Aplicación de las Tecnologías del Habla al Desarrollo del Prelenguaje y el Lenguaje.

W. Ricardo Rodríguez[1], Carlos Vaquero[1], Oscar Saz[1], Eduardo Lleida[1]

[1]Grupo Tecnologías de las Comunicaciones (GTC), Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, Zaragoza, España. {wricardo, cvaquero, oskarsaz, lleida}@unizar.es

Abstract—**This paper shows a study on the use of Speech Technologies to help people with different speech impairments.**

Speech technologies provide methods which can help young people and those who suffer from speech pathologies to develop the Prelanguage and the Language, improving their communication skills. For this purpose, the acquisition of a corpus of pathological speech is needed, in order to analyze the feasibility of the methods developed. As a result of this study two multimedia applications have been developed, to train Prelanguage and Language respectively, and a corpus of pathological speech in Spanish language has been acquired. This corpus will consolidate present and future investigations in Spanish Language.

Palabras claves— **Pre-Lenguaje, señal de voz, logopedia, Córpora.**

I. Introducción

Las Tecnologías del Habla (TH) abarcan diferentes áreas de trabajo y aplicación. Dentro de las tendencias más utilizadas encontramos: Procesado de la voz, Procesamiento del Lenguaje Natural, Sistemas de Diálogo y Lingüística entre otros. En Medicina, las TH apoyan e investigan en campos como: ayudas a la logopedia, estudios sobre fonética acústica y patológica, y psicolingüística. Aplicar algunas de éstas tecnologías para que personas con trastornos del lenguaje se comuniquen de una mejor manera, es el objetivo de éste trabajo.

En algunos casos el lenguaje durante primer año de vida (prelenguaje) no se desarrolla adecuadamente; en otros, debido a diferentes trastornos, el individuo tiene dificultades en la pronunciación, articulación, o generación de palabras, lo que limita su comunicación. Es así como el desarrollo de aplicaciones informáticas basadas en dichas tecnologías, permite que individuos discapacitados con trastornos del lenguaje, puedan comunicarse e interactuar de una mejor manera con su entorno y con ordenadores inclusive.

El presente artículo muestra en el apartado II, una aplicación informática que utiliza animaciones gráficas y tratamiento de la señal de voz, para desarrollar o fortalecer el prelenguaje en niños con trastornos de comunicación. En el apartado III, se comentan cómo las TH pueden utilizarse para facilitar el desarrollo del lenguaje, y en IV, la adquisición de un corpus de habla patológica infantil como apoyo a la investigación y desarrollo de las mimas TH.

II. Prelenguaje

Las primeras manifestaciones de prelenguaje incluyen el llanto, producción de sonidos de carácter vocal modulado, entonación y demás, que tienen lugar durante el primer año de vida. Después de ésta etapa continúa el desarrollo del lenguaje hasta los cinco años, tiempo en donde se adquiere lo esencial del lenguaje para dominarlo con el tiempo [1]. El modelo anterior ocurre en niños sanos pero desafortunadamente existen numerosos casos de niños con trastornos del lenguaje, que ni siquiera desarrollan un prelenguaje adecuado. Para intentar mejorarlo, se desarrolla una aplicación informática que trabaja los principales aspectos del prelenguaje en dos fases. La Fase 1 trabaja la *Presencia-Ausencia de Voz*, y posteriormente la Fase 2 trabajará: el *Control de Tono*, la *Intensidad*, y la *Vocalización*.

Analizando la señal de voz se obtendrán los parámetros para controlar animaciones gráficas (a manera de juego), que retroalimenten en el niño los efectos de su propia voz. La detección, análisis de la señal de voz y animaciones correspondientes se implementan en C++.

Fase 1. *Presencia-Ausencia de Voz.* Es útil porque crea conciencia en el individuo de su existencia y de que la puede utilizar para interactuar con el entorno (comunicarse). La TH utilizada es un VAD (Voice Activity Detector) que analiza la señal de voz y divide la señal en segmentos de voz y silencio, tiene también la propiedad de reconocer el ruido y eliminarlo. Se ha desarrolado un sistema VAD basado en un modelado estadístico del ruido SVAD (Statistical Voice Activity Detector), en general un sistema VAD tiene una estructura como la mostrada en la Fig 1.

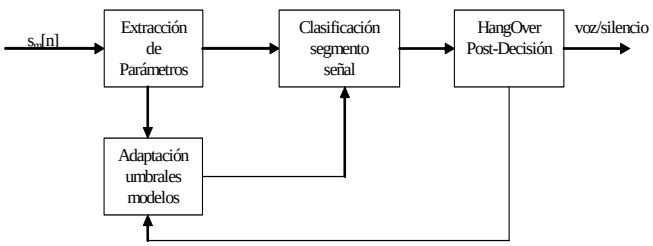

Fig. 1: Diagrama de Bloques de un sistema VAD

El bloque Extracción de Parámetros, extrae los parámetros y medidas necesarias sobre el segmento de señal $s_m[n]$, para la posterior toma de decisión sobre la naturaleza del segmento de señal, se realiza la FFT del segmento y sobre esta información espectral realiza ciertas medidas. La Clasificación segmento Señal es: nivel 0 silencio, nivel 1 posible segmento de voz y nivel 2 segmento de voz. En Adaptación Umbrales y Modelos, se hace la reestimación adaptativa de los umbrales de decisión y de los parámetros de los modelos estadísticos. En HangOver Post-Decisión, se controla posibles errores de clasificación y retarda la toma de decisión del paso de voz a silencio (proceso de hangover) [2].

La Fig 2. muestra la forma de onda y la descisión final del SVAD para una señal de voz. Como el objetivo es crear animaciones gráficas en función de la presencia de voz, se utilizan los segmentos de voz entregados por el SVAD para controlar variables dentro de rutinas en C++ desarrolladas con librerías gráficas Allegro, ya que éstas están diseñadas para juegos. Inicialmente en una ventana vacía (fondo negro) se empiezan a dibujar figuras geométricas básicas en diferentes colores, y se desplazan aleatoriamente por la ventana mientras exista presencia de voz. Lo anterior crea en el niño la conciencia de que con su voz puede dibujar.

Fase 2. Para el *Control de Tono,* se quiere desarrollar en el niño la capacidad de control de la tonalidad, la TH a utilizar es la estimación de Pitch de la señal de voz, es decir la frecuencia de oscilación de las cuerdas vocales al generar sonidos sonoros. En la *Intensidad,* el niño debe aprender a controlar la intensidad de la voz y la respiración. Controlar la intensidad de la voz es importante para comunicarse, imprimir potencia a un soplo o mantenerlo puede sonar fácil pero exige controlar la respiración; la herramienta a utilizar es la estimación de la energía de la señal de voz. En la *Vocalización,* conocer, distinguir y pronunciar correctamente las vocales son los objetivos de ésta etapa; En la generación del habla y específicamente de las vocales, el aire atraviesa la cavidad bucal desde la glotis sin obstáculos, esto hace que sean siempre sonoras; adicionalmente, la cavidad bucal actúa de cavidad resonante (formante) que amplifica la señal de voz y por ende sale con mayor energía; la herramienta a utilizar es el análisis de los formantes (frecuencias) de cada vocal presentes en la señal de voz. La Fase 2 se encuentra en desarrollo y de igual manera se utilizarán animaciones a manera de juegos que motiven al niño a utilizarlas.

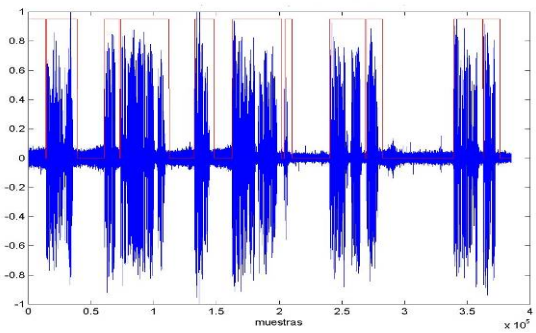

Fig. 2: Forma de onda y decisión final SVAD

III. Desarrollo del lenguaje

Las TH pueden ser muy útiles igualmente para contribuir al desarrollo del lenguaje en aquellas personas que padecen alguna patología en el habla o sencillamente en personas de corta edad que se encuentran en desarrollo. Para ver el alcance de las mismas, conviene analizar el lenguaje desde el punto de vista de la logopedia, que es justamente la disciplina que estudia el desarrollo del lenguaje. Así, existen 4 niveles del lenguaje: nivel fonológico, sintáctico, semántico y pragmático. A continuación se indican una serie de métodos basados en TH para desarrollar los tres primeros niveles del lenguaje.

A. Nivel fonológico:

El nivel fonológico del lenguaje trata la correcta pronunciación y coarticulación de los distintos fonemas y agrupaciones de fonemas pertenecientes a una determinada lengua. Un adecuado tratamiento del nivel fonológico, por tanto, comprendería una primera etapa de identificación de los fonemas que una persona pronuncia de forma incorrecta y una segunda etapa de corrección en la pronunciación de dichos fonemas, que salvo en determinadas patologías degenerativas, puede realizarse mediante entrenamiento continuado, obligando a la persona a realizar una pronunciación adecuada.

Dentro de las TH, existe una técnica conocida como Verificación de Pronunciación [3], que permite obtener una medida de confianza para cada uno de los fonemas pertenecientes a una locución que realice una persona.

Tradicionalmente, esta técnica se utiliza para dotar de robustez a los sistemas de Reconocimiento Automático del Habla (RAH), obteniendo una medida de la fiabilidad que tiene cada una de las palabras que el sistema ha reconocido.

En éste caso, la Verificación de Pronunciación puede utilizarse para obtener una medida de confianza que indique la

calidad de la pronunciación de cada uno de los fonemas pertenecientes a un conjunto de palabras que pronuncie la persona a tratar. Igualmente, esta técnica puede utilizarse para hacer un seguimiento de la evolución de la pronunciación de la persona a tratar.

B. Nivel sintáctico:

El nivel sintáctico del lenguaje trata la correcta combinación de palabras para construir sintagmas y oraciones con sentido completo. Utilizando un modelo del lenguaje adecuado, que contenga las reglas sintácticas que se quieren trabajar, se le puede pedir a la persona a tratar que construya frases con un conjunto reducido de palabras (aquellas incluidas en el modelo del lenguaje del sistema de RAH), de forma que el sistema de RAH reconozca la frase pronunciada utilizando un segundo modelo del lenguaje que admita cualquier orden y combinación de palabras en la oración, y compruebe después si la frase reconocida sigue el conjunto de reglas establecidas en el primer modelo del lenguaje.

C. Nivel semántico

El nivel semántico del lenguaje trata la correcta asociación de significados y/o ideas con palabras, entendidas como un conjunto de fonemas, y por tanto, como un sonido articulado. Una forma muy sencilla de tratar este nivel es mediante actividades o juegos que obliguen a una persona con dificultades a este nivel del lenguaje, a asociar ideas o significados con palabras concretas, por ejemplo, mediante adivinanzas. Otro método sería la clasificación de palabras en grupos semánticos, por ejemplo, obligándole al usuario a pronunciar palabras que pertenezcan a un grupo semántico o que tengan un aspecto semántico en común (por ejemplo, pidiéndole que pronuncie frutas). Un sistema de RAH puede configurarse para que únicamente reconozca una palabra, de forma que proponga al usuario una adivinanza y espere la respuesta del usuario, indicándole si esta es correcta o no. Igualmente, un sistema RAH puede configurarse para que sólo reconozca un conjunto de palabras que tengan en común cierto aspecto semántico, de forma que indique a la persona tratada si la palabra que ha pronunciado pertenece al grupo semántico que se le ha indicado o no.

IV. Adquisición de córpora

Una de las principales limitaciones a la hora de aplicar las TH a personas con trastornos en el lenguaje, es la falta de córpora de este tipo de habla en castellano para investigación. Por esta razón, desde el Instituto de Investigación en Ingeniería de Aragón (I3A), se planteó la adquisición de un corpus o base datos de habla patológica infantil en trabajo conjunto con el Colegio Público de Educación Especial "Alborada" de Zaragoza. La composición de este corpus se divide en dos partes, una parte de habla patológica infantil y juvenil y otra parte de habla normalizada infantil y juvenil.

A. Composición fonológica del corpus

El corpus se compone de grabaciones de las 57 palabras que componen el Registro Fonológico Inducido (RFI) [4]. Este conjunto de palabras es una herramienta de uso muy extendida y reconocida por los logopedas en España y contiene una amplia variedad de todos los fonemas de la lengua española, también en diferentes contextos de coarticulaciones y otras relaciones entre los mismos.

B. Habla patológica en el corpus

La parte de habla patológica del corpus contiene grabaciones de 14 jóvenes de la Alborada (7 chicos y 7 chicas), cada uno de ellos repitiendo 4 veces las palabras del RFI. Esto hace un total de 228 palabras aisladas por cada locutor y un total de 3192 en todo el corpus.

Los locutores tienen entre 11 y 21 años de edad en el momento de la grabación. La mayoría de ellos presentan alteraciones del lenguaje a nivel fonológico, morfosintáctico y semántico-pragmático debido a diferentes niveles y orígenes de retraso mental.

C. Habla no patológica en el corpus

Para la obtención de resultados válidos en la investigación en sistemas de ayuda a la logopedia y de RAH, se hace necesario contar con un conjunto de voces de referencia que permitan estudiar las diferencias a nivel acústico y lingüístico entre la voz patológica y la voz normalizada.

Dada la especificidad del corpus de habla patológica obtenido, donde todos los locutores tienen de 11 a 21 años de edad, se planteó la adquisición de un corpus paralelo de habla normalizada en colaboración con el Colegio de Educación Infantil y Primaria "Río Ebro" y el Instituto de Enseñanza Secundaria "Tiempos Modernos" ambos de Zaragoza. Este corpus consta de 152 locutores desde los 10 a los 18 años de edad, cada uno de ellos realizando una alocución de las 57 palabras del RFI para un total de 8664 palabras aisladas.

D. Parámetros de la adquisición

Todo el proceso de adquisición se llevó a cabo en las instalaciones de los tres centros educativos implicados bajo la supervisión de miembros del I3A y uno de lo educadores de los centros. Las señales fueron adquiridas con una frecuen-

cia de muestreo de 16 kHz y se almacenaron con una resolución de 16 bits.

V. Resultados

Para el Prelenguaje, los resultados obtenidos corresponden a la generación de animaciones en la *Presencia de Voz*. La Fig 3. muestra una imagen cuyas formas y colores se dibujaron con la presencia de voz.

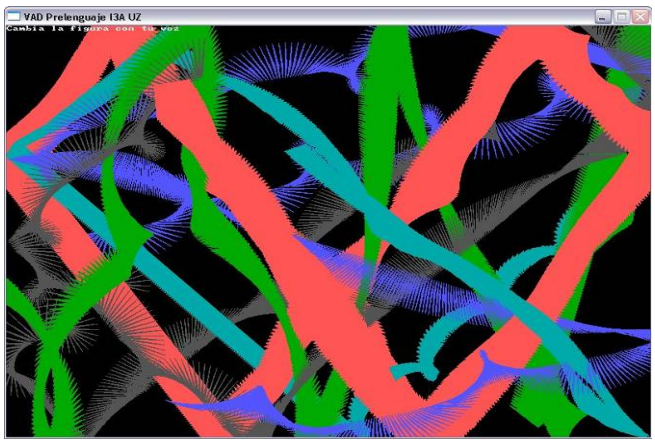

Fig. 3: Imagen dibujada con la voz

Para el Lenguaje, se ha desarrollado una aplicación llamada “Vocaliza” [5], que proporciona distintos juegos que ayudan a desarrollar los niveles fonológico, semántico y sintáctico del lenguaje. Los juegos muestran al usuario un conjunto de imágenes y/o textos que indican al usuario aquello que debe pronunciar. Se utilizan un sistema de RAH que identifica la palabra o frase pronunciada, y a continuación se comprueba si la palabra o frase que el usuario ha pronunciado es aquello que esperaba la aplicación, mostrando al usuario una ventana con dibujos animados en caso de que éste haya tenido éxito en el juego. No se requiere de conocimientos informáticos para utilizar la aplicación, simplemente disponer de un micrófono.

La aplicación utiliza igualmente síntesis de voz para indicar al usuario como se pronuncian las distintas palabras y frases que aparecen en la misma, y también Verificación de Pronunciación para evaluar la calidad de la pronunciación del usuario en los juegos que tratan el nivel fonológico. La Fig. 4. muestra la ventana principal de la aplicación. Como puede apreciarse, la aplicación se ha diseñado para resultar atractiva a los usuarios más jóvenes.

La aplicación Vocaliza está implantada actualmente en dos colegios de educación especial de Zaragoza y previsiblemente se distribuirá por más centros de esta misma ciudad.

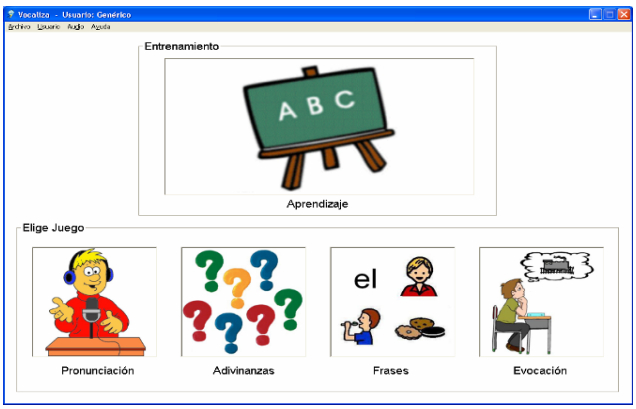

Fig. 4: Ventana principal de la aplicación “Vocaliza”

VI. Conclusiones

Debido al gran variedad de patologías responsables de trastornos de la comunicación, las TH se convierten en una poderosa herramienta aplicable de manera directa y útil; Desarrollar adecuadamente el prelenguaje y el lenguaje en un individuo con discapacidad de comunicación es definitivamente mejorar su calidad de vida y hacerlo partícipe del uso de ordenadores a través de aplicaciones multimedia.

El desarrollo de estas tecnologías requiere el trabajo conjunto de ingenieros y logopedas que posibiliten los mejores resultados; la existencia además de corpus adecuados para investigación aumenta aun más la consecución de estos resultados.

Reconocimientos

Este trabajo ha sido financiado por el Ministerio de Educación y Ciencia de España a través del Proyecto Nacional TIN 2005-08660-C04-01.

Becas Santander Central Hispano – Universidad de Zaragoza.

Referencias

1. Puyuelo S, Miguel (2005) Evaluación del Lenguaje. Capítulo 2. Ed. Masson, España.
2. Julius S. Bendat, Allan G. Piersol (1986) Random Data, Analysis and Measurement Procedures, Wiley-Interscience, 2 Edition.
3. E. Lleida and R.C. Rose, Utterance verification in continuous speech recognition: Decoding and training procedures, IEEE Transactions on Speech and Audio Processing, vol. 8, no. 2, pp. 126–139, 2000.
4. Monfort M., Juárez Sanchez A. (1989) Registro Fonológico Inducido (Tarjetas Gráficas). Ed. Cepe, Madrid.
5. Carlos Vaquero, Óscar Saz, Eduardo Lleida, José Manuel Marcos y César Canalís, “Vocaliza, an Application for Computer-Aided Speech Therapy in Spanish Language”, IV jornadas en tecnologías del habla, Zaragoza, España, pp. 321-326, 2006.