

Tabla de Contenido

1. Introducción	1
1.1. Objetivo General	2
1.2. Objetivos Específicos	2
1.3. Contribución de este trabajo	3
1.4. Estructura del Trabajo de Título	3
2. Marco Teórico y estado del arte	4
2.1. Representaciones de audio	4
2.1.1. Formas de onda	4
2.1.2. Respuestas al impulso en audio	5
2.1.3. Representación de audio en frecuencia	7
2.1.4. Espectrograma	8
2.1.5. Escala logarítmica y escala de Mel	9
2.2. Reverberación en audio	11
2.2.1. Reverberación en la forma de onda	11
2.2.2. Reverberación en el espectrograma	12
2.3. Métodos no supervisados de dereverberación	13
2.3.1. Weighted Prediction Error (WPE)	13
2.3.2. Frequency Domain Normalized Delayed Linear Prediction (FD-NDLP)	15
2.4. Métricas de calidad, inteligibilidad y reverberación	16
2.4.1. Perceptual Evaluation of Speech Quality (PESQ)	17
2.4.2. Log-likelihood Ratio (LLR)	18
2.4.3. Cepstral Distorsion (CD)	18
2.4.4. Frequency-Weighted Segmental Signal to Noise Ratio (fwSNRseg)	19
2.4.5. Speech to Reverberation Modulation Energy Ratio (SRMR)	20
2.5. Machine y Deep Learning	22
2.5.1. Perceptrón Multicapa (MLP)	22
2.5.2. Redes Neuronales Convolucionales (CNN)	23
2.5.2.1. Convolución sobre una imagen	23
2.5.2.2. Conceptos básicos	24
2.5.2.3. Capa Convolutiva	24
2.5.3. Autoencoder	25
2.5.4. Redes Neuronales Recurrentes (RNN)	26
2.5.5. Redes Generativas Adversarias (GAN)	28
2.6. Aplicaciones de Machine y Deep Learning en el problema de dereverberación	29
2.6.1. MLP y LSTM para dereverberación	29
2.6.2. El Autoencoder para dereverberación	31

2.6.3. GAN para dereverberación	32
2.7. Estado del arte de métodos dereverberativos	33
3. Metodología	34
3.1. Dataset de audios de habla inglesa	34
3.2. Datos de respuestas al impulso (RIRs)	35
3.3. Preparación de datos	35
3.4. Respuestas al impulso simuladas	37
3.5. Datos de audio reverberados reales	37
3.6. Modelos de dereverberación	38
3.6.1. Arquitecturas MLP y LSTM	38
3.6.2. Arquitectura U-net	39
3.6.3. Late Reverberation Supression U-net	40
3.6.4. Arquitectura GAN	40
3.7. Métricas	41
3.8. FD-NDLP	41
3.9. Resumen de la metodología	42
4. Resultados y Análisis	43
4.1. Resultados de entrenamiento de modelos neuronales	43
4.2. Evaluación cuantitativa	46
4.2.1. Datos simulados variando el ruido adicionado	46
4.2.2. Datos simulados variando la distancia <i>speaker</i> -micrófono	52
4.2.3. Comportamiento en RIRs simuladas	53
4.2.4. Resultados en datos reales	54
4.2.5. Análisis General	56
4.3. Evaluación cualitativa	57
4.3.1. Visualización en datos simulados	57
4.3.2. Visualización en datos reales	60
5. Conclusiones	64
Bibliografía	66
Anexos	69
A. Modelación con Sistemas LTI	70
A.1. Linealidad	70
A.2. Invarianza en el tiempo	70
A.3. Convolución en sistemas LTI	71
B. TD-NDLP	72