



Dpto. Ingeniería Electrónica
E.T.S.I. Telecomunicación
Universidad Politécnica de Madrid



TESIS DOCTORAL

**MEJORA DE LA FRECUENCIA FUNDAMENTAL
EN LA CONVERSIÓN DE TEXTO A VOZ**

José Ángel Vallejo Pinto
Ingeniero de Telecomunicación

Director de la Tesis

Dr. Ingeniero José Manuel Pardo Muñoz

1998

Resumen

El principal objetivo de este trabajo ha sido, como su título indica, la mejora de la frecuencia fundamental en la conversión de texto a voz, buscando no sólo la obtención de un sistema que genere curvas de frecuencia fundamental de mayor calidad, sino también la definición de una metodología de trabajo que permita la obtención de modelos de entonación de nuevos locutores de forma sencilla y prácticamente automática.

La primera fase del trabajo fue la definición y creación de una base de datos apta para el estudio de los fenómenos prosódicos, con la idea de que fuera útil para cualquier trabajo relacionado con la prosodia.

Una de las partes más importantes de la investigación ha sido la experimentación con redes neuronales como método de generación de curvas de frecuencia fundamental. El trabajo en este campo se ha centrado en la determinación y codificación de los parámetros más relevantes que afectan a la producción de las curvas de frecuencia fundamental, así como en la optimización de la topología y de los parámetros de entrenamiento de las redes neuronales utilizadas.

Otra de las fases del trabajo ha sido la concepción de un sistema de evaluación que permita la comparación de distintos métodos de generación de curvas de frecuencia

fundamental, así como su valoración en términos absolutos.

Por último, se ha abierto una línea de trabajo que consiste en utilizar programación genética como método para obtener la expresión analítica de las curvas de frecuencia fundamental. Aunque los experimentos en este campo pueden considerarse como preliminares, la calidad de los resultados obtenidos en los mismos augura un futuro prometedor para estas técnicas.

Abstract

The main objective of this work has been, as its title implies, the improvement of the fundamental frequency in text to speech conversion, looking for not only obtaining a system that generates fundamental frequency curves of greater quality, but also defining a work methodology that permits the generation of intonation models for new speakers in an easy and almost automatic way.

The first phase of the work was the definition and creation of a database suited for the study of the prosodic phenomena, with the idea of it been useful for any work related with prosody.

One of the most important parts of the investigation has been the experimentation with neural nets as a method for generating fundamental frequency curves. The work in this field has been focused in the determination and coding of the most relevant parameters which affect the production of the fundamental frequency curves, as well as in the optimization of the topology and training parameters of the neural nets used.

Another phase of the work has been the conception of an evaluation system that permits the comparison of different fundamental frequency curves generation methods, as well as their valuation in absolute terms.

Finally, a line of work has been opened that consists in using genetic programming as a method for obtaining the analytic expression of the fundamental frequency curves. Although the experiments in this field can be considered as preliminary, the quality of the results obtained in them omens a promising future for these techniques.

Índice

Encuadre científico y tecnológico	1-1
Introducción	1-1
Asignación automática de contornos de F0	1-2

Base de datos	2-1
Introducción	2-1
Estructura y composición de la base de datos	2-2
Modo oracional	2-3
Casos especiales	2-5
Comentarios sobre los esquemas básicos	2-6
Distribución según los criterios del sistema de categorización y pausado	2-8
Distribución en entrenamiento y reconocimiento	2-9
Grabación	2-9
Segmentación fonética	2-10
Marcado de valores de F0 y revisión	2-12
Categorización y pausado	2-16
Conclusiones y trabajo futuro	2-18

Experimentos con redes neuronales	3-1
Introducción	3-1
Experimento preliminar	3-1
Incorporación del número de sílabas	3-7
Eliminación de la información microprosódica	3-10
Revisión de la base de datos e inclusión de la zona inicial	3-11
Ampliación de la zona inicial	3-12
Simplificación del tipo de terminación	3-13
Otros experimentos	3-20
Conclusiones	3-22

Programación genética	4-1
Introducción	4-1
Algoritmos genéticos	4-1
Funcionamiento	4-2
Justificación	4-7
Programación genética	4-10
Creación de la población inicial	4-12
Operadores genéticos	4-13
Parámetros de control	4-17
Modificaciones de las operaciones	4-18
Aplicación a la obtención de valores de F0	4-19

Codificación de las variables	4-20
Experimentos realizados	4-21
Resultados	4-22
Conclusiones y trabajo futuro	4-25
<hr/>	
Evaluación	5-1
Introducción	5-1
Metodología de evaluación	5-2
Criterios de selección de las frases	5-4
Proceso de evaluación	5-5
Puntuaciones obtenidas por los distintos modelos	5-7
Conclusiones	5-11
<hr/>	
Conclusiones y trabajo futuro	6-1
Introducción	6-1
Base de datos	6-1
Redes neuronales	6-2
Programación genética	6-2
Evaluación	6-3
Conclusión final	6-3
<hr/>	
Bibliografía	7-1
<hr/>	
Apéndices	8-1
Textos de la base de datos	8-1
Grupo 1	8-1
Grupo 2	8-3
Grupo 3	8-5
Grupo 4	8-7
Grupo 5	8-11
Textos utilizados en la evaluación	8-17