

## PHILIPS: Intelligent Speech Interpretation - la tecnología inteligente de reconocimiento de voz

### PHILIPS Speech Recognition Systems

Central: A-1101 Viena  
España: 08007 Barcelona  
info.speech@philips.com  
www.philips.com/speechrecognition

### Javier Viver

Director Regional Iberia & Latino América  
PHILIPS Ibérica SA  
Speech Recognition Systems  
Balmes 22 3º, 08007 Barcelona  
javier.viver@philips.com

**Resumen:** Independientemente de la aplicación, el núcleo de cualquier sistema de reconocimiento de voz sigue siendo la inteligencia que convierte el habla en texto. Al reconocer lo que se dice e interpretarlo de forma razonable, garantiza una intervención mínima en el proceso de transcripción para producir informes útiles.

**Palabras clave:** Philips, SpeechMagic, reconocimiento del habla, tecnología inteligente de reconocimiento de voz, Intelligent Speech Interpretation.

### 1 *Del texto oral al texto escrito*

Philips ha desarrollado SpeechMagic, un conjunto de herramientas de desarrollo de software (SDK) de reconocimiento del habla, para grupos de trabajo profesionales..

Diseñada por profesionales, la tecnología de reconocimiento del habla de Philips se utiliza en instituciones con un gran volumen de dictado para automatizar la creación de documentos, por ejemplo en el sector hospitalario.

Los retos iniciales a los que se enfrenta cualquier reconocedor del habla son los relacionados con el reconocimiento acústico, como el ruido de fondo, las diferencias dialectales y de pronunciación, la velocidad etc., así como los relacionados con el Contexto; el sistema debe reconocer lo que ha dicho el orador y esto empieza por el vocabulario así como la probabilidad asociada a cada palabra concreta y a como estas palabras se encadenan entre ellas.

### 2 *Inteligencia situacional*

Ahora bien, este reconocimiento "literal" de lo que se ha dicho ya no es suficiente.

Los buenos transcripores hacen mucho mas que simplemente reflejar lo que se ha dictado. Para empezar, eliminan las expresiones como 'hum' y 'eh', ignoran diálogos que no forman parte del dictado, incluyen correcciones que se dictan como parte del texto, y también la

información en formularios e incluso expresan las frases de otra manera. Realizan correcciones y comprobaciones para aclarar incoherencias. Y, lo que no es menos importante, dan formato y organizan el texto, añaden cabeceras, numeran listas y bloques de contenidos estándar, entre otros.

En resumen, se aseguran de que el producto final comunica el significado real, en lugar de ofrecer una transcripción literal de lo que se ha dicho.

### 3 *(Re-)Conocimiento contextual*

Reconocer el significado de lo que se ha dicho proporciona una base solida para corregir las frases y las estructuras sintácticas.

Un dictado espontáneo a menudo da lugar a la ausencia de artículos, verbos y puntuación, así como palabras superfluas o repetidas. Una buena comprensión del contexto ayuda a interpretar las palabras para identificar y corregir dichos temas.

La tecnología inteligente de reconocimiento de voz, "Intelligent Speech Interpretation" (ISI), como la ha desarrollado Philips en Viena, Austria, emula las capacidades de los transcripores. La tecnología resulta igual de útil a médicos que prefieren atender por si mismos el proceso de creación de informes, como a los que simplemente quieren dictar y firmar y dejar que otra persona se ocupe de la edición.

#### 4 Ejemplos de interpretaciones de palabras y comprensión de situaciones

En la tabla 1 se muestra algunos de los cambios realizados en el procesamiento de voz inteligente

	Texto dictado	Texto reconocido
Frases redundantes	Fin de dictado. Gracias.	<i>No aparece en el documento final.</i>
Cabeceras de sección (Heading sections)	Condición en/a/sobre/de/ la descarga La siguiente es la condición...	{CONDICIÓN EN LA DESCARGA}
Fechas	Mayo cinco dos mil dos Cinco de mayo de dos mil dos	5 de mayo, 2002
Puntuación automática	Sin escalofríos fiebres sudores nocturnos pérdida de peso...	Sin escalofríos, fiebres, sudores nocturnos, pérdida de peso ...
Silencios / pausas	No se han producido (-----pausa----) grandes cambios...	No se han producido grandes cambios.
Ruido de fondo (Non-speech dictation)	No se han producido (crujido de papel) grandes cambios...	No se han producido grandes cambios.
Vacilaciones (Hesitations)	No se han producido (AAHHMMMMM) grandes...	No se han producido grandes...
Contracción	No s'an producido	No se han producido

Tabla 1: Ejemplos de interpretaciones de palabras y comprensión de situaciones

#### 5 Características avanzadas de reconocimiento de voz en línea

Sin embargo, con la última versión del software, esta comprensión del contexto significa también ampliar el reconocimiento de la voz a la interfase de usuario.

Esto que supone la utilización del reconocimiento de voz no sólo para el dictado del informe sino también para el comando y control de la aplicación; una característica especialmente atractiva para los usuarios que prefieren ocuparse ellos mismos de todo el proceso de generación de informes.

Al diferenciar comandos de los contextos del dictado, las aplicaciones de reconocimiento de voz y de proceso de texto se pueden vincular más estrechamente para eliminar la mayor parte del resto del trabajo de teclear. El usuario puede abrir documentos, activar y desactivar la cursiva, corregir, deletrear nombres de pacientes poco habituales, etc., simplemente mediante su voz.

Para un radiólogo, por ejemplo, esto significa poder dictar y manipular una imagen simultáneamente.

Con una configuración sencilla, un médico puede incluso definir "macros de voz" personales para insertar bloques de texto que suele utilizar a menudo incluso desplazarse a través de formularios utilizados de modo regular y rellenarlos completamente sin necesidad de teclear.

#### 6 Objetivo claro: productividad destacada

El reconocimiento de voz para dictado se puede utilizar para que el facultativo pueda producir inmediatamente los informes para agilizar un servicio de transcripciones tradicional, permitiendo que las secretarías se concentren en el control de calidad, aunque evidentemente el sistema también ayuda en esto al escribir correctamente las palabras.

De cualquier modo, reduce los gastos administrativos generales y disminuye el tiempo entre el dictado y la impresión, algo que numerosos servicios de radiología agradecen, por ejemplo, para complementar un PACS que ya aumenta las expectativas de sus colegas de otras áreas del hospital.

La línea inferior representa tiempos de espera de paciente más cortos y mayor satisfacción del paciente.