

ASELA: Análisis, Síntesis y Evaluación del Lenguaje y la
Audición.

Victoria Marrero^{*}, Andrés de Santos⁺, Santiago Aguilera⁺

^{*} Dpto Lengua Española
Facultad de Filología. U.N.E.D.
⁺ Dpto. Ingeniería Electrónica.
E.T.S.I.Telecomunicación. U.P.M.

El equipo que presenta este trabajo es un grupo interdisciplinar constituido, fundamentalmente, por ingenieros expertos en proceso digital del habla (Departamento de Ingeniería Electrónica de la U.P.M.) y lingüistas especializados en fonética y fonología (Departamento de Lengua Española de la U.N.E.D.), en colaboración con audiólogos (Hospital Ruber Internacional) y logopedas.

Nuestros estudios se han centrado en varios temas que comparten un objetivo prioritario: su utilidad para el tratamiento y ayuda a discapacitados visuales y auditivos.

De los trabajos presentados aquí, los dos primeros (*Conversión Texto-Habla* -con atención a las aplicaciones para invidentes- y *Sistemas para la Representación y el Análisis del Lenguaje Natural* -ayudas para sordos-) son implementables en una única arquitectura hardware adaptable a cualquier PC, y constituyen el proyecto SIPHAD (*Sistema Integral de Procesamiento del Habla para Ayuda a Discapacitados*). El tercero, (*TRD, Test de*

Rasgos Distintivos) está encaminado a proporcionar un conjunto de pruebas para la evaluación del lenguaje sintético, por un lado, y como complemento a las pruebas clínicas de audiometría verbal por otro.

1. Conversión Texto-Habla

Desde el año 1981 hemos trabajado en la elaboración de un conversor texto-habla en tiempo real para el castellano.

Se ha incidido especialmente en tres áreas:

- estudio de los rasgos prosódicos, fonológicos y segmentales de la lengua hablada (duraciones, entonación, transición de formantes, además de la caracterización general del sistema fónico);

- programación de algoritmos que realicen el proceso completo de conversión texto-habla (normalización del texto, conversión de letras en sonidos, asignación de patrones prosódicos, obtención de parámetros y síntesis del habla);

- diseño de sistemas hardware que soporten dichos algoritmos en tiempo real; entre otros se ha desarrollado una placa con un procesador de señales preparada para su conexión a un ordenador personal. Esta placa soporta además otras aplicaciones que se describirán más adelante.

Actualmente estamos trabajando en la mejora de la calidad del habla producida, así como en añadir nuevas posibilidades al sistema.

Según las pruebas de inteligibilidad realizadas, algunos

sonidos presentaban ciertas deficiencias debidas, entre otros factores, a la dirección y velocidad de las transiciones, distribución de áreas de frecuencia, asignación de intensidad, etc. Una vez analizados estos elementos, se está procediendo a su reajuste.

Otro factor que estamos mejorando es el modelo que asigna la duración a cada alófono, especialmente considerando diversas velocidades de producción de habla (*tempo* medio, rápido o lento).

En el momento actual llevamos a cabo también la modificación de determinadas características del sintetizador para permitir la simulación de distintos hablantes, utilizando para ello, tanto modificaciones en la propia estructura del sintetizador (en el modelo de excitación glotal) como cambios en la prosodia y en la caracterización de sonidos.

Entre las nuevas posibilidades del sistema, contamos también con su conexión a diversos programas de uso común (procesadores de textos, bases de datos, etc.), con el fin de proporcionarles una salida hablada, de gran utilidad para deficientes visuales. Las mejoras introducidas (variaciones del tipo de voz y de su velocidad de producción, mejora en la calidad del habla) contribuirán a dar más utilidad y mayores prestaciones a estas aplicaciones.

2. Sistemas para la representación y análisis de lenguaje natural.

Nuestro grupo dispone hoy en día de un sistema potente y versátil para la detección, análisis y posterior representación de diversos parámetros de la voz.

Sobre una única placa hardware, que se conecta al bus de un ordenador personal y se configura por software desde éste, disponemos de dos tipos fundamentales de aplicaciones, además de la ya mencionada de conversión texto-habla:

- Analizador de voz (PCvox). Nos proporciona una representación gráfica en color de los parámetros más representativos de la voz: forma de onda, energía, entonación, cruces por cero y sonogramas; sobre ellos se pueden llevar a cabo numerosas operaciones: segmentación temporal (con posibilidad de escuchar sólo el trozo segmentado, e incluso edición de voz), segmentación frecuencial, inserción de marcas (para la transcripción fónica), modificación del filtrado, modificación de la intensidad de entrada, etc.

El PCvox ofrece las prestaciones más importantes de los sonógrafos tradicionales, pero las integra en un sistema muy manejable, multifuncional y más asequible, por lo que comienza a ser solicitado desde diferentes campos técnicos y profesionales.

- Sistema para el entrenamiento de parámetros suprasegmentales del habla (ISOTON). Sistemas de Ayuda a Sordos

El sistema ISOTON, (acrónimo de *Intensidad, Sonoridad y Tono*, parámetros que permite entrenar), está constituido por un módulo de proceso digital de señal, que extrae estos elementos a partir

de la señal de voz suministrada desde micrófono, y por un programa que los procesa y presenta sobre la pantalla de un ordenador personal.

Es un sistema destinado al entrenamiento del lenguaje que utiliza realimentación visual, permitiendo la retención y superposición de imágenes para analizar con posterioridad los resultados de la pronunciación y compararlos con patrones previamente establecidos. También dispone de diversos videojuegos controlables por la voz.

Permite tres modos de funcionamiento, en cada uno de los cuales se pueden entrenar distintos parámetros del habla:

- a) *Modo Intensidad*: entrenamiento de ritmo, acentuación y duración
- b) *Modo Sonoridad*: entrenamiento de los rasgos oclusión/fricación y sonoridad/sordez.
- c) *Modo Tono*: entrenamiento del tono fundamental.

Actualmente estamos ampliando sus posibilidades para detectar nuevos parámetros del habla, como el rasgo de nasalidad.

Otras aplicaciones en las que estamos trabajando actualmente son un sistema de audiometría tonal informatizada y un compresor de frecuencias personalizado. El primero permitirá la simulación de un audiómetro básico de forma automática. El segundo utilizará técnicas de trasposición y compresión de frecuencias, adaptándolas a las características fónicas peculiares de nuestra lengua, con el fin de aprovechar al máximo los restos auditivos del hipoacúsico.

3. Pruebas para la Evaluación del Lenguaje Sintético.

Desde el comienzo de las investigaciones sobre el tratamiento automático del habla, los equipos que trabajaban en síntesis de voz en nuestro país se enfrentaban al problema de la elaboración de tests para probar la inteligibilidad de sus sistemas. Normalmente esta cuestión se ha solventado acudiendo a las pruebas elaboradas en Estados Unidos y adaptándolas, con mayor o menor fortuna, al castellano.

Nosotros hemos elaborado un Test que, conservando las normas propias de las pruebas más utilizadas internacionalmente, tiene muy en cuenta las peculiaridades de nuestra propia lengua, tanto en el plano fónico como en todos los niveles del sistema lingüístico: el TRD, Test de Rasgos Distintivos.

En su estructura presenta las características esenciales del DRT (*Diagnostic Rhyme Test*¹): un test de respuesta cerrada (determinados pares de palabras) para la diferenciación de las consonantes, que persigue reducir en todo lo posible la información proporcionada por el contexto.

Para delimitar al máximo los factores de respuesta al estímulo, hasta reducirlos a una única variable pertinente (único rasgo distintivo) se han seguido los siguientes pasos:

1) Selección de pares mínimos: parejas de palabras que sólo difieren entre sí en un fonema y gracias a un único rasgo distintivo: *peso/beso* (sonoridad/sordez).

¹El primer trabajo publicado a este respecto fue el de G.Fairbanks: "Test of Phonemic Differentiation: The Rhyme Test". *J.A.S.A.*30-7, 1958, pp.596-600; En los años posteriores, hasta la actualidad, ha ido apareciendo numerosa bibliografía al respecto, con variaciones sobre el material original, con la creación del MRT (*Modified Rhyme Tests*), etc., pero no es éste el momento de entrar en ello.

2) Control de rasgos secundarios. En cada par mínimo, los fonemas que oponen las dos palabras presentan, además del rasgo distintivo diferenciador, otros varios que deben ser controlados para evitar que su presencia falsee los resultados del test.

3) Control del contexto: selección del patrón silábico de las palabras, de la posición del fonema en la sílaba, del entorno en el que aparece, etc.

4) Homogeneización morfológica, con el fin de evitar los desequilibrios consecuencia de la diferente frecuencia relativa de cada clase de palabras.

5) Control de frecuencia y familiaridad, para evitar las distorsiones que produciría la comparación de una palabra muy conocida con otra poco o nada usual para el oyente.

Actualmente estamos adaptando esta prueba para su utilización con lenguaje natural, adecuando su presentación y posibilitando el procesamiento automático de la información que nos proporciona, con el fin de emplearlo en la práctica clínica, dentro del campo de la logaudiometría, como complemento a otras pruebas (de Umbral de Recepción Verbal, de discriminación, audiometría verbal infantil, etc.), elaboradas también por una parte de este equipo.