Yasmina Galve Pastor

Diseño de herramientas de asistencia a la logopedia en una plataforma distribuida

Proyecto Fin de Carrera

Ingeniería de Telecomunicación

Zaragoza, Julio 2012

Director: **Dr. Antonio Miguel Artiaga**

Escuela de Ingeniería y Arquitectura Universidad Zaragoza

Departamento de Ingeniería Electrónica y Comunicaciones Universidad Zaragoza

RESUMEN

Las tecnologías del habla (TH) pueden suponer una herramienta muy poderosa a la hora de facilitar la vida cotidiana a aquellas personas que presentan alguna patología en su capacidad del habla. La utilización de estas tecnologías permite el desarrollo de herramientas que asistan a los profesionales de la logopedia en su labor de ayudar a personas con problemas de dicción. Sin embargo, el alto coste de las herramientas que existen en el mercado y además, la no existencia de herramientas en español, han impedido la utilización de estas tecnologías para este fin.

El presente proyecto tiene como objetivo el diseño de una herramienta libre de asistencia a la logopedia, válida tanto para el español como para otros idiomas, y que pueda ser ejecutada en una plataforma distribuida, evitando problemas de incompatibilidad de sistemas operativos.

La herramienta consiste en un editor de actividades que permita a los logopedas diseñar sus propias actividades sin necesidad de tener conocimientos técnicos. El editor es totalmente configurable con el fin de que pueda trabajar con el número máximo de patologías posible, adaptándolas a cada caso en particular.

Se trabaja el procesado de la señal de voz mediante las técnicas tradicionales y se introducen técnicas que eliminan la influencia del pitch, la cual supone un problema cuando se trata voz de alta tonalidad como es el caso de la voz infantil. Se calcula la longitud del tracto vocal, la cual permite reducir la variabilidad de los formantes mediante su normalización.

Se diseña un motor de simulación física, que transforma los parámetros de la señal de voz obtenidos en distintos tipos de movimiento de un objeto que el usuario final visualizará en la pantalla. Se introduce la posibilidad de añadir otras condiciones que influirán en el movimiento del objeto. De esta manera, los logopedas disponen de mayor flexibilidad en el diseño de las actividades y pueden variar el nivel de dificultad de las mismas.

Por último, se diseña una interfaz gráfica que guíe a los logopedas en la configuración de actividades, proporcionándole numerosas opciones para configurar y variar. La configuración de las actividades se podrá guardar en un archivo XML para poder compartirla con otros profesionales o poder cargarla en el futuro.

AGRADECIMIENTOS

Quiero aprovechar estas líneas para mostrar mi agradecimiento a todas aquellas personas que han colaborado y, de alguna manera, me han ayudado en la realización de este proyecto.

En primer lugar quiero agradecer a mi director, Antonio Miguel Artiaga, la oportunidad de realizar este proyecto tan interesante, así como su dedicación y asesoramiento durante la duración del mismo.

Al colegio Alborada por la disposición e interés mostrados.

A mis compañeros y jefes de Circe quiero agradecer su apoyo y libertad concedida durante la última etapa del proyecto.

A mis padres, Julia y Miguel, a quienes nunca podré agradecer lo suficiente su confianza y apoyo incondicional durante toda la carrera.

Por último, a Carlos y a mis compañeros y amigos.

ÍNDICE DE CONTENIDOS

ÍNDICE DE FIGURAS

1. Introducción

Las tecnologías del habla (TH) pueden suponer una herramienta muy poderosa a la hora de facilitar la vida cotidiana a aquellas personas que presentan alguna patología en su capacidad del habla. Estas patologías pueden deberse a discapacidades motrices que implican dificultad en la pronunciación pero, en otros muchos casos, se deben a problemas más severos como una parálisis cerebral que supone una gran dificultad a la hora de emitir sonidos. Esto impide un adecuado funcionamiento de las TH para este fin.

Las TH pueden ayudar a estas personas con discapacidades motrices o problemas de dicción mediante dos vías. Por un lado, sistemas de ayuda controlados por voz. Estos sistemas permiten controlar dispositivos del entorno con el uso de la voz, mediante reconocedores de patrones orales. Por otro lado, las TH se pueden aplicar al campo de la logopedia. Este tipo de tecnologías permiten desarrollar herramientas que asistan a los profesionales de la logopedia en su labor de conseguir que personas con problemas de dicción mejoren su capacidad del habla, consiguiendo que sea más fácil de comprender.

Sin embargo, los logopedas y profesionales de la educación especial se encuentran con grandes limitaciones a la hora de trabajar con población infantil con discapacidad y voz alterada, debido a que la mayoría de las herramientas disponibles presentaban un alto coste de adquisición y, además, estaban diseñadas para otros idiomas distintos al

español. En estos casos debían trabajar la voz de los pacientes mediante los métodos manuales tradicionales (láminas, globos, imitación de sonidos, etc.).

Este proyecto está enmarcado dentro del Grupo de Tecnologías de las Comunicaciones (GTC) en colaboración con el Colegio Público de Educación Especial Alborada (CPEE Alborada) siguiendo la línea de proyectos realizados con anterioridad en este mismo grupo[1].

También se han utilizado los recursos gráficos que proporciona ARASAAC[2] en forma de pictogramas para facilitar la comunicación de personas con dificultad en esta área.

1.1 Estado del arte

El uso de las TH como herramienta para ayudar a personas con discapacidades motrices y de comunicación, así como su utilización para la asistencia a la logopedia no es una idea reciente. Sin embargo, hasta los últimos años el desarrollo alcanzado era mínimo.

Los profesionales de la educación especial han demandado este tipo de herramientas pero, su gran coste de adquisición y la limitación que estas presentaban al ser válidas únicamente para algunos idiomas, han impedido que las TH se hayan utilizado para este fin.

De la colaboración del GTC con el Colegio Público de Educación Especial Alborada surgió el proyecto "Comunica"[1] [1] [2], pionero en el desarrollo de herramientas libres de asistencia a la logopedia y en español. Dentro de este proyecto, "Vocaliza" [1] [3] es una herramienta que trabaja los niveles fonológico, semántico y sintáctico del lenguaje. La herramienta "PreLingua" [1] [4] se centra en el pre-lenguaje, esto es, las habilidades que adquiere el niño en su primer año de vida y que en ocasiones, no se han desarrollado con normalidad. Otra herramienta es "Cuéntame" [1] [5] que comienza a trabajar con el nivel superior, el pragmático. Este proyecto sigue esta línea, en concreto de "PreLingua", pues su objetivo es tratar las habilidades pre-lingüísticas en población infantil con problemas de dicción.

1.2 Definición de objetivos

El proyecto realizado consiste en el desarrollo de una aplicación cuyo objetivo fundamental es servir de apoyo a la logopedia. Se trata de un editor de actividades que ayuden a los logopedas y profesionales de la Educación Especial a mejorar la capacidad del habla del paciente.

---

1 http://www.vocaliza.es/
2 http://catedu.es/arasaac/

De las distintas reuniones con el director, con el autor de PreLingua y con los profesionales del CPEE Alborada, surgieron una serie de objetivos que el proyecto realizado debía cumplir:

- La aplicación debe de ser de libre distribución.
- Debe poder ejecutarse en una plataforma distribuida, superando el problema de incompatibilidad entre sistemas operativos. Para ello, la aplicación se ha desarrollado en lenguaje Java.
- La herramienta será válida tanto para el español como para otros idiomas.
- La aplicación se diseña con fines logopédicos, es decir, con el objetivo de ayudar a la mejora de las capacidades del habla en personas con discapacidad.
- Puesto que la aplicación está dirigida a población infantil, se deben introducir técnicas que superen las dificultades que este tipo de voz presenta.
- La herramienta debe ser altamente configurable. Debe proveer a los logopedas de suficientes opciones de configuración que les permitan trabajar con el número máximo de patologías posible y adaptarlas a cada caso en particular.
- El diseño del editor debe ser sencillo y que permita a los logopedas diseñar sus propias actividades sin necesidad de poseer conocimientos técnicos.
- La interfaz del usuario final debe ser simple, pues va dirigido a población con discapacidad.

1.3 Tareas realizadas

A continuación se describen las tareas que se han llevado a cabo para la realización de este proyecto.

1.3.1 Documentación

Esta primera fase consistió en el conocimiento del estado del arte y del campo de la logopedia. Una lectura exhaustiva del trabajo previo como Tesis Doctorales y Proyectos Fin de Carrera fue necesaria, así como el estudio y comprensión de las herramientas que en ellos de describían. Varias reuniones con el CPEE Alborada tuvieron lugar durante esta fase del proyecto.

1.3.2 Procesado de señal

La primera parte que se abordó en la ejecución del proyecto fue el procesado de señal. Consistió en el aprendizaje de la utilización de las técnicas de procesado de la señal de voz utilizadas. Para ello, se simuló todo el procesado en MATLAB y se utilizaron como entrada ficheros de voz. Una vez superada esta etapa, se desarrolló esta parte en Java, añadiendo la adquisición de la señal de voz desde el micrófono y consiguiendo un procesado de la señal en tiempo real.

1.3.3 Diseño del motor de simulación física

Con el procesado de señal controlado, se dispuso a diseñar el motor de simulación física del entorno virtual. Aquí, se convierten los parámetros obtenidos en el procesado en el movimiento de un objeto, así como se añaden condiciones adicionales que elevan el nivel de dificultad de la actividad.

1.3.4 Diseño de la interfaz gráfica

Por último, recopilando el trabajo realizado, se diseñó la interfaz gráfica del editor, que permite al logopeda diseñar actividades adaptadas al caso que esté tratando. La interfaz gráfica, así como el resto de partes del proyecto se han programado en Java[3], utilizando el entorno de desarrollo integrado libre NetBeans[4] y su editor gráfico basado en Swing[5] para la parte gráfica.

1.4 Tecnologías empleadas

Varias tecnologías han sido empleadas durante la ejecución del proyecto. En primer lugar, el procesado de señal se simuló en MATLAB[6], lo que permitió la representación de los resultados intermedios que se iban obteniendo.

---

3 http://www.java.com/es/

4 http://netbeans.org/

5 http://netbeans.org/features/java/swing.html

6 http://www.mathworks.es/products/matlab/

La parte de procesado de señal también fue trabajada en C, pues las herramientas diseñadas hasta el momento estaban definidas en este lenguaje.

Por último, tanto esta parte como toda la interfaz gráfica y el motor de simulación física fueron implementados en lenguaje Java. Este lenguaje permite ejecutar la herramienta en una plataforma distribuida, superando la incompatibilidad entre sistemas operativos.

1.5 Organización de la memoria

A continuación, se describe brevemente la organización de la memoria.

En el capítulo 1 se ha desarrollado una breve introducción al proyecto final de carrera y los objetivos principales del mismo.

En el capítulo 2 se introducen los conceptos básicos del procesado de la señal de voz y la extracción de sus parámetros.

El capítulo 3 resuelve la problemática de la voz infantil y describe las técnicas utilizadas para ello.

El capítulo 4 describe el diseño y estructura de la aplicación, así como los algoritmos seguidos para el desarrollo del motor de simulación.

Por último, el capítulo 5 contiene las conclusiones obtenidas junto a una serie de líneas futuras de investigación.

2. Procesado de la señal de voz

2.1 La Voz Infantil

La voz es el principal canal de comunicación entre los seres humanos y por ello requiere una atención especial en la población infantil. La activad vocal experimenta un gran desarrollo durante los primeros meses de vida. Durante este tiempo tienen lugar cambios estructurales que permiten la evolución de la voz y provocan un cambio en su frecuencia que va desde los 400 Hz del llanto a los 110 Hz en niños y 220 Hz en niñas tras la pubertad [6].

2.1.1 Interpretación de la voz

La alteración de la voz es muy común en la población infantil. Si se trata de niños con discapacidad son diversas las causas que pueden influir en la calidad de voz, por ejemplo el retardo mental, el síndrome de Down, la parálisis cerebral, etc.

Se considera como voz alterada o disfonía la alteración de alguna de cualidades acústicas o la combinación de varias de ellas:

- **Intensidad (dB):** volumen de la voz resultante de la presión del aire a su paso por las cuerdas vocales.
- **Tono (Hz):** consecuencia de la vibración de las cuerdas vocales en la fonación. Este valor desciende lentamente desde la infancia hasta la pubertad donde tiene lugar un descenso brusco en el caso de los hombres.
- **Timbre:** es la característica que dota de personalidad a la voz. Está formado por la frecuencia fundamental, sus armónicos y los formantes. Estos últimos dependen de la disposición de los órganos que intervienen en la producción de la voz y varían según el género, talla, y raza del niño.
- **Duración:** tiempo que permanecen las cuerdas vocales en vibración. El tiempo máximo de fonación es de gran interés para el diagnóstico.

2.2 Procesado de voz

2.2.1 Aparato fonador humano y modelo digital de producción de voz

La voz consiste en una onda de presión acústica sonora que se genera por el movimiento de la estructura física del sistema fonador. Esta onda se propaga por el aire a una velocidad de unos 340 m/s. El aparato fonador humano se divide en dos subsistemas principales:

- **Aparato fonatorio:** formado por pulmones, tráquea y laringe hasta las cuerdas vocales.
- **Aparato articulatorio:** formado por paladar, lengua, dientes, labios y mandíbula.

Los sonidos dependen de la posición de estos elementos al paso del aire por ellos.

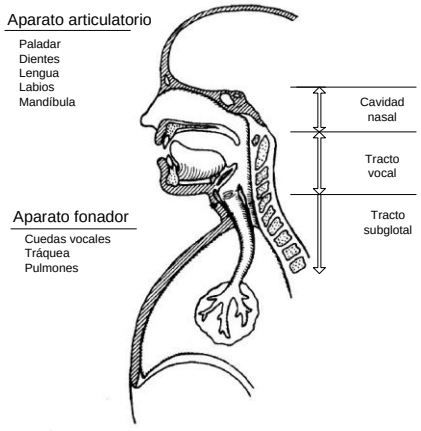


Figura 2.1. Aparato fonador humano.

El aparato fonador humano puede modelarse, de forma bastante aproximada, como un sistema lineal todo polos que representa la posición de los órganos involucrados o tracto vocal, y una señal de entrada o señal de excitación que representa el paso del aire por los pulmones, tráquea y cuerdas vocales. Como señal de salida de obtendrán dos tipos de sonidos dependiendo de la naturaleza de la señal de excitación:

- Sonidos **sonoros**: son de energía elevada y se produce la vibración de las cuerdas vocales. Para generar este tipo de sonidos, la señal de excitación será un tren de impulsos de frecuencia controlada.
- Sonidos **sordos**: no existe vibración de las cuerdas vocales y son de baja energía. Para generar este tipo de sonidos, la señal de excitación será ruido blanco.

Se obtiene un modelo digital de producción de voz (Figura 2.2) que consiste en la combinación de ambas señales de excitación, las cuales modelan el funcionamiento de la glotis y el filtro lineal todo polos que representa el tracto vocal.

El tracto vocal presenta un elevado número de resonancias pero las más importantes son las dos primeras, pues son las que contienen mayor información sobre la producción sonora.

El espectro de la señal de salida se obtendrá a partir del producto del espectro de la excitación y la respuesta frecuencial del filtro que representa el tracto vocal.

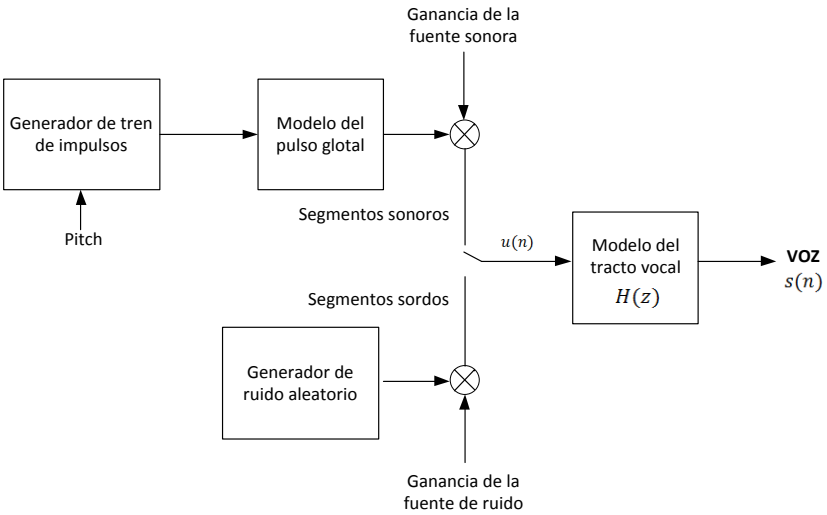


Figura 2.2. Modelo digital de producción de voz.

El procesado de la señal de voz permite extraer sus parámetros más importes.

2.2.2 Pre-procesado

Para procesar la señal de voz digitalmente es necesario convertir la señal acústica en una señal eléctrica con un micrófono. A continuación, se muestrea la señal obtenida para convertirla en una seña digital. En la voz humana, la información frecuencial se concentra en los primeros 4000 Hz así que, una frecuencia de muestreo de 8000 Hz será suficiente para extraer esta información y evitar el aliasing (Teorema de Nyquist).

Tres bloques componen el pre-procesado de la señal de voz: compensación DC, filtro de pre-énfasis y enventanado.

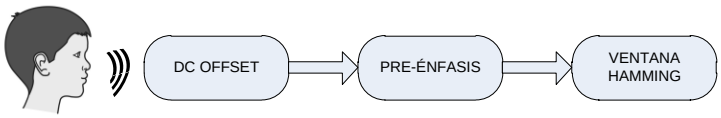


Figura 2.3. Pre-procesado de la señal de voz.

El bloque DC offset elimina la posible componente de continua de la señal mediante un filtro de banda eliminada en la frecuencia 0.

$$\mathrm{H(z)} = \frac{1 - \mathrm{z}^{-1}}{1 - 0.995z^{-1}} \tag{2.1}$$

El filtro de pre-énfasis compensa la caída de -6dB por octava de la señal de voz debido al pulso glotal y la radiación de los labios.

$$H(z) = 1 - az^{-1}, \mathrm{a} = 0{,}95 \tag{2.2}$$

La señal de voz tiene un comportamiento pseudo-aleatorio a corto plazo y, por eso, es necesario llevar a cabo un enventanado que realice un análisis localizado de la señal. La ventana utilizada en el procesado de voz es la ventana de Hamming que, al tener un lóbulo principal ancho y lóbulos secundarios pequeños, produce un suavizado espectral realzando la información central de la ventana y minimizando la información de los extremos.

$$W(n) = \begin{cases} 0.54 + 0.46\cos\left(\frac{2\pi n}{N-1}\right) & 0 \leq n \leq N-1 \\ 0 & otro\ caso \end{cases} \tag{2.3}$$

Para compensar el efecto de minimización de los extremos se solapan las ventanas, de manera que las muestras en los extremos de una ventana correspondan con las muestras centrales de las ventanas consecutivas.

2.2.3 Estimación de energía

La estimación de energía permite distinguir sonidos sonoros cuya energía es elevada, de sonidos sordos cuya energía es menor [7].

$$E_s[m] = \sum_{n=-\infty}^{\infty} (s[n] \cdot w[n-m])^2 = \sum_{n=m-N+1}^{m} s^2[n] \cdot w^2[n-m] \qquad (2.4)$$

Expresando $w^2(n) = h(n)$:

$$E_s[m] = \sum_{n=m-N+1}^{m} s^2[n] \cdot h[n-m] \qquad (2.5)$$

La ventana utilizada para el cálculo de la energía es rectangular.

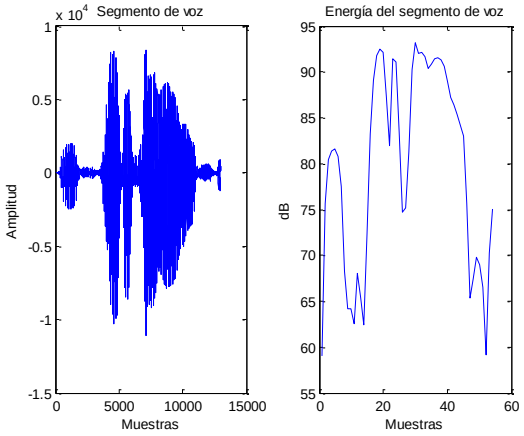


Figura 2.4. Energía de la señal de voz.

En la Figura 2.4 se observa una señal de voz y la energía localizada calculada. Los máximos locales representan los segmentos de sonidos sonoros. Los segmentos de baja energía son sonidos sordos.

La alta energía de los sonidos sonoros permite construir un VAD muy simple que se comporta bien en condiciones de bajo ruido [7]. Este VAD indica la presencia de voz comparando la estimación de energía del segmento con un umbral preestablecido. Si la energía es mayor que ese umbral, el segmento es sonoro. Si es menor, el segmento es sordo.

2.2.4 Autocorrelación

La función de autocorrelación, Γ_x, cuantifica el parecido de una señal consigo misma cuando se le aplica un desplazamiento de k muestras.

$$\Gamma_x[k] = \sum_{n=-\infty}^{\infty} [x(n)x(n-k)] \tag{2.6}$$

La autocorrelación de una señal periódica es también una señal periódica del mismo periodo. Se puede aprovechar esta propiedad para detectar patrones periódicos en la señal de voz e identificar si se trata de un sonido sonoro y, por lo tanto, posee pitch. Este valor de pitch se puede obtener a partir del periodo de la autocorrelación, es decir, de la distancia entre máximos [8].

2.2.5 Análisis de predicción lineal LPC

El análisis de predicción lineal (LPC) permite obtener la función de transferencia del filtro que ha generado la señal de voz y que se corresponde con la función del tracto vocal, así como la señal residual que representa la señal de excitación [7]. Esta técnica se detalla en el Anexo I.

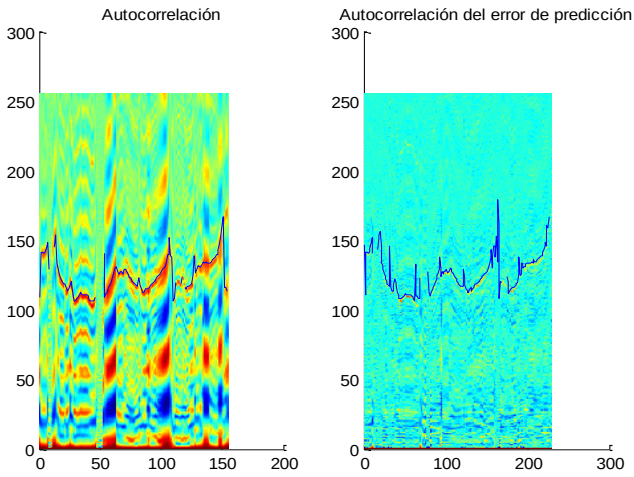
El análisis LPC permite obtener los coeficientes a_k del filtro denominado filtro inverso $A(z)$ cuya entrada es la señal de voz y salida la señal de excitación, a partir de la cual se podrá extraer la frecuencia fundamental o pitch. El filtro que representa el tracto vocal $H(z)$ se modela como $1/A(z)$. De este filtro $H(z)$, será posible extraer los formantes que caracterizan las vocales para cada individuo.

2.2.6 Estimación del Pitch

El análisis LPC descrito en el Anexo I permite obtener la señal de error o residual a partir de una señal de voz. Esta señal residual se corresponde con la excitación vocal y, a partir de ella, se puede extraer más fácilmente el pitch.

Tomando la autocorrelación de esta señal residual, se obtiene la frecuencia de pitch calculando la distancia entre el origen y el primer máximo, pues esta distancia se corresponde con el periodo de pitch [4].

En la Figura 2.5 se observa la diferencia entre el cálculo del pitch mediante la autocorrelación de la señal de voz y la autocorrelación del error de predicción. Se observa como en la primera (a) es más difícil distinguir el pitch del resto de información (formantes). Sin embargo, mediante el cálculo de la autocorrelación del error de predicción (b), donde solo existe información de la señal de excitación, es más fácil determinar el pitch.



(a) (b)

Figura 2.5. Comparación de métodos de estimación del pitch.

Para eliminar datos espurios se utiliza un filtrado de mediana de orden 5. El proceso, pues, para el cálculo de la frecuencia de pitch se resume en la Figura 2.6.

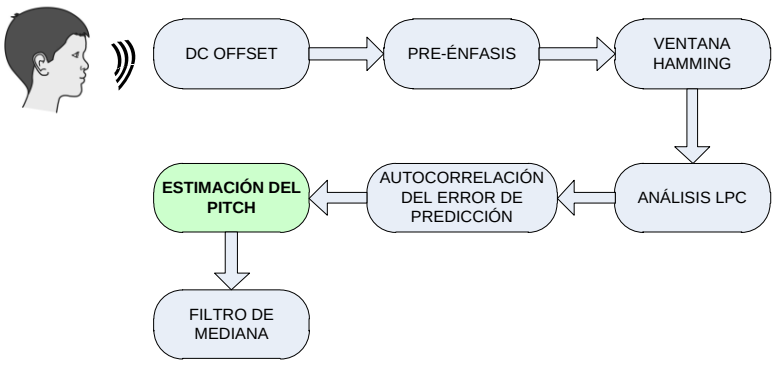


Figura 2.6. Estimación del pitch.

2.2.7 Estimación de Formantes

El análisis LPC descrito en la sección 2.2.5 permite separar la influencia del tracto vocal de la señal de excitación. Por tanto, a partir de éste análisis es posible centrarse en la estructura formántica analizando el polinomio $A(z)$ (2.7).

$$A(z) = 1 + \sum_{i=1}^{p} a_i z^{-i} = \prod (1 - z_k z^{-1}) \tag{2.7}$$

Con un orden de predicción adecuado, aproximadamente $\frac{Fs}{1000}$ de las raíces del polinomio anterior estarán cerca de las frecuencias de resonancia en el plano Z con Fs la frecuencia de muestreo. Los ceros o raíces de $A(z)$, pares complejos conjugados, son los polos de $H(z)$ que modelan los formantes [7].

Los tres primeros formantes se corresponden con las raíces del polinomio $A(z)$, convertidas a frecuencia analógica multiplicando por la frecuencia de muestreo y ordenadas de menor a mayor. Se utiliza de nuevo el filtro de mediana para desechar valores espurios.

El proceso a seguir para el cálculo de los formantes se resume en la Figura 2.7.

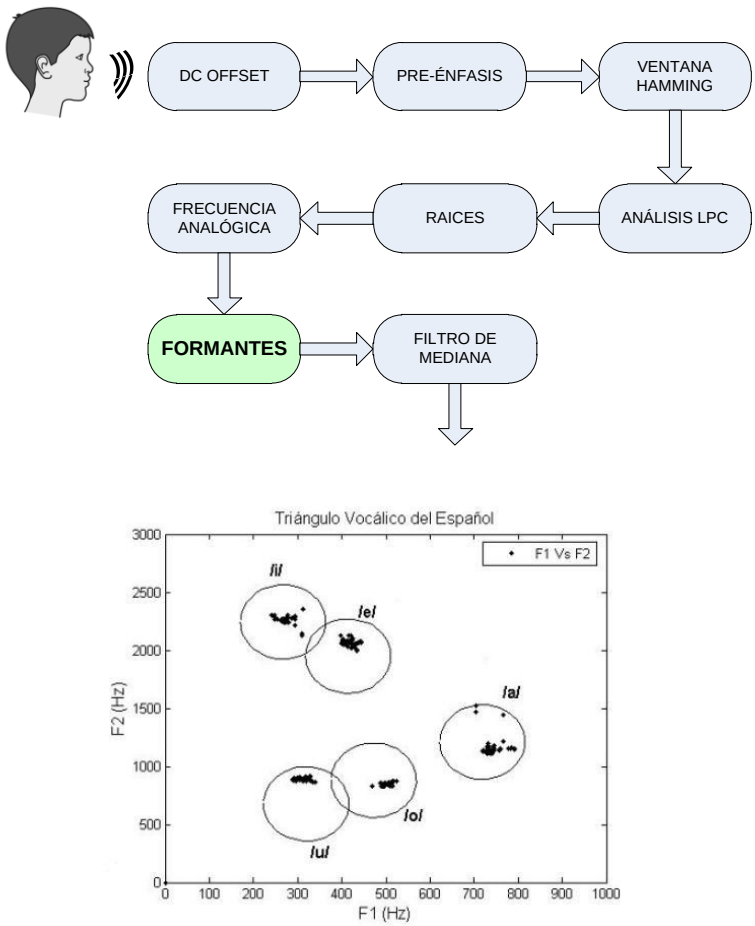


Figura 2.7. Extracción de formantes.

2.2.8 Análisis Homomórfico

El análisis homomórfico se utiliza para convertir un elemento matemático en otro. En el caso del análisis de la señal de voz, se utiliza para separar dos componentes relacionadas mediante una convolución en una suma. De esta manera, la señal de voz en el dominio cepstral, se separa en una combinación lineal de la señal de excitación $\hat{e}[n]$ y el modelo del tracto vocal $\hat{h}[n]$(2.8). Se detallan los fundamentos matemáticos del análisis homomórfico en el Anexo II.

$$s[n] = e[n] * h[n] \rightarrow \hat{s}[n] = \hat{e}[n] + \hat{h}[n] \tag{2.8}$$

Tras su aplicación, se obtiene a la salida la señal de cepstrum (Figura 2.8). La parte baja corresponde con la información del tracto vocal y la parte alta con la excitación [7]. Para separar dicha información se realiza un liftado que es un proceso similar al filtrado pero en el dominio cepstral.

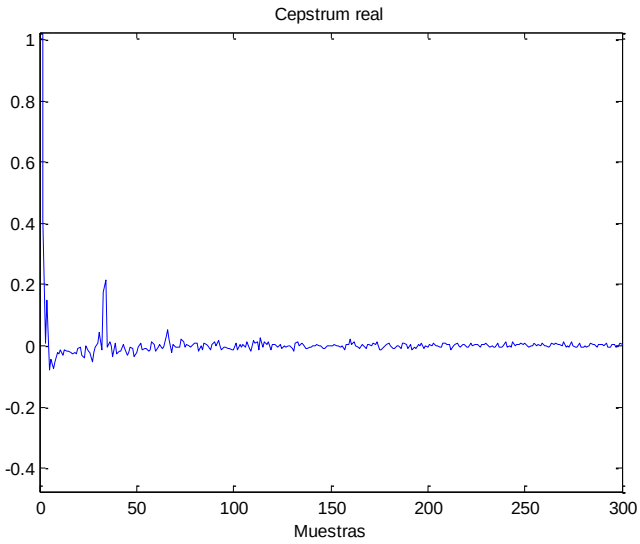


Figura 2.8. Separación de componentes en el dominio cepstral.

3. Estimación robusta de Formantes y Estimación del Tracto Vocal

3.1 Estimación Robusta de Formantes

De las técnicas de procesado de la señal descritas en el capítulo 2, la estimación de formantes es la técnica que más dificultades presenta, pues lleva a estimaciones erróneas en voces con valores altos de pitch.

3.1.1 Dificultad de la voz infantil

Existen diferencias considerables entre la voz de una persona adulta y la de un niño. Se pueden apreciar estas diferencias a través del espectrograma que muestra la evolución temporal de la caracterización espectral de la señal de voz [9].

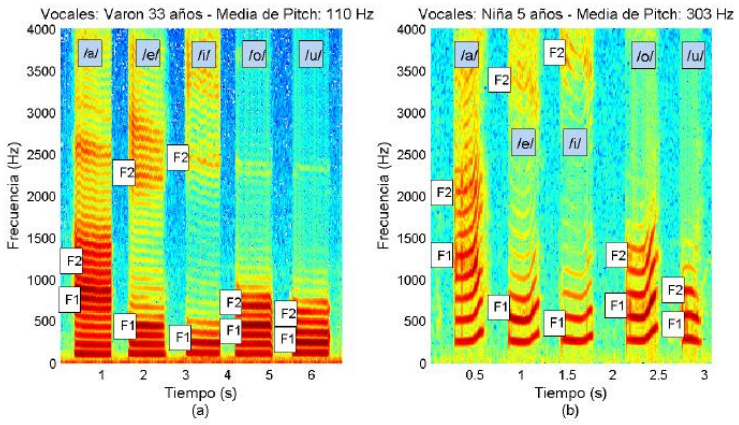


Figura 3.1. Espectrograma para las vocales en voz adulta (a) y voz infantil (b).

La Figura 3.1 muestra dos espectrogramas con las cinco vocales del español pronunciadas de manera aislada. El primer espectrograma (a) corresponde a un varón adulto con un valor de pitch medio de 110 Hz. El segundo espectrograma (b) pertenece a una niña con un valor medio de pitch de 303 Hz. Se aprecian a simple vista las diferencias entre la posición del pitch y sus armónicos y la posición de los formantes. En la voz infantil, el pitch es mayor que para el adulto y sus armónicos están más espaciados y acentuados por lo que los formantes quedan ocultos entre los armónicos.

En cuanto a la información de los formantes, para la voz infantil son superiores a los de la voz adulta debido a que el tracto vocal es más corto y, por tanto, sus frecuencias de resonancia son mayores.

Debido a esto, la estimación de formantes en la voz infantil mediante técnicas tradicionales puede llevar a resultados erróneos, pues los formantes quedan enmascarados bajo el pitch y sus armónicos.

3.1.2 Eliminación de la influencia del pitch

Tras entender la dificultad que muestra la voz con alta tonalidad a la hora de estimar sus formantes, es necesario eliminar la influencia del pitch para calcular los formantes de una manera robusta. Es decir, se desea separar las dos componentes, excitación y tracto vocal, para obtener unos formantes libres de la influencia del pitch.

Esta separación se consigue con el análisis homomórfico descrito en la sección 2.2.8 y Anexo II, llevando la señal al dominio cepstral. Se ha demostrado [10] que no todas las definiciones de cepstrum son apropiadas para la estimación de formantes. El cepstrum complejo (II.2) no es una formulación adecuada para la estimación de formantes debido a su alta sensibilidad a la fase [11]. En el cepstrum real (II.5) no se tiene en cuenta la fase y la magnitud contiene información suficiente sobre la trama para el cálculo de formantes, por lo que resulta más apropiado. Tras aplicar el cepstrum real, la señal está separada linealmente en sus dos componentes y se tratarán independientemente.

$$\hat{s}[n] = \hat{e}[n] + \hat{h}[n] \tag{3.1}$$

La información del tracto vocal $\hat{h}[n]$ se encuentra concentrada en la parte baja del cepstrum mientras que la información del pitch $\hat{e}[n]$ se encuentra en la parte alta como se observa en la Figura 2.8. Se utiliza el valor del periodo de pitch T_{pitch}, estimado anteriormente, para realizar el liftado o filtrado en el dominio del cepstrum y eliminar la parte alta del cepstrum y así, la influencia del pitch.

Es importante elegir una adecuada longitud de la ventana de liftado $w[n]$ porque puede introducir errores en la estimación de formantes ya que, los coeficientes cepstrales cercanos al periodo de pitch pueden ser distorsionados [12]. Algunos autores [12] han propuesto una longitud de $0.5T_{pitch}$ mientras que otros [10] proponen aumentar su longitud a $0.7T_{pitch}$ para voces con pitch superior a 250Hz y $0.6T_{pitch}$ para frecuencias menores.

El pitch varía significativamente dependiendo de la altura y el sexo del hablante. Debido a esto, establecer un valor fijo para la ventana de liftado no tendría los mismos efectos para todas las frecuencias de pitch.

Para adaptar el sistema a cada usuario, se propone establecer una longitud de ventana de liftado variable, αT_{pitch}, donde α se calcula directamente de las características de cada locutor [4].

$$W[n] = \begin{cases} 0 & 0.22 * log(Tpitch)) * Tpitch \leq n \leq N - 1 - 0.22 * log(Tpitch)) * Tpitch \\ 1 & otro \end{cases} \quad (3.2)$$

Tras realizar el liftado, se procede a la obtención de los formantes siguiendo el método tradicional. Calculando la densidad espectral de potencia $S_x(e^{jw})$ de la secuencia liftada $\hat{c}[n]$ y, asumiendo que se trata de un proceso estacionario en sentido amplio (WSS), se obtiene la nueva función de autocorrelación $\Gamma_x[k]$ mediante el teorema de Wiener-Khintchine [8].

$$\Gamma_x[k] = \frac{1}{2\pi}\int_{2\pi} S_x(e^{jw})e^{jwk}dw \quad (3.3)$$

A partir de esta nueva función de autocorrelación se estiman los formantes libres de la influencia del pitch $\hat{F}_k$ con el método LPC descrito en la sección 2.2.5 y con los mismos parámetros (orden de predicción $P = 8$, muestreo $F_s = 8000\ Hz$ y una ventana de análisis de 30 ms).

La estimación de los formantes mejora para todas las vocales con esta técnica [4]. Las vocales /e/ y en especial la /i/ son las que mayor mejora presentan debido a que sus formantes son los más extremos y el primero tiende a estimarse sobre la frecuencia de pitch y el segundo en su primer armónico.

3.2 Estimación de la longitud del tracto vocal y normalización

Existe una alta variabilidad en los formantes para ambos sexos, especialmente en las voces masculinas por el gran cambio que experimentan en la voz durante la pubertad [4]. Es necesario poder llevar los formantes de la voz infantil a un espacio de trabajo más homogéneo, donde exista menor variabilidad, para poder desarrollar aplicaciones independientemente del sexo y la talla del niño. Esto se consigue normalizando los formantes calculados mediante la estimación de la longitud del tracto vocal (VTL) del locutor. Para calcular esta longitud, se parte del modelo del tracto vocal descrito en el Anexo III.

3.2.1 Estimación de la longitud del tracto vocal

Es importante conocer la VTL de los niños en función de su talla para poder relacionarla con los formantes y encontrar un modelo que refleje su comportamiento en función de la talla y sexo.

Pocos estudios relacionan el crecimiento del niño con el crecimiento de su tracto vocal. [13] y [14] relacionan el crecimiento del tracto vocal con la edad, pero lo deseable es relacionar la VTL con la talla y no con la edad.

Se trató pues de estimar la VTL a partir de la emisión de un sonido sonoro, en el que el tracto vocal esté configurado homogéneamente como en el modelo descrito en el Anexo III. Después, se estiman los formantes y se obtiene la VTL según (3.4).

$$l = \frac{c}{4fn}(2n-1) \tag{3.4}$$

Pero el sonido que se debe articular es la vocal francesa /æ/ que se ubica en el centro de masas del triángulo vocálico, el cual se obtiene calculando la media de todos los F1 y todos los F2 de las cinco vocales.

Se utiliza pues, un método que parte de los formantes vocálicos y da como resultado la longitud de un tubo homogéneo, con formantes teóricos que caen próximos al centro de masa del triángulo vocálico ya que se consideran los formantes de las cinco vocales.

Algunos autores [15] proponen la estimación del tracto vocal a partir de la impedancia de los labios o a partir de las áreas del tracto vocal en los modelos de concatenación de tubos como [16], [17] o [18]. El método que se utiliza es el propuesto por [19] donde se estima la longitud del tracto vocal a partir de emisiones vocálicas. En esta técnica se

parte del tubo uniforme del anexo III donde las resonancias (III.13) se encuentran equiespaciadas. El objetivo es ajustar las frecuencias de resonancia medidas $\check{F}_k$, con las frecuencias de resonancia del tubo uniforme que están determinadas por su longitud.

Se trata de reducir al mínimo el error ε:

$$\varepsilon = \sum_{k=1}^{M} D(\check{F}_k, (2k-1)f1) = \sum_{k=1}^{M} D(\check{F}_k, (2k-1)\frac{c}{4l}) \tag{3.5}$$

Donde $\sum_{k=1}^{M} D(\check{F}_k, (2k-1)f1)$ es la distancia entre los formantes medidos $\check{F}_k$ y las resonancias impares del tubo uniforme $(2k-1)f1$:

$$\varepsilon = \sum_{k=1}^{M} \frac{\left(\frac{\check{F}_k}{2k-1} - f1\right)^2}{f1} \tag{3.6}$$

Minimizando (3.6) se obtienen la frecuencia fundamental f_1 del tubo homogéneo.

$$f1 = \left(\frac{1}{M}\sum_{k}\left(\frac{\check{F}_k}{2k-1}\right)^2\right)^{1/2} \tag{3.7}$$

La VTL se obtiene de (3.4) y de (3.7):

$$VTL = \frac{c}{4f1} \tag{3.8}$$

Donde c es la velocidad del sonido ($34000\ cm/s$).

Aplicándose esta técnica, en [4] se estiman las longitudes para cada vocal como el promedio de las longitudes de todas sus tramas. Se calcula la media de las longitudes de todas las vocales para obtener la VTL final de cada locutor.

3.2.2 Normalización de Formantes

Con la VTL calculada en la sección anterior, es posible llevar los formantes a un espacio normalizado donde disminuya su variabilidad.

La técnica de normalización se basa en la hipótesis de que la configuración del tracto vocal durante la emisión de vocales es semejante en todos los individuos y solo se diferencia en la longitud. Se calculan los formantes de un tubo acústico variando su

longitud a una longitud de referencia l_R de valor 17.5 cm. Los formantes normalizados F_{kN} se obtienen a partir de los formantes calculados $\check{F}_k$ siguiendo (3.9).

$$F_{kN} = \frac{LTV}{l_R}\check{F}_k \quad (k = 1, \dots, M) \tag{3.9}$$

En [4] se demuestra que utilizando esta técnica, los formantes se encuentran menos dispersos tras la normalización para ambos sexos, demostrando que el proceso de normalización de formantes reduce considerablemente la variabilidad inter-locutor y dota de robustez a las técnicas de procesado cuando se trabaja con voz infantil.

El diagrama de la Figura 3.2 resume el procesado completo de la señal de voz.

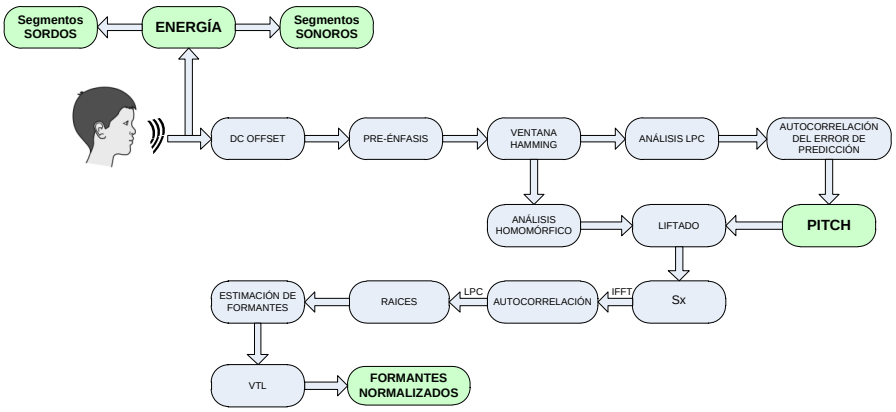


Figura 3.2. Diagrama de bloques para la estimación robusta de formantes.

Es posible diseñar herramientas para la terapia de voz, como se describe en los siguientes capítulos, utilizando las técnicas descritas hasta el momento.

4. Descripción de la Aplicación

El objetivo del proyecto es desarrollar herramientas libres de asistencia a la logopedia que ayuden a los profesionales a educar la voz de pacientes con patologías en el lenguaje debidas a algún tipo de discapacidad o malformación.

Las tecnologías del habla permiten el desarrollo de este tipo de herramientas que han sido demandadas por profesionales de educación especial pero que, la no existencia de aplicaciones libres y en español, han hecho que no se hayan aprovechado este tipo de técnicas para este fin.

Este proyecto está enmarcado dentro del Grupo de Tecnologías de las Comunicaciones (GTC) en colaboración con el Colegio Público de Educación Especial Alborada (CPEE Alborada) siguiendo la línea de proyectos realizados con anterioridad en este mismo grupo. En concreto, sigue la línea de PreLingua [4] ,ya que la aplicación se desarrolla con el objetivo de trabajar los aspectos del pre-lenguaje.

Se ha diseñado un editor de actividades totalmente configurable, que permite a los logopedas diseñar sus propias actividades, con el fin de adaptarlas personalmente a cada paciente y poder compartirlas con otros profesionales para que sean utilizadas en pacientes de características similares.

4.1 PreLingua 2

PreLingua 2 es un editor de actividades que se ha diseñado con el objetivo de ser utilizado como herramienta de asistencia a la logopedia. Es altamente configurable y permite a los profesionales del ámbito diseñar sus propias actividades, sin necesidad de conocimientos técnicos, adaptándolas a cada paciente en particular y con la posibilidad de poder guardar la configuración de la actividad y compartirla a través de la red.

Uno de los problemas que aparecían en aplicaciones anteriores era la incompatibilidad entre sistemas operativos. Estas aplicaciones estaban diseñadas para ser utilizadas en sistemas Windows y, además, era necesario instalarlas en cada uno de los equipos en los que se iba a utilizar. Uno de los objetivos del proyecto es evitar este problema y, para ello, se ha diseñado una aplicación que se pueda ejecutar en una plataforma distribuida, es decir, que se pueda acceder a través de la Web sin necesidad de instalarla en el ordenador. Para conseguirlo, el editor se ha desarrollado en lenguaje Java [20], utilizando el entorno de desarrollo integrado libre NetBeans y su editor gráfico basado en Swing para la parte gráfica.

Otro de los objetivos del proyecto es que el editor sea totalmente configurable. Las aplicaciones desarrolladas anteriormente consistían en actividades que trabajaban en forma de juegos cerrados. La aplicación desarrollada en este proyecto es totalmente configurable, cada logopeda puede elegir qué aspecto de la voz desea trabajar, combinarlos como desee y adaptar los parámetros a las necesidades de cada paciente.

Los aspectos que se trabajan en la aplicación son las habilidades pre-lingüísticas o pre-lenguaje que son las características que se adquieren durante el primer año de vida. El correcto desarrollo de estas habilidades permitirá la evolución del lenguaje del niño para la comunicación en el futuro con total normalidad. Las tecnologías del habla permiten trabajar estos aspectos del pre-lenguaje utilizando las técnicas de procesado de señal de voz descritas en los capítulos anteriores. Las características con las que se va a trabajar son:

- Detección de actividad de voz.
- Control de la intensidad.
- Control del soplo.
- Control del tono.
- Vocalización.

4.2 Estructura de la aplicación

La Figura 4.1 representa la estructura del sistema.

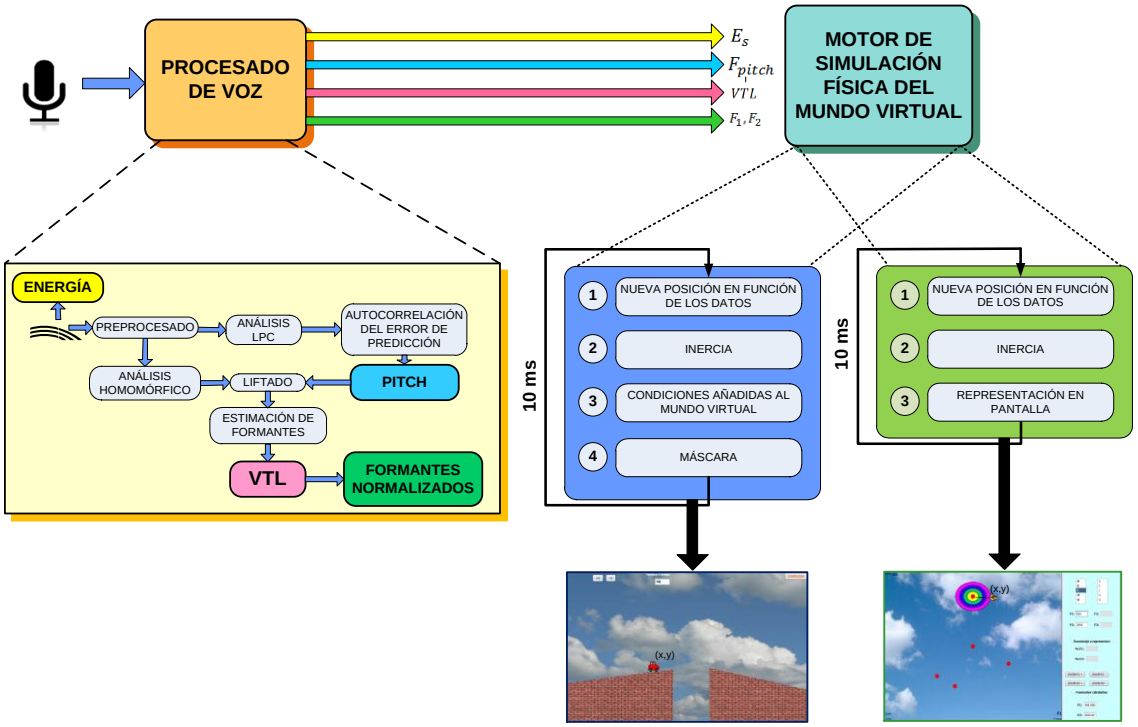


Figura 4.1: Estructura del sistema.

Dos son los grandes bloques de los que consta la aplicación:

Bloque de procesado

El bloque de procesado analiza la señal captada del micrófono y la procesa en tiempo real, extrayendo los parámetros que caracterizan la voz del locutor.

Motor de simulación física

Las actividades finales consisten, básicamente, en un objeto que el niño mueve con la voz sobre una imagen de fondo. El motor de simulación física del entorno virtual o Juego consiste en un bucle, que se repite continuamente y que calcula la posición del objeto en el mundo virtual diseñado, en función de los parámetros obtenidos en el procesado y de las condiciones que configuran la física del mundo.

4.2.1 Bloque de procesado

Este bloque utiliza las técnicas de procesado de voz descritas en los capítulos 2 y 3.

En primer lugar, con una frecuencia de muestreo de 8000 Hz, se obtiene la señal de voz del micrófono en tramas de 80 muestras. Se realiza un pre-procesado de esta señal de entrada para adecuarla a su posterior tratamiento (sección 2.2.2). Este pre-procesado consiste en la preparación de la trama de análisis de 240 muestras con un desplazamiento de 80 muestras, la compensación DC, el filtro de preénfasis y el enventanado de la señal para el análisis localizado (Figura 2.3).

La **energía** se calcula a partir de la señal de entrada y permite la detección de actividad de voz y la discriminación entre sonidos sordos y sonoros (sección 2.2.3).

Mediante el análisis LPC (sección 2.2.5) se obtendrá la señal residual o error de predicción que permitirá, a partir de su autocorrelación, estimar el **pitch** (Figura 2.6).

Para el cálculo de los **formantes** es necesario tener en cuenta que la aplicación está dirigida a población infantil y la problemática que este tipo de voz presenta (sección 3.3.1). Es necesario eliminar la influencia del pitch para obtener una estimación robusta de los formantes (sección 3.1.2). Mediante el análisis homomórfico, se separan las dos componentes de la señal: la información del filtro o tracto vocal y la información de la señal de excitación o pitch (Figura 2.8). Realizando un liftado se elimina la influencia del pitch y se obtiene una estimación robusta de los formantes.

Por último, se calcula la longitud del tracto vocal (**VTL**) para la normalización de los formantes (Figura 3.2) que reduce su alta variabilidad en voz infantil (sección 3.2).

Así pues, a la salida de este bloque, se obtiene un vector $\boldsymbol{P}$ con los parámetros extraídos de la señal de voz (4.1).

$$\boldsymbol{P} = [E, pitch, F1, F2, F3, F4, VTL] \tag{4.1}$$

Este vector es la entrada al siguiente bloque, el motor se simulación.

4.2.2 Motor de simulación física

El segundo gran bloque del sistema es el motor de simulación física del entorno virtual o Juego. Se crea un mundo virtual que simula el movimiento físico de un objeto en tiempo real. Esta simulación consiste en la repetición, cada 10 ms y durante toda la duración de la actividad, de un bucle que calcula las coordenadas (x, y) del objeto en

movimiento. Estas coordenadas serán función de los datos obtenidos en el procesado y de las condiciones añadidas al mundo virtual.

Se pueden configurar dos tipos de actividades y cada una seguirá un algoritmo para calcular la posición del objeto en movimiento.

ACTIVIDADES DE PRELENGUAJE

El bucle sigue el siguiente algoritmo para conseguir el movimiento deseado (Figura 4.2). Cada paso del algoritmo genera un tipo de movimiento como salida.

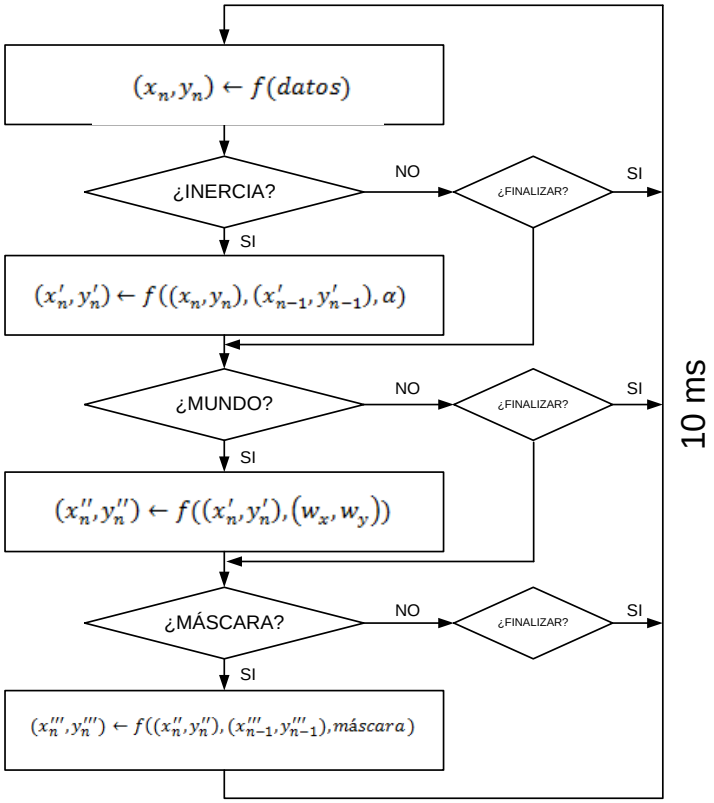


Figura 4.2: Algoritmo del motor de simulación.

PASO 1: Nueva posición en función de los datos.

Este primer paso consiste en la traducción directa de los parámetros de la voz del locutor obtenidos en el procesado (4.1), con un parámetro físico del mundo virtual.

$$(x_n, y_n) \leftarrow f(\boldsymbol{P}) \tag{4.2}$$

Se pueden obtener dos tipos de movimiento en este primer paso, movimiento si existe detección del parámetro de voz especificado y movimiento en función del valor de estos parámetros.

En primer lugar, el movimiento más simple se obtiene incrementando la posición de un objeto en una trayectoria rectilínea, en función de la detección del parámetro de la voz indicado.

$$(x_n, y_n) = \begin{cases} (x_{n-1}, y_{n-1}) + (\Delta x, \Delta y) & si\ P[i],\ P[j]\ activa \\ (x_{n-1}, y_{n-1}) & si\ P[i],\ P[j]\ no\ activa \end{cases} \quad (4.3)$$

Dependiendo del tipo de desplazamiento $(\Delta x, \Delta y)$, este movimiento se puede dividir a su vez en dos subtipos. El primer subtipo consiste en aplicar un desplazamiento constante si se detecta la presencia de algún parámetro de $\boldsymbol{P}$.

Un ejemplo de este tipo sería una actividad en la que un objeto se mueve únicamente si la energía supera cierto umbral definido. Este tipo de actividades presenta buenos resultados en áreas como la estimulación temprana pues es capaz de captar la atención de niños con discapacidades cognitivas importantes.

El segundo subtipo es el que relaciona el valor del desplazamiento $(\Delta x, \Delta y)$ directamente con un parámetro extraído en $\boldsymbol{P}$**.** Se consigue así una dependencia de la voz con la velocidad del movimiento.

$$(x_n, y_n) = \begin{cases} (x_{n-1}, y_{n-1}) + f(P[i]) & si\ P[i], P[j]\ activa \\ (x_{n-1}, y_{n-1}) & si\ P[i], P[j]\ no\ activa \end{cases} \quad (4.4)$$

$P[i]$ y $P[j]$ indican distintos parámetros extraídos durante el procesado. En la actividad, se pueden combinar varios de estos parámetros en su configuración.

Tomando el ejemplo anterior, el objeto se moverá si la energía detectada supera cierto umbral y, además, la velocidad del objeto dependerá de la intensidad de la voz. Si el niño habla más alto, el objeto se moverá más rápido. Este tipo de actividades ayuda a que el niño sea capaz de controlar la intensidad de su voz.

El segundo tipo de movimiento que surge en este paso, se obtiene mediante la correspondencia directa del dato obtenido con las coordenadas (x, y) del objeto en el mundo virtual. Una transformación de las coordenadas será necesaria para adaptarlas a los límites del entorno.

$$(x_n, y_n) \leftarrow f\left(E_s, F_{pitch}, F1, F2\right) \quad (4.5)$$

Como ejemplo se propone una actividad que combina la detección de energía y el valor del pitch (Figura 4.3). El objeto (pez) se mueve horizontalmente ($eje\ x$) en función de la detección de voz (E_s) y la posición en el $eje\ y$ depende del valor del pitch. El locutor entona las notas musicales, lo que se observa en la trayectoria como una especie de escalera ascendente. Se puede comprobar la existencia de valores espurios. Estos valores se pueden reducir, como se verá en el siguiente paso.

Figura 4.3: Actividad de control de tonalidad.

PASO 2: Inercia.

En este segundo paso, la nueva posición calculada depende de la posición en el instante anterior y de un parámetro α que toma valores entre 0 y 1 en función del efecto que se quiera conseguir.

$$(x'_n, y'_n) \leftarrow f((x_n, y_n), (x'_{n-1}, y'_{n-1}), \alpha) \tag{4.6}$$

Para los dos primeros subtipos del paso 1, se aplica esta dependencia con α al desplazamiento del objeto. El objetivo es conseguir un efecto de inercia, de manera que el movimiento se prolonga en ausencia de actividad de voz hasta que se agota, creando el efecto de movimientos más suaves.

$$(\Delta x'_n, \Delta y'_n) = (\Delta x_n, \Delta y_n) \cdot \alpha + (\Delta x'_{n-1}, \Delta y'_{n-1}) \cdot (1 - \alpha) \tag{4.7}$$

$$(x'_n, y'_n) = (x'_{n-1}, y'_{n-1}) + (\Delta x'_n, \Delta y'_n) \tag{4.8}$$

Para valores más bajos de α, el movimiento se prolonga durante más tiempo. Se asignará este valor en función del efecto deseado.

En el caso de que las coordenadas dependan directamente de un parámetro de voz, la dependencia con α se aplica directamente a x e y. Esta dependencia consigue un suavizado del movimiento, evitando datos espurios en la trayectoria. En esta ocasión, se requieren valores de α mayores.

$$(x'_n, y'_n) = (x_n, y_n) \cdot \alpha + (x'_{n-1}, y'_{n-1}) \cdot (1 - \alpha) \tag{4.9}$$

Se propone el mismo ejemplo que en el paso 1 pero añadiendo ahora este efecto de inercia (Figura 4.4). Se asignan distintos valores de α, un valor para el movimiento horizontal de detección y otro para el movimiento vertical en función del valor del pitch. El valor de α para el movimiento horizontal será menor que para el movimiento vertical. En este caso, el locutor aumenta y reduce la tonalidad de su voz. La parte izquierda (a) se corresponde con esta actividad en ausencia del efecto de inercia, se observa la existencia de valores espurios. En la parte derecha (b) se aprecia la mejora que se obtiene al introducir este efecto.

(a) (b)

Figura 4.4: Actividad de control de tonalidad con inercia.

Aunque este efecto ayuda al niño a alcanzar su objetivo, los resultados son positivos ya que también es importante que se sientan motivados en la realización de la actividad.

PASO 3: Condiciones añadidas al mundo virtual.

Conseguido un movimiento más uniforme, se pueden añadir condiciones adicionales al entorno virtual que simulen efectos físicos del mundo real. Este paso dota de más realismo y de más dificultad a la actividad.

$$(x_n'', y_n'') \leftarrow f\left((x_n', y_n'), (w_x, w_y)\right) \tag{4.10}$$

Se puede simular un efecto de gravedad añadiendo un movimiento continuo hacia abajo o un efecto de corriente mediante un movimiento continuo horizontal. El niño deberá superar esta adversidad física con su voz. La dificultad de la actividad recaerá en el valor del paso que se le de al movimiento continuo, el cual se traduce en la velocidad con la que el objeto caerá o se desplazará por el efecto de la corriente.

$$(x_n'', y_n'') = (x_n', y_n') + (w_x, w_y) \tag{4.11}$$

Como ejemplo (Figura 4.5) se muestra un objeto (globo) que cae continuamente hacia abajo. El usuario, deberá impedir que el globo caiga al suelo mediante, por ejemplo, el soplo.

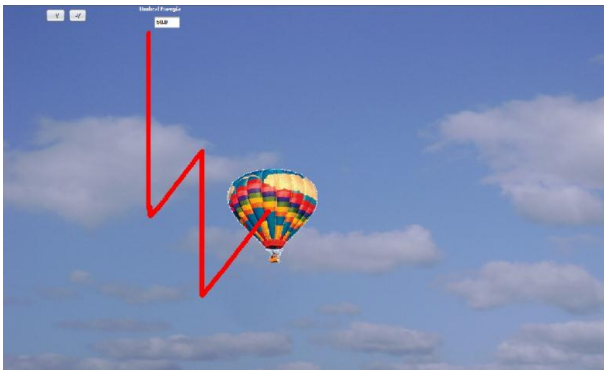


Figura 4.5: Actividad de control de soplo con condiciones física adicionales.

Este tipo actividades de control del soplo ayuda mejorar el control de la respiración, necesario para un hablar fluido.

PASO 4: Máscara.

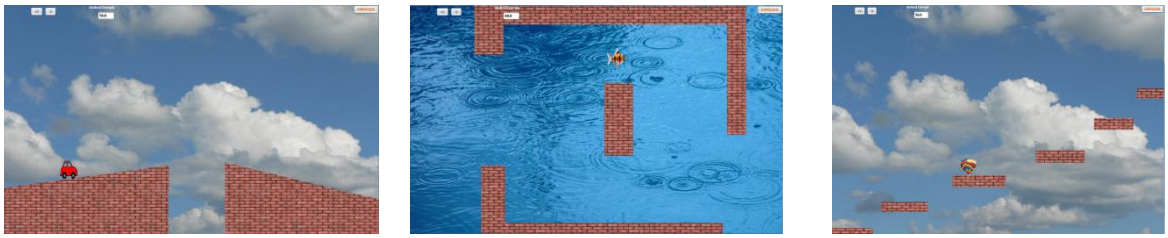
En este último paso se incorpora un nuevo elemento al entorno virtual, la máscara. Una máscara consiste en una imagen que se superpone al fondo. Esta imagen contiene zonas transparentes, donde el objeto podrá moverse, y zonas opacas que actúan de barreras físicas impidiendo el paso del objeto. La inserción de máscaras equivale a una introducción de obstáculos en el mundo virtual.

$$(x_n''', y_n''') \leftarrow f(\,(x_n'', y_n''), (x_{n-1}''', y_{n-1}'''), máscara) \tag{4.12}$$

Antes de calcular la nueva posición del objeto, se comprueba si es una posición válida, esto es, una posición que no esté ocupada por la máscara (m_x, m_y). En el caso de que sea válida, la nueva posición se hará efectiva. Si la coordenada calculada está ocupada por la máscara, el objeto no se moverá (4.13). Se consigue así un efecto de barrera.

$$(x_n''', y_n''') = \begin{cases} (x_n'', y_n'') & si\ máscara(x_n'', y_n'') == ok \\ (x_{n-1}''', y_{n-1}''') & otro\ caso \end{cases} \tag{4.13}$$

El diseño de máscaras hace que las posibilidades a la hora de crear una actividad sean innumerables. El logopeda diseñará estas máscaras dependiendo del efecto que quiera conseguir, la patología de la voz del paciente a tratar y la dificultad que quiera añadir. Se muestran en la Figura 4.6 ejemplos de máscaras.



(a) (b) (c)

Figura 4.6: Actividades con máscara.

En la primera máscara (a), el niño deberá superar el obstáculo que se le presenta impidiendo que el coche caiga por el agujero que se encontrará en su camino. Aplicando el efecto de inercia, el locutor deberá impulsar el coche mediante el soplo o mediante la energía. La segunda máscara es un laberinto, aplicando distintos tipos de movimiento a los diferentes parámetros se consigue que el niño consiga superar el laberinto mediante el control de su voz. Por ejemplo, aplicando un movimiento continuo hacia abajo (w_x, w_y), un movimiento horizontal asignado al soplo (Δx) y un movimiento vertical en función del valor del pitch. La última máscara consiste en unas escaleras que el niño subirá con cada golpe de voz. Este tipo de actividad es muy positiva en problemas de tartamudez pues permite controlar los periodos de actividad de voz y los periodos de respiración del paciente. Además, mediante el diseño de la máscara, se controla el espaciado entre escaleras y el nivel de éstas.

Las máscaras permiten al logopeda una total flexibilidad pues las pueden dibujar con cualquier herramienta de dibujo convencional. Por ejemplo, para el diseño de las máscaras del ejemplo se ha utilizado la herramienta gratuita Gimp[1].

ACTIVIDADES DE VOCALIZACIÓN

La pronunciación de la vocal a trabajar hará que el objeto en movimiento (dardo) se acerque al centro de una imagen. El bucle seguirá los siguientes pasos para calcular la posición del objeto.

PASO 1: Nueva posición en función de los datos.

En el caso de las actividades de vocalización, la posición del objeto será una correspondencia directa con el valor de los dos primeros formantes F_1 y F_2.

$$(x_n, y_n) \leftarrow f(F_1, F_2) \tag{4.14}$$

[1] http://www.gimp.org/

La diferencia entre los distintos tipos de actividades reside en la transformación que se aplica a F_1 y F_2 para calcular la posición del objeto en la pantalla.

PASO 2: Inercia.

Este paso es igual al del algoritmo anterior.

$$(x'_n, y'_n) \leftarrow f((x_n, y_n), (x'_{n-1}, y'_{n-1}), \alpha) \tag{4.15}$$

En estas actividades, el efecto conseguido es un suavizado del movimiento mediante la disminución de la influencia de datos espurios.

PASO 3: Representación en pantalla.

Los distintos tipos de actividades de vocalización se diferencian, técnicamente, en su posición dentro de la pantalla.

Si la actividad es de tipo Imágenes (Figura 4.7), el centro de la pantalla se hace corresponder con los formantes teóricos introducidos por el logopeda. Los valores máximos y mínimos de F_1 y F_2 se hacen corresponder con los límites de la pantalla. Estos máximos y mínimos pueden cambiarse durante la ejecución de la actividad consiguiendo así un efecto de zoom sobre la pantalla.

$$x = (anchura\_ventana)/(F_{1\_max} - F_{1\_min})) * (\hat{F}_1 - F_{1\_min}) \tag{4.16}$$

$$y = (altura\_ventana)/(F_{2\_min} - F_{2\_max})) * (\hat{F}_2 - F_{2\_max}) \tag{4.17}$$

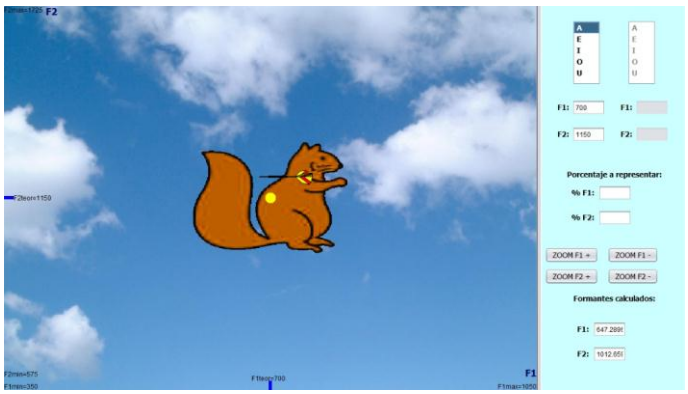


Figura 4.7: Actividad de trabajo vocálico de tipo Imágenes.

Las actividades de tipo Dianas y Transiciones establecerán estos valores de F_1 y F_2 máximos más altos y mínimos mas bajos para poder representar las cinco vocales

sobre la pantalla. Una imagen aparecerá en la posición de la vocal a trabajar según los formantes introducidos por los logopedas. Recorriendo todas las vocales, se consigue definir el triángulo vocálico particular de cada paciente (Figura 4.8).

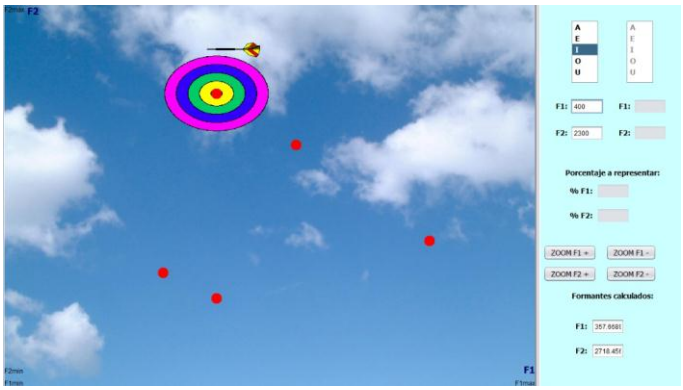


Figura 4.8: Actividad de trabajo vocálico de tipo Dianas.

4.2 Interfaces de usuario

La aplicación funciona principalmente con la voz del paciente, por lo que será necesario un micrófono que convierta la señal de voz en una señal eléctrica y, así, poder capturarla y almacenarla para su posterior procesado. Además, serán necesarias las interfaces habituales para controlar cualquier aplicación como el monitor, el ratón y el teclado para la configuración y visualización de la actividad.

Dos son los tipos de usuarios que van a interactuar con la aplicación. Por un lado, el logopeda o profesional del habla que se encarga de diseñar la actividad en función de las necesidades del paciente. Por otro lado, el usuario final que es la persona con discapacidad o con dificultad en la producción de voz y a la cual va dirigida la actividad.

Para la parte de configuración, la interfaz gráfica consiste en una ventana formada por varios paneles numerados que guían al logopeda en el diseño de una actividad (Figura 4.9).

Como objetivo del proyecto, el editor debía de ser altamente configurable. Dentro de cada panel, se han introducido controles sencillos (botones, checkbox, combobox, tablas y listas) para todas las variables que se pueden controlar dentro del mundo virtual. Además, cada panel tiene una pestaña adicional (Figura 4.10) que permite un diseño avanzado de la actividad mediante la asignación de valores concretos a estas variables controlables.

La interfaz final está formada, básicamente, por una imagen de fondo y un objeto que se mueve sobre él. El usuario será un niño con discapacidad y se deben tener en cuenta las limitaciones que éstos presentan. Como cada paciente tendrá sus propias limitaciones, el logopeda puede elegir convenientemente las imágenes a utilizar. Las figuras 4.4, 4.5, 4.6, 4.7 y 4.8 son ejemplos de esta interfaz.

Figura 4.9: Interfaz de configuración.

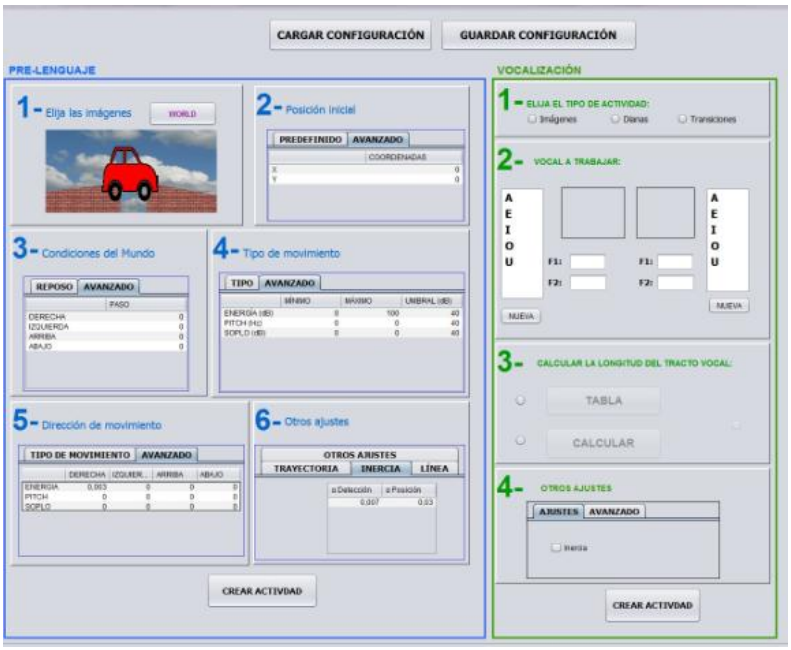


Figura 4.10: Interfaz de configuración avanzada.

4.3 Estructura de clases Java

La aplicación se ha programado en Java. La Figura 4.11 es un esquema de las clases que se han definido para desarrollar el editor de actividades. Se han agrupado según su funcionalidad.

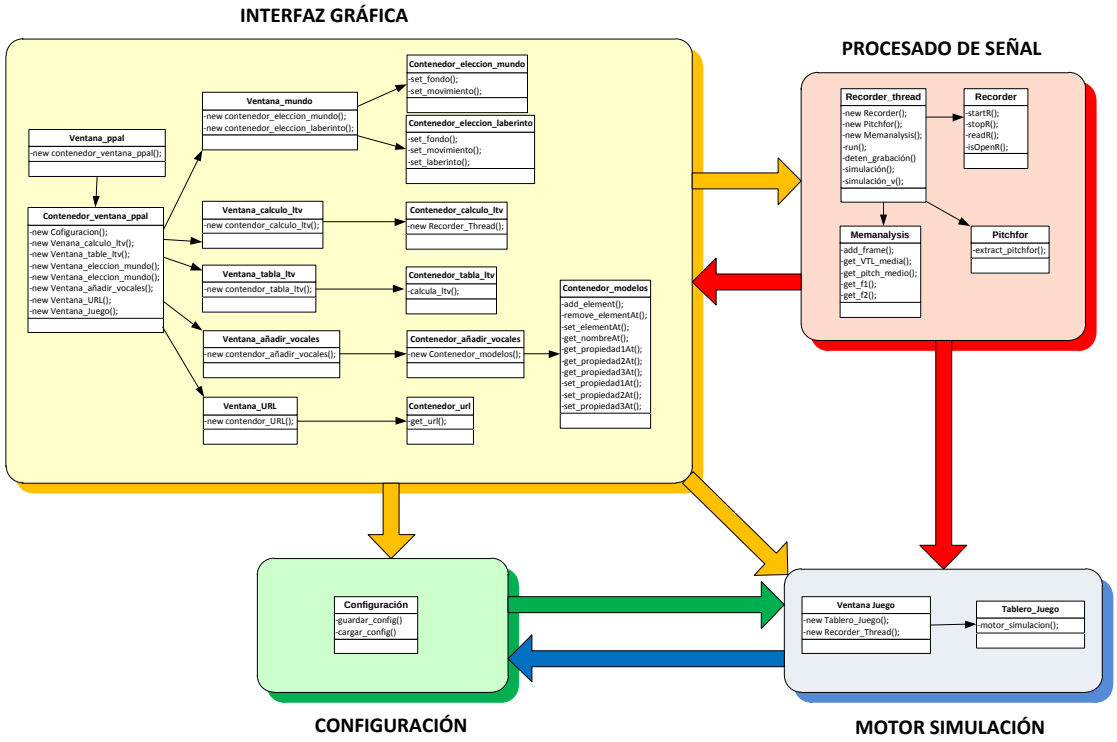


Figura 4.11: Estructura de clases.

4.3.1 Clases de la interfaz gráfica

Estas clases son las relacionadas con la interfaz gráfica y permiten la configuración de una nueva actividad.

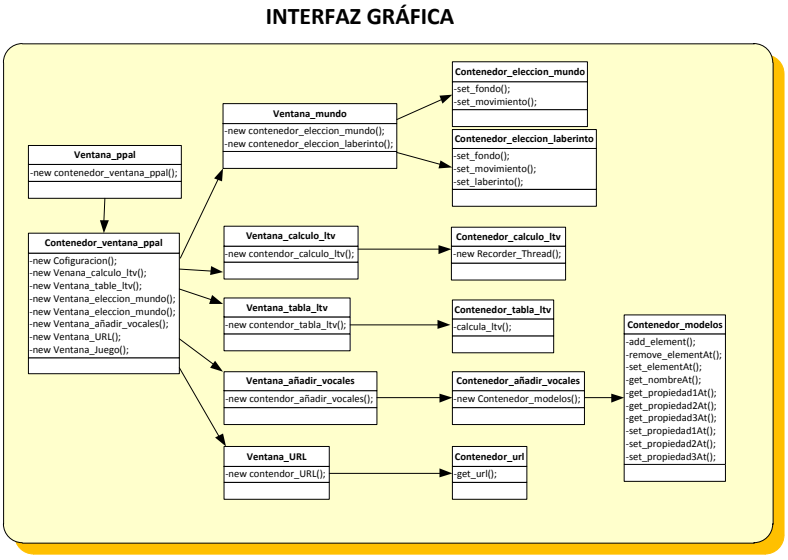


Figura 4.12: Clases de la interfaz gráfica.

Ventana_ppal y Contenedor_ventana_ppal

Forman la ventana principal de configuración donde el logopeda introducirá los parámetros para el diseño de la actividad (Figura 4.9). Desde aquí se podrá cargar una configuración y guardar la que se ha diseñado.

La clase Contenedor_ventana_ppal llama a la clase Configuración para cargar una configuración que esté almacenada en un archivo XML o guardar la configuración diseñada.

Una vez se ha diseñado la actividad o se ha cargado una configuración, se crea la ventana de juego (Clase Ventana_Juego).

Esta clase se apoya en las siguientes para la obtención de todos los parámetros que componen el diseño de una nueva actividad.

Ventana_eleccion_mundo y Contenedor_eleccion_mundo

Permiten la elección de las imágenes utilizadas en la actividad mediante las clases Elección_mundo y Elección_laberinto. Estas imágenes se pueden elegir entre las propuestas en la aplicación o cargar nuevas imágenes desde archivo o desde Internet.

Ventana_URL y Contenedor_URL

Son las que se encargan de cargar la imagen desde internet. Solo es necesario introducir la URL de la imagen y esta clase la almacena para su utilización.

Ventana_calculo_ltv y contenedor_ calculo_ltv

Para las actividades de vocalización es necesario normalizar los formantes obtenidos en el procesado para reducir su variabilidad. Esta normalización se realiza utilizando la longitud del tracto vocal (VTL) (sección 3.2). Se han propuesto dos métodos para la obtención de este parámetro, calcularlo directamente de la voz del niño u obtenerlo de una tabla. Para el primer caso, esta clase graba la voz del niño (clase Recorder) que tiene que emitir las 5 vocales para que el cálculo sea válido. Esta señal es procesada (clase Pitchfor) y se obtiene la LTV como la media de este valor durante todo el procesado (clase Memanalysis).

Ventana_tabla_ltv y contenedor_tabla_ltv

La segunda opción para calcular la VTL es obtenerlo de una tabla. Basándose en el sexo del niño y su altura se calcula la VTL [4].

Ventana_añadir_vocales y Contenedor_añadir_vocales

Durante el diseño de la herramienta, surgió la posibilidad de hacer que ésta no fuera válida únicamente para el español. Definiendo nuevas vocales, se permite que esta herramienta se adapte a otros idiomas. Esta clase permite añadir nuevas vocales introduciendo los formantes que las definen y almacenándolos.

4.3.2 Clase configuración

Este bloque solo contiene la clase Configuración. Es la clase que almacenan todos los datos seleccionados en la pantalla principal de configuración y los pone a disposición del motor de simulación para calcular la posición del objeto en movimiento.

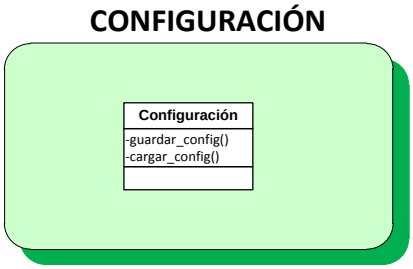


Figura 4.13: Clase Configuración.

Esta clase permite la serialización de los datos, es decir, permite guardar una configuración diseñada por el logopeda en un archivo XML. Asimismo, permite la carga de archivos en este formato, de manera que es posible guardar una configuración y cargarla en cualquier otro momento, así como compartirla con otros usuarios.

4.3.3 Clases para el procesado de señal

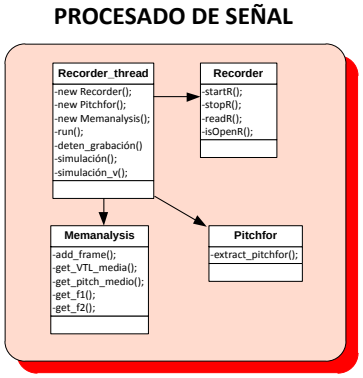


Figura 4.14: Clases para el procesado de señal.

Recorder_thread

Esta clase es un hilo de ejecución que permite la obtención de la señal de micrófono, su procesado y la simulación del movimiento físico mientras se ejecuta toda la parte gráfica. Esto hará que todo el proceso se realice en tiempo real.

Recorder

La clase Recorder se encarga de obtener la señal de audio del micrófono en tramas de 80 muestras y con una frecuencia de muestreo de 8000 Hz. La señal obtenida se adapta y almacena para su posterior procesado.

Pitchfor

Esta clase procesa la señal obtenida en la clase anterior siguiendo el esquema de la Figura 3.2 y extrae los parámetros característicos de la voz. La trama de análisis está formada por 240 muestras y un desplazamiento de 80 que permite el análisis localizado de la señal de voz de entrada. Por cada trama que se procesa, esta clase proporciona un vector de salida con los parámetros de voz obteniéndose un cálculo de éstos en tiempo real.

Memanalysis

La clase Memanalysis almacena los parámetros extraídos de las tramas procesadas. Permite también la obtención de sus valores medios en el caso de ser necesarios.

4.3.4 Clases para el motor de simulación

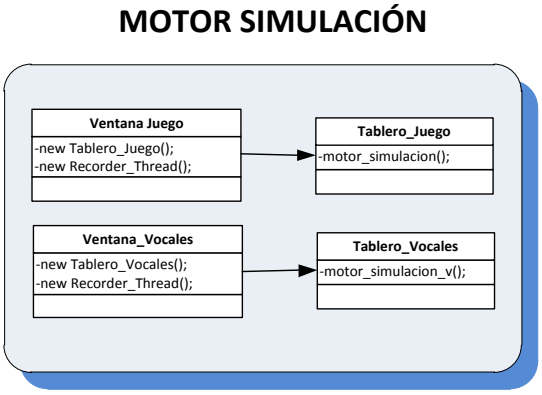


Figura 4.15: Clases del motor de simulación física.

Ventana_Juego y Tablero_Juego

La clase Ventana_Juego es la que crea el hilo de ejecución de la clase Recorder_thread descrita anteriormente.

El contenedor de esta ventana es de la clase Tablero_Juego y es el motor de simulación física propiamente dicho. Esta clase calcula la posición del objeto en movimiento siguiendo el algoritmo de la Figura 4.2 y basándose en los datos de la clase Configuración y los parámetros obtenidos en la clase Pitchfor.

Las tareas que lleva a cabo esta clase son:

- Obtención de la posición inicial.
- Obtención de la máscara.
- Cálculo de la siguiente posición.
- Aplicación del efecto de inercia.
- Aplicación de las condiciones adicionales del mundo virtual.
- Comprobación de que la posición calculada es una posición correcta según la máscara.
- La nueva posición se hace efectiva.

Ventana_Vocales y Tablero_Vocales

En el caso de que la actividad sea de tipo Vocalización, estas dos clases compondrán el motor de simulación. Como en las clases anteriores, Ventana_Vocales iniciará el hilo de ejecución y Tablero_Vocales, su contenedor, se encarga de calcular el movimiento.

4.4 Configuración de actividades

El editor diseñado es altamente configurable. Se ha diseñado de tal manera que un logopeda puede programar sus propias actividades sin necesidad de tener conocimientos técnicos, siguiendo los pasos que se indican en los paneles numerados de la ventana principal (Figura 4.9). En esta sección se detallan los pasos a seguir en la configuración de los dos grandes tipos de actividades, las del pre-lenguaje y las de vocalización.

4.4.1 Actividades de Pre-lenguaje

Este tipo de actividades se configura en la parte izquierda de la ventana principal (Figura 4.9). Estos son los pasos a seguir:

PASO 1: Elección del mundo.

Figura 4.16: Elección del mundo.

En este primer paso se eligen las imágenes que se van a utilizar. Se debe seleccionar una imagen de fondo y otra que será el objeto en movimiento. En caso de que la actividad utilice máscaras también se seleccionará en este paso.

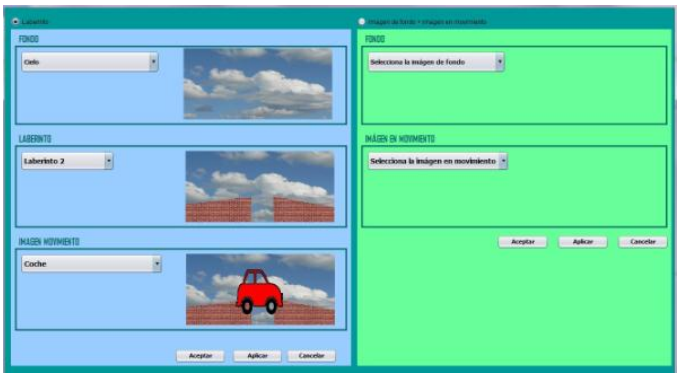
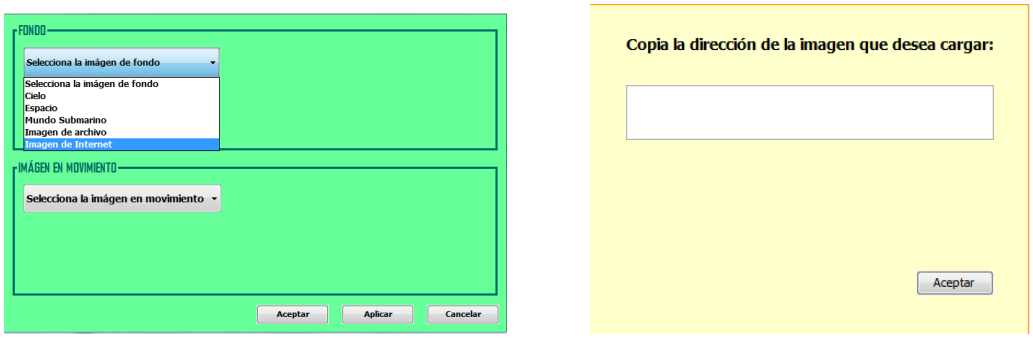


Figura 4.17: Elección de imágenes.

Además de las imágenes que se proponen en la aplicación, es posible cargarlas desde archivo (Figura 4.18 (a)) o desde Internet mediante su URL (Figura 4.18 (b)).

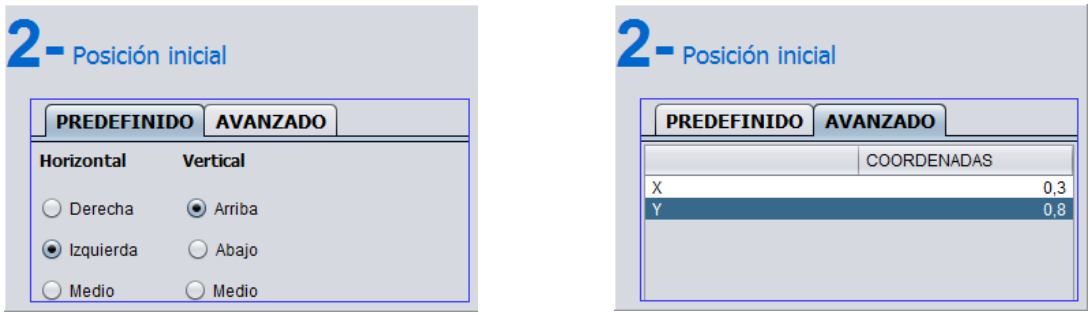


(a) (b)

Figura 4.18: Carga de nuevas imágenes.

PASO 2: Posición inicial.

Una vez elegidas las imágenes, se decidirá la posición en la que el objeto comenzará su movimiento (Figura 4.19). Además de poder elegir entre las posiciones sugeridas (a), es posible indicar la coordenada (x_0, y_0) donde iniciar la actividad en la pestaña "AVANZADO" (b). Todos los valores introducidos en las tablas toman un valor entre 0 y 1.

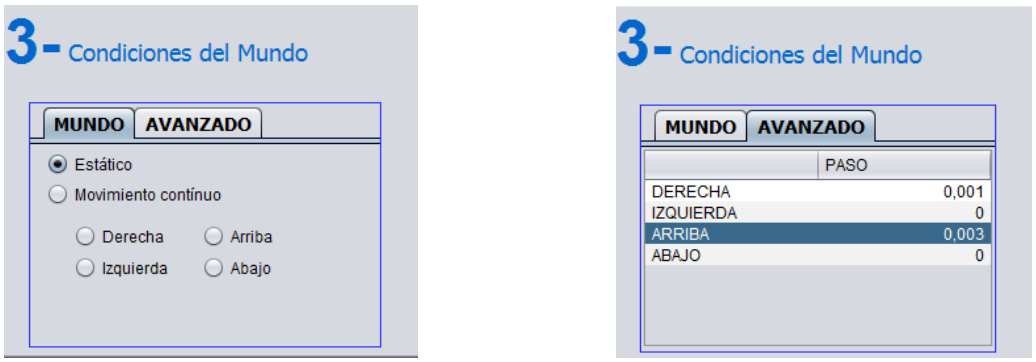


(a) (b)

Figura 4.19: Elección de la posición inicial.

PASO 3: Condiciones del mundo.

En este paso se añaden las condiciones externas que tendrá el mundo, las cuales aportarán un movimiento adicional al provocado por la voz. Aquí se le da valor a (w_x, w_y)(4.11).



(a) (b)

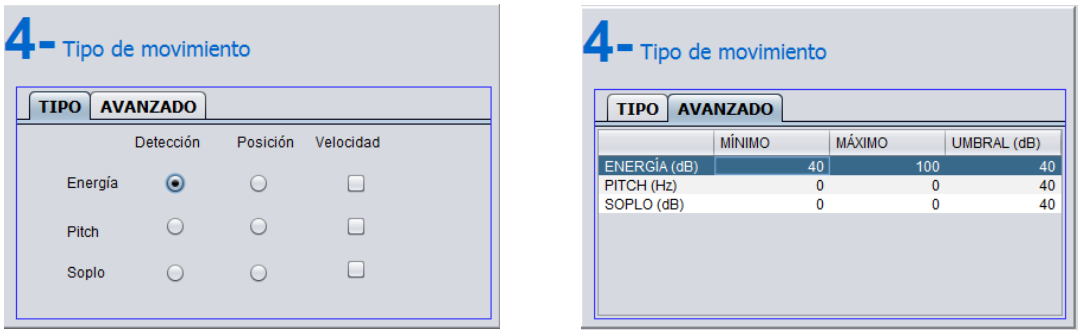
Figura 4.20: Elección de las condiciones adicionales del Mundo.

Mediante la opción "Estático", el objeto no se moverá en ausencia de voz ($(w_x, w_y) = (0,0)$). Dotando al objeto de un movimiento continuo hacia abajo se simula un efecto de gravedad y un efecto de corriente con un movimiento horizontal (a). Se pueden combinar varias direcciones para crear trayectorias oblicuas.

En la pestaña "AVANZADO" (b) se asigna el valor de (w_x, w_y), lo que se traduce en la elección de la velocidad con la que el objeto caerá o se moverá horizontalmente. En el caso de combinar varias direcciones, variar los valores de w_x y w_y consigue decidir la pendiente de la trayectoria oblicua.

PASO 4: Tipo de movimiento.

Este paso de la configuración se corresponde con la elección del tipo de movimiento del paso 1 del algoritmo de la Figura 4.2.



(a) (b)

Figura 4.21: Elección del tipo de movimiento.

Para cada parámetro $P[i]$, $i\epsilon$ (energía, pitch, soplo), se decide si se desea un tipo de movimiento de "Detección" en el que objeto se mueve si se activa $P[i]$ (4.3), o un movimiento de "Posición" en el que las coordenadas del objeto se corresponden con el valor de $P[i]$ (4.5). La casilla "Velocidad" solo se puede seleccionar para un movimiento de "Detección" y se corresponde con un movimiento del tipo descrito por (4.4). El objeto se mueve si se activa $P[i]$ y su velocidad dependerá del valor que tome $P[i]$.

En la pestaña "AVANZADO" se asignan los valores mínimos y máximos para cada parámetro $P[i]$ y el umbral de detección de voz.

PASO 5: Dirección del movimiento.

Este paso también se corresponde con el primer paso del algoritmo de la Figura 4.2. Es donde se asignan los valores $(\Delta x, \Delta y)$ de la expresión 4.3.

(a) (b)

Figura 4.22: Elección de la dirección del movimiento.

Para cada parámetro seleccionado en el paso anterior, se decide la dirección que tomará el objeto en su movimiento (Figura 4.22(a)). En la pestaña "AVANZADO" (b) se asigna un valor a Δx y Δy que fijan la velocidad con la que el objeto se moverá. Como en el paso 3 se pueden definir trayectorias oblicuas cuya pendiente podrá variar en función de los valores asignados a la dirección horizontal y a la dirección vertical.

PASO 6: Otros ajustes.

Permite añadir funcionalidades adicionales a la actividad configurada (Figura 4.23(a)). La primera opción, "Efecto de inercia", se corresponde con el paso 2 del algoritmo de la Figura 4.2.

(a) (b) (c)

Figura 4.23: Otros ajustes en actividades de Pre-lenguaje.

La pestaña "INERCIA" permite asignar los valores de α de (4.7) y (4.9). Para valores menores de α se consigue una mayor prolongación del movimiento, es lo deseable para "α Detección". Valores más altos son suficientes para "α Posición", con el fin de conseguir un suavizado del movimiento en el caso de movimientos del tipo (4.5).

Otras opciones de visualización se pueden añadir. El dibujo de la trayectoria permitirá al logopeda evaluar la ejecución de la actividad. La línea vertical y horizontal, a elegir dependiendo de la dirección del movimiento, fijan unos objetivos que el paciente debe alcanzar. La posición de las líneas se puede variar con el teclado. Por ejemplo, en la Figura 4.24 niño tiene que hacer que el cohete traspase la línea empujándolo con su soplo.

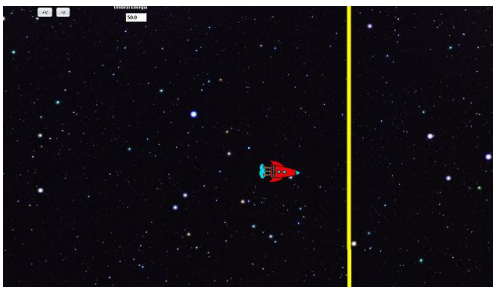


Figura 4.24: Actividad con objetivo.

Tanto para la trayectoria como para las líneas, se puede cambiar su color para distinguirlas sobre el fondo elegido.

4.4.2 Actividades de Vocalización

Este tipo de actividades se configura en la parte derecha de la ventana principal (Figura 4.9). Estos son los pasos a seguir:

PASO 1: Elección del tipo de actividad.

En este paso se elige entre los tres tipos de actividades.

- "Imágenes" consiste en la representación en el centro de la pantalla de una imagen que comenzará por la vocal que se desea trabajar. El centro de la imagen se corresponde con los formantes teóricos establecidos y el paciente debe acertar con un dardo en el centro de esta imagen mediante la pronunciación de dicha vocal (Figura 4.25 (a)).
- "Dianas" amplia el espacio de representación y, dependiendo de la vocal a trabajar, se situará una diana en la coordenada de la pantalla que se corresponda con los formantes teóricos de dicha vocal Figura 4.25 (b)).
- "Transiciones" es similar a "Dianas" pero, en este caso se trabaja con dos vocales. Dos imágenes que representarán estas vocales se situarán en las coordenadas de la pantalla fijadas por sus formantes. El niño deberá realizar la transición de una vocal hacia la otra y un coche recorrerá este camino Figura 4.25 (c)).

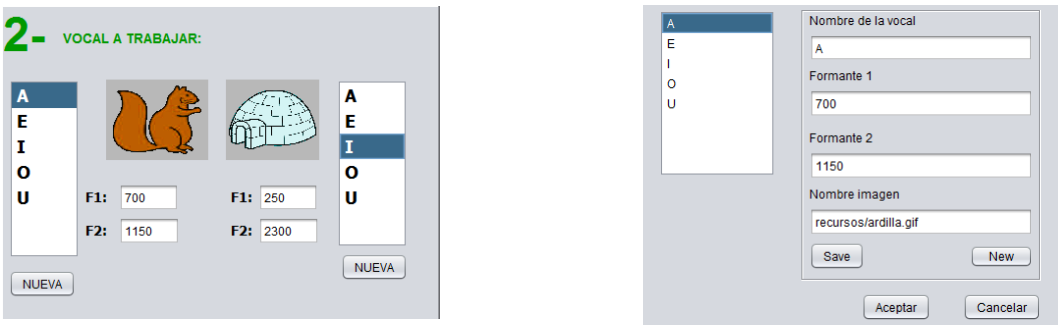
El tipo de movimiento en las tres actividades es el que definen (4.16) y (4.17).

(a) (b) (c)

Figura 4.25: Elección del tipo de actividad.

PASO 2: Selección de las vocales a trabajar.

En este paso se seleccionan las vocales a trabajar (Figura 4.26 (a)). Las actividades "Imágenes" y "Dianas" solo necesitan una vocal mientras que la actividad "Transiciones" necesita dos.



(a) (b)

Figura 4.26: Selección de vocales y adición de nuevas vocales.

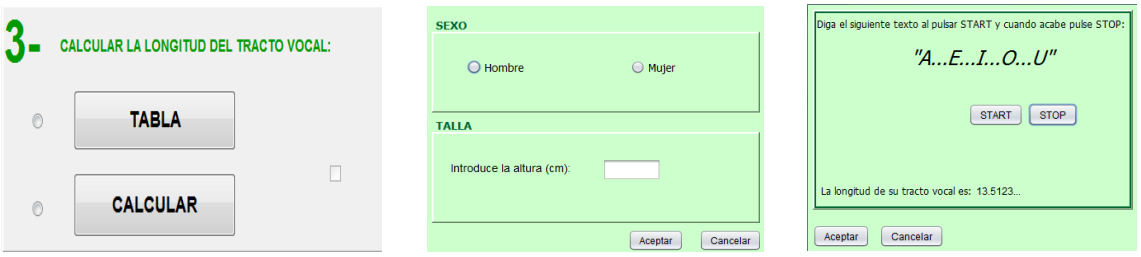
Nuevas vocales podrán ser añadidas a la lista con el fin de que la herramienta no sea válida solo para el español, sino también para otros idiomas (Figura 4.26 (b)). Introduciendo los formantes que caracterizan la nueva vocal y la imagen que se desea asignar a ésta, quedará definida una nueva vocal.

PASO 3: Cálculo de la longitud del tracto vocal.

Dos vías se han definido para el cálculo de la longitud del tracto vocal (Figura 4.27(a)):

- Obtención de la VTL de una tabla introduciendo el sexo y la talla del niño (Figura 4.27(b)) [4].

- Cálculo de la propia VTL a partir de la voz del niño (Figura 4.27(c)). Se calcula la VTL (sección 3.2) mediante la grabación de la voz del niño pronunciando las cinco vocales.



(a) (b) (c)

Figura 4.27: Cálculo de la VTL.

PASO 4: Otros ajustes.

Este último paso permite añadir un efecto de inercia similar al de las actividades de Pre-lenguaje (Figura 4.28 (a)). En la pestaña "AVANZADO" (Figura 4.28 (a)) se asigna el valor concreto para este parámetro.

(a) (b)

Figura 4.28: Otros ajustes en actividades de Vocalización.

En el Anexo IV se proponen algunos ejemplos de actividades que se pueden configurar con el editor diseñado.

5. Conclusiones y Líneas Futuras

5.1 Conclusiones

En el presente proyecto se ha desarrollado una herramienta libre de asistencia a la logopedia. Se ha diseñado un editor de actividades que sirve de apoyo a los profesionales de la educación especial en su labor de ayudar a personas con alguna discapacidad motriz o mental a mejorar su capacidad de comunicación.

En el capítulo 1 se definían una serie de objetivos que se debía intentar cumplir a la hora de diseñar la herramienta. Se ha diseñado una herramienta libre, que será utilizada en el CPEE Alborada y en cualquier centro público que la considere necesaria, para ayudar a personas con patologías del lenguaje a mejorar su capacidad del habla.

Como se ha programado en Java, será posible ejecutarla en una plataforma distribuida evitando la incompatibilidad entre sistemas operativos que aparecía en aplicaciones anteriores.

Aunque en un primer momento la herramienta iba a ser desarrollada para el español, durante el diseño de la misma surgió la posibilidad de definir nuevas vocales, lo que hace que sea válida para otros idiomas.

Se han implementado técnicas que extraen, en tiempo real y en Java, la energía, el pitch y los formantes de la señal de voz. El análisis localizado se ha utilizado para la extracción de la energía y durante todo el procesado. El pitch se puede estimar mediante la autocorrelación de la señal de voz o mediante la autocorrelación del error de predicción obtenido del análisis LPC, siendo la segunda opción la elegida por sus mejores resultados.

Se han introducido técnicas que resuelven la problemática de la voz en población infantil, proporcionando resultados que se aproximan más a la realidad. Se ha implementado en Java la técnica del liftado que elimina la influencia del pitch en la estimación de los formantes, y la estimación de la longitud del tracto vocal que permite la normalización de los formantes.

Por otro lado, el editor diseñado es altamente configurable, de manera que el logopeda puede crear actividades sin necesidad de tener conocimientos técnicos. Estas actividades se pueden adaptar a las necesidades de cada paciente. Con la funcionalidad de las máscaras, la versatilidad de la herramienta se incrementa considerablemente.

En un primer momento del proyecto se meditó la posibilidad de incluir una cámara que registrara los movimientos del paciente y sirviera de realimentación en la actividad. Sin embargo, la rigurosidad de la confidencialidad de datos hizo que esta idea se desechara por el momento, ya que se pretende que esta herramienta esté disponible en la Web continuando el entorno Vocaliza 2.0 comenzado en un anterior PFC [21].

5.2 Difusión del proyecto

La herramienta diseñada será utilizada en el CPEE Alborada y en cualquier institución pública que la considere necesaria.

La versión anterior de la herramienta ha tenido una gran aceptación entre la comunidad de logopedas por la novedad que presenta y, además, por tratarse de una herramienta libre. Se han contabilizado numerosas visitas a su Web[1] y descargas. A principios del 2011 contaba con más de 7340 descargas. También cuenta con publicaciones en revistas como [22], [23] y [24].

Así mismo, se recibieron aportaciones de logopedas, fonoaudiólogos y profesionales de la Educación Especial con recomendaciones y casos de éxito. Se han organizado cursos a nivel internacional para la utilización de la herramienta, así como se han elaborado fichas de trabajo realizadas por terapeutas.

---

1 http://www.vocaliza.es/

Con todo este impacto, existe mucha expectación por el lanzamiento de la versión Web, de la que este proyecto constituye un prototipo.

5.3 Líneas futuras

Una aplicación de este tipo siempre se puede mejorar tanto la parte gráfica como la técnica, sobre todo porque se trata del primer editor que permite configurar actividades propias. Estas son algunas mejoras que se proponen.

- Mejorar los algoritmos de estimación de pitch y formantes mediante técnicas de seguimiento como el algoritmo de Viterbi o filtros de partículas.

- Dotar al editor de mayores opciones de configuración. Añadir nuevas funcionalidades que den todavía más libertad al logopeda en el diseño de sus actividades.

- Introducir nuevos objetos con los que el usuario pueda interactuar.

- Permitir la interacción de varios usuarios en una misma actividad.

- Añadir informes tras la ejecución de la actividad que resuman los resultados obtenidos y que permita a los profesionales evaluarlos.

- Incluir niveles más altos del lenguaje como el semántico o el sintáctico.

BIBLIOGRAFÍA

[1] O. Saz, W.-R. Rodríguez, E. Lleida, C. Vaquero, and A. Escartín, "Comunica - Plataforma para el desarrollo, distribución y evaluación de herramientas logopédicas asistidas por ordenador," in *V Jornadas en Tecnología del Habla*, 2008, pp. 37-40.

[2] W. Rodríguez, O. Saz, E. Lleida, C. Vaquero, and A. Escartín, "COMUNICA - Tools for Speech and Language Therapy," in *Workshop on Chlid Computer and Interaction, ICMI08*, 2008.

[3] C. V. Avilés-Casco, "Reconocedor de Comandos Orales para Eliminar Barreras de Comunicación y Movilidad en Personas con Discapacidades Motrices y de Comunicación," Proyecto Fin de Carrera, Dirigido por O.Saz ,Departamento de Ingeniería Electrónica y Comunicaciones, Universidad de Zaragoza, 2006.

[4] W. R. Rodríguez Dueñas, "Aplicación de las Tecnologías del Habla en la Educación de la Voz Infantil Alterada.," Tesis Doctoral, Departamento de Ingeniería Electrónica y Comunicaciones, Universidad de Zaragoza, 2010.

[5] A. Escartín Villellas, "Gestión de 'COMUNICA : Conjunto de herramientas para la logopedia' y ampliación de sus herramientas a los niveles semántico y pragmático del lenguaje," Proyecto Fin de Carrera, Departamento Ingeniería Electrónica y Comunicaciones, Universidad de Zaragoza.

[6] J. Vila, *Guía de Intervención Logopédica en la Disfonía Infantil*. Editorial Síntesis, 2009.

[7] R. W. S. Lawrence R. Rabiner, *Digital processing of speech signals*. Prentice-Hall, 1978.

[8] J. G. P. DIMITRIS G. MANOLAKIS, *Tratamiento digital de señales*. Prentice-Hall, 1998.

[9] H. Traunmüller and A. Eriksson, "A method of measuring formant frequiencies at high fundamental frequencies," in *Proceedings of Eurospeech*, 1997, vol. 0.

[10] M. S. Rahman and T. Shimamura, "Formant frequency estimation of high-pitched speech by homomorphic prediction," *Acoustical Science and Technology*, vol. 26, no. 6, pp. 502-510, 2005.

[11] T. Quatieri, “Minimum and mixed phase speech analysis-sinthesis by adaptative homomorphic deconvolution,” *IEEE Transaction on Acoustics, Speech and Signal Processing*, vol. 27, no. 4, 1979.

[12] W. Verhelst and O. Steenhaut, “A new model for the short-time complex cepstrum of voiced speech,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 1, pp. 43-51, Feb. 1986.

[13] U. Goldstein, “An articulatory model for the vocal tracts of growing children,” PhD thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1977.

[14] Houri K. Vorperian, “Development of vocal tract length during early childhood: A magnetic resonance imaging study,” *Journal of Acoustical Society America*, vol. 117, no. 1, pp. 338-350, 2005.

[15] a. Paige and V. Zue, “Calculation of vocal tract length,” *IEEE Transactions on Audio and Electroacoustics*, vol. 18, no. 3, pp. 268-270, Sep. 1970.

[16] H. Wakita, “Normalization of vowels by vocal-tract length and its application to vowel identification,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 25, no. 2, pp. 183 - 192, 1977.

[17] R. Kirlin, “A posteriori estimation of vocal tract length,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26, no. 6, pp. 571 - 574, 1978.

[18] M. R. Schroeder, “Determination of the Geometry of the Human Vocal Tract by Acoustic Measurements,” *Journal of the Acoustical Society of America*, vol. 41, no. 4B, pp. 1002-1010, 1967.

[19] B. F. Neciog, M. A. Clements, and T. Barnwell, “Unsupervised estimation of the human vocal tract length over sentence level utterances,” in *Acoustics, Speech, and Signal Processing. ICASSP*, 2000, pp. 1319-1322.

[20] B. Eckel, *Piensa en Java*. Prentice-Hall, 2007.

[21] J. M. Escartín Villellas, “Diseño de una plataforma web de docencia distribuida aplicada a la logopedia y comunicación en educación especial,” Universidad de Zaragoza, 2011.

[22] O. Saz, S.-C. Yin, E. Lleida, R. Rose, C. Vaquero, and W. R. Rodríguez, “Tools and Technologies for Computer-Aided Speech and Language Therapy,” *Speech Communication*, vol. 51, no. 10, pp. 948-967, Oct. 2009.

[23] O. Saz, J. Simón, W.-R. Rodríguez, E. Lleida, and C. Vaquero, “Analysis of Acoustic Features in Speakers with Cognitive Disorders and Speech Impairments,” *EURASIP Journal on Advances in Signal Processing*, no. 1, 2009.

[24] O. Saz, “Consideraciones en el desarrollo de herramientas informáticas para logopedia en educación especial,” *Maremagnum: publicación galega sobre os trastornos do espectro autista*, vol. 13, pp. 131-138, 2009.