

Tabla de Contenido

1. Introducción	1
1.1. Hipótesis	1
1.2. Objetivos	2
1.2.1. Objetivo General	2
1.2.2. Objetivos específicos	2
1.3. Estructura de tesis	2
2. <i>Visual Servoing</i> con arreglo de micrófonos	4
2.1. Introducción	4
2.2. Trabajos relacionados y marco teórico	7
2.2.1. Reconocimiento automático de la voz e interacción humano-robot	7
2.2.2. <i>Weighted delay-and-sum</i> y arreglos lineales de micrófonos	8
2.2.3. MVDR	11
2.2.4. Rastreo de objetos y <i>visual servoing</i>	12
2.3. Disposición del escenario de grabación	12
2.3.1. Movimientos de robot PR2	14
2.3.2. Fuentes de voz y de ruido	14
2.3.3. Reconocimiento y rastreo de objetos	15
2.3.4. <i>Beamforming</i> y uso de información visual para mejorar su desempeño	16
2.4. Estimación de la directividad de un arreglo de micrófonos lineal	21
2.5. Reconocimiento automático de la voz - ASR	23
2.6. Resultados y discusión	24
2.7. Conclusiones del capítulo	33
3. <i>Speech enhancement</i> con aprendizaje profundo	34
3.1. Introducción	34
3.1.1. El problema de separación de fuentes clásico	35
3.1.2. El problema de las fuentes múltiples y la reverberación en espacios interiores	36
3.1.3. Aprendizaje profundo, separación de fuentes y <i>speech enhancement</i>	38
3.1.4. <i>Compact bilinear pooling</i>	39
3.1.5. Contribución de este capítulo	39
3.2. <i>speech enhancement</i> y escenarios reverberantes y variables en el tiempo	40
3.2.1. Solución propuesta	43
3.3. Experimentos	45
3.3.1. Generación de los conjuntos de datos	46
3.3.2. Generación de <i>beamforming</i> sin reverberación	46

3.3.3.	Generación de <i>beamforming</i> con reverberación	46
3.3.4.	Generación de <i>beamforming</i> variable en el tiempo	47
3.3.5.	Posible escenario experimental	47
3.4.	Entrenamiento del sistema TCN/CBP	48
3.5.	Sistema ASR	48
3.6.	Resultados y discusión	49
3.7.	Conclusiones del capítulo	53
4.	Conclusiones	56
	Bibliografía	58