

## DEFINICIÓN DEL PROBLEMA

El Reconocimiento del Habla parece tan natural y sencillo para las personas que se pensó que podría ser fácilmente realizado por las máquinas. Sin embargo, cuando se empezó a profundizar en el tema, se comprobó que esto no es así. De hecho, es un tema que se ha revelado más complicado que la producción automática de voz.

Ya la historia lo ha demostrado: las primeras y rudimentarias máquinas parlantes aparecieron en la segunda mitad del siglo XVIII, mientras que los primeros intentos en máquinas capaces de reconocer la voz no aparecieron hasta principios del siglo XX, con la máquina de Flower, capaz de escribir el alfabeto fonográfico pronunciado por una persona. Cinco son los factores que determinan la complejidad del Reconocimiento del Habla:

### A. *El Locutor*

Es quizás el aspecto que introduce mayor variabilidad en la forma de onda entrante, y por tanto requiere que el sistema de reconocimiento sea altamente robusto. Una persona no pronuncia siempre de la misma forma, debido a distintas situaciones físicas y psicológicas (es la llamada variabilidad intra-locutor). Existe además gran variedad entre distintos locutores (hombres, mujeres, niños), diferencias según la edad o la región de origen (variabilidad interlocutor). Es mucho más sencillo que un sistema funcione para un determinado locutor y que este lo haya entrenado previamente (se dice que el sistema es dependiente del locutor), a que un sistema funcione para cualquier locutor (sistema independiente del locutor).

### B. *La forma de hablar*

Es el segundo factor que determina la complejidad de un reconocedor de habla. El hombre pronuncia las palabras de una forma continua, y debido a la inercia de los órganos articulatorios, que no pueden moverse instantáneamente, se producen efectos coarticulatorios. Ello, unido a las variaciones introducidas por la prosodia, hace que una palabra al principio de una frase sea diferente que cuando se dice en medio, o que sea diferente dependiendo de que es lo que le precede o le sigue. Un reconocedor es relativamente sencillo si sólo tiene que reconocer una palabra dicha de forma aislada (reconocedor de palabras aisladas) y es más complejo si debe reconocer las palabras de una frase, pero introduciendo una pausa entre cada dos de ellas (habla conectada). El sistema más complicado es aquel que debe funcionar reconociendo habla continua, que es la forma natural de hablar.

### C. *El Vocabulario*

Se conoce por tal el número de palabras diferentes que debe reconocer el sistema. Mientras mayor es el número de palabras más difícil es el reconocedor, por dos motivos. El primero porque al aumentar el número de palabras es más fácil que aparezcan palabras parecidas entre sí, y el segundo porque el tiempo de tratamiento aumenta al aumentar el número de palabras con las que comparar. Una solución posible a este problema sería el utilizar unidades lingüísticas inferiores a la palabra (alófonos, sílabas, etc.) que en principio tienen un número limitado, e inferior al de posibles palabras. Sin embargo, la dificultad de reconocer estas unidades es aun mayor debido a que su duración es muy corta, la frontera entre dos unidades sucesivas es muy difícil de establecer y los efectos coarticulatorios son mucho más fuertes que entre palabras.

#### *D. La Gramática*

Es el conjunto de reglas que limita el número de combinaciones permitidas de las palabras del vocabulario. En general la existencia de una gramática en un reconocedor ayuda a mejorar la tasa de reconocimiento, al eliminar ambigüedades y puede ayudar a disminuir la necesidad de cálculo, al limitar el número de palabras en una determinada fase del reconocimiento ("perplejidad" de la gramática). En sistemas de palabras aisladas en los que no existe una gramática en el sentido estricto del término, se puede entender por tal el número de palabras a reconocer. Si, por ejemplo, el sistema debe reconocer un número telefónico urbano, la gramática de este sistema dice que el vocabulario son los diez dígitos, y debe reconocer un conjunto de siete dígitos, de forma que si el sistema reconoce más o menos, es que hay algún error.

#### *E. El Entorno físico*

Es una parte tan importante como las anteriores para definir el reconocedor. No es lo mismo un sistema que funciona en un ambiente poco ruidoso, como puede ser el despacho de un médico, o el que tiene que funcionar en un coche o en una fábrica. O por ejemplo, el que debe de funcionar a través de la línea telefónica, con la consiguiente reducción de banda o el que recibe la voz a través de un micrófono, que tiene mayor ancho de banda que la línea telefónica.