

Tabla de Contenido

| | |
|---|-----------|
| Introducción | 1 |
| 1. Marco Teórico | 5 |
| 1.1. Audio y Musica | 5 |
| 1.2. Procesamiento Digital de Señales de Audio | 7 |
| 1.2.1. Audio Enhancement | 7 |
| 1.2.1.1. Singing Voice Detection (SVD) | 7 |
| 1.2.1.2. Singing Voice Separation (SVS) | 8 |
| 1.2.2. Descriptores de Audio | 10 |
| 1.3. Aprendizaje de Máquinas | 14 |
| 1.3.1. Aprendizaje Supervisado y Clasificación | 14 |
| 1.3.1.1. Clasificación de Audio | 15 |
| 1.3.1.2. Gaussian Mixture Models | 16 |
| 1.3.2. Deep Learning | 17 |
| 1.3.2.1. Redes Neuronales Artificiales | 20 |
| 1.3.2.2. Redes Neuronales Convolucionales | 22 |
| 1.3.2.3. Wavenet | 24 |
| 1.3.2.4. Redes Neuronales Recurrentes | 26 |
| 1.3.2.4.1. Long Short-Term Memory | 27 |
| 1.3.2.5. Transformer y Mecanismos de Auto-Atención | 29 |
| 2. Metodología y Datos | 32 |
| 2.1. Hipótesis y Metodología | 32 |
| 2.2. Datos | 33 |
| 2.2.1. Análisis Exploratorio | 34 |
| 2.2.2. Obtención de datos de audio | 34 |
| 2.2.3. Preprocesamiento de Datos | 36 |
| 2.3. Evaluación de Desempeño y Métricas | 38 |
| 3. Solución | 40 |
| 3.1. Extracción de Características para Métodos Alternativos | 40 |
| 3.2. Métodos Alternativos | 41 |
| 3.3. End-to-end Waveform Classifier (EWC) | 41 |
| 3.3.1. Waveform Encoder | 42 |
| 3.3.1.1. WE-CNN | 43 |

| | | |
|-----------|--|-----------|
| 3.3.1.2. | WE-CNN-LSTM | 44 |
| 3.3.1.3. | WE-WN | 44 |
| 3.3.1.4. | WE-WN-LSTM | 44 |
| 3.3.1.5. | WE-WN-TF | 44 |
| 3.3.2. | Pooler | 45 |
| 3.3.3. | FCNN Classifier | 45 |
| 4. | Resultados | 46 |
| 4.1. | Generalidades | 46 |
| 4.2. | Análisis de Resultados | 47 |
| 4.3. | Utilización de Hardware | 52 |
| 5. | Conclusiones | 54 |
| 5.1. | Generalidades | 54 |
| 5.2. | Competitividad en la Tarea SID | 54 |
| 5.3. | SID en Auto-Etiquetado | 55 |
| 5.4. | <i>Encoding</i> de Formas de Onda | 56 |
| 5.5. | <i>Deep Learning</i> sobre Formas de Onda | 56 |
| 5.6. | Escalabilidad y Uso de Recursos | 57 |
| 5.7. | Trabajo Futuro | 58 |
| | Bibliografía | 60 |
| | Apéndices | 66 |
| | A. Matrices de Confusión | 66 |
| | B. Regímenes de Entrenamiento No-Convergentes | 71 |