



KOMPUTER SAPIENS

Revista de Divulgación de la Sociedad Mexicana de Inteligencia Artificial

Año 7
Volumen 3
Sep-Dic 2015

\$ 50.00

ROBOTS

MANIPULADORES

APRENDIZAJE
DE LAS MATEMÁTICAS

SILLA DE
RUELAS
CON MANOS LIBRES

PRODUCCIÓN
ARTIFICIAL

DEL HABLA



9772007069007



©Komputer Sapiens, Año VII Volumen III, septiembre-diciembre 2015, es una publicación cuatrimestral de la Sociedad Mexicana de Inteligencia Artificial, A.C., con domicilio en Ezequiel Montes 56 s/n, Fracc. los Pilares, Metepec, Edo. de México, C.P. 52159, México, <http://www.komputersapiens.org>, correo electrónico: editorial@komputersapiens.org, tel. +52 (833)357.48.20 ext. 3024, fax +52 (833) 215.85.44. Impresa por Sistemas y Diseños de México S.A.

de C.V., calle Aragón No. 190, colonia Álamos, delegación Benito Juárez, México D.F., C.P. 03400, México, se terminó de imprimir el 30 de diciembre de 2015, este número consta de 1000 ejemplares.

Reserva de derechos al uso exclusivo número 04-2009-111110040200-102 otorgado por el Instituto Nacional de Derechos de Autor. ISSN 2007-0691.

Los artículos y columnas firmados son responsabilidad exclusiva de los autores y no reflejan necesariamente los puntos de vista de la Sociedad Mexicana de Inteligencia Artificial. La mención de empresas o productos específicos en las páginas de Komputer Sapiens no implica su respaldo por la Sociedad Mexicana de Inteligencia Artificial.

Queda estrictamente prohibida la reproducción total o parcial por cualquier medio, de la información aquí contenida sin autorización por escrito de los editores.

Komputer Sapiens es una revista de divulgación en idioma español de temas relacionados con la inteligencia artificial. Creada en \LaTeX , con la clase **papertex** disponible en el repositorio *CTAN*: Comprehensive TeX Archive Network, <http://www.ctan.org/>

Indizada en el IRMDCT de CONACYT y en Latindex.

	Directorio SMIA	Directores Fundadores
Presidente	Grigori Sidorov	Carlos Alberto Reyes García
Vicepresidente	Miguel González Mendoza	Ángel Kuri Morales
Secretario	Félix Castro Espinoza	
Tesorero	Ildar Batyrshin	
Vocales:	Rafael Murrieta Cid	
	Maya Carillo Ruiz	
	Sofía Natalia Galicia Haro	
	Luis Villaseñor Pineda	
	Gustavo Arroyo Figueroa	
	Hugo Terashima Marín	
	Oscar Herrera Alcántara	
	Obdulia Pichardo Lagunas	
	Sabino Miranda Jiménez	
	Enrique Muñoz de Cote	
	Antonio Marín Hernández	
	Noé Alejandro Castro Sánchez	
	Ma. de Lourdes Martínez Villaseñor	
	Omar Montaña Rivas	
	Francisco Viveros Jiménez	
	Komputer Sapiens	
Director general	Grigori Sidorov	Comité Editorial
Editora en jefe	Laura Cruz Reyes	Félix A. Castro Espinoza
Editores asociados	Elisa Schaeffer	Jesús Favela Vara
	Claudia Gómez Santillán	Sofía Natalia Galicia Haro
	Marco A. Aguirre Lam	Miguel González Mendoza
Coordinadora de producción e-Tlakuilo	Viridiana Mena Gómez	Oscar Herrera Alcántara
	Jorge A. Ruiz-Vanoye	Raúl Monroy Borja
	Ocotlán Díaz-Parra	Eduardo F. Morales Manzanares
Estado del IArte	Ma del Pilar Gómez Gil	Leonardo Garrido Luna
	Jorge Rafael Gutiérrez Pulido	Carlos Alberto Reyes García
Sakbe	Héctor Gabriel Acosta Mesa	Angélica Muñoz Meléndez
	Claudia G. Gómez Santillán	Antonio Sánchez Aguilar
IA & Educación	María Yasmín Hernández Pérez	Luis Enrique Sucar Succar
	María Lucía Barrón Estrada	Ángel Kuri Morales
	J. Julieta Noguez Monroy	José A. Martínez Flores
Deskubriendo Konocimiento	Alejandro Guerra Hernández	Juan Manuel Ahuactzin Larios
	Leonardo Garrido Luna	Manuel Montes y Gómez
Asistencia técnica	Irvin Hussein López Nava	Ofelia Cervantes Villagómez
	Alan G. Aguirre Lam	Alexander Gelbukh
Corrección de estilo	Sandra Giovanna Reyes Raya	Grigori Sidorov
	Claudia L. Díaz González	Laura Cruz Reyes
	Denisse Alvarado Castillo	Elisa Schaeffer
	Jose Antonio Martínez Flores	Ramon Brena Pinero
	Silvia Clementina Guzmán Ortiz	Juan Humberto Sossa Azuela
	Karen Daniela Cruz Hernández	
	J. David Terán Villanueva	
	Lucila Morales Rodríguez	
	Marcela Quiroz Castellanos	
	Gilberto Rivera Zárate	
	Laura Gómez Cruz	
Edición de imagen	Daniel Rubio Badillo, Altera Diseño	Árbitros
Portada		Elisa Schaeffer
		David J. Rios
		Tania Turrubiates-López
		Sandra Nava-Muñoz
		María Lucila Morales Rodríguez
		Héctor Hugo Avilés Arriaga
		Juan H. Sossa Azuela
		Ofelia Cervantes
		Sara Elena Garza Villarreal
		Carlos Soubervielle Montalvo
		César Torres



Contenido

ARTÍCULO ACEPTADO

Producción artificial de habla: Síntesis de voz

por Marvin Coto Jiménez, John Goddard-Close y Fabiola Martínez-Licona

pág. 6 ⇒ El habla: La nueva interacción entre el ser humano y las máquinas.

ARTÍCULO ACEPTADO

Controlando una silla de ruedas mediante expresiones faciales y movimientos de cabeza

por Ericka Janet Rechy-Ramírez, Huosheng Hu, Antonio Marín-Hernández y Homero Vladimir Ríos-Figueroa

pág. 11 ⇒ La silla de ruedas controlada en modo manos libres es una tecnología que facilita el desplazamiento para personas con movilidad limitada.

ARTÍCULO ACEPTADO

El arte de diseñar robots manipuladores

por Dan-El N. Vila-Rosado y Axel Domínguez-López

pág. 17 ⇒ El buen diseño de un brazo manipulador es cuestión de conocer previamente su objetivo y su área de acción.

ARTÍCULO ACEPTADO

Uso de tangibles robóticas en el aprendizaje colaborativo de las matemáticas

por Sergio Hernández-González, Genaro Rebolledo-Méndez, N. Sofía Huerta-Pacheco y Luis G. Montané-Jiménez

pág. 24 ⇒ En México un campo poco explorado es el empleo de tecnologías de la información como apoyo en el aprendizaje significativo y a largo plazo.

Columnas

Sapiens Piensa. Editorial [pág. 2](#)

e-Tlakuilo [pág. 3](#)

Estado del IArte [pág. 5](#)

Sakbe [pág. 4](#)

IA & Educación [pág. 29](#)

Deskubriendo
Konocimiento [pág. 31](#)

ARTÍCULO ACEPTADO

Producción artificial de habla: Síntesis de voz

Marvin Coto Jiménez, John Goddard-Close y Fabiola Martínez-Licona

Introducción

A finales de la década de los 80, el estadounidense Raymond Kurzweil, inventor del primer sistema comercial que es capaz de tomar texto de un libro y pronunciarlo mediante habla artificial, escribió la obra "La era de las máquinas inteligentes" [1]. En esta realiza una serie de predicciones sobre lo que esperaba del desarrollo tecnológico en las siguientes décadas. Entre sus pronósticos se encuentra que para principios de la primer década del año 2000, iba a ser posible sostener conversaciones con otras personas que hablaran un idioma distinto por medio del teléfono, el cual tendría la capacidad de traducir automáticamente y pronunciar el mensaje en el otro idioma. En el presente año se ha anunciado una versión beta (que estará disponible en los próximos meses) de un sistema que promete iniciar esta posibilidad de comunicación. Previamente, esfuerzos como el de la Quinta generación japonesa [2] han propuesto esta posibilidad de comunicación sin obtener resultados exitosos.

Kurzweil, en una obra posterior, "La era de las máquinas espirituales", [3] predice que para el 2019 las personas se comunicarán con las máquinas usando habla en ambos sentidos, gracias a asistentes virtuales con personalidades que podrán ser seleccionadas o ajustadas según el gusto del usuario. Así, la relación de personas con dispositivos como computadoras y robots se parecería cada vez más a la forma de relacionarse entre seres humanos.

Si bien, algunas de sus predicciones en otros aspectos se han acercado mucho a la realidad, y la tecnología hoy en día parece capaz de llevarla a cabo, en el tema de la interacción con las computadoras por medio del habla, ha parecido mostrar un optimismo en sus obras que no ha sido, y posiblemente no será, satisfecho con las características que esperaba. No sería el primer autor en realizar predicciones en el tema de habla artificial que no han sido satisfechas pasado el año pronosticado.

Podemos atribuir estos vaticinios no consumados a dos razones principales:

- La enorme complejidad del habla humana, producto de la multitud de órganos y fenómenos que se dan en el aparato fonador, además de razones evolutivas y culturales que han producido nuestra forma de hablar tal como la emitimos y comprendemos en la actualidad.
- La sensibilidad humana en la percepción de la voz, donde pequeñas desviaciones de lo que percibimos como usual puede notarse como una gran diferencia. Por ejemplo, si la conocida prueba de Turing fuera desarrollada mediante conversaciones habladas, en lugar de mensajes de texto, seguramente en la actualidad sentiríamos estar más lejos de lograr el desarrollo de sistemas que pasen la prueba.

Es posible comparar las expectativas que ha generado la producción de robots y computadoras que puedan comunicarse de forma natural con seres humanos utilizando voces artificiales, con las generadas alrededor de las primeras computadoras que jugaban ajedrez, que parecían muy lejos del nivel de los campeones mundiales de aquella época. Aún en la década de los 80, algunos expertos consideraban que las máquinas jamás podrían vencer a los mejores expertos humanos en ajedrez [4]. Sin embargo hoy, 60 años después de los primeros programas que fueron capaces de jugar partidas completas, se considera que las computadoras han superado las capacidades humanas en este juego [5]. Para el caso de imitar las capacidades humanas de hablar, el periodo de espera se está haciendo mucho más largo.

En este trabajo se abordan las técnicas principales que han llegado a nuestros días pretendiendo resolver el problema de producir habla artificial, así como las perspectivas a futuro reciente que puede tener esta tecnología.

Las dificultades para producir habla de forma artificial han resultado inesperadas, tomando en cuenta las expectativas generadas décadas atrás

Síntesis de voz

El objetivo de la síntesis de voz se puede establecer como el convertir un texto cualquiera en una señal acústica que sea indistinguible del habla humana. Este texto puede provenir de una computadora o del escaneo de un

texto, tal como el mencionado sistema desarrollado por Kurzweil en la década de los 70.

En la actualidad, quizás la mayoría de personas habrán tenido algún contacto con voces artificiales, por la interacción con teléfonos inteligentes que han implemen-

tado aplicaciones con esta tecnología, la proliferación de videos en Internet narrados con voces artificiales (usualmente para mantener el anonimato de quien lo produce), o bien escuchando al científico Stephen Hawking, quien se comunica utilizando un sintetizador de voz (Figura 1). La ciencia ficción también ha puesto en la mente de muchas personas lo que podría ser una interacción con computadoras utilizando puramente lenguaje natural, como en las películas *2001: Space Odyssey* de Stanley Kubrick (1968), o más recientemente *Her* de Spike Jonze (2013).



Figura 1. Stephen Hawking con el dispositivo que le permite comunicarse utilizando síntesis de voz. Fotografía: NASA/Paul Alers.

Existen antecedentes de la reproducción de sonidos humanos por medios mecánicos desde el siglo XVIII, con los sistemas de Christian Kratzenstein y Wolfgang von Kempelen, los cuales consistían en un conjunto de fuelles y dispositivos que modificaban los caminos del aire, emulando la producción de sonidos en el aparato fonador. En el siglo XX se inicia la producción de habla artificial utilizando dispositivos electrónicos, dentro de lo que destacó como primer avance el VODER, presentado por Homer Dudley en la Feria Mundial de 1939[6], del cual se encuentran grabaciones de audio en Internet.

Unos años después, la Compañía Telefónica del Reino Unido, presentó el primer dispositivo comercial que hizo uso de una técnica de síntesis de voz. La finalidad de éste era pronunciar la hora, al unir fragmentos de la pronunciación de la hora, minutos y segundos con grabaciones distribuidas entre cuatro discos de registro.

En el camino hasta nuestros días se ha pasado por una larga serie de dispositivos y técnicas, con diversas aspiraciones sobre lo que la síntesis de voz puede aportar. En nuestros días visualizamos muchas aplicaciones de esta tecnología, y podrían ser aún más si los resultados sonoros fueran más satisfactorios dentro de las necesidades de los usuarios. Además del mencionado sistema creado por Raymond Kurzweil de asistencia para personas con discapacidad visual, aplicaciones actuales como contestadoras en centrales telefónicas, sistemas de navegación GPS,

robots utilizados en investigación, y asistentes virtuales en teléfonos celulares o tabletas van formando parte de la interacción cotidiana de las personas con la tecnología.

Sin embargo, existen muchas otras posibilidades de aplicación si la voz obtenida por una computadora dentro de un robot o dispositivo pudiera ser indistinguible de una voz humana y estar dotada de mayor flexibilidad, por ejemplo [7-8]:

- Sistemas de entretenimiento, donde se espera que voces de alta calidad puedan despertar imágenes o sensaciones semejantes a la música.
- Asistentes virtuales personalizables (como los descritos por Kurzweil), en los cuales el género de la voz, el estilo y la emoción puedan ser modificados a gusto del usuario o adaptarse de forma dinámica de acuerdo con el contexto de la conversación o información que está brindando.
- Aprendizaje de nuevos idiomas, pues un sistema de síntesis de voz puede repetir sin descanso y con precisión la pronunciación de lo que se desee en un idioma que se esté aprendiendo.
- Reconstrucción de voces, para personas con problemas degenerativos del habla, como el mencionado caso de Stephen Hawking.

Cualquier aplicación que contemple interacción humano-computadora puede establecerse como área potencial de interés para implementar síntesis de voz, incluyendo dispositivos portátiles inteligentes y la robótica.

Requerimientos

De manera semejante al proceso que realizamos las personas al leer un texto en voz alta, un sistema computacional de síntesis de voz debe contener varios procedimientos, especialmente porque la manera en que escribimos en la mayoría de los idiomas no tiene una correspondencia perfecta con la pronunciación del texto. Es decir, no basta conocer el nombre y sonido de cada letra aislada, pues existen variantes importantes, en mayor o menor grado de cada idioma. En el español, por ejemplo, la letra “h” en determinadas posiciones no representa ningún sonido (como al principio de una palabra), pero lo adquiere si se encuentra después de la “c”. Otras excepciones importantes son la combinación de la letra “g” o la “q” con la “u”.

Por otra parte, se debe prestar atención a aspectos prosódicos, relacionados con las características de energía, entonación y duración que producimos a lo largo de las frases, gracias a las cuales distinguimos si se trata de una pregunta, una exclamación o se ha llegado a una pausa ortográfica, como un punto. En la emisión del habla humana con estas características intervienen una

gran cantidad de órganos del aparato fonador, además de aspectos evolutivos y culturales, lo que hace que el proceso y su emulación artificial sean muy complejos. Tanto los aspectos fonéticos como prosódicos se incluyen en lo que se llama la especificación lingüística de un texto por pronunciar.

Ya sea por medio de modelos matemáticos o algoritmos que tomen en cuenta los aspectos mencionados, un sistema de síntesis de voz debe tener necesariamente como paso final una forma de generar audio en correspondencia con el texto que se está pronunciando. La Figura 2 ilustra el proceso general descrito, que constituiría un

sistema texto a voz (TTS por las siglas en inglés de Text to Speech).

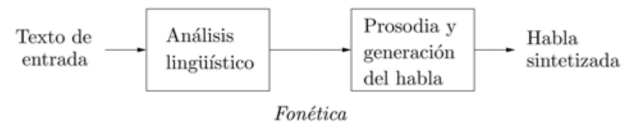


Figura 2. Esquema general de un sistema de transcripción de texto a habla.

Se ha presentado una revisión a las principales tecnologías utilizadas en la actualidad para producir voz de forma artificial

Los objetivos que podemos mencionar que se persiguen en la actualidad son la flexibilidad y la naturalidad, es decir, un parecido suficiente con las características humanas que permitan extender la utilización a más usuarios y las posibilidades de más aplicaciones. En las siguientes secciones se describirán las dos técnicas principales que han llegado a nuestros días como dominantes en implementaciones e investigación: la síntesis basada en concatenación de unidades, y la síntesis estadística paramétrica.

Síntesis basada en concatenación de unidades

Este tipo de síntesis es considerada como la más sencilla de producir, más comprensible y usualmente la de sonido más natural [9]. Parte de un conjunto grande de grabaciones de habla real, de la cual selecciona y concatena los segmentos más adecuados para generar las nuevas frases. Estos segmentos pueden ser de distinto tamaño, como fonemas, sílabas o incluso palabras.

La selección de estos segmentos depende de las características del idioma que se esté tratando, y el determinar cuáles son los óptimos es uno de los aspectos principales que pueden determinar los mejores resultados. Teniendo varias opciones en la base de datos, los algoritmos actuales más utilizados seleccionan entre aquellas unidades que resulten mejor evaluadas en dos funciones calculadas con determinados pesos asignados [10]. La Figura 3 ilustra la concatenación de unidades seleccionadas entre muchas opciones para producir una palabra.

A pesar de la gran calidad lograda en las voces artificiales producidas de esta manera en términos de su naturalidad, uno de sus mayores inconvenientes es su alto requerimiento de grabaciones para generarla. Para obtener resultados semejantes a los mejores sistemas actuales se requieren decenas o hasta centenas de horas de habla grabada con condiciones adecuadas, lo cual acarrea costos de producción considerables.

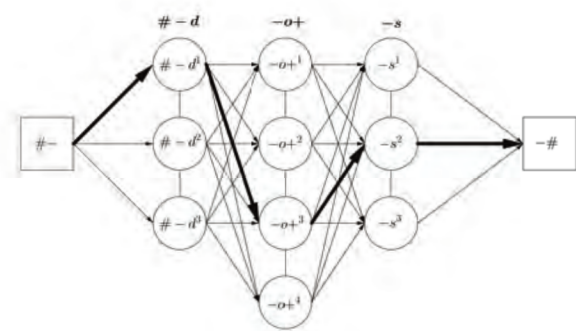


Figura 3. Ejemplo de selección de unidades de habla para sintetizar la palabra “dos” en la síntesis basada en concatenación. El símbolo # representa un silencio. Adaptada de [10].

Síntesis estadística paramétrica

A partir de la primer década del año 2000, surge el método llamado síntesis estadística paramétrica de voz [11], también llamada síntesis basada en Modelos Ocultos de Markov (HMM) por ser este el modelo matemático principal que la sustenta. A diferencia de la síntesis concatenativa, este método no manipula directamente grabaciones de audio, sino que extrae en primer lugar una representación paramétrica de este (es decir, un conjunto de parámetros) con la cual entrena modelos matemáticos que puedan aprender a reproducirlos cuando se requiere sintetizar nuevas frases.

Este método contempla los siguientes pasos [12]:

- Partir de una base de datos de audio y su correspondiente transcripción textual.
- Extraer del audio de habla una representación paramétrica, que incluya coeficientes espectrales, de tono y duración, conformando vectores característicos que representan segmentos de audio.

- Realizar una identificación de los vectores que representan cada fonema, para definir un HMM por cada uno de estos que tendrá en correspondencia un conjunto de vectores.
- Entrenar cada HMM con los vectores que le corresponden, de manera que pueda reproducirlos.
- En la parte de síntesis, utilizar los HMM de forma generativa, para extraer la representación paramétrica de las nuevas frases que se desean sintetizar y generar el audio a partir de los parámetros con un filtro.

Como se puede observar, los procedimientos y requerimientos son más complejos que en el método descrito previamente, lo cual contribuye a hacer de la síntesis estadística paramétrica una técnica de difícil implementación. Sin embargo, tiene ventajas importantes, por ejemplo, el no tener que referirse directamente al audio en cada emisión permite que las voces pasen de requerir varios Gigabytes de almacenamiento y procesamiento a unos pocos Megabytes, lo cual puede hacerla más apta para aplicaciones móviles.

Por otra parte, el tener toda la información en términos de coeficientes hace teóricamente posible modificar cualquier característica de la voz (desde su género, acento, estilo) si se logra descifrar la manera de hacerlo transformando estos coeficientes. Esto permite una gran flexibilidad, y crear por ejemplo, voces promedio entre varias personas, incluso de distintos géneros o edades.

La principal desventaja es que aún no se alcanza la calidad de voz artificial de los mejores métodos basados en concatenación de audio. Sin embargo, una intensa investigación está siendo realizada para buscar mejoras en todas las partes del proceso.

La evaluación de voces artificiales

Un aspecto fundamental en el tema de síntesis de voz es su evaluación, es decir, contar con métodos que permitan discriminar entre aquellas voces que tengan mayor calidad de aquellas que resulten menos convenientes dentro de las aplicaciones en que se espera implementar. El contar con evaluaciones de calidad tiene ventajas importantes tanto para desarrolladores de sistemas como para usuarios finales [13].

No existe un consenso en la actualidad sobre cuáles sean las metodologías y métricas más convenientes para establecer la mayor o menor calidad de una voz artificial. Con el fin de promover el desarrollo de la síntesis de voz a partir de datos y evaluaciones estandarizadas, el Blizzard Challenge, competencia convocada anualmente, se ha constituido como un espacio destacado de muestra e intercambio científico para promover el desarrollo y estandarización de métodos de evaluación de nuevas propuestas.

Se trata de una convocatoria abierta para desarrollar voces a partir de una base de datos común, con métodos de evaluación también comunes a todas las voces que se desarrollan. Por esta razón, sus resultados no evalúan voces ya establecidas como las de AT&T Natural Voices, CereProc, Loquendo, IVONA entre otras desarrolladas con fines comerciales. Han sido usuales en las distintas convocatorias anuales sintetizadores que utilizan selección de unidades, así como métodos estadísticos paramétricos e incluso híbridos de ambos, los cuales pretenden aprovechar lo mejor de cada técnica.

Desarrollo futuro

Como se ha mencionado, el uso más extendido de la síntesis de voz depende de la calidad que pueda lograrse en sus resultados, así como su flexibilidad. El aporte de la síntesis estadística paramétrica y su potencial mayor control sobre las voces hace posible imaginar a un mediano plazo aplicaciones donde pueda cambiarse el estilo de habla y personalidad, tal como imaginó Kurzweil para el año 2019. Sin embargo, el avance en el tema sigue presentando dificultades importantes, aunado al hecho de requerir especialistas de muy diversas áreas, como la lingüística, el procesamiento digital de señales, los sistemas inteligentes y la matemática.

Los desarrollos más importantes en el tema del habla artificial han sido originados en Japón, Inglaterra, China y Estados Unidos, por lo que es en los idiomas que hablan estos países donde se pueden escuchar los mejores resultados hasta el momento. En el caso de América Latina, es incipiente el desarrollo de voces artificiales. Por ejemplo, si bien los sistemas estadísticos paramétricos tienen quince años de desarrollo, en el año 2013 se iniciaron las primeras experiencias reportadas con esta técnica en esta región, en México [14,15].

Un mayor interés en el área podría facilitar que nuestros dispositivos electrónicos, todo tipo de sistemas inteligentes y robots puedan comunicarse a futuro utilizando voces que nos sean más cercanas y comprensibles, incluso reconstruyendo una voz cualquiera de personajes famosos, de familiares o incluso la propia. Además de esta deseada propiedad, hay mucho trabajo pendiente para incorporar lenguajes autóctonos latinoamericanos en sintetizadores, y contribuir así a su estudio y rescate.

Conclusiones

Se ha presentado una visión general sobre la síntesis de voz y las técnicas dominantes en la actualidad para llevarla a cabo. A pesar de las intensas investigaciones llevadas a cabo en muchas partes del mundo, las predicciones realizadas sobre su avance han sido frecuentemente sobre estimadas con respecto a los logros. Esto se debe en gran parte a la complejidad del habla humana, además de la conjugación de varias áreas que deben involucrarse en el desarrollo de voces artificiales.

Hay una vasta área de investigación abierta, y en América Latina tenemos retos importantes para obtener voces en las variantes del español que manejamos, además de las lenguas autóctonas de esta región. De esta manera, nuestros avances en aplicaciones, robótica y sistemas inteligentes podrán incorporar el medio de interacción entre seres humanos y tecnología que parece ser la tendencia a futuro: el habla. Esto con estilos y variantes que nos sean más convenientes y cercanos, para lograr a futuro que esta interacción sea la forma más natural de comunicarnos con la tecnología que utilizamos, tal como es la forma más natural y representativa de comunicación entre seres humanos.*

REFERENCIAS

1. Kurzweil R. (1990) "The age of intelligent machines". Vol. 579. Cambridge: MIT press.
2. Feigenbaum E.A. y McCorduck P. (1993) "The fifth generation". Addison-Wesley Pub.
3. Kurzweil R. (1999) "The Age of Spiritual Machines: When Computers Exceed Human Intelligence". New York, NY: Penguin Books.
4. Stinson C. (1982) "Chess Championship: Machines Play, People Watch". *Softline*. p. 6. Recuperado el 20 de septiembre de 2014, de: <http://www.cgwmuseum.org/galleries/index.php?year=1982&pub=6&id=3>.
5. Rohrer F. (2014) "The unwinnable game". *BBC News Magazine*. Recuperado el 21 de septiembre de 2014 de: <http://www.bbc.com/news/magazine-25032298>.
6. Cunningham C. (2009) "Speech Synthesis in the year 1939". Recuperado el 25 de septiembre de 2014, de: <http://makezine.com/2009/02/04/speech-synthesis-in-the-year-1939/>.
7. Bailly G., Campbell N. y Möbius B. (2003) "ISCA special session: hot topics in speech synthesis". En *INTERSPEECH*.
8. Keller E. y Zellner-Keller B. (2000) "New uses for speech synthesis". *The Phonetician*, Vol. 81, pp. 35-40.
9. Narendra N.P. y Rao K.S. (2013) "Optimal weight tuning method for unit selection cost functions in syllable based text-to-speech synthesis". *Applied Soft Computing*, Vol. 13, No.2, pp. 773-781.
10. Schroeter J. (2004) "Fifty years of progress in speech synthesis". *The Journal of the Acoustical Society of America*, Vol. 116, No.4, pp. 2497-2497.
11. Black A.W., Zen H. y Tokuda K. (2007) "Statistical parametric speech synthesis". En *Acoustics, Speech and Signal Processing, IEEE International Conference*, Vol. 4, pp. IV-1229).
12. Zen H., Tokuda K. y Black A.W. (2009) "Statistical parametric speech synthesis". *Speech Communication*, Vol. 51, No. 11, pp. 1039-1064.
13. Van Heuven V.J.J.P. y Van Bezooijen R.L. (1995) "Quality evaluation of synthesized speech".
14. Camacho A.H. y Ávila F.D.R. (2013) "Development of a Mexican Spanish Synthetic Voice Using Synthesizer Modules of Festival Speech and HTS-Straight". *International Journal of Computer and Electrical Engineering*, Vol. 5, No.1, p. 36.
15. Coto-Jiménez M. (2014) "Síntesis estadística paramétrica de voz". Tesis de Maestría. Universidad Autónoma Metropolitana, México.

SOBRE LOS AUTORES



Marvin Coto-Jiménez es Ingeniero Electricista y Maestro en Matemática por la Universidad de Costa Rica. Actualmente es estudiante del programa de doctorado en Ciencias y Tecnologías de la Información en la Universidad Autónoma Metropolitana, Unidad Iztapalapa, en México D.F.



John C.H. Goddard recibió una B.Sc (1st Class Hons) de la Universidad de Londres, y un Ph.D en Matemáticas de la Universidad de Cambridge. Es Profesor en el Departamento de Ingeniería Eléctrica de la Universidad Autónoma Metropolitana. Sus intereses incluyen reconocimiento de patrones, reconocimiento de emociones en habla y síntesis del habla.



Fabiola Martínez-Licona estudió la Licenciatura y la Maestría en Ingeniería Biomédica en la UAM-Iztapalapa. Como profesora de la UAM, fue Coordinadora de la Licenciatura en Ingeniería Biomédica y ha impartido cursos en todos los programas académicos de licenciatura y posgrado del Departamento de Ingeniería Eléctrica. Ha dirigido proyectos relacionados con la Gestión de Tecnologías Médicas, Ingeniería Clínica, Reconocimiento de Patrones, Procesamiento de Señales y Aprendizaje Maquinal. Actualmente es responsable del proyecto línea de investigación de Análisis y Aplicaciones del Habla, investigación que incluye reconocimiento, síntesis y análisis de emociones en español, análisis y síntesis de lenguas en peligro de extinción y reconocimiento de música.