



**Universidad Politécnica
de Madrid.**



**Escuela Técnica
Superior de Ingenieros de Telecomunicación**

Tesis doctoral

**Contribución al reconocimiento del
habla con redes neuronales y
algoritmos de programación
dinámica**

Autor:

Xavier Menéndez-Pidal Sendrail

Tutor:

José Manuel Pardo Muñoz

Febrero 1995

RESUMEN

Desde que en 1985 se introdujo el algoritmo de retropropagación, hemos asistido al renacimiento de los métodos conexionistas como alternativa o apoyo a las técnicas clásicas estadísticas de reconocimiento de patrones. En reconocimiento de habla se han realizado muchos esfuerzos tanto en el diseño de nuevas arquitecturas de redes neuronales artificiales (NN) más aptas para esta tarea, como en los distintos métodos posibles de integración entre las redes neuronales y los algoritmos de programación dinámica (DP).

La contribución de esta Tesis ha ido dirigida por una parte hacia la mejora de la calidad de las actuales redes neuronales en reconocimiento del habla, y por otra parte también hacia la mejora del diseño de un reconocedor de habla basado en técnicas híbridas DP-NN. Fruto de todos estos trabajos son los tres sistemas completos de reconocimiento de habla que aquí presentamos dónde el objetivo perseguido era realizar sistemas con un coste de desarrollo mínimo sin perder prestaciones en reconocimiento:

1.- El 1^{er} sistema es un reconocedor de dígitos aislados independientes del locutor basado exclusivamente en redes neuronales, realizado a finales de 1990 y 1991.

2.- El 2^o sistema es un reconocedor de habla aislada de grán vocabulario dependiente del locutor donde se integra las salidas de la red neuronal en un algoritmo de programación dinámica de acceso léxico (NN-AcLx), realizado en 1992.

3.- El 3^{er} sistema es un sistema híbrido neuronal Markoviano (NN-HMM) realizado sobre una tarea de habla continua dependiente del locutor (DARPA-RM1), realizado en 1993 y 1994.

ABSTRACT

Since 1985, when the retro-propagation algorithm was introduced, we have contributed to the rebirth of the connectionist methods as an alternative to or a support of the classic statistical techniques of speech recognition. Many efforts have been realized both in the design of new architectures of artificial neural networks (ANNs) and in the possible distinct integration methods between neural networks and dynamic programming (DP) algorithms.

The contribution of this dissertation is twofold. It is directed on the one hand by improving the quality of the actual neural networks in speech recognition, and on the other hand, toward improving the design of the speech recognition system based on the DP-NN hybrid techniques. Resulting from these works, we present three complete speech recognition systems where the pursued objective was to achieve systems with low cost development without losing performance in recognition.

1. The first system is an isolated digit, speaker independent voice recognizer exclusively based on neural networks. This work was fulfilled by the end of 1991.
2. The second system is an isolated, speaker-dependent speech recognizer with a large vocabulary where neural network outputs are integrated with a lexical access dynamic programming algorithm (NN-AcLx). This work was achieved in 1992.
3. The third system is a hybrid neural-Markov (NN-HMM) system realized as a continuous speech, speaker dependent (DARPA-RM1) task. Fulfillment of this system was achieved in 1994.

CONTENIDO

Reconocimientos	viii
Resumen	ix
Abstract	x
Lista de Figuras	xi
 Capítulo 1 Introducción	 1
1.1 Marco y objetivos de la tesis	1
1.2 Complejidad del reconocimiento del habla	2
1.3 Módulos típicos de un sistema de reconocimiento de habla	3
1.4 Estructura general de la tesis	4
 Capítulo 2 Redes MLP utilizadas en reconocimiento del habla	 6
2.1 Introducción	6
2.2 Arquitecturas de tipo MLP utilizadas en reconocimiento del habla	7
2.2.1 Unidades de base utilizadas en los clasificadores neuronales	7
2.2.2 Arquitecturas clásicas MLP utilizadas en reconocimiento del habla	10
2.3 Algoritmo de entrenamiento	15
2.3.1 Algoritmos de entrenamiento alternativos y mejoras propuestas	22
2.4 Interpretación probabilística de las redes neuronales de entrenamiento supervisado	23
2.4.1 La guerra de dos mundos	29
2.4.2 Construcción de un sistema de reconocimiento de habla MLP-DP	35
2.4.3 Tipos de funciones de error utilizadas	37
2.5 Estado del arte de los sistemas neuronales utilizados en reconocimiento de habla	42
2.5.1 Sistemas híbridos basados en clasificadores o cuantificadores neuronales y módulos clásicos de programación dinámica independientes	43

2.5.2	Otros tipos de sistemas neuronales utilizados en reconocimiento del habla	47
2.6	Experimentos preliminares	48
2.6.1	Sistema de reconocimiento de dígitos aislados independientes del locutor	48
2.6.2	Tarea de clasificación de vocales	52
2.6.3	Conclusiones	56
Capítulo 3 Sistema híbrido neuronal y acceso léxico en habla aislada dependiente del locutor		58
3.1	Introducción	58
3.2	Definición de los módulos del sistema	59
3.2.1	Módulo de procesamiento de la señal	60
3.2.2	Módulo de clasificación acústico-fonética	60
3.2.3	Módulo de conversión de las salidas de la red a una transcripción acústico-fonética	61
3.2.4	Módulo de reconocimiento y evaluación, Algoritmo de Acceso Léxico	64
3.2.4.1	Algoritmo básico de Acceso Léxico (AcLx)	65
3.2.4.2	Variante del algoritmo de AcLx	67
3.2.4.3	Variante del algoritmo de AcLx con modelado contextual independiente del contexto léxico-fonético	70
3.3	Comparación de los distintos módulos neuronales	73
3.3.1	Condiciones de entrenamiento	73
3.3.2	Descripción de las 3 arquitecturas empleadas	74
3.3.3	Evaluación de las cadenas fonéticas no compactadas	77
3.3.4	Evaluación de la calidad acústico-fonética de la TSNN completa	79
3.4	Comparación de los distintos Algoritmos de Acceso Léxico	81
3.5	Comparación con otros sistemas sobre la misma tarea	82
3.6	Optimización y mejoras propuestas	85
Capítulo 4 Sistema híbrido neuronal y de programación dinámica en habla continua americana		90

4.1	Introducción	90
4.2	Descripción global del sistema	93
4.2.1	Estrategia general de entrenamiento y diseño de la base de datos	94
4.3	Descripción de la arquitectura neuronal modular	97
4.3.1	Introducción al diseño modular neuronal	97
4.3.2	Estrategias de integración modular	102
4.3.3	Arquitectura modular jerárquica utilizada	104
4.3.4	Arquitectura TSNN utilizada en DARPA-RM1	104
4.3.5	Dimensionamiento de los distintos módulos neuronales	106
4.3.6	Distintas codificaciones de los vectores de entrenamiento utilizadas	108
4.4	Descripción de los distintos métodos de integración HMM-NN	113
4.4.1	Integración probabilística discriminante HMM-NN (PDI)	115
4.4.2	Integración HMM-NN con modelado continuo de las salidas de la red (FCI)	121
4.4.3	Integración HMM-NN con modelado probabilístico de las salidas de la red (FPI)	134
4.5	Experimentación preliminar sobre el locutor JWS04	145
4.5.1	Entrenamiento trama a trama de la red neuronal	146
4.5.2	Evaluación de distintas parametrizaciones	148
4.5.3	Evaluación de los módulos neuronales	150
4.5.4	Evaluación del número de estados óptimos	154
4.5.5	Evaluación de distintas inicializaciones de la matriz de transición de estados	157
4.5.6	Evaluación de distintas inicializaciones del factor K en la integración FCI	158
4.5.7	Evaluación de los distintos vectores de probabilidad que se pueden modelar en la integración FCI	160
4.5.8	Evaluación de distintos criterios de error introducidos en la integración FCI	162
4.5.9	Evaluación final de las dos parametrizaciones de 19 y 63 coeficientes	163
4.5.10	Conclusiones de la primera fase de puesta a punto del sistema	163

4.6	Experimentación final sobre tres locutores de la base de datos DARPA	167
4.6.1	Condiciones generales de la experimentación final	168
4.6.2	Estrategia de evaluación de la red neuronal como clasificador bayesiano	173
4.7	Experimentación final con el locutor JWS04	175
4.7.1	Evaluación del módulo neuronal	175
4.7.2	Evaluación de todos los tipos de integración sobre JWS04	177
4.7.2.1	Experimentación con FPI sobre JWS04	178
4.7.2.2	Resultados finales con FPI sobre JWS04	186
4.8	Resultados sobre los otros dos locutores con el mejor sistema	187
4.8.1	Resultados con CMR0 y BEF0	187
4.8.2	Resultados medios sobre los 3 locutores con la parametrización de 22 coeficientes	190
4.9	Mejoras finales del sistema	191
4.9.1	Mejoras en CMR02	191
4.9.2	Mejoras en BEF03	193
4.9.3	Conclusiones sobre la re-segmentación con FPI y el uso de la matriz A_{ij}	198
4.9.4	Resultados finales	201
4.10	Comparación del sistema con otros similares sobre la misma tarea	201
	Capítulo 5 Conclusiones	204
5.1	Conclusiones del presente trabajo	204
5.2	Líneas futuras de investigación propuestas	208
	Apéndices	214
A1	Lista de símbolos entrenados en habla aislada	214
A2	Lista de símbolos entrenados en habla continua	215
A3	Test de validación estadística	216
A4	Estimación del número de parámetros	217
	Bibliografía	219