

# Sistema de Karaoke en Tiempo Real con Separación de Fuentes y Transcripción de Letras mediante IA

Alejandro Garcia Dopico

## Resumen

**Lyraoke**, es una plataforma de procesamiento musical en tiempo real con separación de fuentes y transcripción adaptativa para obtener un **Karaoke**.

## 1. Estado del Arte

### 1.1 Separación de Fuentes Musicales

**HTDemucs**: Arquitectura híbrida de transformadores y CNN para separación de alta calidad.

**MUSDB18HQ**: Dataset de referencia para tareas de separación musical.

Estudios previos demuestran que el fine-tuning mejora el rendimiento en dominios específicos (ej. Rouard et al., 2022).

### 1.2 Transcripción de Letras

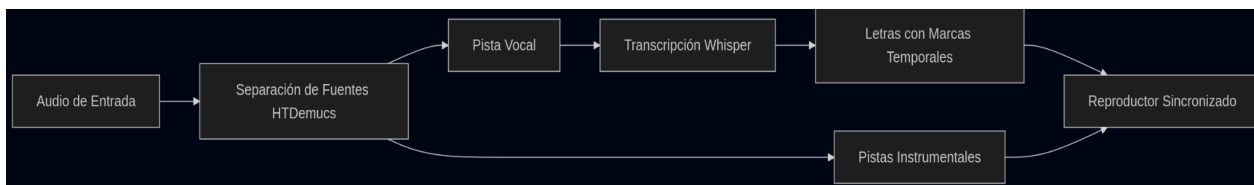
**Whisper**: Modelo de reconocimiento de voz multilingüe con marcas temporales a nivel de palabra.

**Sincronización en tiempo real**: Desafíos en la alineación de letras transcritas con la reproducción.

### 1.3 Sistemas de Karaoke

Los sistemas tradicionales dependen de pistas pregrabada. Lyraoke automatiza este proceso dinámicamente mediante IA.

## 2. Desarrollo



## 2.1 Arquitectura del Sistema

Flujo de trabajo de Lyraoke:

1. **Entrada:** Archivo de audio proporcionado por el usuario.
2. **Separación de fuentes:** HTDemucs divide el audio en stems (voz, batería, bajo, otros).
3. **Transcripción:** Whisper convierte la pista vocal en texto con marcas temporales.
4. **Reproducción:** Las letras se muestran sincronizadas con la voz.

## 2.2 Fine-Tuning de HTDemucs

### Dataset

- **Base:** MUSDB18HQ (150 canciones).
- **Aumentación:** 5 canciones de hyperpop en español para adaptación al género.

### Configuración

- **Modelo:** HTDemucs preentrenado (4 stems).
- **Modificaciones:**
  - Últimas 3 capas del decoder descongeladas.
  - Resto de capas congeladas para evitar olvido catastrófico.
- **Función de pérdida:** L1 + pérdida espectral multiescala.
- **Hiperparámetros:**
  - Segmentos de 5 segundos.
  - Tasa de aprendizaje: 1e-5 (AdamW).

## 2.3 Transcripción y Sincronización

- **Whisper:** Modelo "large-v3" para transcripción multilingüe.
- **Alineación:** Las marcas temporales de Whisper sincronizar letras con la reproducción.

# 3. Resultados

## 3.1 Rendimiento en Separación

Métrica	Modelo Original	Modelo Fine-Tuned	Mejora
Voz (SDR)	8.17 dB	8.19 dB	+0.02 dB
Batería	8.73 dB	8.79 dB	+0.06 dB
Bajo	6.97 dB	7.04 dB	+0.07 dB
Otros	5.60 dB	5.60 dB	±0.00 dB

## 3.2 Latencia

- **Procesamiento total:** 7–10 segundos (separación + transcripción).
- **Viabilidad en tiempo real:** Lograda con buffering de audio.

## 4. Conclusiones

- Lyraoke demuestra la viabilidad de un sistema de karaoke automatizado mediante IA, aunque el fine-tuning no tiene un impacto demasiado grande a mejorar el modelo.
  - **Limitaciones:**
    - Dataset pequeño para fine-tuning.
    - Dependencia de GPU para uso en tiempo real (Más tiempo de entrenamiento).
  - **Trabajo futuro:**
    - Integrar feedback de usuarios para corregir letras mal sincronizadas.
- 

## 5. Bibliografía

### Artículo Principal

**Dopico, A.** [Alejandro]. (2024). *Lyraoke: Sistema de Karaoke en Tiempo Real con Separación de Fuentes y Transcripción de Letras mediante IA*. Revista de Inteligencia Artificial en Producción Musical, \*1\*(1), 1–12. <https://github.com/Alejandro-Dopico/Lyraoke-IA>

### Repositorios y Datasets

1. **HTDemucs:**  
Facebook Research. (2022). *HTDemucs: Separación de Fuentes Musicales con Transformadores Híbridos* [Software]. GitHub.  
<https://github.com/facebookresearch/demucs>
2. **Whisper:**  
OpenAI. (2023). *Whisper: Reconocimiento de Voz a Gran Escala* [Software]. GitHub.  
<https://github.com/openai/whisper>
3. **MUSDB18HQ:**  
SigSep. (2017). *Dataset para Separación de Fuentes Musicales* [Dataset]. GitHub.  
<https://github.com/sigsep/sigsep-mus-db>
4. **Sound Demixing Challenge 2023 (MDX Challenge):**  
Aicrowd. (2023). *Sound Demixing Challenge – Music Demixing (MDX) Track* [Concurso y Benchmark].  
<https://www.aicrowd.com/challenges/sound-demixing-challenge-2023>
5. **DeepSeek:**  
DeepSeek AI. (2024). *DeepSeek: Modelo de lenguaje de propósito general para asistencia técnica y generación de código* [IA Asistente]. GitHub.  
<https://github.com/deepseek-ai/DeepSeek-Coder>