporizador de 5 segundos. Una vez agotado su tiempo comprueba si hay palabras en la lista **palabras** y llamará al proceso de generar subtítulos si existen palabras. Esto limita el tiempo máximo de persistencia en 5 segundos para permitir una comunicación en directo. La Figura 3.10 muestra este proceso.

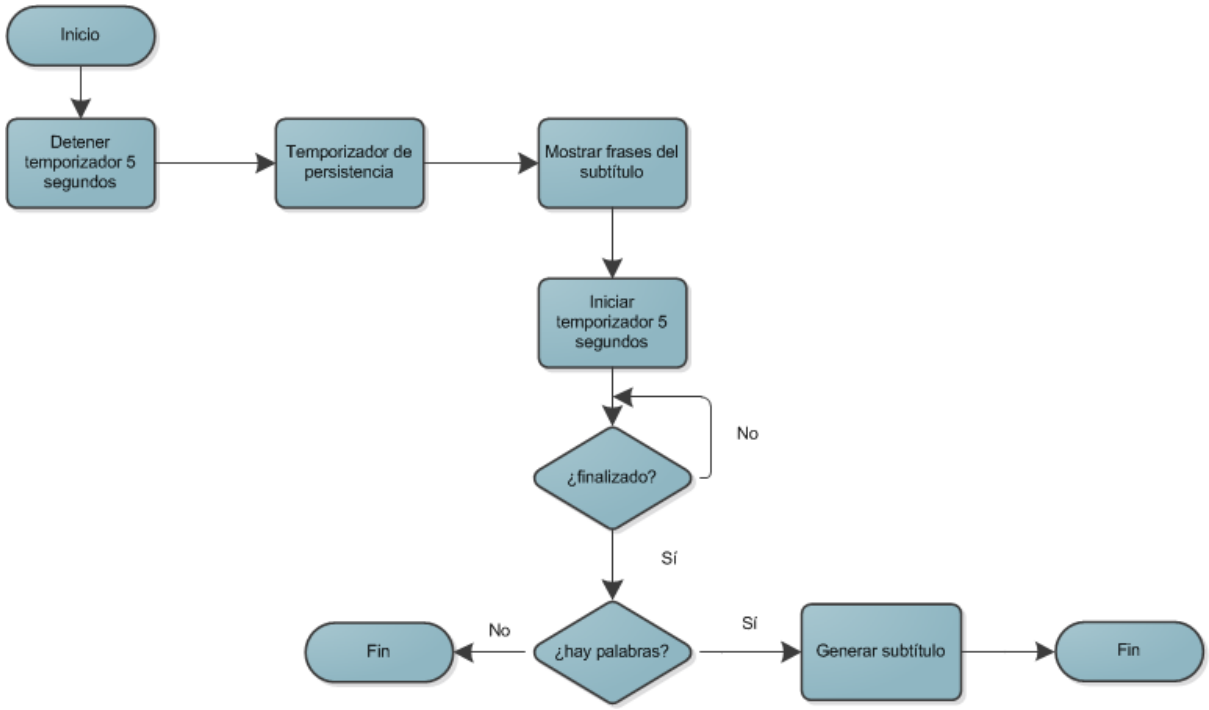



Figura 3.10: Mostrar subtítulos

### Interfaces gráficas de usuario (GUI)

A continuación se resumen los requisitos mínimos que han de cumplir las interfaces gráficas de usuario.

- Interfaz principal (tripulante):

  Es la interfaz con la que se inicia la herramienta de forma automática. El contenido mínimo que presentará esta interfaz y las condiciones que ha de cumplir son:

  - Permitirá ver la lista de perfiles existentes en la herramienta y seleccionar uno de ellos para su uso.
  - Permitirá salir de la herramienta.
  - Permitirá el acceso al administrador para la gestión de la herramienta.
  - Presentará un diseño sencillo y fácil de manejar.

- Interfaz secundaria (administrador):

  Es la interfaz que permite al administrador gestionar la herramienta. El contenido mínimo que presentará esta interfaz y las condiciones que ha de cumplir son:

  - Permitirá ver la lista de perfiles con sus características.
  - Permitirá crear, modificar, importar, exportar y eliminar perfiles.
  - Permitirá importar, exportar y eliminar modelos.
  - Permitirá salir de la herramienta.
  - Permitirá volver a la interfaz principal.
  - Presentará un diseño sencillo y fácil de manejar.

- Interfaz pantallas:

  Es la interfaz que presentará el texto en las pantallas conectadas al terminal donde se encuentre la herramienta. El contenido mínimo que presentará esta interfaz y las condiciones que ha de cumplir son:

  - Mostrará el subtítulo en un tamaño adecuado al tamaño de la pantalla.
  - El resto de la pantalla será un fondo negro.

- Interfaz web:

  Es la interfaz que presentará el texto en la web, ver Figura 3.11. El contenido mínimo que presentará esta interfaz y las condiciones que ha de cumplir son:

  - Mostrará el subtítulo en un cuadro de texto con fondo negro que estará en el centro de la página web.
  - Permitirá aumentar o disminuir el tamaño de letra.


Figura 3.11: Página web

## 3.4. Consideraciones del sistema

Consideraciones operativas y tecnológicas para el correcto funcionamiento del sistema.

### 3.4.1. Consideraciones operativas

Debido al motor de RAH utilizado, los siguientes requisitos de hardware para que la herramienta opere serán:

- CPU: procesador Intel Pentium ©de 1GHz como mínimo o procesador equivalente.
- Caché de procesador: 512 KB como mínimo.
- Espacio libre en el disco duro: 500 MB mínimo. Se necesitará más a medida que se vayan añadiendo nuevos perfiles y modelos de locutor.
- Sistema operativo: Windows 7 (32 bits ó 64 bits).
- RAM: 2 GB mínimo.
- Tarjeta de sonido: tarjeta de sonido compatible con captura a 16 bits.
- Unidad de DVD-ROM si la instalación se realiza utilizando un disco.

### 3.4.2. Consideraciones tecnológicas

Dada la solución existen restricciones tecnológicas que han de considerarse para un completo funcionamiento de la herramienta.

La principal limitación de esta tecnología es la utilización del motor de RAH. Esta tecnología no es óptima en el entorno de operación, debido al ruido del interior de la aeronave. Para garantizar una alta tasa de reconocimiento coherente se requieren entornos sin ruido o con mínimas perturbaciones.

Por otra parte, si se utiliza para habla continua, la tasa de acierto baja notablemente. Actualmente no existe un motor capaz de transcribir en habla continua con alta tasa de acierto, y se necesita un perfil adaptado al locutor y al entorno para mejorar dicha tasa. Estas consideraciones son importantes y si se mejoran los motores de RAH se debería actualizar la herramienta.

# CAPÍTULO 4

# Pruebas y resultados

En este capítulo se procede a describir todas las pruebas realizadas en la herramienta, las cuales han servido para determinar las prestaciones de la herramienta caracterizando el RAH.

## 4.1. Descripción de las pruebas

Se pretende caracterizar el motor de reconocimiento utilizado en la herramienta[1], en un entorno ruidoso, en este caso el interior de una aeronave, y comprobar qué circunstancias ofrecen mejores resultados para así mejorar las prestaciones de la herramienta.

### 4.1.1. Caracterización del reconocedor

La utilización óptima de los reconocedores sería cuando el usuario hablase con una clara vocalización, marcando las pausas y a velocidad media. Ésta no es la situación real en la que se encuentra el usuario. En el escenario de este trabajo la tripulación realizaría un comunicado a ritmo de habla normal, a veces acelerada, pero sin dictar y con una pobre vocalización.

#### Datos utilizados

Debido a la falta de grabaciones pertenecientes al interior de la aeronave, se ha decidido utilizar un audio-libro, el cual ofrece una grabación de voz a un ritmo normal y con diferentes niveles de vocalización (gracias a los cambios de voz que realiza el lector en los diálogos de los personajes para diferenciarlos de la narración de la historia), como grabación de habla continua. Y para el ruido sí se ha obtenido una grabación del interior de la cabina, pero el micrófono se encontraba en el asiento del piloto, por lo que es ruido ambiental.

Los archivos de audio utilizados en las pruebas se dividen en dos bloques: los datos de adaptación (train) y los datos para las pruebas (test) más el archivo de audio de ruido. El audio-libro seleccionado para las pruebas es *Alicia en el país de las maravillas* [5] en lengua inglesa. Se han obtenido dos lecturas,

[1] En concreto el reconocedor utilizado ha sido Dragon Naturally Speaking 10.

la primera (en adelante LibroA) corresponde a una lectura del libro completo por un mismo locutor. Y una segunda lectura (en adelante LibroB) [6] realizada por 12 locutores distintos (uno por cada capítulo).

Finalmente el archivo de ruido procedente del interior de una aeronave y los distintos niveles SNR a los que han sido mezclados los archivos de audio con el ruido.

Todos los archivos de audio han sido remuestreados a una frecuencia de 16kHz y han sido convertidos a formato .Wav.

1. **Train** [2] - Capítulo 1 y 4 del LibroA, estos datos se encuentran resumidos en la Tabla 4.1. El motor de RAH DNS necesita mínimo 15 minutos de audio para las adaptaciones.

| **Train: LibroA** | |
|---|---|
| Capítulo | Duración (min:seg) |
| 1 | 11:20 |
| 4 | 12:46 |
| Total | 24:07 |

Tabla 4.1: Audio para adaptación (Train)

2. **Test** - Para la prueba 1) LibroA - libro completo excepto el capítulo 1 y 4 (marcados con *), y el LibroB completo. Estos datos se encuentran resumidos en la Tabla 4.2 ( H = Hombre M = Mujer).

| **Test: LibroB** | | | **Test: LibroA** |
|---|---|---|---|
| Capítulo | Duración (min:seg) | Género | Duración (min:seg) |
| 1 | 10:27 | M | 11:20* |
| 2 | 11:59 | H | 11:10 |
| 3 | 17:08 | H | 08:46 |
| 4 | 19:20 | M | 12:46* |
| 5 | 13:06 | H | 11:49 |
| 6 | 12:39 | M | 14:45 |
| 7 | 17:01 | M | 11:47 |
| 8 | 13:03 | M | 13:02 |
| 9 | 14:45 | M | 13:11 |
| 10 | 22:17 | H | 12:51 |
| 11 | 09:56 | H | 10:17 |
| 12 | 12:38 | H | 11:43 |
| Total | 2:54:19 | | 2:23:27 |

Tabla 4.2: Audio para pruebas (Test)

3. **Ruido** - grabación de ruido procedente del interior de aeronave.

[2] El motor de RAH DNS denomina a la adaptación de locutor como entrenamiento (train) y esta nomenclatura será utilizada en adelante en este documento.

4. **Niveles SNR utilizados** - la mezcla de los archivos de audio Test LibroA con el Ruido se ha realizado con los siguientes niveles de **SNR = [-18, -12, -6, -3, 3, 6, 12, 18]**.

### Software de apoyo utilizado

Para la realización de las pruebas se han utilizado las siguientes herramientas de apoyo:

1. **Matlab** [31]
   - Mezclar ruido dado un nivel de SNR: Add Noise [48]
     Esta herramienta permite mezclar los archivos de audio Test con el archivo Ruido a distintos niveles de SNR.
   - Substracción espectral: SSBol79 [54]
     Esta herramienta permite realizar una substracción espectral a un archivo de audio y recuperarlo en el dominio del tiempo.
2. **CMUSphinx** [10]
   - Comparar resultados: NISTAlign [9]
     Esta herramienta permite comparar dos textos (original e hipótesis resultante de la transcripción realizada por DNS) y devuelve como resultado el porcentaje de palabras correctas una vez realizada una alineación entre textos. Requiere que ambos textos no tengan ningún signo de puntuación ni formato para obtener el mejor resultado.

## 4.1.2. Conjunto de pruebas realizadas

Esta sección describe cada una de las pruebas realizadas para comprobar la viabilidad y robustez del RAH para esta herramienta.

El idioma de las pruebas es inglés estadounidense[3] y por tanto el modelo acústico de DNS utilizado es el modelo US English, adaptado o no según la prueba.

1. **Prueba de independencia de locutor**

   Esta prueba consiste en demostrar el comportamiento de un RAH cuando se utiliza de forma genérica para todo tipo de personas. Para ello se han transcrito los audio Test LibroB con un modelo sin adaptar, es decir con el básico ofrecido por DNS.

2. **Prueba de dependencia de locutor**

   Esta prueba consiste en demostrar que el RAH es una herramienta de transcripción dependiente del locutor. Para ello se han transcrito los audio Test LibroA y LibroB con un modelo adaptado con el audio Train.

3. **Robustez al ruido y al preprocesado**

   Esta prueba consiste en comprobar la capacidad de acierto en palabras que tiene el motor de RAH para transcribir utilizando un modelo adaptado con el audio Train.

   Para ello se transcriben los audios Test del LibroA limpio, mezclados a los niveles SNR mencionados anteriormente y preprocesados[4].

[3] Se realiza sólo en este idioma debido a la existencia de numerosos archivos de audio transcrito en la red, la facilidad para poder utilizarlos y que los reconocedores actuales poseen un mayor número de horas de entrenamiento para este idioma.

[4] Audio preprocesado: corresponde a la mezcla del LibroA con Ruido a los niveles SNR mencionados anteriormente y a los cuales se les ha realizado la substracción espectral con la herramienta de Matlab.

4. **Adaptación al ruido**

   Esta prueba consiste en comprobar la robustez de los RAH cuando se adaptan con un audio obtenido en el mismo entorno de aplicación utilizando un modelo adaptado con el audio Train que previamente ha sido mezclado con Ruido a un SNR = -6 dB.

   Para ello se transcriben los audios Test del LibroA limpio, mezclados a los niveles SNR mencionados anteriormente y preprocesados.

5. **Adaptación al preprocesado**

   Esta prueba consiste en comprobar la robustez de un RAH cuando se realiza un preprocesado al audio antes de utilizar el motor de RAH utilizando un modelo adaptado con el audio Train que previamente ha sido preprocesado pero no mezclado con ruido.

   Para ello se transcriben los audios Test del LibroA limpio, mezclados a los niveles SNR mencionados anteriormente y preprocesados.

6. **Adaptación al ruido y al preprocesado**

   Esta prueba consiste en comprobar si la robustez de un RAH cuando se realiza preprocesado al audio obtenido en el mismo entorno de la aplicación utilizando un modelo adaptado con el audio Train mezclado con Ruido a un SNR = -6 dB y posteriormente preprocesado.

   Para ello se transcriben los audios Test del LibroA limpio, mezclados a los niveles SNR mencionados anteriormente y preprocesados.

7. **Simulación de la herramienta**

   Consiste en una simulación completa de la herramienta en la que se ha hablado en directo realizando un mensaje de bienvenida corto para los pasajeros. El audio ha sido transcrito en directo con un modelo adaptado al locutor, en este caso el autor del documento.

**Nota**: Las pruebas 3, 4, 5 y 6 utilizan un modelo adaptado al locutor (Tabla 4.1) y para Test el LibroA (Tabla 4.2) mezclado o no con ruido según el nivel SNR que se indique en las tablas (en adelante Mezclado con Ruido) y ese mismo audio preprocesado (en adelante Preprocesado SS).

### Tabla resumen de las pruebas

Se resumen todas las pruebas realizadas en la Tabla 4.3 mediante su identificador, su nombre, la condición de que estén o no adaptadas al locutor y una breve descripción de la prueba.

| **Pruebas realizadas** | | | |
|---|---|---|---|
| **Id** | **Nombre de la prueba** | **Adaptado al locutor** | **Breve descripción** |
| 1 | Independencia del locutor | No | Utiliza un modelo no adaptado y transcribe el LibroB |
| 2 | Dependencia del locutor | Sí | Utiliza un modelo adaptado al locutor (Train) y transcribe el LibroA y LibroB |
| 3 | Robustez al ruido y al preprocesado | Sí | Utiliza un modelo adaptado al locutor (Train) y transcribe el LibroA sin ruido, mezclado con ruido y mezclado con ruido + preprocesado |

*Continúa en la siguiente pág.*

Tabla 4.3 – *Proviene de la pág. anterior*

| Id | Nombre de la prueba | Adaptado al locutor | Breve descripción |
|---|---|---|---|
| 4 | Adaptación al ruido | Sí | Utiliza un modelo adaptado al locutor (Train) con ruido a SNR = -6 dB y transcribe el LibroA sin ruido, mezclado con ruido y mezclado con ruido + preprocesado |
| 5 | Adaptación al preprocesado | Sí | Utiliza un modelo adaptado al locutor (Train) preprocesado y transcribe el LibroA sin ruido, mezclado con ruido y mezclado con ruido + preprocesado |
| 6 | Adaptación al ruido y al preprocesado | Sí | Utiliza un modelo adaptado al locutor (Train) con ruido a SNR = -6 dB y preprocesado y transcribe el LibroA sin ruido, mezclado con ruido y mezclado con ruido + preprocesado |
| 7 | Simulación | Sí | Mensaje corto de bienvenida realizado en directo |

Tabla 4.3: Pruebas realizadas sobre la herramienta

## 4.2. Resultados obtenidos y discusión

Esta sección muestra los resultados obtenidos en porcentaje de acierto de palabras comparando el texto obtenido de la transcripción con el original.

Por decisión del autor todos los resultados han sido redondeados con 2 decimales.

Los resultados se han resumido en las siguientes tablas y gráficas.

1. **Prueba de independencia de locutor**

   Los resultados obtenidos en esta prueba se muestran en la Tabla 4.4 (Los capítulos marcados con * no se tienen en cuenta para la media).

| **Resultados independencia de locutor** | | | |
|---|---|---|---|
| **Capítulo** | **%Acierto LibroB** | **Género** | **%Acierto LibroA** |
| 1 | 50.10* | Mujer | no |
| 2 | 61.30 | Hombre | 85.50 |
| 3 | 39.80 | Hombre | 13.95 |
| 4 | 43.90* | Mujer | no |
| 5 | 67.70 | Hombre | 83.01 |
| 6 | 51.80 | Mujer | 83.90 |
| 7 | 70.60 | Mujer | 77,68 |
| 8 | 53.40 | Mujer | 84.20 |
| 9 | 68.50 | Mujer | 79.46 |
| 10 | 31.50 | Hombre | 71.61 |
| 11 | 25.40 | Hombre | 81.90 |

*Continúa en la siguiente pág.*

| Tabla 4.4 – *Proviene de la pág. anterior* | | | |
|---|---|---|---|
| **Capítulo** | **%Acierto LibroB** | **Género** | **%Acierto LibroA** |
| 12 | 51.10 | Hombre | 85.70 |
| media | 52.11 | | 52.60 |

Tabla 4.4: Resultado: independencia de locutor

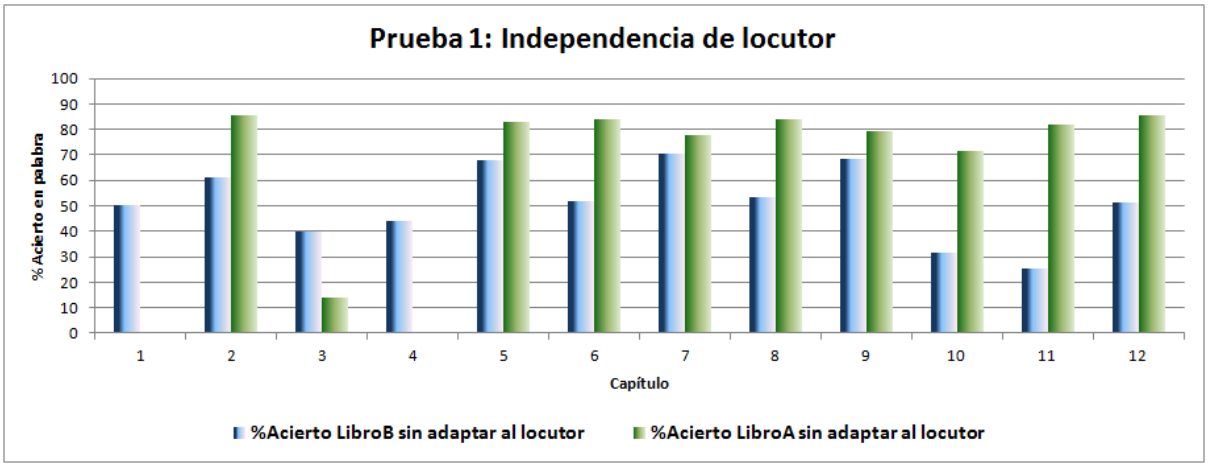



Figura 4.1: Gráfica: Independencia de locutor

Estos resultados obtenidos muestran que ante un modelo básico no adaptado el porcentaje de acierto es muy variable, no hay ninguna forma de saber con certeza ni preaviso la calidad (entendiéndose por calidad la obtención de una transcripción correcta) a obtener.

La Figura 4.1 muestra claramente la diferencia del peor resultado obtenido en el capítulo 11 frente al mejor en el 7. No se puede garantizar una comprensión del texto. La Tabla 4.5 muestra de los capítulos 7 y 11 un extracto de la transcripción obtenida por el reconocedor frente al original.

| **Muestra de texto** | |
|---|---|
| **Original** | ... with some curiosity what a funny watch she remarked it tells the day of the month and doesn't tell what o'clock it is ... |
| **Capítulo 7** | ... with some curiosity is funny **why is** she remarked it tells the day of the month and doesn't **help that bookmark** it **as** ... |
| **Original** | ... read as follows the Queen of hearts she made some tarts all on a summer day ... |
| **Capítulo 11** | ... **road** as follows the Queen of hearts she **moves in clients** all on a summer day ... |

Tabla 4.5: Texto: independencia de locutor

2. **Prueba de dependencia de locutor**

Los resultados obtenidos en esta prueba se muestran en la Tabla 4.6.

| Resultados dependencia de locutor | | | |
|---|---|---|---|
| **Capítulo** | **%Acierto LibroB** | **Género** | **% Acierto LibroA** |
| 1 | 26.80* | Mujer | no |
| 2 | 64.60 | Hombre | 89.80 |
| 3 | 59.10 | Hombre | 82.00 |
| 4 | 24.50* | Mujer | no |
| 5 | 77.70 | Hombre | 86.50 |
| 6 | 59.70 | Mujer | 86.30 |
| 7 | 63.10 | Mujer | 80.00 |
| 8 | 46.40 | Mujer | 86.00 |
| 9 | 63.50 | Mujer | 79.70 |
| 10 | 49.70 | Hombre | 77.80 |
| 11 | 59.20 | Hombre | 84.50 |
| 12 | 35.60 | Hombre | 86.00 |
| media | 57.86 | | 83.86 |

Tabla 4.6: Resultado: dependencia de locutor

Los capítulos marcados con * no se tienen en cuenta para la media.

Estos resultados muestran las mejoras obtenidas cuando el locutor adapta el modelo acústico y lo utiliza para sí mismo, como se observa en la Figura 4.2. En este caso se puede garantizar una comprensión del texto. Cuanto mejor se adapte el modelo al locutor, mejor será el porcentaje de acierto de palabra y por lo tanto la comprensión del texto (para los lectores).

Para el LibroB sigue sin poder obtenerse resultados de confianza, no se garantiza una transcripción correcta ni una comprensión de lo transcrito.

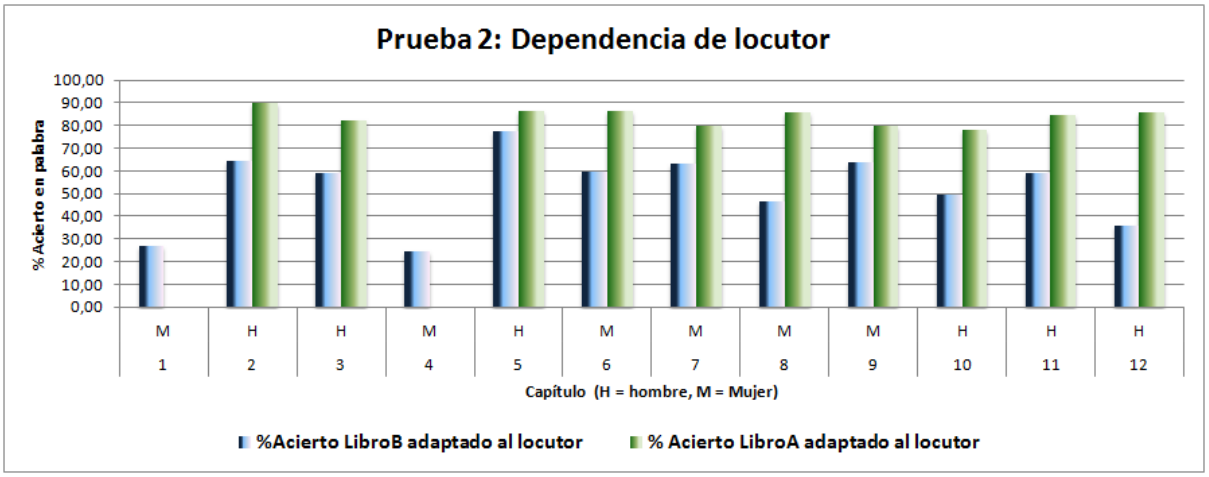



Figura 4.2: Gráfica: Dependencia de locutor

**Nota**:

Las pruebas 3, 4, 5 y 6 utilizan un modelo adaptado al locutor (Tabla 4.1) y para Test el LibroA (Tabla 4.2) mezclado o no con ruido según el nivel SNR que se indique en las tablas (Mezclado con Ruido) y ese mismo audio preprocesado (Preprocesado SS).

Los resultados corresponden a la media del porcentaje de acierto por palabra obtenido de los 10 capítulos que componen el Test LibroA.

3. **Robustez al ruido y al preprocesado**

   Los resultados obtenidos en esta prueba se muestran en la Tabla 4.7.

| Resultados de robustez al ruido y preprocesado | | |
|---|---|---|
| **SNR** | **Mezclado con Ruido** | **Preprocesado SS** |
| -18 | 14.91 | 11.00 |
| -12 | 34.49 | 30.98 |
| -6 | 60.12 | 58.55 |
| -3 | 67.88 | 67.01 |
| 3 | 76.26 | 75.46 |
| 6 | 78.66 | 77.60 |
| 12 | 81.34 | 79.92 |
| 18 | 82.65 | 81.41 |
| sin ruido | 83.86 | 83.17 |

Tabla 4.7: Resultado: robustez al ruido y al preprocesado

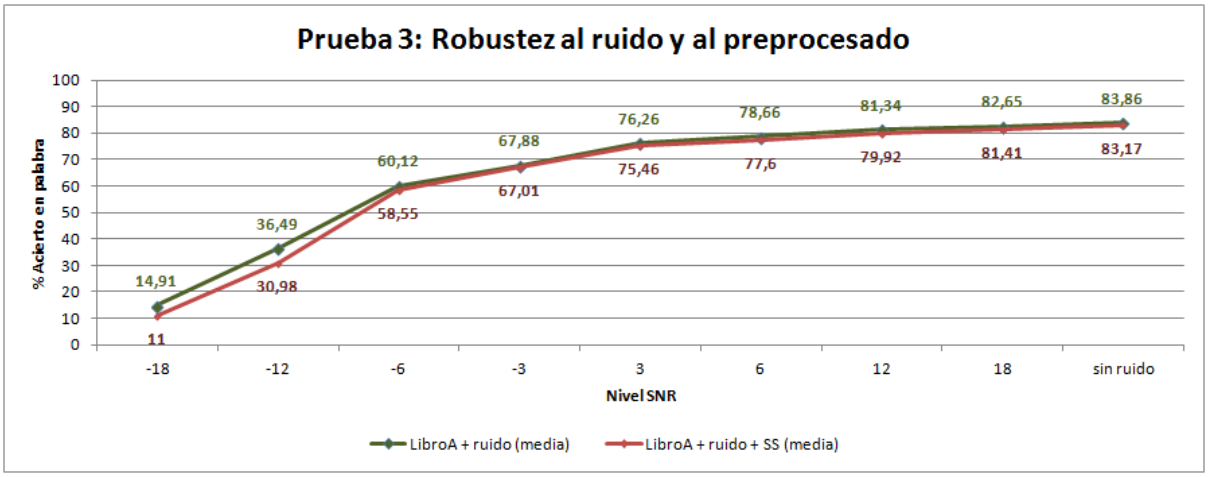



Figura 4.3: Gráfica: Adaptación al locutor

Observando los resultados (Figura 4.3) se concluyen dos hechos triviales en los actuales motores de RAH y son que:

*a*) A mayor ruido en la señal de audio menor será la calidad de transcripción.

*b*) Un modelo adaptado al locutor ofrece mejores transcripciones.

Comparando los resultados obtenidos de la transcripción de archivos mezclados con ruido y los mismos preprocesados para un modelo adaptado al locutor (sin ruido) se observa un 0.97 % menos de acierto en el preprocesado.

4. **Adaptación al ruido**

   Los resultados obtenidos en esta prueba se muestran en la Tabla 4.8.

| Resultados adaptación al ruido | | |
|---|---|---|
| **SNR** | **Mezclado con Ruido** | **Preprocesado SS** |
| -18 | 21.05 | 16.85 |
| -12 | 49.81 | 46.01 |
| -6 | 67.09 | 62.49 |
| -3 | 71.56 | 62.86 |
| 3 | 76.71 | 71.08 |
| 6 | 78.02 | 70.41 |
| 12 | 79.13 | 69.21 |
| 18 | 79.12 | 70.14 |
| sin ruido | 76.87 | 76.80 |

Tabla 4.8: Resultado: adaptación al ruido

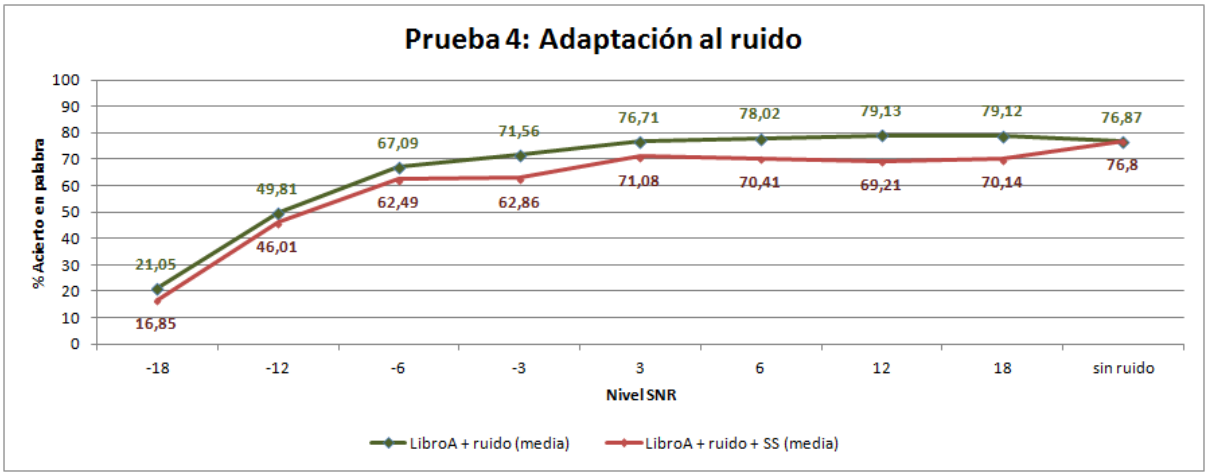



Figura 4.4: Gráfica: Adaptación al ruido

Estos resultados muestran una gran mejora en la transcripción de archivos en entornos ruidosos cuando el modelo está adaptado al ruido. Tanto para los mezclados con ruido como para los preprocesados.

Sin embargo, a medida que el ruido disminuye la capacidad de transcripción también lo hace ya que entre el modelo y el audio existe una gran desigualdad. La Figura 4.4 muestra de forma gráfica este suceso.

5. **Adaptación al preprocesado**

Los resultados obtenidos en esta prueba se muestran en la Tabla 4.9.

| Resultados adaptación al preprocesado | | |
|---|---|---|
| **SNR** | **Mezclado con Ruido** | **Preprocesado SS** |
| -18 | 13.67 | 9.66 |
| -12 | 35.32 | 30.41 |
| -6 | 58.68 | 58.44 |
| -3 | 66.86 | 66.00 |
| 3 | 75.40 | 73.69 |
| 6 | 77.89 | 77.82 |
| 12 | 80.72 | 80.46 |
| 18 | 82.18 | 81.84 |
| sin ruido | 83.28 | 82.20 |

Tabla 4.9: Resultado: adaptación al preprocesado

El proceso de realización de una substracción espectral previa a la utilización del motor de RAH puede ser causa de empeoramiento de las transcripciones obtenidas.

En este caso, los resultados muestran que utilizando este modelo de adaptación (adaptación a la herramienta de preprocesado sin ruido) obtiene un porcentaje de acierto de palabra alto para audio con muy poco o nada de ruido y empeora aquellos que tienen alta contaminación de ruido.

Observando la Figura 4.5 se obtiene que no sería necesario la utilización de un preprocesado del audio.

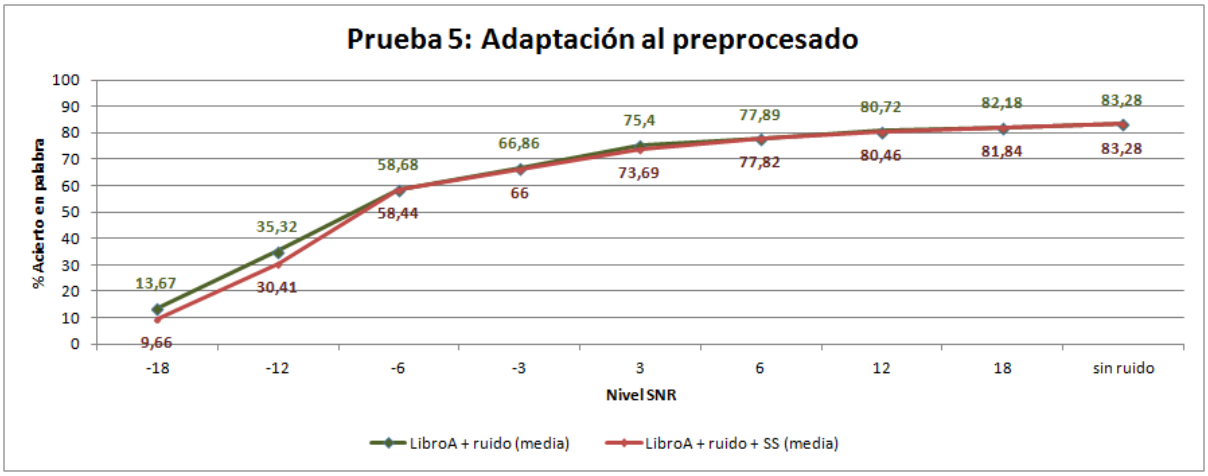



Figura 4.5: Gráfica: Adaptación al preprocesado

6. **Adaptación al ruido y al preprocesado**

Los resultados obtenidos en esta prueba se muestran en la Tabla 4.10.

| Resultados adaptación al ruido y al preprocesado | | |
|---|---|---|
| **SNR** | **Mezclado con Ruido** | **Preprocesado SS** |
| -18 | 16.96 | 18.70 |
| -12 | 41.91 | 43.20 |
| -6 | 63.69 | 64.90 |
| -3 | 70.13 | 69.18 |
| 3 | 76.24 | 74.37 |
| 6 | 77.94 | 75.11 |
| 12 | 78.50 | 75.97 |
| 18 | 79.77 | 77.44 |
| sin ruido | 79.18 | 78.60 |

Tabla 4.10: Resultado: adaptación al ruido y al preprocesado

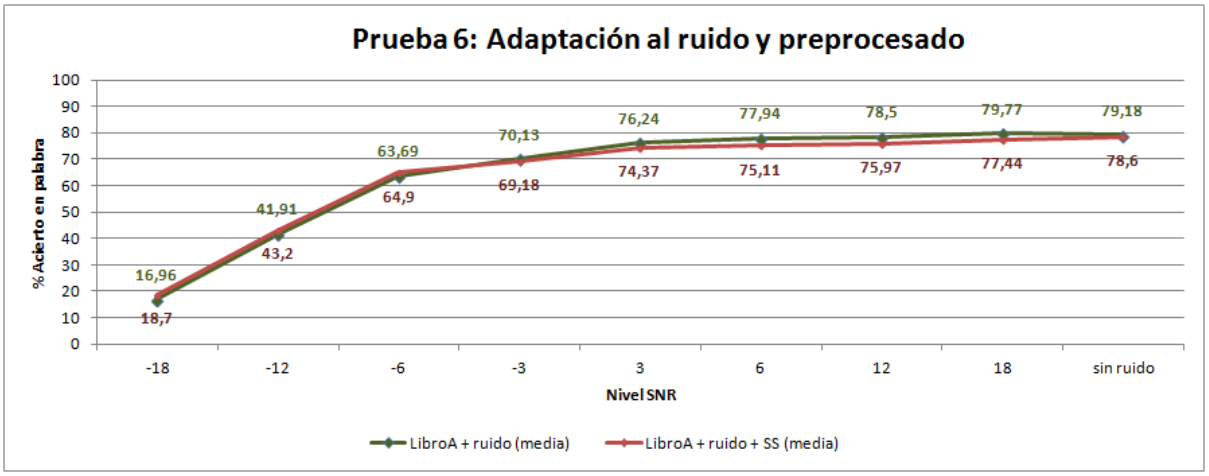



Figura 4.6: Gráfica: Adaptación al ruido y al preprocesado

Finalmente se combinan las dos pruebas anteriores adaptando el modelo al ruido y al preprocesado, es decir el audio Train (Tabla 4.1) ha sido contaminado con ruido y posteriormente se le ha realizado una substracción espectral antes de utilizarse para adaptar.

Estos resultados, representados en la Figura 4.6, muestran mejoras tanto para audio mezclado con ruido como el preprocesado cuando el nivel de ruido es alto pero a medida que el ruido disminuye el porcentaje de acierto no aumenta más que en las circunstancias de las pruebas 3 y 5.

7. **Simulación de la herramienta**

   Para comprobar las anteriores pruebas se han realizado 5 simulaciones distintas (Tabla 4.11) comunicando un mensaje corto de bienvenida (Tabla 4.12).

   - La primera (0) corresponde a la captura del texto mostrado por la utilización de la herramienta en directo. Se muestra el texto en un formato subtítulo de dos líneas.
   - Las siguientes (1, 2, 3 y 4) corresponden a la captura de audio en directo con la locución del mensaje completo. Este archivo de audio ha sido mezclado con ruido a un nivel SNR = -9 dB, según se indique y transcrito con distintos modelos adaptados al locutor o al ruido (adaptado al locutor con un SNR = -9 dB).

     Para estas pruebas se muestra la transcripción completa donde cada dos líneas corresponden al formato subtítulo.

| **Conjunto de simulaciones realizadas** | | |
|---|---|---|
| **Id** | **Modelo adaptado** | **Audio** |
| 0 | Al locutor | Directo |
| 1 | Al locutor | Sin ruido |
| 2 | SNR -9 | Sin ruido |
| 3 | Al locutor | SNR -9 |
| 4 | SNR -9 | SNR -9 |

Tabla 4.11: Simulaciones realizadas

| **Mensaje corto de bienvenida** |
|---|
| Buenos días señoras y señores, les habla el comandante. Cumpliendo con el horario previsto nos dirigimos hacia Londres. Durante este vuelo que durará dos horas aproximadamente, alcanzaremos una altitud de crucero de 30 000 pies y una velocidad de 800 km/h. Estimamos aterrizar en Londres a las diez hora local donde el tiempo es nublado y la temperatura es de 30ºC. Durante el vuelo recorreremos una distancia de 3000 km. Es un placer para nosotros tenerles a bordo y esperamos que disfruten de este vuelo. Gracias por su atención. |

Tabla 4.12: Texto: mensaje corto de bienvenida

A continuación se muestran los resultados de las simulaciones realizadas. Los resultados han sido redondeados sin decimales por decisión del autor.

0 **Simulación 0**

La Figura 4.7 es una muestra de la utilización de la herramienta en directo (prueba 0). Se captura el audio en directo por el hablante y se muestra en formato subtítulo el texto obtenido. Este formato de lectura facilita la comprensión para el lector.

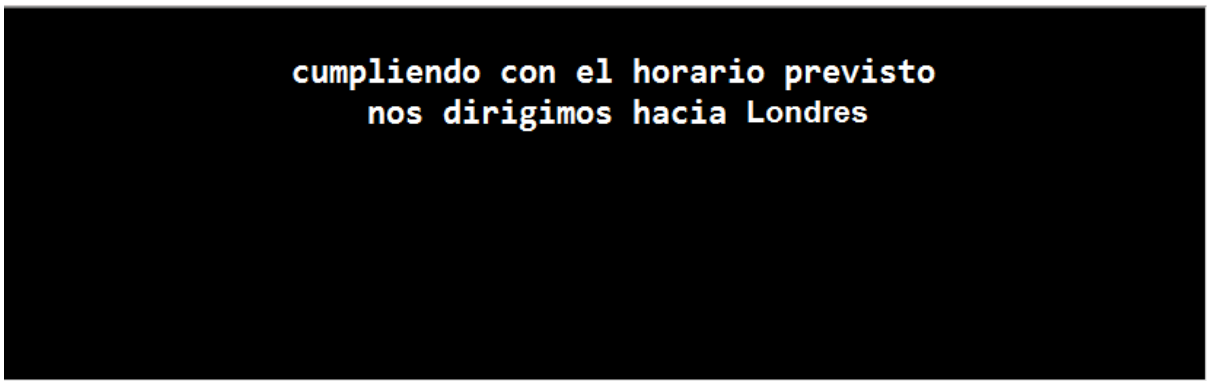



Figura 4.7: Directo

1 **Simulación 1**

Utilizando un modelo adaptado al locutor y a un entorno sin ruido, ofrece resultados altos de acierto en palabras aunque no es un 100 %.

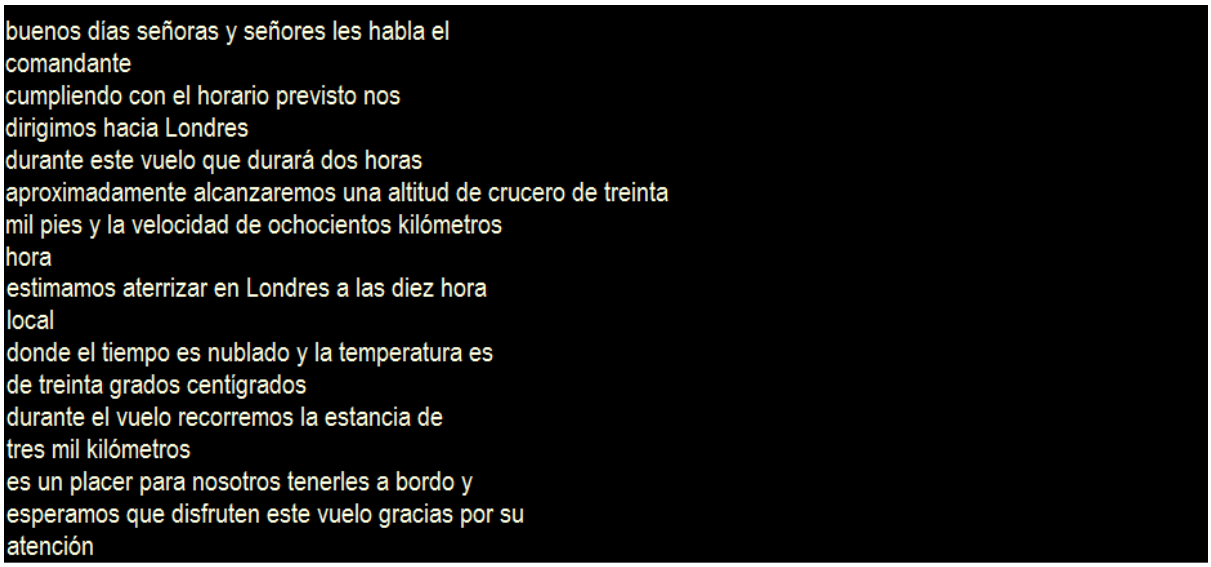



Figura 4.8: Adaptado al locutor y audio sin ruido

La Figura 4.8 muestra la transcripción obtenida corresponde a un 97 % de acierto con la original, cometiendo un error en 2 palabras pero con el contexto es fácil sobreentenderlo.

2 **Simulación 2**

La Figura 4.9 es una muestra de la utilización errónea de la herramienta. Utilizando un perfil adaptado a un entorno ruidoso en un entorno sin ruido.


Figura 4.9: Adaptado al ruido y audio sin ruido

Debido a la diferencia de las características de los datos obtenidos en la adaptación al ruido frente a los que se obtienen en un entorno sin ruido el reconocedor no es capaz de obtener una transcripción correcta. En este caso la tasa de acierto es del 0 %.

3 **Simulación 3**

Del mismo modo que en la simulación anterior, la Figura 4.10 es una muestra de la utilización errónea de la herramienta. Utilizando un perfil no adaptado a un entorno ruidoso en un entorno con ruido.

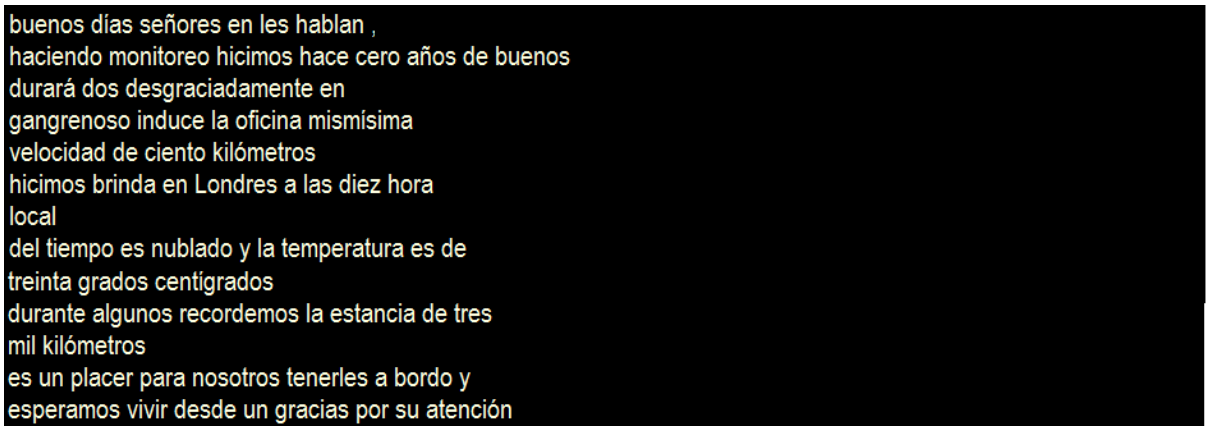



Figura 4.10: Adaptado al locutor y audio con ruido

La transcripción obtenida no es correcta debido al mismo caso que en la simulación anterior. Aunque en este caso el reconocedor puede transcribir parte de la locución, con una tasa de acierto del 42 %, no garantiza que sea correcta.

4 **Simulación 4**

Finalmente la Figura 4.11 muestra el resultado de una simulación del escenario del trabajo, interior de una aeronave, con un modelo adaptado al ruido y un audio mezclado con ruido al mismo nivel SNR que el de la adaptación.

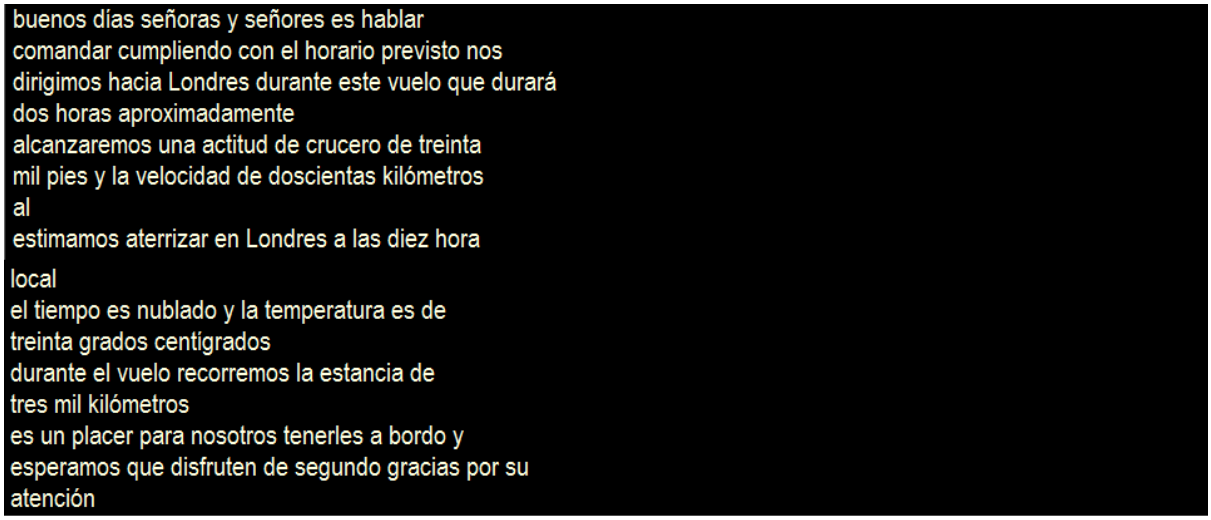



Figura 4.11: Adaptado al ruido y audio con ruido

Ofrece unos resultados de transcripción altos de un 88 % de acierto en palabra. El lector puede comprender el texto aún con palabras erróneas.

### 4.2.1. Resumen comparativo de las pruebas

A continuación se agrupan los resultados obtenidos y se muestran de forma resumida (mostrando su media).

**Resumen de resultados pruebas 1 y 2:**

Como se puede comprobar en las pruebas 1 y 2, resumidas por su media en la Tabla 4.13 y la Figura 4.13, la adaptación al locutor mejora de forma notable la calidad de transcripción y garantiza que ésta se corresponda con el audio. Por lo que se concluye que es mejor utilizar un modelo acústico adaptado al locutor.

| Resumen de resultados pruebas 1 y 2 | | |
|---|---|---|
| **Id** | | **Media** |
| **1** | **LibroA** | 70.73 |
| | **LibroB** | 51.26 |
| **2** | **LibroA** | 85.01 |
| | **LibroB** | 52.49 |

Tabla 4.13: Resumen de resultados pruebas 1 y 2

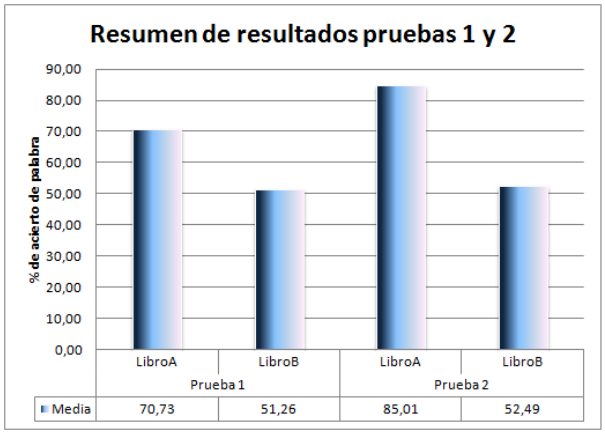



Figura 4.12: Gráfica: Resumen de pruebas 1 y 2

**Resumen de resultados pruebas 3, 4, 5 y 6:**

Se resume los resultados obtenidos para las pruebas (3, 4, 5 y 6) en las que se adapta al locutor siguiendo distintos criterios de adaptación, recordando:

1. Mezclado con ruido: Test LibroA mezclado con el ruido del interior de la aeronave al SNR indicado.

2. Mezclado con ruido y preprocesado: Test LibroA mezclado con el ruido del interior de la aeronave al SNR indicado y al que se le ha realizado una substracción espectral.

Los resultados de las pruebas realizadas, según los criterios previamente descritos, se muestran en la siguiente Tabla Resumen 4.14.

| Resumen de resultados pruebas 3, 4, 5 y 6 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Id** | **SNR** | | | | | | | | | **Media** |
| | -18 | -12 | -6 | -3 | 3 | 6 | 12 | 18 | sin | |
| **Audio mezclado con ruido** | | | | | | | | | | |
| **3** | 14.91 | 34.49 | 60.12 | 67.88 | 76.26 | 78.66 | 81.34 | 82.65 | 83.86 | 64.69 |
| **4** | 21.05 | 49.81 | 67.09 | 71.56 | 76.71 | 78.02 | 79.13 | 79.12 | 76.87 | 66.60 |
| **5** | 13.67 | 35.32 | 58.68 | 66.86 | 75.40 | 77.89 | 80.72 | 82.18 | 83.28 | 63.78 |
| **6** | 16.96 | 41.91 | 63.69 | 70.13 | 76.24 | 77.94 | 78.50 | 79.77 | 79.18 | 64.92 |
| **Audio mezclado con ruido y preprocesado** | | | | | | | | | | |
| **3** | 11.00 | 30.98 | 58.55 | 67.01 | 75.46 | 77.6 | 79.92 | 81.41 | 83.17 | 62.79 |
| **4** | 16.85 | 46.01 | 62.49 | 62.86 | 71.08 | 70.41 | 60.21 | 70.14 | 76.8 | 60.65 |
| **5** | 09.66 | 30.41 | 58.44 | 66 | 73.69 | 77.82 | 80.46 | 81.84 | 82.20 | 62.28 |
| **6** | 18.70 | 43.20 | 64.90 | 69.18 | 74.37 | 75.11 | 75.97 | 77.44 | 78.60 | 64.16 |

Tabla 4.14: Resumen de resultados pruebas 3, 4, 5 y 6

La media, mostrada en la Figura 4.13, servirá para, en caso de desconocer el entorno en el que operaría la herramienta, la condición que mejores resultados ofrecerá el reconocedor. En este caso, adaptar al ruido y al locutor sin preprocesar el audio.

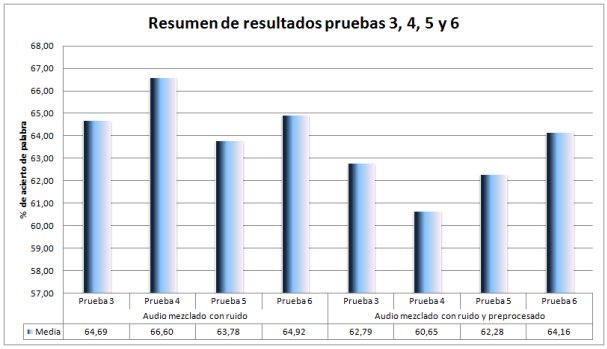



Figura 4.13: Gráfica: Resumen de pruebas 3, 4, 5 y 6

Las gráficas 4.14 y 4.15 representan los datos anteriormente mencionados, destacando la prueba 3 y 4.

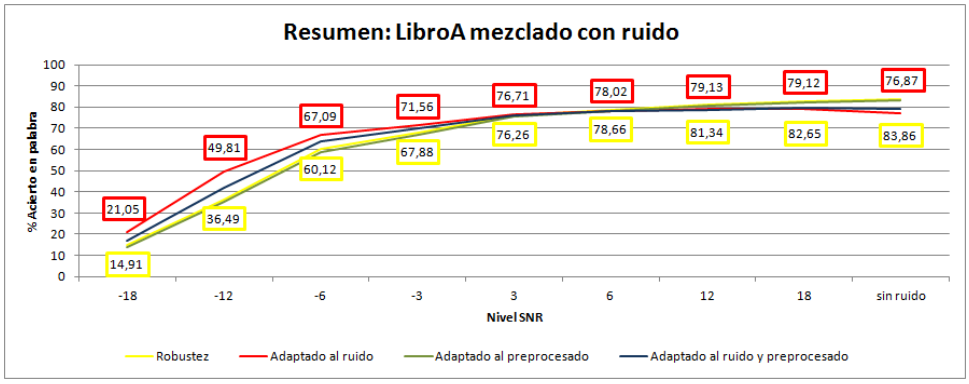



Figura 4.14: Gráfica: Resumen transcripción de audio ruidoso

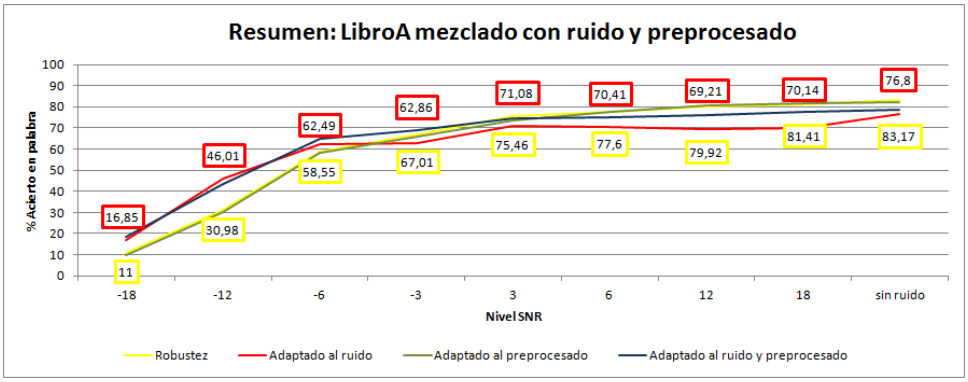



Figura 4.15: Gráfica: Resumen transcripción de audio ruidoso preprocesado

**Resumen de resultados de la simulación:**

Se muestra la Tabla 4.15 y la Figura 4.16 con los resultados de las simulaciones según los distintos casos previamente descritos.

| Resumen de resultado de simulaciones | | |
|---|---|---|
| **Simulación** | **Caso** | **Tasa de acierto** |
| 1 | Sin ruido | 97 % |
| 2 | Audio sin ruido - Modelo con ruido | 0 % |
| 3 | Audio con ruido - Modelo sin ruido | 42 % |
| 4 | Con ruido | 88 % |

Tabla 4.15: Resumen de resultados de simulaciones

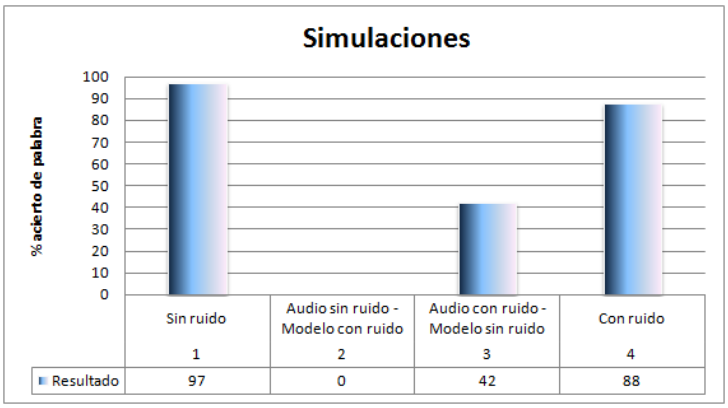



Figura 4.16: Gráfica: Resumen resultados de simulaciones

Las simulaciones confirman los resultados ya obtenidos por las pruebas anteriores, y muestran que el mejor caso es aquel en el que se adapta al locutor y al entorno (puede ser con ruido o sin él).

### 4.2.2. Discusión de resultados

El motor de reconocimiento utilizado es una herramienta eficiente para obtener una transcripción de un dictado de palabras. En este caso se ha realizado la prueba con un audio-libro, una narración con cambios de tonos de voz y velocidad de pronunciación (en el mismo audio el locutor cambia su forma de hablar según los personajes y la situación).

Para obtener una tasa de reconocimiento alta es mejor crear un modelo adaptado al locutor, en caso contrario la tasa de reconocimiento puede variar de forma aleatoria entre un 25 % (incluso menos) y un 60 %. Adaptando al locutor y al entorno, ya sea con ruido o sin él, podemos obtener una mejora en el reconocimiento de más de un 80 %. Este acierto puede mejorar con una readaptación (es decir más tiempo de adaptación del modelo al locutor) y también si la calidad del audio mejora (mejor micrófono y locución, mejor posicionamiento del micrófono para la captura...).

En el escenario de entorno ruidoso y en este caso en particular, el interior de una aeronave, se ha procedido a realizar la contaminación del Test LibroA con el ruido a distinto SNR y un preprocesado de estos mismos archivos realizándoles una substracción espectral. En total se obtienen los archivos Test resumidos en la Tabla 4.16:

**Archivos Test LibroA utilizados**

| **Estado** | **Número de archivos** |
|---|---|
| Originales (sin ruido) | 10 (uno por cada capítulo) |
| Con ruido a distinto SNR (-18, -12, -6, -3, 3, 6, 12, 18) | 80 (10 por cada SNR) |
| Preprocesados (substracción espectral de los archivos con ruido) | 80 (10 por cada SNR) |
| Total | 170 archivos de audio |

Tabla 4.16: Archivos Test LibroA utilizados

Con todos los archivos anteriormente mencionados se han realizado las pruebas y se deducen las siguientes conclusiones:

1. Siempre es necesario realizar una adaptación al locutor para un alto porcentaje de acierto.
2. Si la adaptación se realiza en el mismo entorno en el que será utilizada la herramienta, el porcentaje de acierto de palabra en ese entorno será mayor que si se utilizase un modelo adaptado sólo al locutor, pero fuera de él empeora.
3. Para este caso, realizar un preprocesado de audio no supone una mejora notable en el reconocimiento del habla, y por lo tanto se descarta su implementación en la herramienta.

# CAPÍTULO 5

# Gestión del proyecto

Este capítulo presenta el plan de comunicación interna a la organización, el cual incluye la descripción de la gestión del proyecto que permitirá su planificación, estimación de costes y recursos humanos, temporales y materiales, implementación y seguimiento [1].

## 5.1. Voz del cliente y voz del sector

El cliente de este proyecto será la empresa que integrará la herramienta en su aeronave. Se distinguen dos tipos, aquellos que integrarán la herramienta en la construcción de la aeronave y aquellos que utilizarán un dispositivo externo y lo colocarán en la aeronave para su utilización.

La voz del cliente permite dar una definición de las necesidades de cliente y las características del proyecto que cumplirán con los requisitos de cliente. Con el fin de conocer cuáles son los aspectos con mayor importancia para el cliente, se realiza un despliegue funcional de la calidad, entendida como un enfoque de los requisitos que el sistema ha de cumplir de acuerdo a las necesidades del usuario. Para ello utilizamos un QFD.

Así mismo, la voz del sector realiza un estudio de factores internos (Debilidades y Fortalezas) y factores externos (Oportunidades y Amenazas) con el fin de obtener un mayor conocimiento sobre el entorno operacional y el proyecto. Este análisis permite tomar mejores decisiones y fijar objetivos realistas conociendo los puntos fuertes y débiles.

[1] Estas herramientas de planificación estratégica, calidad, etc. han sido utilizadas por el autor y los datos son valoraciones generadas por el propio autor.

### 5.1.1. QFD (Quality Function Deployment)

La Tabla 5.1 muestra la trazabilidad de requisitos del cliente [2] con las características del proyecto [3]. Incluye la importancia que el cliente da a cada uno de sus requisitos principales.

Esta información permite enfocar el proyecto y su planificación potenciando las características más valoradas por el usuario, obteniendo así una mayor rentabilidad y eficacia del servicio prestado.

**QFD (Quality Function Deployment)**

| **Requisitos de usuario** | **Características del proyecto** | | | | | | | | **Importancia** |
|---|---|---|---|---|---|---|---|---|---|
| | **Motor ASR** | **Selección de idioma** | **Transcripción** | **Interfaz de usuario** | **Presentación** | **Control** | **Manejo sencillo** | **Precio** | |
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | |
| **Económico** | 1 | 0 | 0 | 3 | 3 | 0 | 0 | 9 | 5 |
| **Fácil de utilizar** | 9 | 1 | 1 | 9 | 3 | 9 | 9 | 0 | 4 |
| **Subtitular habla continua** | 9 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 5 |
| **Idiomas (Español/ Inglés)** | 1 | 9 | 3 | 0 | 0 | 0 | 0 | 0 | 3 |
| **Configurable** | 0 | 3 | 1 | 3 | 9 | 3 | 3 | 0 | 2 |
| **Robustez** | 9 | 3 | 1 | 1 | 0 | 1 | 1 | 0 | 3 |
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | |
| **Evaluación(E)** | 134 | 46 | 63 | 60 | 45 | 45 | 45 | 45 | |
| **Índice de dificultad (ID)** | 5 | 1 | 5 | 2 | 3 | 2 | 3 | 1 | |
| **Características a mejorar(CM)** | 670 | 46 | 315 | 120 | 135 | 90 | 135 | 45 | |

Relación MUY FUERTE = 9
Relación FUERTE = 3
Relación DÉBIL = 1
Relación NULA = 0
Importancia(índice de dificultad): 1 (baja) - 5 (alta)

Tabla 5.1: QFD Trazabilidad de requisitos

Las ecuaciones 5.1 y 5.2 muestran el cálculo de los datos:

- Evaluación (E) de características del proyecto

$$E_i = \sum_{j=1}^{8} (R_{j,i} + I_j) \tag{5.1}$$

[2] Los requisitos han sido determinados en función de los realizados en proyectos similares y de las necesidades del cliente.

[3] Las características han sido determinadas en función de las obtenidas en proyectos similares y por el propio autor.

donde:

| | |
|---|---|
| E_i | Resultado de la evaluación de la característica i |
| R_j,i | Relación del requisito j con la característica i |
| I_j | Importancia del requisito j |

- Características a mejorar (CM)

$$CM\_i = E\_i * ID\_i \tag{5.2}$$

donde:

| | |
|---|---|
| CM_i | Valor de la característica a mejorar |
| E_i | Resultado de la evaluación de la característica i |
| ID_i | Índice de dificultad de la característica i |

La Tabla 5.1 muestra las características del proyecto que se han de potenciar, en este caso las más notables son:

1 Motor ASR: la herramienta estará compuesta de un motor robusto de RAH.

4 Transcripción: la herramienta tendrá que ser capaz de transcribir en directo habla continua.

Las características 5 Presentación y 7 Manejo sencillo también son importantes ya que permitirán que la herramienta sea accesible, y se incluirán en el despliegue del proyecto.

### 5.1.2. DAFO

Este apartado permite conocer la situación real en que se encuentra la organización, así como el riesgo y oportunidades que brindan el entorno en que opera.

El análisis DAFO no gestiona costes.

- **Debilidades**

  Aspectos que limitan o reducen la capacidad de desarrollo. Constituyen una amenaza y deben ser controladas y superadas.

  **D1** Falta de experiencia en el sector de RAH.

  **D2** Limitaciones de los motores de RAH.

  **D3** Entorno de operación (interior de una aeronave).

  **D4** Personal limitado.

- **Amenazas**

  Fuerza del entorno que puede impedir la implantación de una estrategia, o los recursos que se requieren para su implantación o bien reducir su rentabilidad.

  **A1** Existen empresas establecidas con más experiencia en el sector de RAH.

  **A2** Limitación de presupuesto para la realización del proyecto.

  **A3** Limitación de material para la realización de pruebas.

- **Fortalezas**

  Capacidades, recursos, posiciones alcanzadas y ventajas competitivas propias de la organización [4].

  **F1** Personal con interés y motivación.

  **F2** Proyecto perteneciente al ámbito de la organización (especializada en subtitulado y audio-descripción).

  **F3** Precio final asequible y competitivo.

  **F4** Permite continuar con el desarrollo e investigación de motores de RAH para la creación de herramientas de accesibilidad.

- **Oportunidades**

  Posibilidad para mejorar la eficacia de la organización.

  **O1** Creciente demanda de sistemas con motores de RAH en entornos públicos y hogares para accesibilidad de gente con discapacidad sensorial.

  **O2** Existencia de motores de RAH OpenSource que abaratan los costes y se convierten en una herramienta de investigación y aprendizaje fácil de obtener y manejar.

  **O3** Avance y desarrollo de nuevas tecnologías.

  **O4** Avances tecnológicos en equipos y sistemas que simplifican las infraestructuras para el desarrollo del proyecto.

### 5.1.3. Valoración de factores internos y externos

Estudio de los puntos fuertes y débiles del proyecto. Las Tablas 5.2 y 5.3 muestran una valoración del análisis DAFO previamente realizado.

Definición de columnas:

| | |
|---|---|
| Peso | Relevancia relativa de cada factor |
| Valor | Valoración de la organización en cada uno de los factores (escala de 0 a 3, siendo 0 mal y 3 excelente) |
| Peso * Valor | Producto de la columna Peso por su Valoración |

| **Debilidades** | | | | **Amenazas** | | | |
|---|---|---|---|---|---|---|---|
| **Debilidades** | **Peso** | **Valor** | **Peso * Valor** | **Amenazas** | **Peso** | **Valor** | **Peso * Valor** |
| **D1** | 40 | 1 | 40 | **A1** | 60 | 1 | 60 |
| **D2** | 15 | 1 | 15 | **A2** | 25 | 2 | 50 |
| **D3** | 25 | 2 | 50 | **A3** | 15 | 1 | 15 |
| **D4** | 20 | 3 | 60 | | | | |
| **Total** | 100 | | 165 | | 100 | | 125 |

Tabla 5.2: DAFO: Debilidades y Amenazas

[4] Podemos utilizar, también, el término "capacidades esenciales" para referirnos a las fortalezas.

| Fortalezas | | | | Oportunidades | | | |
|---|---|---|---|---|---|---|---|
| **Fortalezas** | **Peso** | **Valor** | **Peso * Valor** | **Oportunidades** | **Peso** | **Valor** | **Peso * Valor** |
| **F1** | 20 | 2 | 40 | **O1** | 30 | 2 | 60 |
| **F2** | 40 | 3 | 120 | **O2** | 30 | 2 | 60 |
| **F3** | 25 | 2 | 50 | **O3** | 30 | 3 | 90 |
| **F4** | 15 | 1 | 15 | **O4** | 10 | 1 | 10 |
| **Total** | 100 | | 225 | | 100 | | 220 |

Tabla 5.3: DAFO: Fortalezas y Oportunidades

La Tabla 5.4 resume los resultados anteriormente obtenidos y sirve para generar la gráfica de diagnóstico Gráfica 5.1.

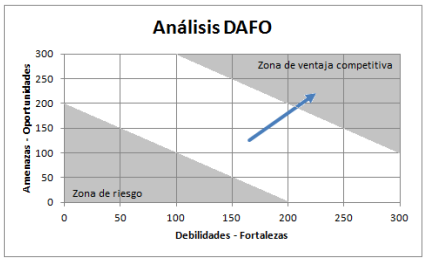



Figura 5.1: Gráfica: Análisis DAFO

| Resumen DAFO | |
|---|---|
| | **Puntos** |
| **D,A** | (165,125) |
| **F,O** | (225,220) |

Tabla 5.4: Resumen DAFO

Con este instrumento se pretende representar de forma fácil, qué relación mantienen las oportunidades y las amenazas con las debilidades y fortalezas. Como se observa en la Figura 5.1 se han destacado dos zonas, la zona de riesgo para la realización del proyecto y la zona de ventaja competitiva.

Para la realización de este proyecto no existe nivel de riesgo y entra en la zona competitiva del mercado, por lo tanto se prosigue con su planificación e implementación. En concreto habrá que aprovechar las oportunidades, sobre todo O1 y O3, para transformar las debilidades en fortalezas, empezando por las que tengan más confluencia con las oportunidades (D2 y D3); mantener las fortalezas relevantes (F1, F2 y F3); determinar las amenazas que se pueden afrontar con menor éxito para potenciar las capacidades relacionadas (A1) y atender la mayor de nuestras debilidades por ser la más afectada por las amenazas (D1 y D4).

## 5.2. Planificación

En esta sección se muestra la planificación del proyecto, el desglose de tareas a realizar y la duración estimada de cada una de ellas.

### 5.2.1. Listado de tareas y duración estimada

A continuación se muestra en la Tabla 5.5 el listado de tareas, agrupadas por funciones, su duración estimada en días[5] y sus tareas predecesoras. Incluye los hitos (H) del proyecto que servirán como puntos de control.

**Listado de tareas y duración estimada**

| Id. | Tarea | | Duración (días) | Predecesora |
|---|---|---|---|---|
| 1 | H0 | Inicio del proyecto | 0 | |
| 2 | **Planificación** | | 8 días | |
| 3 | P0 | Plan de comunicación interna | 5 | H0 |
| 4 | P1 | Listado y planificación de tareas | 3 | P0 |
| 5 | **Financiación** | | 6 días | |
| 6 | F0 | Estimación de personal y recursos | 3 | P1 |
| 7 | F1 | Estimación de costes | 3 | F0 |
| 8 | H1 | Decisión de realización del proyecto | 0 | F1 |
| 9 | **Desarrollo** | | 135 días | |
| 10 | D0 | Especificación funcional: requisitos de usuario y técnicos | 20 | H1 |
| 11 | D1 | Diseño: arquitectura y detallado | 25 | D0 |
| 12 | H2 | Control de especificación y diseño | 0 | D1 |
| 13 | D2 | Implementación de la herramienta | 60 | H2 |
| 14 | D3 | Pruebas y simulaciones | 30 | D2 |
| 15 | H3 | Prototipo de la herramienta | 0 | D3 |
| 16 | **Documentación** | | 22 días | |
| 17 | M0 | Manuales | 7 | H3 |
| 18 | M1 | Redacción del proyecto | 15 | P0, P1, F0, F1, D0, D1, M0 |
| 19 | H4 | Fin del proyecto | 0 | M2 |
| **Total** | | | 171 días | |

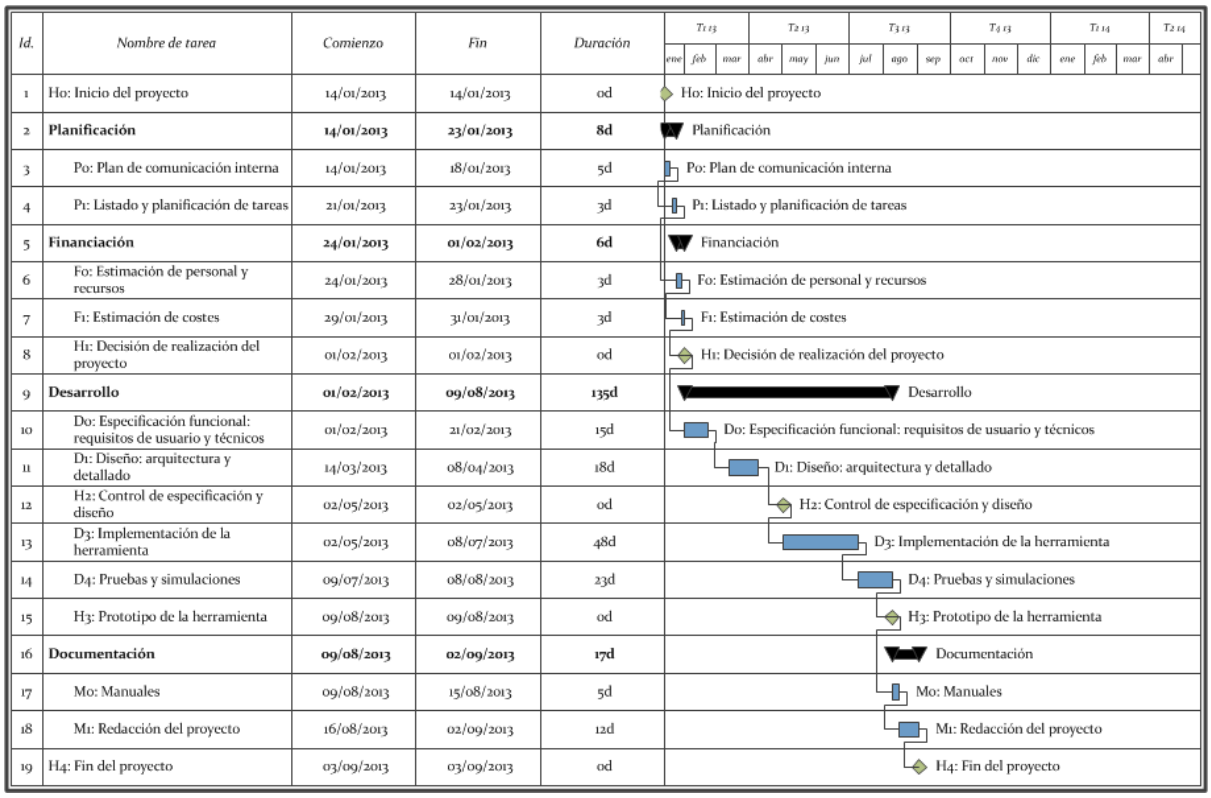
Tabla 5.5: Listado de tareas y duración estimada

En base a la planificación se sitúa el desarrollo del proyecto con una duración estimada de 171 días.

[5]Definiéndose para la realización de este proyecto: 1 día = 4 horas de trabajo.

### 5.2.2. Calendario de tareas del proyecto: Gantt

A tenor del listado de tareas de la Tabla 5.5 se muestra, de una forma ordenada, la sucesión de cada una de las tareas que componen el desarrollo del proyecto en el siguiente Gantt (Figura 5.2), iniciado para el 14 de enero de 2013[6].



| Id. | Nombre de tarea | Comienzo | Fin | Duración |
|---|---|---|---|---|
| 1 | Ho: Inicio del proyecto | 14/01/2013 | 14/01/2013 | 0d |
| 2 | **Planificación** | **14/01/2013** | **23/01/2013** | **8d** |
| 3 | Po: Plan de comunicación interna | 14/01/2013 | 18/01/2013 | 5d |
| 4 | P1: Listado y planificación de tareas | 21/01/2013 | 23/01/2013 | 3d |
| 5 | **Financiación** | **24/01/2013** | **01/02/2013** | **6d** |
| 6 | Fo: Estimación de personal y recursos | 24/01/2013 | 28/01/2013 | 3d |
| 7 | F1: Estimación de costes | 29/01/2013 | 31/01/2013 | 3d |
| 8 | H1: Decisión de realización del proyecto | 01/02/2013 | 01/02/2013 | 0d |
| 9 | **Desarrollo** | **01/02/2013** | **09/08/2013** | **135d** |
| 10 | Do: Especificación funcional: requisitos de usuario y técnicos | 01/02/2013 | 21/02/2013 | 15d |
| 11 | D1: Diseño: arquitectura y detallado | 14/03/2013 | 08/04/2013 | 18d |
| 12 | H2: Control de especificación y diseño | 02/05/2013 | 02/05/2013 | 0d |
| 13 | D3: Implementación de la herramienta | 02/05/2013 | 08/07/2013 | 48d |
| 14 | D4: Pruebas y simulaciones | 09/07/2013 | 08/08/2013 | 23d |
| 15 | H3: Prototipo de la herramienta | 09/08/2013 | 09/08/2013 | 0d |
| 16 | **Documentación** | **09/08/2013** | **02/09/2013** | **17d** |
| 17 | Mo: Manuales | 09/08/2013 | 15/08/2013 | 5d |
| 18 | M1: Redacción del proyecto | 16/08/2013 | 02/09/2013 | 12d |
| 19 | H4: Fin del proyecto | 03/09/2013 | 03/09/2013 | 0d |



Figura 5.2: Gráfica: Calendario de tareas del proyecto Gantt

## 5.3. Estimación de recursos y costes

Una vez descritas las tareas y su duración se procede a estimar los recursos humanos y materiales necesarios en tiempo y coste.

### 5.3.1. Lista de recursos humanos y materiales necesarios

Los recursos humanos necesarios para el desarrollo del proyecto se enumeran en la Tabla 5.6, especificando el nombre del trabajo que desempeñan, la cantidad de personal necesario, salario por hora que cobra y cantidad de horas de trabajo necesarias.

[6]Esta fecha es una referencia de inicio.

| Listado de recursos humanos | | | |
|---|---|---|---|
| **Id.** | **Nombre** | **Cantidad** | **Coste (Euros/ hora)** |
| RH1 | Director de proyecto | 1 | 36.18 |
| RH2 | Jefe de proyecto | 1 | 23.19 |
| RH3 | Desarrollador (becario) | 1 | 5 |

Tabla 5.6: Listado de recursos humanos

Los recursos materiales necesarios para el desarrollo del proyecto se enumeran en la Tabla 5.7, especificando el nombre del recurso, la cantidad necesaria y coste por unidad [7] al cual se le calculará la amortización según se indica.

| Listado de recursos materiales | | | |
|---|---|---|---|
| **Id.** | **Nombre** | **Cantidad** | **Coste (Euros/ unidad)** |
| RM1 | Ordenador portatil + O.S Windows 7 | 1 | 500 |
| RM2 | Dragon Naturally Speaking | 1 | 100 |
| RM3 | Microsoft Office 2010 (Hogar y estudiantes) | 1 | 119 |
| RM4 | Matlab Student Version 2012a | 1 | 111 |
| RM5 | Material de oficina | 1 | 20 |
| RM6 | CMU Sphinx | 1 | 0 |

Tabla 5.7: Listado de recursos materiales

Para la estimación de costes se realiza un cálculo de la amortización (ecuación 5.3) de los recursos materiales:

$$Amortiza = \frac{Factura}{Deprecia} * Usa \tag{5.3}$$

Factura = coste de facturación del recuso material.
Deprecia = tiempo de depreciación, en este caso se estima en 36 meses (3168 horas[8]).
Usa = tiempo de uso del recurso material en el proyecto.

### 5.3.2. Estimación de costes

Estimación de costes (Tabla 5.8) para la realización del proyecto según su duración estimada y la tasa estándar mostrada en los recursos humanos y materiales. El coste se desglosa por resumen de tareas.

[7] El coste del producto es redondeado respecto al coste real en diciembre del 2013
[8] Teniendo en cuenta 22 días por mes y 4 horas por día.

| Estimación de costes | | |
|---|---|---|
| **Recurso** | **Horas** | **Coste (Euros)** |
| | | *Costes de personal* |
| RH1 | 68.4 | 2611.51 |
| RH2 | 273.6 | 6344.78 |
| RH3 | 684 | 3420.00 |
| | | *Costes de amortización* |
| RM1 | 684 | 107.95 |
| RM2 | 360 | 11.36 |
| RM3 | 684 | 25.69 |
| RM4 | 120 | 4.20 |
| RM5 | 684 | 4.32 |
| RM6 | 120 | 0.00 |
| **Total costes personal y amortización** | | **12529.83 Euros** |
| | | *Costes de oficina* |
| Local | | 1200 |
| **Total** | | **13729.83 Euros** |
| | | *Base imponible* |
| I.V.A. | | 21 % |
| | | *Coste total estimado* |
| **Total** | | **16640.55 Euros** |

Tabla 5.8: Estimación de costes

En total se estima un coste de **16640.55 Euros** para el proyecto. Este coste podrá verse alterado por el incremento de los precios en los recursos materiales o si se ve alterada la planificación.

## 5.4. Control

Reuniones semanales de 2 horas para comentar progresos, dificultades, retrasos, cambios y cualquier tema relacionado con el proyecto.

Se han fijado 5 hitos que se utilizarán como mecanismo de control y muestra de los resultados obtenidos hasta la fecha. También servirán para realizar cambios en la planificación y en el proyecto.

### 5.4.1. Mecanismos de seguimiento y comunicación

Se informará por email sobre los progresos del proyecto durante el resto de la semana, fuera de las reuniones. Cualquier cambio será informado por un documento escrito, previo aviso al jefe de proyecto, por email o en persona.

# CAPÍTULO 6

# Conclusiones

Este Proyecto Fin de Carrera ha presentado una solución para el problema de accesibilidad para discapacitados sensoriales, en particular con deficiencia auditiva, en el interior de una aeronave.

El objetivo era implementar una herramienta capaz de generar una transcripción en directo del comunicado de la tripulación al pasaje aéreo y presentar el texto en formato subtítulo, pudiendo visualizarlo en cualquier dispositivo a través de su navegador web o, en caso de existir, los monitores del avión.

Las pruebas han sido realizadas en un entorno controlado en directo y diferido. Simulando las condiciones del interior de una aeronave debido a la imposibilidad de trabajar en un entorno real durante el desarrollo del proyecto.

## 6.1. Conclusiones

El proyecto ha desarrollado una solución cumpliendo con el objetivo de resolver el problema existente presentando de forma textual mensajes orales al pasaje aéreo. Resulta importante recordar que el proyecto fue desarrollado en un escenario acotado para la investigación y no ha sido llevado al entorno real, lo cual limita los resultados obtenidos.

La mayor dificultad del presente trabajo fue la integración del motor de RAH en la herramienta y la comprobación de su funcionamiento en el entorno solicitado. En primer lugar, porque el motor de RAH limita a la herramienta, ya que no es capaz de transcribir con un acierto de palabra del 100 % en directo. Para conseguir una mayor tasa de acierto se requeriría una locución pausada y clara. Esto no es posible obtenerlo en el entorno real en el cual se pretende utilizar la herramienta. En segundo lugar, porque no ha sido posible llevar a cabo las pruebas en el interior de un avión, realizándose en su lugar simulaciones con el equipo de laboratorio.

En el escenario en el que opera la herramienta, el ruido es un factor muy importante. Tal y como muestran las pruebas, se obtienen mejores tasas de acierto si se adapta al locutor y al entorno en el que será utilizado. Aunque para poder conseguirlo sería necesario realizar esas adaptaciones en el mismo lugar. Lo cual no es factible y la mejor aproximación es obtener unas grabaciones del ruido con el micrófono que sería utilizado y mezclar los archivos a adaptar con ese ruido. Que es la aproximación que se ha realizado en el proyecto.

Sabiendo que la señal de audio depende de factores cambiantes como el locutor, tono de voz, estrés, ruido, etc. la señal podría ser mejorada mediante un aumento de su ganancia y filtrando el ruido, para

obtener una mejor tasa de acierto en el reconocedor. Pero en esta herramienta se ha optado por adaptar al locutor y al entorno, sin filtrar el ruido ya que el reconocedor obtiene el audio directamente de la tarjeta de sonido.

Las pruebas requirieron para su realización bastante trabajo. Primero obtener todos los archivos de audio con sus respectivas transcripciones, convertirlos a formato .wav y misma frecuencia de muestreo. A continuación, mezclarlos con los distintos niveles de ruido y adaptar el reconocedor creando modelos acústicos por cada nivel de ruido. Después, transcribir todos los archivos con cada uno de los modelos acústicos y finalmente comparar el texto resultante con el original. Este trabajo fue parte automatizado y parte manual, con esto se consiguió caracterizar la herramienta de una forma más sencilla.

Las pruebas han sido realizadas sólo para habla inglesa, debido a que los reconocedores están mejor entrenados en este idioma. Además es más fácil encontrar numerosos archivos de audio con su correspondiente transcripción. El motor de RAH utilizado posee una buena tasa de acierto tanto para español como inglés. Pero debido a la falta de bases de datos de audio con transcripción en español se decidió optar sólo por el otro idioma. Es importante destacar que en ningún momento se ha utilizado una base de datos de audios correspondientes a mensajes generados en el interior de una aeronave, en ninguno de los dos idiomas para los que se ha generado la herramienta.

Se concluye que dado el objetivo del proyecto, la herramienta desarrollada realiza correctamente su función dentro del escenario definido, considerando que los motores de RAH son una tecnología en desarrollo y no perfecta cuyos resultados acotan la calidad de la transcripción obtenida.

La viabilidad de esta herramienta hace posible la integración en aeronaves en un futuro, mejorando así la accesibilidad universal y por lo tanto la calidad de vida.

## 6.2. Líneas futuras de trabajo

Esta herramienta permite asentar una base en este ámbito que motivará la búsqueda de otras soluciones en este tipo de situaciones. Toda mejora supone un cambio muy favorable.

Realizar simulaciones en el entorno real y realizar pruebas con grabaciones de audio en español, pertenecientes al entorno y con un micrófono situado correctamente en el hablante, acotarían mejor los resultados obtenidos y caracterizarían de forma completa la herramienta.

También, utilizar un motor de reconocimiento capaz de transcribir en directo con una muy alta tasa de acierto y, si fuese posible, independiente del locutor. Esto sería una gran mejora ya que simplificaría aún más la utilización de la herramienta, puesto que no habría que crear ni perfiles ni modelos acústicos y sólo se tendría que iniciar directamente la herramienta. Lo cual limitaría y haría más sencilla la interacción de la tripulación con la herramienta.

Otra línea de trabajo sería obtener un reconocimiento robusto en entornos ruidosos. Se han mencionado varias técnicas de robustez que podrían ser aplicadas a motores ya existentes y comprobar si se consiguen o no las mejoras en tasas de acierto. Otra mejora sería adaptar el modelo de lenguaje y limitarlo al contexto en el que se ha de utilizar.

Finalmente, este proyecto puede motivar a realizar más herramientas de ayuda a personas con discapacidad sensorial en distintos entornos.

# Anexo A: Acrónimos

| | |
|---|---|
| **ASR** | **A**utomatic **S**peech **R**ecognition |
| **CESyA** | **C**entro **E**spañol del **S**ubtitulado y la **A**udiodescripción |
| **CMN** | **C**epstral **M**ean **N**ormalization |
| **CVN** | **C**epstral **V**ariance **N**ormalization |
| **DNS** | **D**ragon **N**aturally **S**peaking |
| **EMT** | **E**mpresa **M**unicipal de **T**ransporte |
| **GUI** | **G**raphic **U**ser **I**nterface |
| **HEQ** | **H**istogram **EQ**ualization |
| **HMM** | **H**idden **M**arkov **M**odel |
| **INE** | **I**nstituto **N**acional de **E**stadística |
| **LGCA** | **L**ey **G**eneral de **C**omunicación **A**udiovisual |
| **LPC** | **L**inear **P**redictive **C**oding |
| **MAP** | **M**aximum **A** **P**osteriori |
| **MEL** | **MEL** scale cepstral analysis |
| **MFCC** | **M**el **F**requency **C**epstral **C**oefficient |
| **MLLR** | **M**aximum **L**ikelihood **L**inear **R**egression |
| **OMS** | **O**rganización **M**undial de la **S**alud |
| **ONU** | **O**rganización de las **N**aciones **U**nidas |
| **PFC** | **P**royecto **F**in de **C**arrera |
| **PLP** | **P**erceptual **L**inear **P**rediction |
| **PMC** | **P**arallel **M**odel **C**ombination |
| **RAH** | **R**econocimiento **A**utomático del **H**abla |
| **SS** | **S**pectarl **S**ubstraction |
| **UC3M** | **U**niversidad **C**arlos **III** de **M**adrid |
| **VTS** | **V**ector **T**ailor **S**eries |

# Bibliografía


[1] AENOR. *UNE 153010:2012. Subtitulado para personas sordas y personas con discapacidad auditiva*. AEN/CTN 153 - Productos de apoyo para personas con discapacidad, May 2015.

[2] M.A. Anusuya and S.K. Katti. Front end analysis of speech recognition: a review. *International Journal of Speech Technology*, 14(2):99–145, 2011.

[3] Research at Google. *Speech Processing*. Google, 2014. Available at: http://research.google.com/pubs/SpeechProcessing.html.

[4] S. Boll. Suppression of acoustic noise in speech using spectral subtraction. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 27(2):113–120, Apr 1979.

[5] L. Carroll. *Alice's Adventures in Wonderland*. Internet Archive, 2013. Read by Cory Doctorow. Available at http://archive.org/details/AliceInWonderlandReadByCoryDoctorow.

[6] L. Carroll. *Alice's Adventures in Wonderland*. Librivox, 2013. Read by LibriVox Volunteers. Available at https://librivox.org/alices-adventures-in-wonderland-by-lewis-carroll.

[7] CERMI. Discapacidad. AENA instala 250 equipos de inducción magnética para orientar a las personas sordas en los aeropuertos. *Semanal.CERMI*, 2012. Available at: http://semanal.cermi.es/noticia/AENA-instala-equipos-inducci%C3%B3n-magn%C3%A9tica-personas-sordas-aeropuertos.aspx.

[8] CESyA. *Centro Español del Subtitulado y la Audiodescripción*, 2014. Information available at: http://www.cesya.es/.

[9] CMUSphinx. *NISTAlign*. Carnegie Mellon University. Available at: http://cmusphinx.sourceforge.net/sphinx4/javadoc/edu/cmu/sphinx/util/NISTAlign.html.

[10] CMUSphinx. *Open Source Toolkit For Speech Recognition*. Carnegie Mellon University. Information available at: http://cmusphinx.sourceforge.net/.

[11] Biswajit Das, Sandipan Mandal, Pabitra Mitra, and Anupam Basu. Aging speech recognition with speaker adaptation techniques: Study on medium vocabulary continuous bengali speech. *Pattern Recognition Letters*, 34(3):335 – 343, 2013.

[12] Subdirección General de Difusión Estadística INE. Encuesta sobre discapacidades, deficiencias y estado de salud. *INEbase*, 2000. Available at: http://www.ine.es/prodyser/pubweb/discapa/disctodo.pdf.

[13] Gobierno de España. Accesibilidad,TIC y Educación. In Centro Nacional de Información y Comunicaciń Educativa (CNICE-MEC), editor, *Desarrollo, transferencia y difusión social de la investigación en TIC para la Educación*, volume 17. Instituto Nacional de Tecnologías Educativas y de Formación al Profesorado. Ministerio de Educación. Available at: http://ares.cnice.mec.es/informes/17/contenido/20.htm.

[14] Gobierno de España. *Constitución Española*. Cortes Generales, 1978.

[15] Gobierno de España. Ley 51/2003, de 2 de diciembre, de igualdad de oportunidades, no discriminación y accesibilidad universal de las personas con discapacidad. In *Boletín Oficial del Estado*, volume 289, chapter BOE-A-2003-22066. Ministerio de la Presidencia, December 2003.

[16] Gobierno de España. Ley 7/2010, de 31 de marzo, general de la comunicación audiovisual. In *Boletín Oficial del Estado*, volume 79, chapter BOE-A-2010-5292. Ministerio de la Presidencia, March 2010.

[17] EMT. *La EMT trata de facilitar el acceso al autobús, a los usuarios con discapacidades auditivas*, November 2012. Available at: http://blog.emtmadrid.es/2012/11/05/la-emt-trata-de-facilitar-el-acceso-al-autobus-a-los-usuarios-con-discapaci

[18] Parlamento Europeo. *La Carta de los Derechos Fundamentales de la Unión Europea*. 2010.

[19] M. J F Gales and S.J. Young. Robust continuous speech recognition using parallel model combination. *Speech and Audio Processing, IEEE Transactions on*, 4(5):352–359, Sep 1996.

[20] A. Gallardo Antolín and C. Peláez Moreno. *Panorámica de las Tecnologías del Habla*. Dpto. de Teoría de la Señal y Comunicaciones. Universidad Carlos III de Madrid, 2013.

[21] H. Höge, S. Hohenner, B. Kämmerer, N. Kunstmann, S. Schachtl, M. Schönle, and P. Setiawan. Automotive speech recognition. In *Automatic Speech Recognition on Mobile Devices and over Communication Networks*, Advances in Pattern Recognition, pages 347–373. Springer. London, 2008.

[22] iSpeech. *iSpeech*, 2014. Available at: www.iSpeech.org.

[23] Julius. *Julius. Open-Source Large Vocabulary CSR Engine*, 2013. Available at: http://julius.sourceforge.jp/en_index.php.

[24] Nam Soo Kim, June Sig Sung, and Doo Hwa Hong. Factored MLLR Adaptation. *Signal Processing Letters, IEEE*, 18(2):99–102, Feb 2011.

[25] H. D. Kopald, A. Chanen, S. Chen, E. C. Smith, and R. M. Tarakan. Applying automatic speech recognition technology to air traffic management. In *Digital Avionics Systems Conference (DASC), IEEE/AIAA*, volume 32nd, Oct 2013.

[26] H. Kurtulus Ozcan and S. Nemlioglu. In-cabin noise levels during commercial aircraft flights. In *Journal of the Canadian Acoustic Organization*, volume 34. Canadian Acoustics, 2006.

[27] K. Lagus and M. Kurimo. Language model adaptation in speech recognition using document maps. In *Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on*, pages 627–636, 2002.

[28] J. Li, L. Deng, Y. Gong, and R. Haeb-Umbach. An overview of noise-robust automatic speech recognition. Number 99, pages 1–1, 2014.

[29] A. Lozano Torrijos. *Proyecto fin de carrera: Sistema para el alineamiento de subtítulos y audio en escenarios de rehablado*. Universidad Carlos III de Madrid, June 2012.

[30] J. Lumsden, S. Durling, and I. Kondratova. A comparison of microphone and speech recognition engine efficacy for mobile data entry. *NRC Publications Archive.*, pages pp. 519–527, November 2008.

[31] MathWorks. *MATLAB*. Mathworks, 1994-2013. Information available at: http://www.mathworks.es/.

[32] Microsoft. *Microsoft Speech Platform*. Available at: http://msdn.microsoft.com/en-us/library/hh361572.aspx.

[33] Microsoft and Cambridge University. *Hidden Markov Model Toolkit (HTK)*. Available at: http://htk.eng.cam.ac.uk/.

[34] N. Morgan and H. Bourlard. Continuous speech recognition. *Signal Processing Magazine, IEEE*, 12(3):24–42, May 1995.

[35] A. Moriano Roldán. *Proyecto fin de carrera: Servicio Web de Subtitulado en Diferido. APEINTA*. Universidad Carlos III de Madrid, October 2013.

[36] Nuance. *Dragon Naturally Speaking*. Available at: http://www.nuance.es/dragon/index.htm.

[37] OMS. Deafness and hearing loss. *OMS Media centre*, no. 300, February 2013. Available at: http://www.who.int/mediacentre/factsheets/fs300/en/index.html.

[38] ONU. *48/96. Normas Uniformes sobre la igualdad de oportunidades para las personas con discapacidad*. December 1993.

[39] S. Preeti and K. Parneet. Automatic speech recognition: A review. In *International Journal of Engineering Trends and Technology*, volume 4, page 132. February 2013.

[40] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, Feb 1989.

[41] Michael Raj.T.F., B. RajaKumar, S. Swaminathan, and M. Ramkumar. A novel approach: Voice enabled interface with intelligent voice response system to navigate mobile devices for visually challenged people. In *Emerging Trends in VLSI, Embedded System, Nano Electronics and Telecommunication System (ICEVENT), 2013 International Conference on*, pages 1–4, Jan 2013.

[42] Douglas A. Reynolds, Thomas F. Quatieri, and Robert B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10(1–3):19 – 41, 2000.

[43] Koichi Shinoda. *Speaker Adaptation Techniques for Automatic Speech Recognition*. APSIPA ASC 2011, Tokyo, Japan, Tokyo Institute of Technology edition, 2011.

[44] G. Shires and H. Wennborg. *Web Speech API Specification*. Google Inc. Contributors to the Web Speech API Specification, 2012. Available at: https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html.

[45] Khe Chai Sim. Approximated parallel model combination for efficient noise-robust speech recognition. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 7383–7387, May 2013.

[46] Z-H. Tan and B. Lindberg. Network, distributed and embedded speech recognition: An overview. In *Automatic speech recognition on mobile devices and over communication networks*. London : Springer,, 2008.

[47] R. Wenxia, Z. Huili, and L. Wenzhe. Realization of isolated-words speech recognition system. In *Circuits, Communications and Systems, 2009. PACCS '09. Pacific-Asia Conference on*, pages 353–355, May 2009.

[48] K. Wojcicki and K. Florian. *Add Noise*. MATLAB, 2011-2013. BSD License. Available at: http://www.mathworks.com/matlabcentral/fileexchange/32136-add-noise/content/addnoise/addnoise.m.

[49] P.C. Woodland, J. J. Odell, V. Valtchev, and S. J. Young. Large vocabulary continuous speech recognition using HTK. In *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, volume ii, pages II/125–II/128 vol.2, Apr 1994.

[50] Xiong Xiao, Jinyu Li, Eng-Siong Chng, and Haizhou Li. Maximum likelihood adaptation of histogram equalization with constraint for robust speech recognition. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 5480–5483, May 2011.

[51] H. Xuedong and K.F. Lee. On speaker-independent, speaker-dependent, and speaker-adaptive speech recognition. *Speech and Audio Processing, IEEE Transactions on*, 1(2):150–157, Apr 1993.

[52] S. Yoshizawa, N. Hayasaka, N. Wada, and Y. Miyanaga. Cepstral gain normalization for noise robust speech recognition. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, volume 1, pages I–209–12 vol.1, May 2004.

[53] B. Zafarifar, Jingyue Cao, and P.H.N. de With. Instantaneously responsive subtitle localization and classification for tv applications. In *Consumer Electronics (ICCE), 2011 IEEE International Conference on*, pages 165–166, Jan 2011.

[54] E. Zavarehei. *SBoll79*. MATLAB, 2005. No BSD License. Available at http://www.mathworks.com/matlabcentral/fileexchange/7675-boll-spectral-subtraction/content/SSBoll79.m.