

J. M. Pardo



**Universidad Politécnica
de Madrid.**



**Escuela Técnica Superior de Ingenieros de
Telecomunicación**

Dpto. de Ingeniería Electrónica

**Técnicas Eficientes para
Reconocimiento de Habla en
Español**

Autor:

**Manuel Antonio Leandro Reguillo
Ingeniero Superior de Telecomunicación**

Director:

**José Manuel Pardo Muñoz
Catedrático de Universidad**

Mayo 1995

Resumen

La evidencia de la existencia de un alto porcentaje de zonas quasi-estacionarias en las producciones de habla del español nos ha conducido, por una parte, a investigar el modelado estático de fonemas como base de un sistema de hipótesis-verificación para habla aislada y, por otra, a utilizar aquellos fonemas más estacionarios como puntos de anclaje que generen posibles segmentaciones en producciones de habla continua. Los sistemas resultado de nuestra investigación están contruidos sobre los pilares siguientes: adaptación a las características acústicas del español; uso de conocimiento experto para dirigir los algoritmos; y estrategias modulares y algoritmos eficientes que generen poca carga computacional.

El sistema de habla aislada combina el modelado estático con información de estabilidad espectral y conocimiento heurístico para generar una cadena fonética con la que se accede al léxico mediante programación dinámica. Los primeros candidatos son posteriormente verificados generando para cada uno un modelo que vuelve a utilizar modelos estáticos además de información heurística de duraciones de fonemas.

El sistema de habla continua sigue una estrategia ascendente en la cual el primer módulo establece posibles segmentaciones, apoyándose en puntos de anclaje previamente detectados y clasificados, y el segundo reconoce los segmentos utilizando modelos de Markov. El léxico se expresa como un grafo en el que se combinan los modelos de Markov y los eventos acústicos de anclaje.

Aún cuando los resultados los consideramos todavía muy preliminares, los dos sistemas resultantes presentan muy baja carga computacional en comparación con otros algoritmos utilizados actualmente, y su tasa de reconocimiento es competitiva. Se han comparado con otros sistemas, también para español, desarrollados en nuestro laboratorio. Además, el sistema de habla aislada se ha implementado en tiempo real.

Las conclusiones de esta tesis apoyan la investigación de algoritmos alternativos en aquellos idiomas que, como el español, se pueden denominar "cuasi-fonéticos". Creemos demostrar que el uso de técnicas adaptadas a un idioma concreto, y el uso de conocimiento experto en puntos estratégicos, es una solución potente a la vez que eficiente, en comparación con otros algoritmos "de moda" que no aplican estos principios.

Abstract

Evidence of a high rate of quasi-stationary segments in Spanish speech utterances has led us, in one hand, to the investigation of phoneme static modeling for an hypothesis-verification system for isolated word recognition and, in the other hand, to use the most stationary phonemes as anchor points for possible segmentations in continuous speech recognition. The systems coming out from our research are built on the following principles: adaptation to Spanish acoustic characteristics; use of expert knowledge to drive algorithms; and efficient modular strategies and algorithms for a low computational load.

The isolated word system combines static modeling, spectral stability information and heuristic knowledge to generate a phonetic string used to access the lexicon through dynamic programming. First candidates are then verified generating for each one a model that uses phoneme static models again, and heuristic information of phoneme duration.

The continuous speech system follows a bottom-up strategy in which the first module resolves a tree of possible segmentations, using anchor points previously detected and classified, and the second one recognizes the segments using hidden Markov models. Lexicon is expressed as a graph in which Markov models and acoustic events for anchoring are combined.

Even if we consider that our results are preliminar, both systems present a low computational load in comparison to other algorithms currently used, and their recognition rate is competitive. They have been compared to other systems, also for Spanish, developed in our laboratory. Besides, the isolated word systems has been real time implemented.

The conclusion of this thesis reinforces research in alternative algorithms for those languages that, as Spanish, can be considered "quasi-phonetic". We think we prove that the use of technics adapted to a particular language, and the use of expert knowledge in strategic points is a powerful and efficient solution, in comparison to other state of the art algorithms that do not apply these principles.

Índice de contenidos

Índice de contenidos	xiii
Resumen	xvii
Abstract	xviii
Lista de abreviaturas	xix
Glosario (Palabras clave)	xxi
<i>Parte I: Introducción</i>	1
Capítulo 1: Introducción	3
<i>Parte II: Encuadre y fundamentos científico-tecnológicos</i>	5
Capítulo 2: Descodificación acústico-léxica	7
2.1 INTRODUCCIÓN	7
2.2 DEFINICIÓN	11
2.3 ESTRATEGIAS ARQUITECTURALES	13
2.4 TECNOLOGÍA	16
2.4.1 Enfoque general	16
2.4.2 Modelos o rasgos: el dilema	18
2.4.3 Estacionariedad de la señal. Caracterización estática o dinámica	20
2.4.4 Modelos estocásticos	24
2.4.5 Redes neuronales	29
2.4.6 Conocimiento experto	31
2.5 CONSIDERACIONES PARA HABLA CONTINUA	33
Capítulo 3: Sistemas de reconocimiento	35
3.1 SISTEMAS	35
3.2 EVALUACIÓN	41
3.2.1 Problemática	41
3.2.2 Consideraciones	43
3.2.3 Sistemas de dictado. Metodología de evaluación	44
Capítulo 4: Los conceptos claves de la tesis	51
4.1 EFICIENCIA	51
4.1.1 Oposición eficacia-eficiencia	51
4.1.2 El por qué de la eficiencia	52
4.1.3 Diferentes aproximaciones para conseguir eficiencia	52
4.2 ADAPTACIÓN AL IDIOMA	55

4.3 PLANTEAMIENTO DE TESIS	56
4.3.1 Sistema de habla aislada	57
4.3.2 Sistema de habla continua	57
4.3.3 Evaluación	58
Parte III: Trabajo realizado	59
Capítulo 5: Características acústicas del español	61
5.1 ALÓFONOS DEL ESPAÑOL. CARACTERÍSTICAS ESPECTRALES	61
5.1.1 Vocales	61
5.1.2 Oclusivas	65
5.1.3 Fricativas sonoras	65
5.1.4 Fricativas sordas	
5.1.5 Nasaes	
5.1.6 Vibrantes	69
5.1.7 El sonido lateral 'l'	71
5.1.8 Africadas	72
5.1.9 Otros alófonos	73
5.1.10 Semivocales y semiconsonantes	74
5.1.11 Articulación al final de locución	75
5.2 DETECCIÓN DE EVENTOS ACÚSTICOS	76
5.2.1 Rasgos, parámetros y eventos acústicos	76
5.2.2 Principio y fin de elocución	77
5.2.3 Oclusivas sordas	78
5.2.4 Fricativas sordas	79
5.2.5 Africadas	79
5.2.6 Nasaes	79
5.2.7 Ejemplo gráfico	81
5.3 CORRESPONDENCIA ACÚSTICA DE LOS GRAFEMAS	83
Capítulo 6: Reconocedor de habla aislada	85
6.1 DIAGRAMA DE BLOQUES DEL SISTEMA	85
6.2 MODELADO Y CONOCIMIENTO EXPERTO	87
6.2.1 Modelado acústico	87
6.2.2 Modelado léxico	88
6.2.3 Heurística en el postproceso de la cadena fonética	94
6.3 PREPROCESO	95
6.4 SUBSISTEMA DE GENERACIÓN DE HIPÓTESIS	97
6.4.1 Generador de cadena fonética	98
6.4.2 Acceso léxico	104
6.5 SUBSISTEMA DE VERIFICACIÓN DE HIPÓTESIS	107
6.5.1 Separador de lóbulos	108
6.5.2 Alineación de lóbulos	108
6.5.3 Generador de modelo	108
6.5.4 Comparador	108
6.5.5 Ejemplo	112
6.6 CONSIDERACIONES PARA TIEMPO REAL	113
6.7 EVALUACIÓN	114
6.7.1 Evaluación del generador de cadena fonética	114

6.7.2 Evaluación del acceso léxico	118
6.7.3 Evaluación del subsistema de hipótesis	120
6.7.4 Evaluación del subsistema de verificación	122
6.7.5 Ponderación de scores	123
6.7.6 Robustez del modelado	125
6.7.6 Comparación con otros sistemas de reconocimiento	125
6.8 CONCLUSIONES Y MEJORAS	126
6.8.1 Integración de modelado estático de fonemas estables y transiciones	128
Capítulo 7: Reconocedor de habla continua	131
7.1 DIAGRAMA DE BLOQUES DEL SISTEMA	131
7.2 MODELADO	132
7.2.1 Definición de eventos acústicos	132
7.2.2 Modelado para detección de eventos	133
7.2.3 Modelado para clasificación de eventos	137
7.2.4 Modelado lingüístico	141
7.2.5 Modelado acústico	144
7.3 SEGMENTACIÓN	145
7.3.1 Detección de rasgos	145
7.3.2 Aislamiento de eventos	145
7.3.3 Clasificación de eventos	146
7.3.4 Ejemplo de segmentación	147
7.4 RECONOCIMIENTO	148
7.5 EXPERIMENTOS	153
7.5.1 Segmentación	153
7.5.2 Reconocimiento	155
7.5.3 Carga computacional	157
7.6 CONCLUSIONES Y MEJORAS	159
Parte IV: Conclusiones	163
Capítulo 8: Conclusiones	165
Parte V: Apéndices	168
Apéndice A: Bibliografía	169
Apéndice B: Algoritmo de entrenamiento de modelos estáticos	179
Apéndice C: Base de datos para evaluación en habla aislada	183
Apéndice D: Obtención de los rasgos binarios para detección de eventos	187
D.1 PROCESO DEL RASGO ENER	188
D.2 PROCESO DEL RASGO TCC	190
D.3. PROCESO DEL RASGO NAS	191
Apéndice E: Codificación de eventos para su clasificación	195
E.1 CODIFICACIÓN INICIAL	195

E.1.1 Codificación de la secuencia de momentos principales	196
E.1.2 Codificación de momentos asociados	196
E.1.3 Codificación del contexto	196
E.2 ADAPTACIÓN A LA ENTRADA NEURONAL	197
Apéndice F: Modelo de lenguaje para reconocimiento de dígitos conectados	201
F.1 EXPECTATIVAS DE ACTIVACIÓN PARA SEG1	201
F.2 EXPECTATIVAS DE ACTIVACIÓN PARA SEG2	202
F.3 GRAFO LINGÜÍSTICO	203
Apéndice G: Base de datos para evaluación en habla continua	207