

SÍNTESIS DE VOZ

Las personas ciegas y con discapacidad visual no funcional utilizan, para acceder a sus dispositivos electrónicos, un producto de apoyo conocido como Lector de pantalla. Este tipo de software emplea habitualmente una voz sintética para transmitir al usuario ciego la información del dispositivo.

La síntesis de voz es el proceso mediante el cual se busca que una máquina genere voz o simule el habla humana, a partir de un texto, mediante el uso de procesos matemáticos y el estudio de la fonética y la [teoría lingüística](#), produciendo, de forma artificial, expresiones vocales por parte de la máquina, sin que éstas sean pregrabadas.

Las cualidades más importantes de los sistemas de síntesis de voz son la "naturalidad" (grado de semejanza a la voz humana) y la "inteligibilidad" (grado de entendimiento). Los sistemas de síntesis de voz tratan usualmente de maximizar estas características, por lo que la calidad de un sintetizador de voz se evalúa por la similitud que tenga con la voz humana y su habilidad para ser entendido de manera clara.

Los sistemas operativos de PC integran sintetizadores de voz desde principios de los 90.

Existen varios tipos de síntesis de voz:

- Voz pregrabada (codificación) (por ejemplo, las de alguna compañía de telefonía).
- Conversión de texto a voz: Text-To-Speech (TTS) (por ejemplo, la mayor parte de las que llevan los lectores de pantalla).

A lo largo de los años, mediante diferentes investigaciones y el desarrollo de la tecnología, se ha llegado a diversas formas de sintetizar voz en las máquinas:

- Sintetizador Articulatorio: se basa en hacer una analogía entre los parámetros de los órganos articulatorios del hombre con parámetros circuitales que le permitan a la máquina emularlos. La calidad que ofrecen es muy alta pero son muy complejos a la hora de configurarlos.
- Sintetizadores por formantes: buscan imitar el tracto vocal humano mediante una serie de filtros que se activan por fuentes que simulan las cuerdas vocales.
- Sintetizadores derivados de las técnicas de predicción lineal: son sintetizadores basados en los dos anteriores.
- Sintetizadores por concatenación de forma de onda: poseen una base de datos con unidades o fonemas pregrabados, de modo que después del correcto análisis, concatenan las unidades adecuadas para generar nuevas frases. Su grado de complejidad es alto y se limita a las características de un solo hablante, el que realiza las grabaciones, pero

mejora la calidad y se obtienen muy buenos resultados, siendo el método más utilizado en la actualidad para diversas aplicaciones.

A continuación, se profundiza un poco más en este tipo de síntesis.

Síntesis por concatenación

Este método está basado en bloques de voz pregrabada de forma tal, que dependiendo del sistema, se unen adecuadamente para poder llegar a formar las frases o palabras que se introdujeron por texto.

Dentro de la síntesis concatenativa, existe la síntesis por selección de unidades. Este tipo de síntesis es la que más suele gustar, ya que son las voces que suenan más naturales y más se asemejan a la voz humana porque utilizan una base de datos (que será mayor cuanto más se parezca a la voz humana) de letras, fonemas, sílabas, palabras, frases y hasta oraciones que graban diferentes locutores/personas. Este tipo de síntesis es la que usa Loquendo, Acapela, Ivona, Capstral, Verbio, AT&T, etc.

Síntesis por difonos

Usa una base de datos de voz mínima que contiene todos los difonos. En ésta se graban todos los difonos de un idioma determinado para una voz. En la programación posterior hay que concatenar bien para conseguir un buen tono y una buena pronunciación. Síntesis de este tipo son, por ejemplo, Festival, Elan Speech, la voz de Javier de Acapela.

Síntesis de dominio específico

Concatena palabras y frases pre-grabadas para crear enunciados completos. Se usa en aplicaciones donde la variedad de los textos del sistema está limitada a una salida de audio en un dominio particular, como los anuncios en un calendario de tránsito, calculadoras, relojes, ascensores parlantes o audio juegos.

Síntesis de formantes:

No utiliza muestras de voz humana durante el tiempo de ejecución. Este tipo de síntesis, que tiene muchas ventajas, se usa mucho. Es la que utiliza Eloquence, DecTalk o Fonix Talk. Son voces que al no usar voz grabada pesan muy poco. La tecnología de síntesis de formantes genera una voz artificial con sonido robótico que no podría ser confundida con la voz humana. El objetivo de este tipo de sistemas de síntesis de voz no es la naturalidad pero puede ser inteligible, inclusive a grandes velocidades. Este tipo de voces permite modificaciones en sus parámetros: velocidad, tono, frecuencia, timbre analógico, entonación, etc. En los sistemas basados en la voz humana no se pueden modificar estos parámetros.

Síntesis basada en modelos HMM

Se basa en modelos ocultos de Markov, también llamada síntesis estadística paramétrica. Este tipo de síntesis se utiliza mucho en los móviles. Es la voz, por ejemplo de Vocalizer, Acapela para las voces Colibrí, Svox Pico, Google TTS o Samsung TTS.

USOS, MOTORES Y LIBRERÍAS

Las voces digitales dependen de un motor de síntesis de voz, capaz de comunicarse con la tarjeta de sonido del equipo informático para reproducir los diversos sonidos encadenados que formarán las frases habladas. Además, este motor de voz tendrá en cuenta reglas lingüísticas y de entonación. Necesita unos patrones sonoros o de pequeñas grabaciones de pronunciación que permitan definir el tono, el acento y la sonoridad final que tendrá la grabación y a este conjunto de datos se le conoce como una voz para un motor de síntesis.

En los sistemas operativos pertenecientes a la familia Microsoft Windows, existen unas librerías para poder utilizar diversos motores de voz de forma sencilla para aumentar la compatibilidad entre software. Estas librerías son conocidas como SAPI (Speech Application Program Interface).

En las versiones actuales de Microsoft Windows, más concretamente, a partir de la versión de Windows XP, SAPI 5 ya viene incluida en el sistema.

En idioma castellano hay multitud de voces gratuitas para el formato SAPI 4. Para Microsoft Windows XP/2003/Vista (últimas versiones de este sistema operativo) se utiliza, de serie, el formato SAPI 5.1. Hay compañías como Loquendo o Nuance que se dedican a crear voces de calidad para terminales de atención al cliente, puntos de venta automatizados o sistemas de megafonía para estaciones de transporte y aeropuertos.

En los equipos con MacOS X, el sistema operativo de Apple para sus equipos portátiles y de sobremesa, está integrado en el sistema un motor de voz genérico que puede ser utilizado por diversos programas y servicios del sistema. Actualmente se incluyen varias voces con diversos acentos y efectos sonoros aunque el problema es que todas las voces están en inglés. Se pueden adquirir voces para otros idiomas.

En GNU-Linux existen varios motores o librerías de síntesis de voz como Festival, Mbrola o una versión para GNU-Linux de Eloquence. Estos motores se utilizan en el sistema a través de un interfaz que comunique las aplicaciones del sistema con el motor de voz, un ejemplo de este tipo de interfaz es GNOME Speech.

LIBRERÍAS

NUANCE VOCALIZER



Vocalizer Expressive

Tecnología de síntesis de voz en forma de librería a disposición de los desarrolladores que quieren incorporar esta característica a sus aplicaciones.

Autor: Nuance.

Idioma: Muchos, entre los que se encuentra español y catalán.

Licencia: Para el lector de pantalla gratuito, NVDA se dispone de un complemento de pago aunque existen versiones de prueba.

Para el lector de pantalla JAWS, así como para el magnificador de pantalla Magic, Freedom Scientific permite descargar [Vocalizer Direct Voices](#), que sólo funcionará con estos softwares.

Viene incluido con la compra del lector de pantalla Nuance [Talks & Zoom](#) para equipos con sistema operativo Symbian. También está incluido con la compra del entorno cerrado de usuario [Mobile Accessibility](#) de Code Factory diseñado para teléfonos Android, con la compra del reproductor / grabador [BookSense XI](#) y con la compra del reconocedor óptico de caracteres (OCR) [OmniPage](#).

ETI-ELOQUENCE TTS



Es un motor TTS muy ágil que consume pocos recursos. Está disponible en muchos idiomas para distintos sistemas operativos y posee una amplia variedad de aplicaciones como:

Lectores de pantalla y aplicaciones para ciegos: Talkback o Mobile Accessibility.

Software GPS (Google Maps).

Lectores de libros electrónicos (Play Books).

Traductores (Traductor de Google).

Autor: [Code Factory](#).

Idioma: Se encuentra disponible en muchos idiomas (10) entre ellos el español.

Licencia: En Windows viene incluida con el lector de pantalla JAWS.

En Android se puede comprar en la Play Store de Google. No es un motor gratuito

La Licencia de la versión para Android incluye 10 idiomas: alemán, español (de España y México) finlandés (Finlandia) francés (de Francia y Canadá) inglés (de Reino Unido y Estados Unidos) italiano y portugués (de Brasil).

REQUISITOS

- Sistemas operativo Windows XP, 7 y 8 (32 y 64 Bits) y Android 4.0 o superior.
- Altavoces o auriculares.

USO

En Windows la síntesis de voz que viene incluida con el lector de pantalla JAWS muestra un rendimiento muy óptimo que resulta muy ágil al manejar los diferentes programas. Además, dispone de multitud de opciones que se entremezclan con las propias del lector.

Si se quiere hacer uso de este motor con cualquier otro lector, es necesario comprarla.

En Android dispone de muy buenas posibilidades de configuración como son el diccionario personalizado, la optimización de audio y el ajuste de frecuencia de muestreo para optimizar el rendimiento del terminal, especialmente los de menores prestaciones.

LOQUENDO TTS



Tecnología de síntesis de voz en forma de librería a disposición de los desarrolladores que quieren incorporarlo a sus aplicaciones para crear sus

propios textos para convertirlos a voz, ajustarlos y guardar las particularidades para su utilización futura.

Disponible en más [de 30 idiomas con más de 70 voces](#). Loquendo TTS es compatible con los principales sistemas operativos y estándares de voz y está disponible en una amplia gama de configuraciones para cumplir los requisitos de cualquier aplicación.

Voz muy expresiva que permite sonidos de toses, risas o llanto. Permite definir la pronunciación de los acrónimos, los nombres propios, las abreviaturas, etc. Con posibilidades de configuración para modificar la velocidad, el tono del habla, la frecuencia y la duración de las pausas.

Autor: Nuance.

Idioma: Más de 30 idiomas. En español cuenta con múltiples voces: Jorge y Carmen. En catalán Jordi y Montserrat. Tiene voces en gallego (Carmela) en valenciano (Empar) y muchas voces en español de América latina: Ximena, Carlos, Esperanza (México) Soledad (Colombia) Francisca (Chile) Diego (Argentina).

Licencia: Se puede usar en los siguientes sistemas operativos: Server, Microsoft Windows (Server 2008, Server 2003), Linux (Red Hat Enterprise, SuSE, CentOS), Microsoft Windows (7, Vista, XP, 8, 8.1, 10), Linux (Debian, Ubuntu), Mac OS X (SnowLeopard), Genivi, Linux Embedded, Microsoft Auto, QNX, VxWorks, Android, iPhone OS, LiMo, Maemo, Moblin, MeeGo, PalmOS, Symbian OS, Windows Mobile, Linux Embedded, Windows Embedded Standard, Windows Embedded CE, RTEMS.

Requisitos:

CPU:Procesador 32 y 64 bits y cualquier procesador Pentium o equivalente.

Todos los sistemas operativos.

256 MB de memoria RAM.

OTROS SINTETIZADORES

PICO TTS



En Android el motor TTS viene incorporado por defecto aunque es posible instalar y personalizar varios motores, eligiendo siempre uno como motor TTS principal.

El sintetizador de voz por defecto en la mayoría de dispositivos Android es PICO TTS, desarrollado por la compañía SVOX. La calidad de voz no es mala pero resulta muy monótona.

Autor: Google/SVOX.

Idioma: Muchos, entre ellos español.

Coste: Gratuito.

Disponibilidad: Sólo con algunos dispositivos Android.

IVONA TTS



Uno de los mejores sintetizadores de voz para Android junto con SVOX es IVONA TTS. Tiene una voz muy clara, realista y con buen acento y dispone de muchas voces en múltiples idiomas. En español, disponemos de [Conchita](#) y [Penélope](#).

Autor: IVONA.

Idioma: Muchos, entre ellos español.

Coste: Gratuita la versión beta.

Disponible: En IVONA TTS en Google Play.

SVOX CLASSIC TTS



Al igual que IVONA tiene un amplio surtido de [voces e idiomas](#), siendo además una de las empresas con más tiempo y experiencia en el sector. En español tenemos la voz de [Noelia](#) y [Pablo](#) y en América latina, las voces de Angélica y Juan, todas ellas de coste. Se pueden instalar las versiones de prueba desde Google Play.

Este sintetizador se puede configurar: aumentar o reducir el volumen de voz, el tono o mejorar el rendimiento.

Autor: SVOX.

Idioma: Muchos, entre ellos español.

Coste: Se pueden probar las voces gratuitamente pero con coste.

Disponible: En Google Play SVOX Classic TTS.

SAMSUNG TTS



Samsung incorpora un sintetizador de voz propio en los dispositivos Android de su marca basados en los sintetizadores de [HTS Working Group](#).

Entre las voces que incorpora, se encuentran **Ji-Soo (coreano)** **Alice (inglés EEUU)**, **Rosa (español)** **Hanna (alemán)** y **Marie (francés)** todas voces femeninas.

Además de poder configurarse como otras voces (tono, velocidad, volumen...) incorpora la posibilidad de activar ciertos efectos de voz (**profunda, alta, fina, gruesa, robótica o con efecto de helio**).

Autor: Samsung.

Idioma: Muchos, entre ellos español.

Coste: Gratuito.

Disponible: Sólo en dispositivos de la marca Samsung.

BIBLIOGRAFÍA, MANUALES, RECURSOS Y ENLACES DE INTERÉS.

- Síntesis de voz:

<http://www.programaraciegas.net/?tag=sintesis-de-voz>

Tipos de síntesis de voz:

- Wikipedia:

https://es.wikipedia.org/wiki/S%C3%ADntesis_de_habla

- Tipos de síntesis: ventajas y desventajas

<http://klango.net/pl/forum/thread/tid/939569>

Librerías:

- Compartolid (Vocalizer):

<http://www.compartolid.es/vocalizer/>

- Compartolid:

<http://www.compartolid.es/eti-eloquence-tts/>

- Sintetizadores de voz para Android:

<http://www.emezeta.com/articulos/10-sintetizadores-de-voz-tts-para-android>

- Monografías:

<http://www.monografias.com/trabajos89/sintesis-voz/sintesis-voz.shtml>

- CIDAT:

www.once.es/cidat/