El PROCESO DE RECONOCIMIENTO DEL HABLA CON UN AMPLIO LEXICO

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 una- dicción discreta. Actualmente se están realizando las versiones española, francesa, alemana, italiana y neerlandesa, cuya finalización está prevista para 1992.

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 de Markov Escondidos (H.M.M. : Hidden Markov Models) cuyos parámetros pueden ser re-evaluados a base de una información fonética mínima. Se ha alcanzado ya la fase de comparación entre prototipos en las distintas lenguas mencionadas : operan con un vocabulario reducido de unas siete mil palabras logrando un rendimiento comparable al de

El léxico básico del DragonDictate está compuesto de 30.000 unidades fonéticas, de las cuales 5.000 son escogidas por el usuario, que las añade al núcleo de las 25.000 unidades mas frecuentes de una lengua particular. Así~, por ejemplo, para el inglés americano se ha recurrido a un análisis estadístico de materiales facilitados por la agencia UPI. El indicio de frecuencia acompaña a la ortografía de cada unidad así como un modelo acústico que el ordenador construye a partir del análisis de la voz tal como fue pronunciada por un hablante nativo. Estos 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.

Son un algoritmo de verificación rápida, otro algoritmo de programación dinámica y el ya referido modelo lingüístico estadístico.

EL LEXICO BASICO

En cualquier momento DragonDictate dispone de un conjunto de 30.000 palabras caracterizadas en cuadros de 20 milisegundos por medio de 8 parámetros acústicos. Esto no implica que un hablante de referencia haya tenido que decirlas todas ni para el usuario que haya que hacer lo mismo a fin de personalizar el léxico porque un vocablo 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 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.

EL 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 en los que ha reunido previamente palabras que empiezan de la misma forma. Así el número de palabras eventualmente dichas se reduce a unos doscientas.

El algoritmo de programación dinámica hace uso de H.M.M. estableciendo una compleja red de probabilidades entre diversos PELs acompañados por la expresión en milisegundos de su duración calculada en base al entorno lingüístico.

De momento la función del modelo lingüístico se limita a la estimación de la probabilidad de una palabra, teniendo en cuenta la palabra anterior. Se basa en un cálculo de frecuencias, que podrá incluir también frecuencias de pares de palabras.

EL ESPAÑOL

En cuanto a reconocimiento bastan para la caracterización fonética del español 11 vocales y diptongos, y 22 consonantes. Inicialmente se tiene prevista la inclusión de 3 tipos de acentuación por analogía con, entre otros, el neerlandés y el inglés. Para facilitar la comparación entre los sistemas en idiomas diferentes, se han tomado como punto de partida los capítulos 1, 2, 3, 7 y 8 del texto de A. de Saint-Exupéry "Le Petit Prince" y sus traducciones a las lenguas respectivas. A estos se les van añadiendo textos auténticos en cada una de las lenguas.

Pruebas con el DragonDictate para el inglés americano pusieron de relieve que tras una adaptación de aproximadamente 2.000 palabras, entre el 85% y el 90% de las palabras son reconocidas correctamente. En cuanto a los errores se distinguen los tipos siguientes :

- "opciones" (o) : la palabra correcta figura en una lista de hasta 8 opciones alternativas que aparece en la pantalla de modo que el usuario puede corregir el error pulsando una tecla

- "nueva palabra" : (N) todavía no existe un modelo acústico para la palabra puesto que no forma parte del léxico básico del sistema. Por tanto, el usuario habrá de escribir toda la palabra : a continuación se elabora un modelo acústico que el sistema podrá relacionar con la grafía en cuestión.

- "error" (E) la palabra no figura entre las opciones y el usuario tiene que introducir uno o más caracteres antes de que sea reconocida correctamente.

Por lo que respecta al prototipo español se han llevado a cabo varios experimentos.

En una primera prueba intervenían tres locutores : el hablante de referencia que habla el español peninsular (1), otro hablante nativo procedente de Venezuela (2) y, finalmente, una

persona cuya lengua materna no es el español (3). Se indica el sexo de los hablantes entre paréntesis. En el desarrollo de la prueba se sucedían tres etapas : lectura del primer capítulo de "El Principito" - fase de iniciación -, lectura de los capítulos 2 y 3 - fase de adaptación -, lectura de los capítulos 7 y 8 - fase de evaluación del rendimiento -.

Como se desprende del esquema, aunque inicialmente el rendimiento es mejor para el hablante de referencia el sistema se adapta al acento peculiar del usuario. Además, pruebas en otras lenguas demuestran que incluso la divergencia de sexo entre el hablante de referencia y otros usuarios no plantea mayores dificultades. (Las cifras indican porcentajes)

	cap.1			cap.2,3			cap.7,8		
	O	N	E	O	N	E	O	N	E
1. AD (f)	83	12	5	86	10	4	88	7	5
2. KK (f)	76	9	15	85	7	8	86	6	8
3. KH (f)	76	12	12	84	7	9	87	6	7
POR MEDIO	78	11	11	85	8	9	87	6	7

Durante la fase inicial el hablante de referencia obtiene, claro está, el mejor rendimiento ; sin embargo, basta con dictar unas 2000 palabras para que el sistema se haya adaptado a la fonética de cada usuario. No hace falta que la pronunciación sea 'correcta', tan solo habrá de ser consistente.

En el caso de los dos usuarios otros que el hablante de referencia, el porcentaje de verdaderos 'errores' - en que la unidad dictada ni siquiera figuraba en la lista de opciones - es el doble del porcentaje análogo para el hablante de referencia. Esto sugiere que existan cualidades propias de los modelos de verificación rápida fuera del alcance de la adaptación.

Tan solo para la versión inglesa del DragonDictate disponemos ya de la versión con un amplio vocabulario de reconocimiento (30.000 unidades). La comparación de los resultados obtenidos, por una parte, con un prototipo reducido y, por otra, con la versión de léxico amplio pone de relieve que un aumento considerable del vocabulario no afecta al rendimiento.