

# SUBTITULADO EN TIEMPO REAL DE INFORMATIVOS EN DIRECTO PARA LA TELEVISIÓN MEDIANTE RECONOCIMIENTO AUTOMÁTICO DEL HABLA

Alfonso Ortega<sup>1</sup>, José Enrique García<sup>1</sup>, Eduardo Lleida<sup>1</sup>,  
Emiliano Bernués<sup>2</sup>, Daniel Sánchez<sup>3</sup> y Miguel Ferrer<sup>2</sup>

<sup>1</sup> Aragón Institute of Engineering Research (I3A).  
University of Zaragoza, Spain.

<sup>2</sup> Corporación Aragonesa de Radio y Televisión, Spain.

<sup>3</sup> Aranova, Spain.

# RESUMEN

- Introducción
- Descripción del Sistema
- Módulo de Obtención del Texto
- Módulo de Alineamiento temporal Habla-Texto
- Detector de Noticias
- Evaluación de Prestaciones
- Conclusiones

# INTRODUCCION

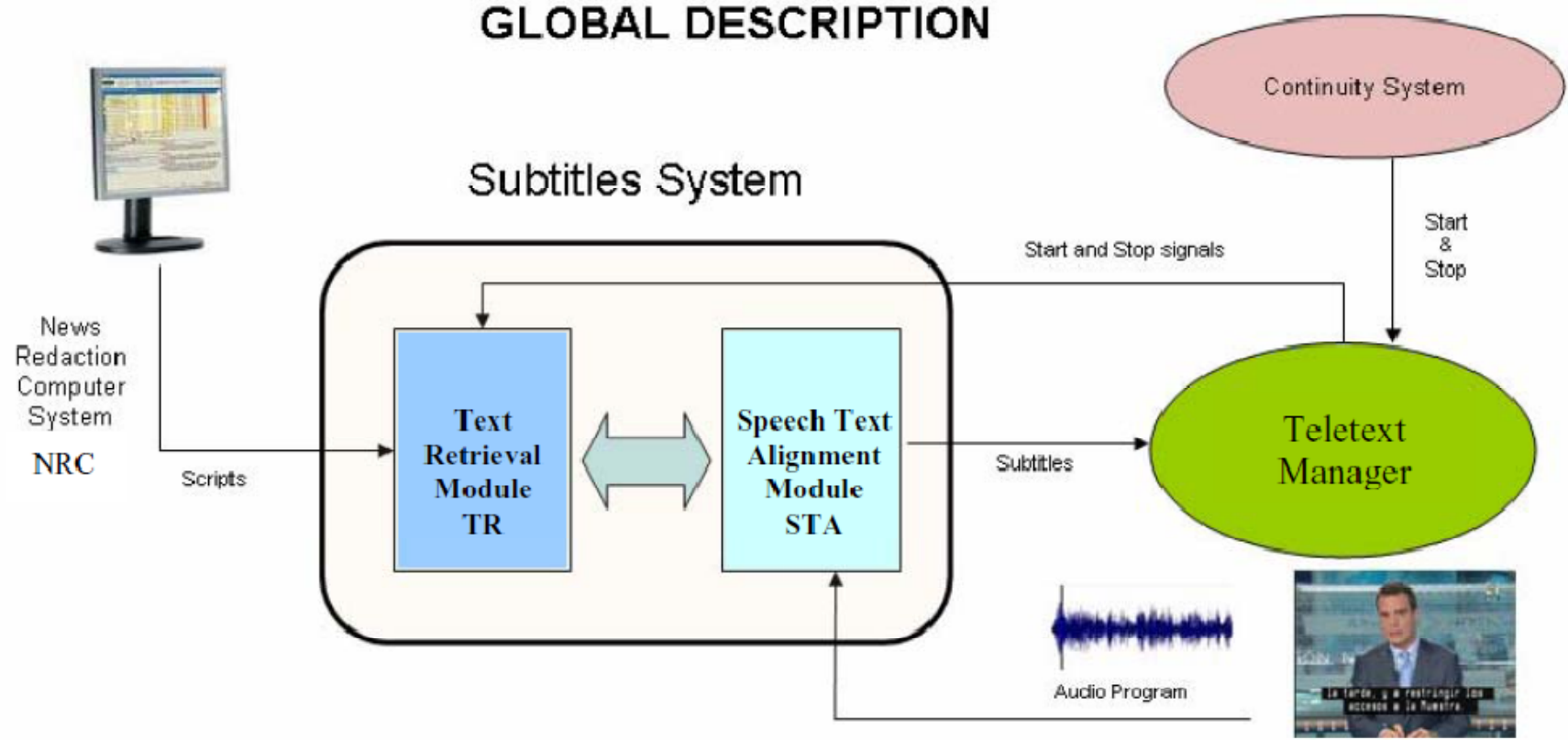
- Subtitulado en Tiempo real de Informativos en TV:
  - Necesidad para personas sordas o con problemas de audición
  - Útil para todo el mundo en lugares ruidosos.
- Subtítulos
  - Normalmente disponibles para programas grabados
  - Tarea difícil y costosa para programas en directo
    - Necesidad de personal cualificado (Estenógrafos)
    - Sistemas Semi-Automáticos
      - Locutores en la sombra u otro tipo de asistencia humana.
    - Sistemas completamente automáticos
      - Uso de motores de reconocimiento automático del habla
        - » Latencia
        - » Precisión

# INTRODUCCION

- Objetivo:
  - Bajo coste
    - Sin necesidad de asistencia humana.
  - Alta precisión y baja latencia
    - Evitar el uso de transcripciones generadas libremente por módulos de RAH.
    - Uso de los guiones previamente elaborados.
- Nuestra propuesta:
  - Obtener el texto de las noticias.
  - Usar el Reconocimiento Automático del Habla para alinear temporalmente texto y voz.

# DESCRIPCIÓN DEL SISTEMA

## GLOBAL DESCRIPTION



# DESCRIPCIÓN DEL SISTEMA

- Módulos:
  - Text Retrieval Module. (TRM)
    - Encargado de recuperar los textos de las noticias
    - Conectado a los ordenadores de la redacción.
    - Envía los textos al módulo de alineamiento.
  - Speech-Text Alignment. (STA)
    - Encargado de alinear temporalmente la voz y el texto.
    - Motor de Reconocimiento Automático del Habla.
  - Detector de Noticias.
    - Notifica automáticamente cuál es la siguiente noticia en entrar en caso de pérdida momentánea de sincronismo.

# DESCRIPCIÓN DEL SISTEMA

- Conexiones Externas:
  - Ordenador de la Redacción del Informativo (News Redaction Computer, NRC)
    - Contiene los guiones de las noticias.
  - Sistema de Gestión del Teletexto
    - Recibe los subtítulos desde el módulo de alineamiento temporal voz-texto.
    - Muestra estos subtítulos a través de la página 888.
  - Sistema de Continuidad
    - Asiste al sistema de subtitulado enviando las señales de comienzo y fin del informativo.

# TEXT RETRIEVAL MODULE

- Funciones:
  - Obtiene los guiones de las noticias del ordenador de la redacción del informativo.
    - Elimina la información no necesaria
    - Formatea la información relevante (XML)
  - Envía los textos al módulo de alineamiento temporal (STA Module)
  - Monitoriza continuamente el contenido del ordenador de la redacción y envía cada nuevo cambio al módulo de alineamiento en tiempo real.
  - Redirige las señales de comienzo y fin al módulo de alineamiento.



# SPEECH-TEXT ALIGNMENT

- Funciones:
  - Recibe los guiones del módulo TRM.
  - Recibe el audio desde producción.
  - Alinea automáticamente los textos contenidos en los guiones con la voz del locutor.
  - Da formato y envía los subtítulos al gestor del teletexto.

# SPEECH-TEXT ALIGNMENT

- Propiedades:
  - Motor de Reconocimiento Automático del Habla.
    - Basado en HMM
      - Unidades acústicas incontextuales
      - GMM de 16 componentes
      - Modelos acústicos independientes del locutor.
    - Vectores de Características Acústicas
      - 12 MFCC + logEnergy
      - Delta y Delta-Delta
    - Modelo de Lenguaje
      - Gramática de Estados Finitos
        - » Construida haciendo uso de los textos recibidos.
        - » Emplea redes de fonemas para detener el progreso de la gramática cuando hay fragmentos de audio sin transcripción asociada

# SPEECH-TEXT ALIGNMENT

- Detector de Noticias:
  - Determina cuál es la noticia actualmente en el aire.
    - Tarea clave
    - El orden con el que llegan las noticias puede variar, el locutor modifica el contenido de una noticia, la noticia anterior no finaliza correctamente, ...
    - La siguiente noticia no estará lista para ser subtitulada cuando se comience a locutar.

# SPEECH-TEXT ALIGNMENT

- Detector de Noticias:
  - Basado en un motor de Reconocimiento Automático del Habla.
  - Estima la noticia que está siendo emitida actualmente mediante una decodificación acústica del audio que recibe.
  - La noticia en el aire será aquella cuyo camino en el decodificador haya progresado más allá de un límite de palabras establecido.

# SPEECH-TEXT ALIGNMENT

- Detector de Noticias:
  - De forma secundaria, puede usarse cierta información de status para algunas noticias.
  - El módulo TRM envía el estado de algunas de las noticias.
    - Sólo aquellas que contengan videos llevarán asociado el campo status.
    - Este campo cambiará de CUED a PLAY cuando la noticia vaya a entrar.
    - Si el cambio se produce y la noticia no está siendo subtitulada, el sistema detectará que ha perdido el sincronismo y saltará automáticamente a esa noticia.

# EVALUACIÓN DE PRESTACIONES

- Estudio estadístico sobre los datos obtenidos durante un mes de emisiones.
- Medidas de Prestaciones Consideradas:
  - Tasa de NOTICIAS SUBTITULADAS CON ÉXITO.
    - El 100% de sus subtítulos correspondientes han sido correctamente enviados.
  - Tasa de NOTICIAS PARCIALMENTE SUBTITULADAS.
    - Algunos de sus subtítulos no han sido enviados.
  - Tasa de NOTICIAS PERDIDAS.
    - Ninguno de sus subtítulos ha sido enviado.

# EVALUACIÓN DE PRESTACIONES

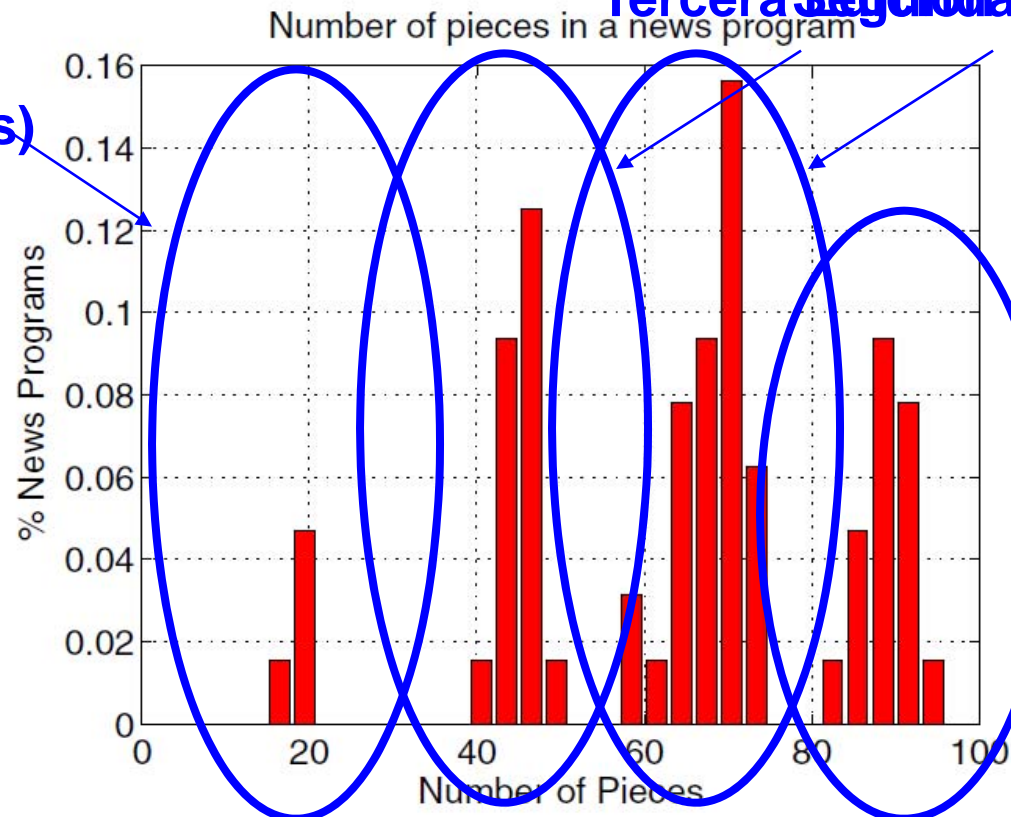
- Informativos en Aragón Televisión:
  - Tres Ediciones:
    - Primera Edición
      - 14:00, 60 minutos de duración.
      - Sobre 90 noticias.
      - Cambios muy frecuentes en las noticias o en su orden.
    - Segunda Edición
      - 20:30. 45 minutos de duración.
      - Sobre 75 noticias.
      - El número de cambios es alto pero no tanto como en la primera edición.
      - Edición especial reducida los domingo (15 minutos de duración, 20 noticias)
    - Tercera Edición.
      - 00.00 , 30 minutos de duración.
      - Sobre 50 noticias.
      - Bajo número de cambios.

# EVALUACIÓN DE PRESTACIONES

- Informativos en Aragón Televisión:
  - Histograma del número de noticias por informativo:

Segunda Edición (Domingos)

Tercera Edición

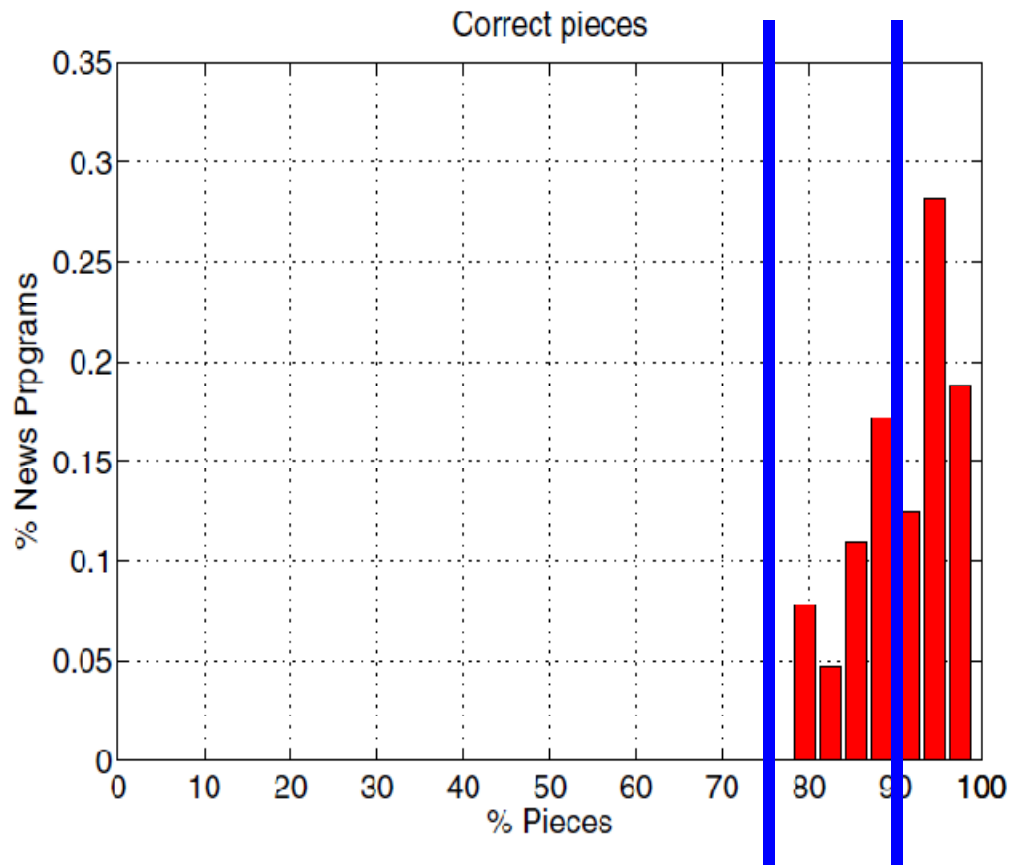


Primera Edición



# EVALUACIÓN DE PRESTACIONES

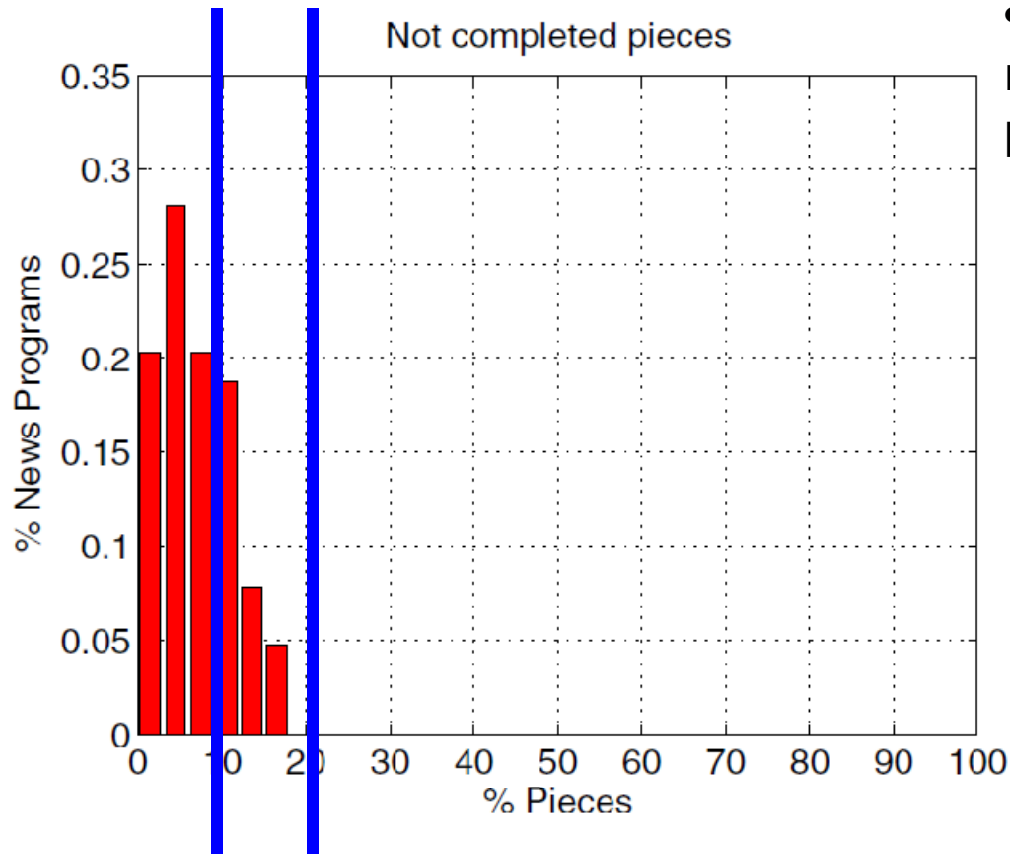
## • NOTICIAS CORRECTAMENTE SUBTITULADAS



- Ningún Informativo contiene menos del 75 % de sus noticias correctamente subtuladas.
- Más del 55% de los informativos finaliza con más del 90% de sus noticias correctamente subtuladas

# EVALUACIÓN DE PRESTACIONES

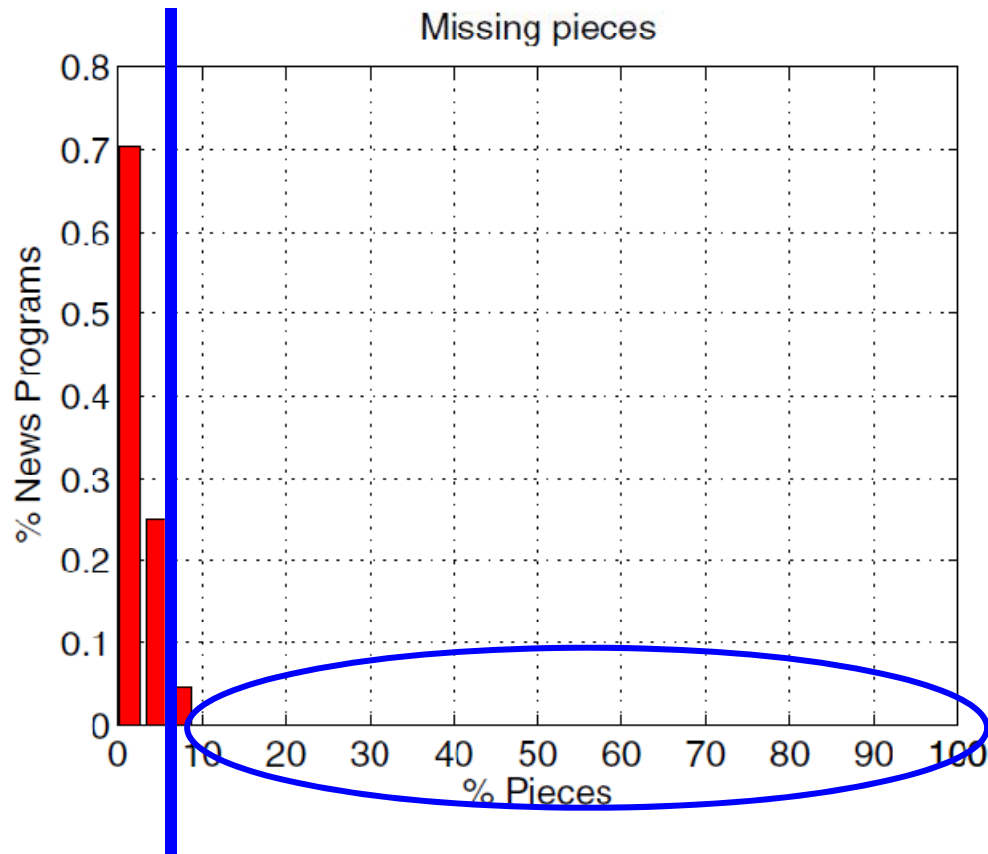
## • NOTICIAS PARCIALMENTE SUBTITULADAS



- Ningún informativo termina con más del 20% de las noticias parcialmente subtituladas.
- Alrededor del 70% de los informativos finaliza con menos del 10% de noticias parcialmente subtituladas.

# EVALUACIÓN DE PRESTACIONES

## • NOTICIAS PERDIDAS



- Más del 90% de los informativos finaliza con menos del 5% de sus noticias sin subtítular.
- Ningún informativo contiene más del 7% de noticias perdidas

# EVALUACIÓN DE PRESTACIONES

- Resumen de prestaciones

EDICIÓN	COMPLETAS	PARCIALES	PERDID.
1ª ED	88.64%	8.55%	2.81%
2ª ED	91.50%	6.90%	1.60%
3ª ED	94.42%	4.65%	0.93%
MEDIA	91.52%	6.70%	1.78%

- Más del 90% de las noticias en un informativo se completan con éxito.

- Menos del 6.70% se subtitulan parcialmente

- Menos del 2% se pierden

- La precisión es menor en la 1ª Ed. puesto que el número de cambios es muy alto. Sin embargo se sitúa alrededor del 90%.

- La precisión es mayor en la 3ª Ed. puesto que los cambios son menos frecuentes. Alrededor del 95% son correctas y un 1% perdidas

# EVALUACIÓN DE PRESTACIONES

- Latencia:
  - Los subtítulos se preprocesan cuando se reciben los guiones.
  - Se envían inmediatamente después de que la primera palabra ha sido pronunciada.
    - Solamente se añade un pequeño retardo intencionadamente al comienzo de cada noticia para evitar falsos comienzos.

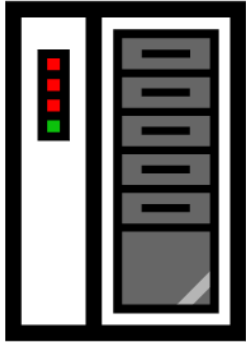
– LATENCIA NO APRECIABLE

# CONCLUSIONES

- Sistema de subtulado en tiempo real, totalmente automático para informativos televisivos en castellano.
- Basado en Reconocimiento Automático del Habla.
- Realiza un alineamiento temporal entre la voz y el texto.
- Evaluación de Prestaciones:
  - Más del 90% de las noticias de un informativo se subtitulan correctamente.
  - Menos de un 2% se pierden.
  - Latencia no apreciable.
- Usado por Aragón Televisión desde Junio de 2008.

# DEMO

- Demostración



TELETEXT  
MANAGER



SPEECH-TEXT  
ALIGNMENT  
MODULE



TEXT  
RETRIEVAL  
MODULE



TV SREEN