

*Desde hace varios años, un grupo multidisciplinario de profesores de la ULA, formado por ingenieros y lingüistas, viene trabajando en las Tecnologías del Habla, campo de investigación y aplicación dedicada a la comunicación oral hombre-máquina.*

## Beneficios directos para minusválidos visuales y auditivos

# Síntesis de habla en español venezolano: usos prácticos y de investigación

*Elsa Mora\* y Manuel Rodríguez\*\**

Recientemente, este grupo logró implementar un sistema de conversión texto a voz de alta calidad para el español, capaz de convertir cualquier texto en formato electrónico a voz sintetizada en español venezolano. La capacidad de lectura se hace universal al combinarlo con un sistema OCR (*optical character recognition*) de escáner con conversión de caracteres a texto.

Dentro de sus múltiples usos este sistema tiene una utilidad social inmediata para quienes sufren una minusvalía que les impida la comunicación oral, ya que permite por ejemplo darle capacidad de «lectura» a los ciegos e incapacitados visuales y darle «voz» a través de un computador a personas sordomudas, mudas o con incapacidades mayores que restringen la producción de voz. Por otra parte, el sistema es una

herramienta de investigación muy poderosa para estudiar la entonación de nuestra lengua, una materia que apenas empieza a desenmascarar sus secretos.

Los profesores de la ULA más directamente involucrados en el desarrollo de este sistema de síntesis de voz son la lingüista Elsa Mora de la Facultad de Humanidades y Educación y el ingeniero electrónico Manuel Rodríguez de la Facultad de Ingeniería.

### Antecedentes

El profesor Manuel Rodríguez tuvo sus primeras experiencias con síntesis de voz en 1983, durante el año sabático que realizó en la Universidad Politécnica de Madrid. En esa época, la técnica utilizada era Síntesis por Formantes, por

medio del sintetizador *Klatt*, que sintetizaba fonema por fonema los sonidos del habla en base a una descripción de los parámetros acústicos de estos sonidos, específicamente los formantes, que son las resonancias naturales del tracto vocal. Con esta técnica salieron al mercado los primeros sistemas comerciales, sobre todo el *DecTalk* de la *Digital Equipment Corporation*.

El grupo de la Universidad Politécnica de Madrid, bajo la coordinación del Jefe de Departamento de Electrónica, Elías Muñoz, llevó a cabo los análisis extensos de señal de voz para extraer los datos acústicos de la versión española de este sistema de síntesis. Aunque este sistema fue un gran logro para la época, había problemas de inteligibilidad, atribuibles a deficiencias de la codificación de la voz con la técnica LPC (*Linear Predictive Coding*), y por la necesidad de rellenar matemáticamente la transición de los formantes entre los fonemas.

Por su parte, la profesora Elsa Mora realiza sus estudios doctorales de 1993 a 1996 en la *Université de Provence* en Francia y regresa a Mérida consciente de la importancia de los sintetizadores de voz para la investigación lingüística. De allí que se propone un convenio EcosNord - CONICIT (hoy FONACIT) entre la *Université de Provence* y la ULA, en Francia representado por Daniel Hirst y Christian Cavé, y en Mérida, por Elsa Mora y Manuel Rodríguez, teniendo como meta el desarrollo de un sistema de síntesis de voz en la variedad del español venezolano, basado en una técnica que el grupo de la *Université de Provence* aplicó exitosamente al francés, que es el sistema *mbrola* de concatenación de difonos con procesado de señal.

### Aspectos técnicos

**¿Qué es un difono?** Un difono representa el sonido que abarca desde la mitad de la realización de un fonema hasta la mitad de la realización del fonema siguiente. El propósito de esta unidad de sonido es incorporar a la unidad de síntesis la transición de sonido entre fonemas, hecho que había causado tanta dificultad en los sistemas iniciales.

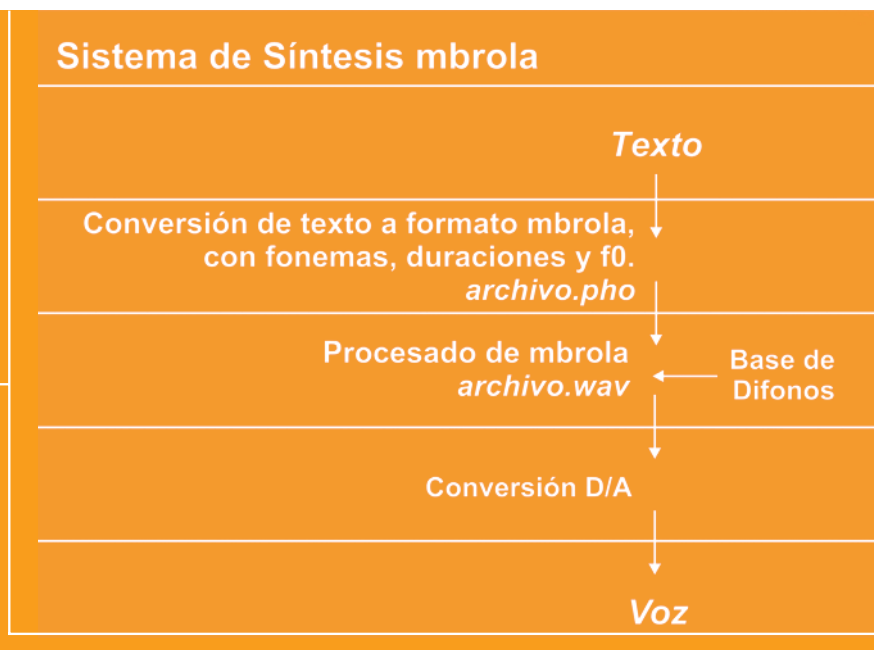
La síntesis consiste, entonces, en la concatenación de segmentos de señal en el tiempo, siendo los segmentos difonos. Esta técnica de concatenación de difonos se vino perfeccionando durante los años 80 y 90 hasta lograr una calidad excelente con la propuesta de Thierry Dutoit, de la *Faculté Polytechnique de Mons* en Bélgica, con la técnica llamada *mbrola*. Para respetar patrones de prosodia naturales, los segmentos de señal son alterados en su duración y en su frecuencia fundamental ( $f_0$ ) (entonación) con el procesamiento de la señal conocido como PSOLA (*Pitch Synchronous Overlap and Add*), que logra estas transformaciones manteniendo otros aspectos del espectro de la señal sin desmejorar la calidad del sonido.

### Desarrollo del sistema

La primera tarea cumplida por el convenio *Université de Provence* - ULA fue la creación de una base de difonos para el dialecto venezolano. Se determinó que en el habla venezolana hay 34 fonos, incluyendo pausa, de modo que el universo de difonos venezolanos sería en teoría  $34 \times 34 = 1156$ , pero eliminando las combinaciones imposibles o inexistentes, se redujo el conjunto a 793 difonos. Se buscó luego un venezolano con excelente pronunciación (el profesor Enrique Obediente de la Facultad de Humanidades y Educación) y se grabó su voz pronunciando 793 frases, procurando que cada difono del conjunto se pronunciara con la misma intensidad y una entonación neutra; estas precauciones auguran



Figura 1. Diagrama en bloques del sistema de síntesis mbrola. En el centro y a la izquierda se hace notar el bloque correspondiente a la base de datos de difonos.



mejor calidad en el procesado de señal. La segmentación de los difonos fue realizada fundamentalmente por estudiantes de doctorado en Fonética Experimental Funcional y Aplicada de la *Université de Provence*, y del Doctorado en Lingüística de nuestra Universidad.

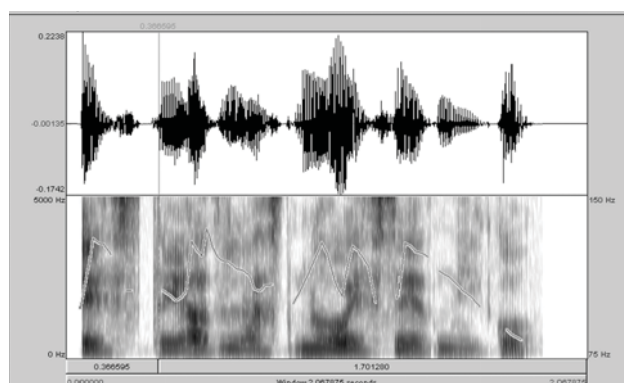
Después de segmentar los difonos se enviaron al centro de administración de la técnica *mbrola* en Bélgica, donde empaquetan la base de datos en el formato adecuado. Hoy en día esta base de datos está disponible al mundo entero, junto con las bases de datos de otros 33 dialectos en varias lenguas, a través de la página web de *mbrola* (<http://tcts.fpms.ac.be/synthesis/mbrola.html>).

Este centro, a su vez, facilita el juego de programas *mbrola* de procesado de señal que, en combinación con la base de datos de difonos, permite sintetizar una frase.

En la Figura 1 se muestra un diagrama en bloques del sistema de síntesis *mbrola*. En el centro y a la izquierda se hace notar el bloque correspondiente a la base de datos de difonos. En el centro está el bloque *mbrola* encargado de realizar el procesado de señal. En la parte superior hay un bloque que tiene como entrada un texto y que genera como salida un archivo .pho, el cual contiene la información de los fonemas a sintetizar, junto con los valores de los parámetros de duración y de entonación.

Ahora bien, para sintetizar una frase, se debe armar este archivo .pho donde se incorpora toda la información necesaria. Este archivo .pho puede armarse manualmente, con cualquier programa editor de texto. El bloque central *mbrola* se alimenta de este archivo, en función del cual llama a los segmentos de señal requeridos y almacenados en la base de datos de difonos, transforma, luego, los segmentos para darles la duración y la f0 idóneas para la frase y, partiendo de cero, los va concatenando en la señal de voz creando así un archivo de voz, hasta terminar la frase. Finalmente, el bloque inferior representa la reproducción de la señal por medio de una corneta de salida del computador.

Figura 2, se muestra la forma de onda sintetizada de la frase «*Este atlas no es étnico*», y abajo el espectrograma y la función de f0 correspondiente.



## Procesado automático

El último logro del grupo *Université de Provence* - ULA fue automatizar el proceso de transformación del texto de entrada en el formato del archivo de fonemas .pho, por medio de un programa llamado tafv.pl (Texto a Fonema Venezolano). El programa está escrito en el lenguaje de programación PERL, un lenguaje particularmente apto para procesado de texto. Igualmente es capaz de funcionar en el sistema operativo Linux, Windows 98, Windows NT y Windows XP.

El programa ejecuta una cantidad de funciones tales como transcripción ortográfica-fonética, segmentación en sílabas, colocación de acento prosódico, asignación de duración, asignación de entonación simbólica de acuerdo al sistema de transcripción ToBI (Tones and Break Indices), asignación de entonación numérica, manejo de abreviaciones, siglas, acrónimos, números enteros, decimales y romanos.

Hoy en día es capaz de manejar cualquier artículo de prensa de una página Web, que por medio del ratón se copia y se pega a la entrada del programa tafv.pl, que procede a generar automáticamente el archivo .pho. Este programa también puede obtenerse de la página web de *mbrola*.

## Aplicaciones del sistema

En este momento los miembros del grupo *Université de Provence* - ULA están concentrados en promover el conocimiento y el uso del sistema, sobre todo para personas con alguna minusvalía que impida la comunicación verbal. En la Facultad de Humanidades y Educación se está promocionando su uso entre el grupo de personas que esperan ingresar a la Licenciatura para Sordos. También se ha llevado al Centro de Atención Integral de Ciegos y Deficientes Visuales (CAIDV) de Mérida, quienes lo han recibido con entusiasmo, pues este sistema vendría a reemplazar productos similares provenientes de España, suplantando así un acento peninsular con un acento venezolano, de mayor aceptación entre la población local. De igual forma, se piensa extender entre la población de nuestro país con problemas de visión y de audición. Una segunda parte prevista es hacer las adaptaciones necesarias para minusvalías más severas.

Finalmente, el programa también permitirá realizar trabajos de investigación en el estudio de la prosodia de nuestra lengua, incluso ya se está trabajando sobre el programa, por parte de algunas estudiantes tanto del Postgrado de Computación, como del postgrado de Lingüística.

*\*Profesora de la Fac.de Humanidades y Educación.*

*Grupo de Investigaciones en Ciencias Fónicas.*

*E-mail: elsamora@ula.ve*

*\*\*Profesor de la Fac. de Ingeniería.*

*E-mail: hourcade@ula.ve*



*En este momento los miembros del grupo están concentrados en promover el conocimiento y el uso del sistema, sobre todo para personas con alguna minusvalía que impida la comunicación verbal.*