

# DETECCIÓN DE ACTIVIDAD DEL HABLA EN VÍDEOS

Autor: José Miguel Acosta Triana

Tutor: Carlos David Martínez Hinarejos

Director experimental: David Gimeno Gómez



GENERALITAT  
VALENCIANA

# Contenidos

- Introducción
- Trabajos relacionados
- Conjunto de datos
- Procesado de datos
- Experimentación
- Funcionamiento e interfaz
- Conclusiones y trabajos futuros

# Introducción y motivación

- Identificar hablante en vídeo en base al audio

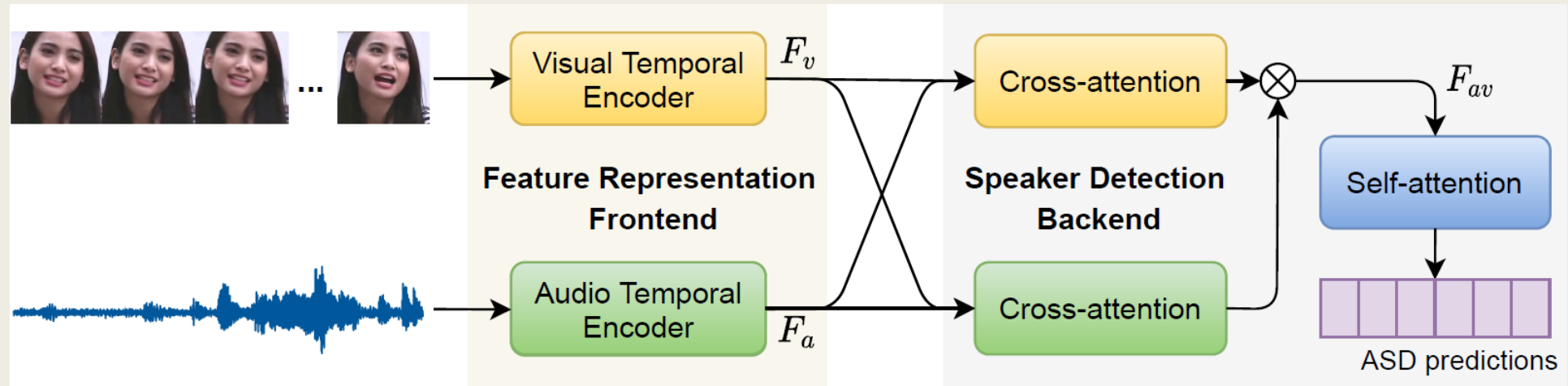
- Múltiples aplicaciones (entre otras):

- Detección Deepfakes
- Enfoque automático
- Interacción humano-robot
- Seguridad y vigilancia



# Nuestro objetivo

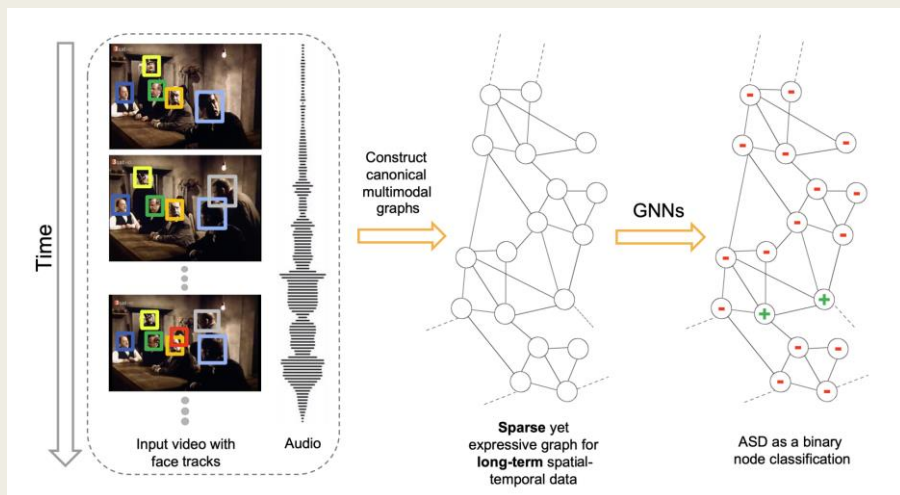
- Herramienta para la creación de base de datos
- *Fine-tuning* → TalkNet-ASD



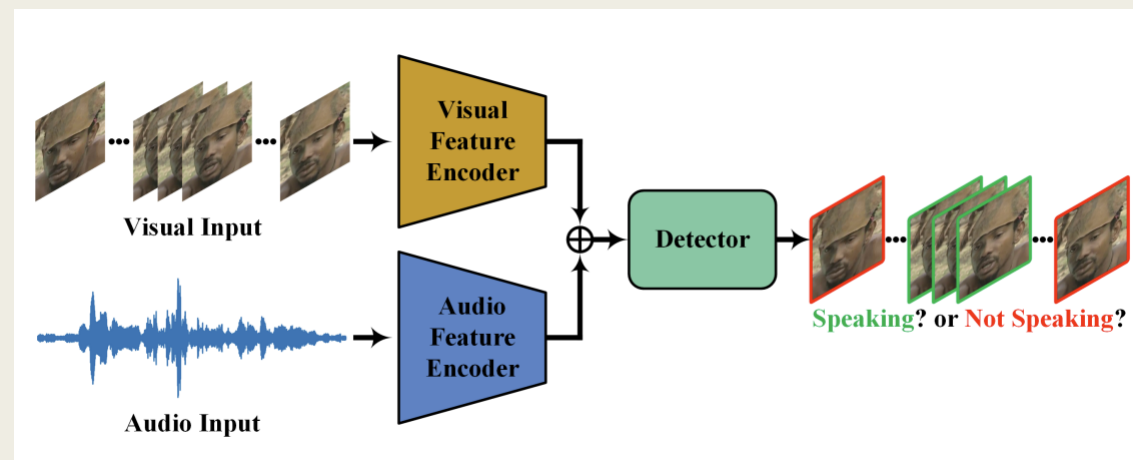
# Trabajos relacionados

- Conformers
- Mecanismos de auto-atención
- *Graph Neural Networks* (GNN)

SPELL (2021)



Light-ASD (2023)



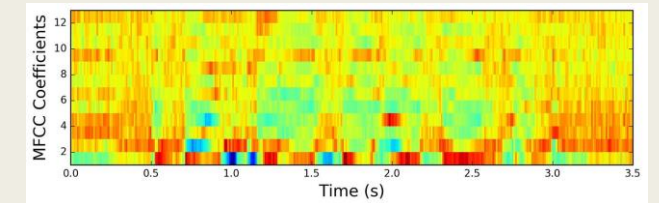
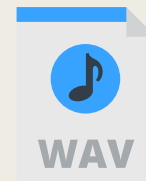
# Conjunto de datos: LIP-RTVE

- Partición independiente del locutor: Train (7142) – Dev (1638) – Test (1572)
- 323 hablantes
- 10353 vídeos
- ~13 horas de vídeo
- Resolución 480x270



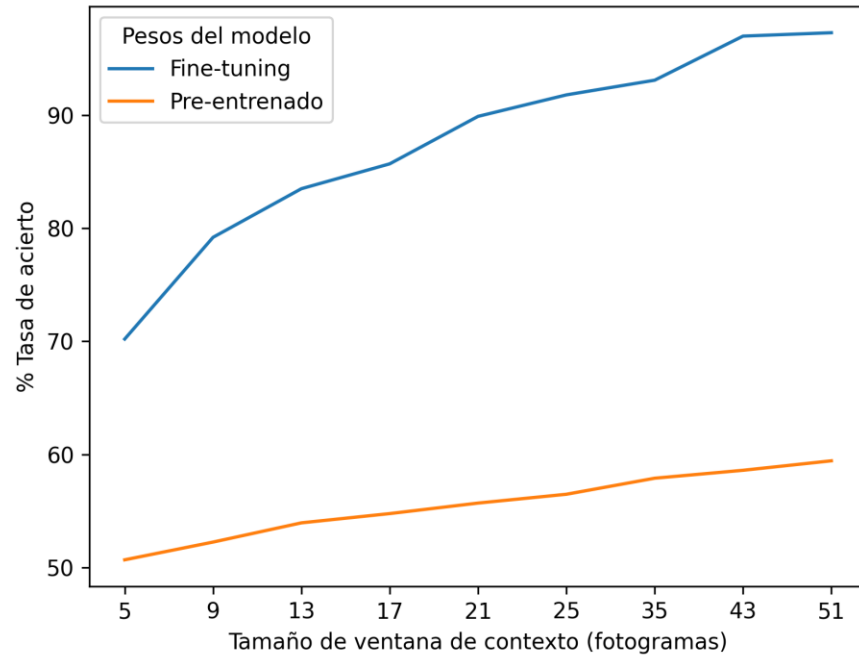
# Procesado de los datos

- CSV con (video,audio,etiqueta,centro)
- Tipos de muestras:
  - Positivas
  - Negativas:
    - Desfase de audio
    - *Mismatch* parcial (mismo hablante)
    - *Mismatch* completo (distinto hablante)



# Experimentación

Tasa de acierto en validación de modelo pre-entrenado vs. fine-tuning



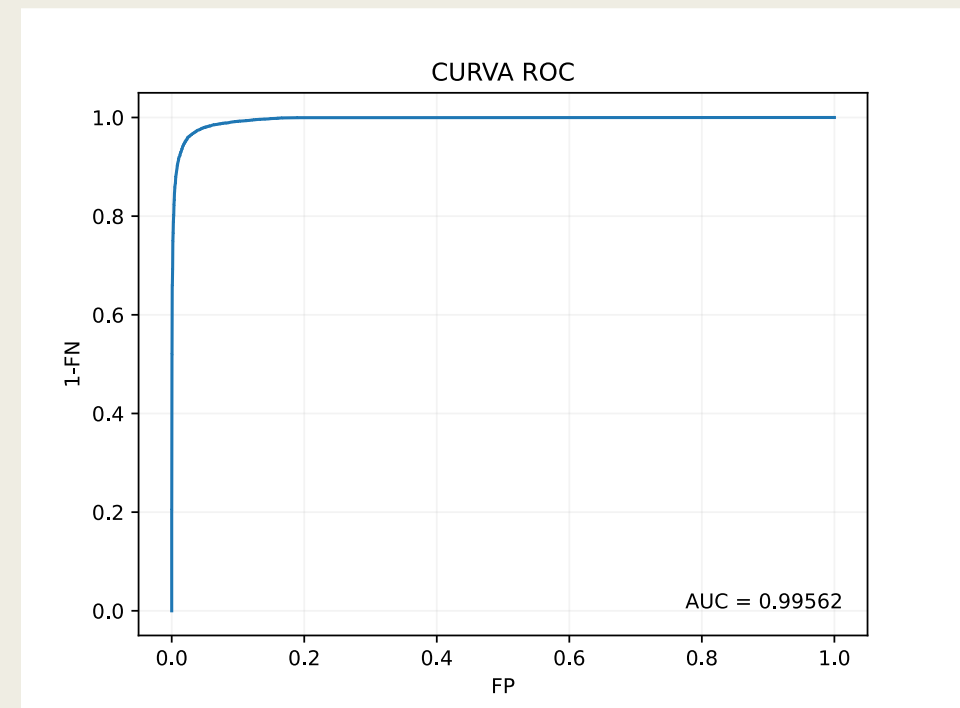
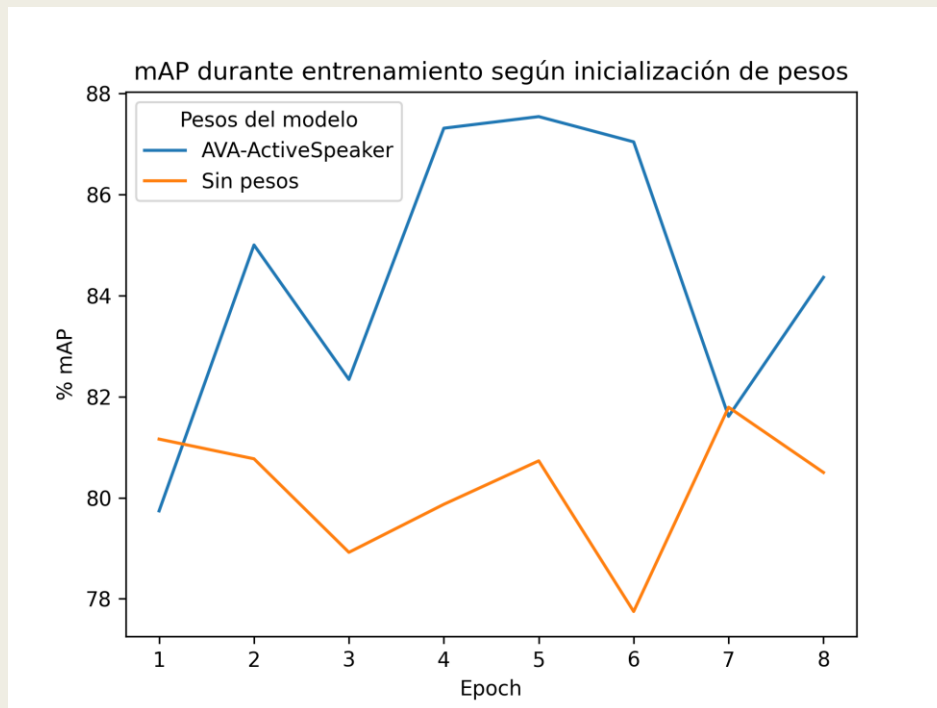
Conjunto de validación

Tamaño ventana (frames)	Tasa de acierto (accuracy)	mAP
5	70.2±0.5	60.0
13	83.5±0.4	77.7
17	85.7±0.4	78.4
25	91.8±0.3	87.5
35	93.1±0.3	90.7
51	97.3±0.2	95.8

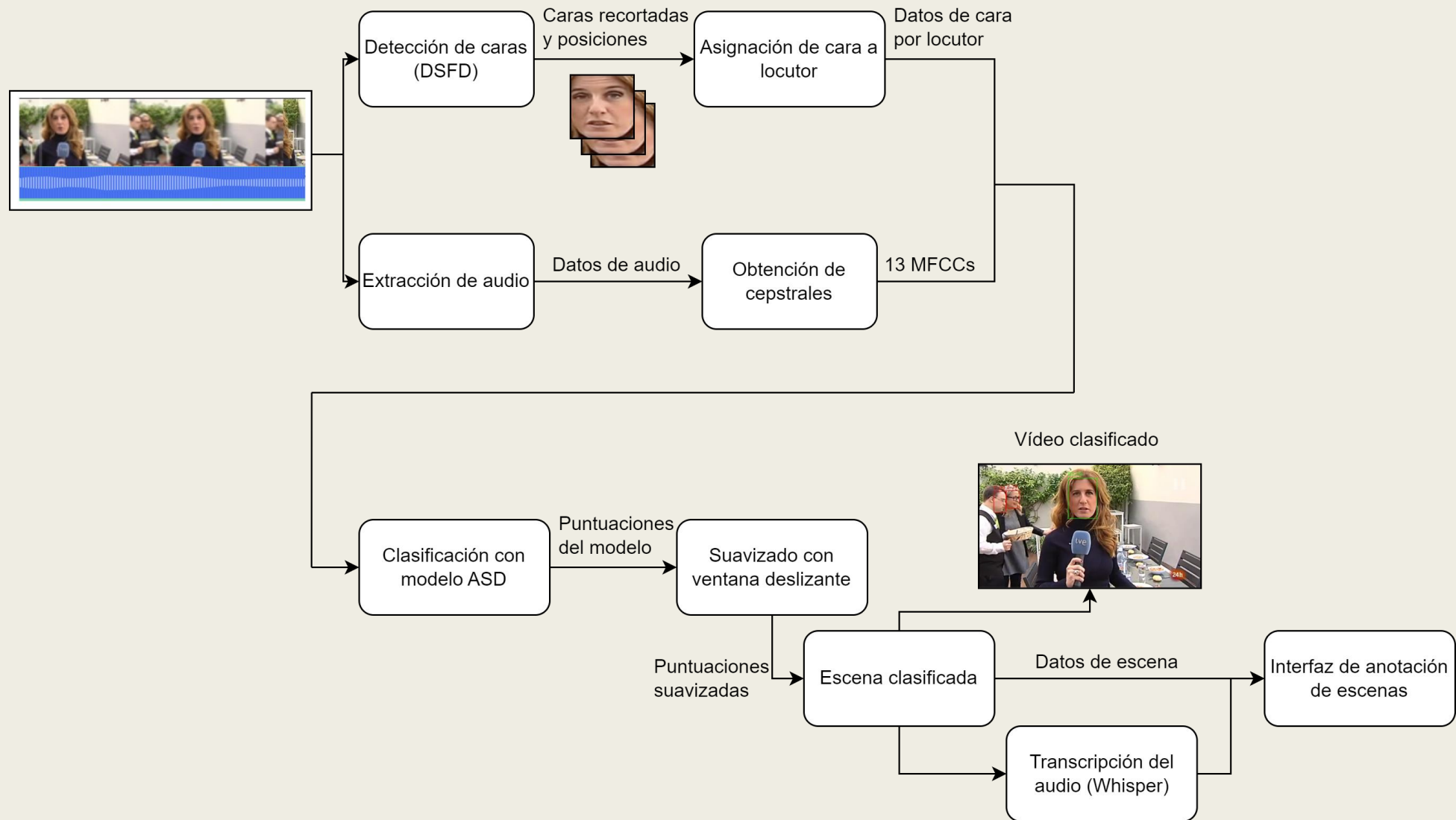


# Experimentación

## ROC 51 fotografías



# Funcionamiento del *pipeline*



# Interfaz de anotación



# Conclusiones

- Mejores resultados con ventanas largas
- Buena precisión en datos con baja resolución
- Capaz de generalizar (*speaker independent*)
- Contribución a la facilidad de anotar datos

# Trabajos futuros

- Instalación automática de dependencias
- Histogramas de color para cambios de escena
- Probar otros modelos
- Reetiquetado de escenas
- Unificar herramienta