

Tabla de Contenido

1. Introducción	1
1.1. Motivación	1
1.2. Hipótesis de trabajo	2
1.3. Objetivo general	2
1.4. Objetivos específicos	2
2. Marco teórico y antecedentes bibliográficos	3
2.1. Robots sociales y perfilamiento de usuario en HRI	3
2.2. Reconocimiento automático de emociones con voz	4
2.2.1. Antecedentes del reconocimiento automático de emociones utilizando la voz	4
2.2.2. Reconocimiento dimensional de emociones	6
2.2.2.1. <i>Concordance Correlation Coefficient</i> (CCC)	6
2.2.3. Inteligencia artificial: conceptos y modelos asociados a SER	7
2.2.3.1. Red neuronal	7
2.2.3.2. Perceptrón multicapa	7
2.2.3.3. Red neuronal convolucional	8
2.2.3.4. Autoencoder	9
2.2.3.5. Ladder network	10
2.2.3.6. Redes atencionales	10
2.2.3.7. Redes neuronales tipo <i>transformer</i>	12
2.2.3.8. Wav2vec 2.0	13
2.3. Herramientas y aplicaciones que utilizan la voz en HRI	14
2.3.1. Antecedentes de reconocimiento de emociones en HRI	14
2.3.2. <i>Beamforming</i>	15
2.3.2.1. <i>Direction of Arrival</i> (DOA)	15
2.3.2.2. <i>Delay and Sum</i>	15
2.3.2.3. <i>Minimum Variance Distortionless Response</i> (MVDR)	17
2.3.3. Modelamiento del canal acústico	17
2.3.3.1. Estimación de RIRs: método exponential swept-sine	18
3. Sistema de reconocimiento de emociones utilizando voz en HRI	20
3.1. Plataforma robótica y grabación de base de datos	20
3.1.1. Escenario HRI estático	22
3.1.2. Escenario dinámico	23
3.2. Sistema de SER en HRI propuesto	23
3.2.1. Modelamiento del canal de voz <i>indoor</i>	24

3.2.2.	Creación de base de datos simulada	25
3.2.3.	Aplicación de técnicas de <i>beamforming</i>	25
3.2.3.1.	Implementación de D&S	25
3.2.3.2.	Implementación de MVDR	26
3.2.4.	Módulo de SER	27
3.2.4.1.	Arquitectura del modelo Ladder Network utilizado	27
3.2.4.2.	Arquitectura del modelo Wav2vec 2.0	28
3.3.	Descripción experimental	30
3.3.1.	Descripción de bases de datos de entrenamiento	31
3.3.2.	Descripción de bases de datos de prueba	31
3.3.3.	Procedimiento de entrenamiento de Módulo de SER	31
3.3.3.1.	Entrenamiento semi-supervisado de Ladder network	31
3.3.3.2.	Procedimiento de <i>Fine-tuning</i> de Wav2vec	32
3.3.4.	Métricas de rendimiento	32
3.3.4.1.	<i>Signal-to-Noise Ratio</i> (SNR)	32
3.3.4.2.	<i>Concordance Correlation Coefficient</i> (CCC)	32
4.	Resultados y discusión	33
4.1.	Resultados de <i>beamforming</i> para HRI	33
4.2.	Resultados de simulación	36
4.3.	Resultados de reconocimiento de emociones con voz	37
4.3.1.	Entrenamiento con base de datos original y evaluación en base de datos HRI	37
4.3.2.	Modelos entrenados con base de datos simulada y evaluados en base de datos simulada	40
4.3.3.	Modelos entrenados con base de datos simulada y evaluados en HRI real	41
4.3.4.	Discusión	43
5.	Conclusiones	46
5.1.	Trabajo a futuro	47
	Bibliografía	48
	Anexo	52
	Anexo A: Lista de acrónimos	52