

Las técnicas de reconocimiento de voz se van haciendo operativas

¿Entiendes lo que te digo?

Joan García Pol. Dr. Ingeniero Industrial. Lcdo. en Informática

La comunicación entre la persona y el ordenador ha pasado por muchas etapas adoptando numerosas formas: desde perforaciones en una cinta de papel hasta impulsos sobre un radioenlace. Una vez que el mensaje está introducido en el sistema, todo lo demás es fácil porque el sistema sabe hacerlo de manera automática. La problemática (que siempre ha sido más complicada) es la entrada inicial de un dato, puesto que implica la adaptación de un lenguaje humano a códigos y reglas impuestos por la máquina.

Como es lógico, las soluciones establecidas normalmente se han basado en la ley del mínimo esfuerzo: ya que la persona tiene unas capacidades muy superiores a la máquina, ha sido más fácil adaptar la persona a la máquina que lo contrario. Por eso se usó el teclado de la máquina de escribir y más tarde se ha empleado el "lápiz selector", que evolucionó en el ratón como forma de apuntar sobre una pantalla.

Si bien se piensa, no ha habido otros medios de relación entre el sistema humano y el de la máquina precisamente en la dirección indicada. La relación en sentido contrario (desde la máquina hacia la persona, salida de información en definitiva) es mucho más fácil por las mismas razones: la persona tiene posibilidad de adaptarse a variados soportes de salida (se-

an visuales en papel, pantalla, escritos o en gráfico, incluso sonoros) mediante terminales acústicos o sencillamente teléfonos.

Sin embargo, una de las formas de comunicación entre personas es a través del lenguaje hablado. No es la única y quizá tampoco la más importante, pero sí la más asequible y sencilla para la gran mayoría de la gente. La moderna implantación de los sistemas informáticos aspira a conseguir su difusión masiva y, en consecuencia, se preocupa de conseguir una facilidad de uso cada vez mayor. La evolución en esta dirección ha llevado a poner sobre la mesa la búsqueda de solución a la comunicación hablada entre persona y máquina. La tecnología que ha permitido esto es la llamada reconocimiento de voz.

Para implantar esta solución se necesita un PC bastante potente con un micrófono y una tarjeta de sonido, un programa adecuado y un par de altavoces si queremos que el PC nos conteste.

El software de reconocimiento de voz es suficientemente inteligente para interpretar lo que se le dice y ejecutar la orden correspondiente o

bien convertir las palabras en texto escrito. Para ello, usa métodos de inteligencia artificial para determinar qué palabras se le están diciendo.

Tres grupos de posibilidades definen un software de este tipo:

1.- Amplitud del vocabulario conocido: cuanto mayor sea, más libertad de expresión del usuario pero más lentitud en la interpretación del PC.

2.- Uso de habla discreta o continua según sea necesaria una pausa de separación entre palabras o no. Naturalmente, hablar sin pausas exige un programa más potente.

3.- Dependencia o independencia del que habla: un sistema dependiente del que habla debe ser entrenado para reconocer la voz de cada usuario, mientras que si es independiente puede ser usado sin entrenamiento previo por

distintos usuarios. El "entrenamiento" se hace hablándole al sistema las distintas personas para que éste "aprenda" los modelos de habla de cada uno. La complejidad del programa es también diferente.

La primera ventaja que aparece al usar un sistema de reconocimiento de voz es la constatación de tener las manos libres. Tan importante como

«Naturalmente, hablar sin pausas exige un programa más potente».

ésta es la de tener los ojos libres ya que no necesita mirar al teclado ni a la pantalla. Todo ello permitirá hacer otras cosas mientras se dicta.

El sistema también es capaz de incorporar nuevas palabras para ir enriqueciendo su diccionario según las preferencias del usuario. Puede reconocer diferencias de contexto para distinguir expresiones de sonido parecido como porque y por qué. Evidentemente se ha de disponer de un vocabulario por cada idioma. Han aparecido también diccionarios ampliados que se orientan a los vocablos y expresiones particulares de ciertas profesiones.

En cuanto a la velocidad de dictado, existen sistemas que proclaman alcanzar no sólo las 120 palabras por minuto, sino las 140 e incluso las 160, superando en mucho las marcas de mecanógrafos experimentados. En cuanto a la extensión del vocabulario a efectos de obtener buenos rendimientos, se suele empezar con unas 40 o 50.000 palabras con posibilida-

des de ampliar con 20.000 términos que se van incorporando automáticamente según las costumbres del usuario. Están disponibles también amplios diccionarios de hasta 500.000 términos con los que no podremos exigir un rendimiento tan alto.

Después de un dictado, es lógico reparar el texto y corregirlo. Para ello se suele incluir una conversión desde el texto escrito a sonido, así como un corrector ortográfico corriente.

Empleando un sistema de reconocimiento de voz nos daremos cuenta de la perfección que tiene el ser humano en el arte de entender el lenguaje hablado. Estudios realizados han probado que el margen de error de las personas al escuchar la lectura de números dígitos es de un error cada 10.000 mientras que el ordenador está en siete errores cada 1.000, margen aceptable para algunas aplicaciones pero que no deja de ser enorme para otras.

En un dictado continuo la persona tiene un margen de error del

0,4% y en los sistemas actuales es de un 8 al 10%. Si consideramos una conversación, las dificultades son considerables para el ordenador. La persona es capaz de seguir una conversación aunque pierda trozos de la misma ya que su capacidad de deducción y de intuición le facilita las cosas. También es capaz de seguir varias conversaciones a la vez entre múltiples interlocutores.

Los sistemas informáticos están todavía lejos de igualar al ser humano aunque las soluciones presentes pueden ser muy útiles. Y de lo que no cabe la menor duda es de que la evolución será notable en el futuro aun quedando mucho por hacer, tanto en los sistemas de voz como en los lingüísticos. Se necesitará también un fuerte desarrollo tecnológico para tener más potencia de proceso a coste accesible.

No creemos que con todo ello podamos prescindir del teclado, pero nuestro trabajo será más cómodo y fácil. ■