

UN SISTEMA DE RECONOCIMIENTO DEL ESPAÑOL CON UN LEXICO DE 30.000 UNIDADES

Anne DEMEDTS

Dragon Systems, Inc en Boston

En 1989 Dragon Systems, Inc. presentó la primera versión del DragonDictate para el inglés americano en Speech Tech en Nueva York : se trata de un sistema de reconocimiento del habla que se adapta a la fonética particular del usuario y que reconoce en tiempo real un amplio léxico de 30.000 unidades en palabra desconectada. Actualmente se están finalizando las versiones española, francesa, italiana y neerlandesa ; el sistema alemán ya está en fase de «beta-testing» en Alemania.

Según esta tecnología, primero hay que crear modelos acústicos de todos los fonemas de una lengua determinada, considerados en la variedad de todos sus contextos posibles. A partir de ahí se elaboran Modelos Ocultos de Markov (Hidden Markov Models) cuyos parámetros pueden ser re-evaluados a base de una información fonética mínima. En el léxico, pues, la ortografía de una unidad fonética va acompañada del correspondiente modelo acústico y, también, de un indicio de frecuencia. Los datos estadísticos son actualizados constantemente de acuerdo con el idiolecto del usuario.

En el proceso de reconocimiento intervienen tres componentes básicos, que suponen una aproximación gradual entre un conjunto de probabilidades y la palabra dicha por el usuario : un algoritmo de verificación rápida, un conjunto de modelos acústicos de palabras basados en Modelos Ocultos de Markov y el ya referido modelo lingüístico estadístico.

ADAPTACIÓN FONÉTICA

En cualquier momento DragonDictate dispone de un conjunto de 30.000 palabras caracterizadas en tramas de 20 milisegundos por medio de 8 parámetros acústicos. Esto no implica que un hablante de referencia haya tenido que decir las todas ni para el usuario que haya que hacer lo mismo a fin de personalizar el léxico porque un vocable se compone de cierto número de 'fonemas en contexto' (PIC : Phonemes In Context) que no le son privativos. A su vez un fonema siempre consta de, hasta 6 diferentes, elementos fonéticos (PEL : Phonetic Element) que se encuentran compartidos por varios alófonos del mismo. De esta manera se entablan relaciones, a menudo muy complejas, entre diferentes unidades. También explica que el DragonDictate 'aprende' a medida que el usuario lo utilice, gracias a esta extrapolación de datos.

A esta información fonética básica se añade otra de tipo durativo relacionada con tanto la ubicación del acento como la estructura de la palabra. Considérese, por ejemplo, que en español la vocal acentuada es relativamente larga en palabras agudas que no terminen en 'n' o 'l' («papá»), mientras la vocal inacentuada es generalmente breve. Si bien es cierto que el papel de estos factores varía, las mismas herramientas informáticas son válidas para cada lengua.

En cuanto a reconocimiento bastan para la caracterización fonética del español 11 vocales y diptongos, y 22 consonantes. Para las vocales se distinguen, además, la ocurrencia en sílaba tónica y la realización átona.

PROCESO DE RECONOCIMIENTO

Tampoco los componentes integrantes del mecanismo de reconocimiento dependen fundamentalmente de la lengua utilizada.

El algoritmo de verificación rápida efectúa una primera selección dentro del conjunto léxico comparando la secuencia inicial de la unidad detectada con la de unos cientos o, aun, miles de grupos

Tf 1-617-965-5200

Fax# 617-527-0372

- A.Juilland, «Frequency Dictionary of Spanish Words», 1964, The Hague : Mouton.
- Bamberg P.G., «Adaptable Phoneme-based Models for Large-Vocabulary Speech Recognition», Proceedings of the ESCA Tutorial and Research Workshop on Speaker Characterization, University of Edinburgh, U.K., June, 1990.
- Bamberg P. et al., «Incorporating natural-language information into a statistical language model for large-vocabulary recognition», Proceedings of SpeechTech, New York, 1992.