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 construidos 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 conclusion de esta tesis apoya 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