

**DISEÑO Y DESARROLLO DE UNA HERRAMIENTA SOFTWARE COMO SOPORTE A  
LA COMUNICACIÓN VÍA CELULAR MEDIANTE SÍNTESIS DE VOZ**

**CLAUDIA VICTORIA CORREA PUGLIESE  
HOOVER FABIÁN RUEDA CHACÓN**



**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA  
2009**

**DISEÑO Y DESARROLLO DE UNA HERRAMIENTA SOFTWARE COMO SOPORTE A  
LA COMUNICACIÓN VÍA CELULAR MEDIANTE SÍNTESIS DE VOZ**

**CLAUDIA VICTORIA CORREA PUGLIESE  
HOOVER FABIÁN RUEDA CHACÓN**

**Trabajo de grado presentado para optar por el título de  
Ingeniero de Sistemas**

**Director**

**MPE. HENRY ARGUELLO FUENTES**

**Profesor**

**Escuela de Ingeniería de Sistemas e Informática – UIS**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA**

**2009**

## **AGRADECIMIENTOS**

Los autores expresan su agradecimiento:

Al profesor HENRY ARGÜELLO FUENTES, director del proyecto por sus enseñanzas y acompañamiento; porque aún en la distancia estuvo siempre presente.

A los compañeros de maestría del Grupo de Investigación en Ingeniería Biomédica (GIIB), por su apoyo y compañía. Especialmente a SERGIO ANTONIO PINO GALLARDO e IRENE LIZETH MANOTAS GUTIERREZ porque desinteresadamente nos prestaron su invaluable colaboración y nos compartieron parte de su conocimiento y experiencia.

A los compañeros de pregrado que compartieron con nosotros la estadía en el grupo GIIB, especialmente a HUGO ARMANDO GUALDRÓN COLMENARES por aportar al proyecto desde su conocimiento.

Al señor JORGE ESTUPIÑÁN quien colaboró prestando su voz para la realización de las grabaciones asociadas al corpus de voz.

Al DEPARTAMENTO DE COMUNICACIONES UIS, en especial a RADIO UIS, por permitirnos utilizar sus cabinas para realizar las grabaciones.

A la UNIVERSIDAD INDUSTRIAL DE SANTANDER por habernos ofrecido las herramientas necesarias para nuestra formación profesional.

*A Dios, que ha sido muy generoso conmigo y con mi familia. Porque ha puesto personas valiosas en mi camino y me ha permitido cumplir satisfactoriamente mis objetivos.*

*A mis papás, Raúl y Claudia, que con amor y esfuerzo me han ofrecido las mejores oportunidades y han sido los principales patrocinadores de mis logros. Gracias por su constante apoyo; esta meta conseguida también es de ustedes.*

*A mi hermano, que ha sido mi compañero en la vida, y que me ha hecho sentir como su otra “mamá”; a su manera me ha demostrado su cariño y apoyo.*

*A mis abuelitas, Alicia mi angelito en el cielo, cuyo recuerdo guardo siempre en mi corazón. Y Anita mi angelito en la tierra, gracias por las oraciones, el amor y el apoyo.*

*A Hoover por haber recorrido gran parte de esta etapa conmigo. Por ser mi compañero, mi amigo, y sobre todo por el amor y el respeto que nos une. Porque con su alegría y humor hace que un momento cualquiera sea memorable.*

*A mi familia en general, todos y cada uno de aquellos que de una u otra manera han contribuido con mi formación personal y académica, y han hecho de mí la persona que hoy soy.*

*Claudia*

*A Dios por la inteligencia, la sabiduría, el entendimiento y la serenidad para afrontar los retos de cada día, por la salud y todos los favores recibidos.*

*A mi mamita Cecilia por su apoyo, amistad y consejos, especialmente por su amor, comprensión y acompañamiento a lo largo de mi vida. Por sus enseñanzas, su nobleza, lucha y fortaleza para salir adelante.*

*A mi hermano Fernando por su apoyo incondicional, por ser la compañía en pasajes importantes de mi vida, siendo mi amigo y confidente.*

*A mi hermosa novia Claudia, por todo su amor, comprensión, apoyo, inteligencia, belleza y por su espectacular sonrisa, motivo de inspiración y olvido de dificultades. Por compartir sus ideales conmigo y dejarme ser parte de ellos.*

*A mi familia, por el entorno que me han otorgado, por el orgullo que siento de formar parte de ésta, por los valores inculcados y todo el apoyo dispensado.*

*A la familia Correa Pugliese, por la colaboración y apoyo ofrecido, además de las alegrías compartidas.*

*Y a todas las personas que de una u otra forma han estado conmigo a lo largo de este camino.*

*Hoover*

## TABLA DE CONTENIDO

INTRODUCCIÓN.....	1
1. ESPECIFICACIONES DEL PROYECTO .....	3
1.1 DEFINICIÓN DEL PROBLEMA .....	3
1.2 JUSTIFICACIÓN .....	5
1.3 OBJETIVOS .....	6
1.3.1 Objetivo General.....	6
1.3.2 Objetivos Específicos .....	6
2. MARCO TEÓRICO .....	7
2.1 CONCEPTOS FUNDAMENTALES.....	7
2.2 SÍNTESIS DE VOZ.....	8
2.2.1 Parámetros de caracterización de los sistemas de síntesis de voz.....	10
2.2.2 Técnicas de Síntesis .....	11
2.2.2.1 Síntesis por concatenación.....	11
2.2.2.1.1 Síntesis por selección de unidades.....	11
2.2.2.1.2 Síntesis de difonemas.....	11
2.2.2.1.3 Síntesis específica para un dominio.....	12
2.2.2.2 Síntesis de formantes.....	12
2.2.2.3 Síntesis articulatoria.....	12
2.2.2.4 Síntesis por derivación de técnicas de predicción lineal (LPC).....	13
2.3 CORPUS DE VOZ.....	13
2.4 FREETTS .....	14
2.5 JAVA MEDIA FRAMEWORK (JMF).....	15
2.5.1 Captura de Medios (MediaLocator y DataSource) .....	16
2.5.2 Procesamiento de Medios (Processor) .....	17

2.5.3	Reproducción de Medios (Player).....	18
2.5.4	Almacenamiento de Medios (DataSink).....	20
2.6	EL ESPAÑOL EN COLOMBIA.....	21
3	HERRAMIENTA CHSPEECH.....	23
3.1	SINTETIZADOR DE VOZ.....	23
3.1.1	Procesadores .....	24
3.1.1.1	Tokenizador.....	25
3.1.1.2	Normalizador.....	25
3.1.1.3	Divisor de palabras.....	27
3.1.1.4	Agrupador de fonemas.....	28
3.1.1.5	Buscador.....	29
3.1.1.6	Concatenador.....	30
3.2	CORPUS DE VOZ.....	31
3.2.1	Corpus de Frases .....	33
3.2.2	Corpus de Difonemas .....	34
3.2.2.1	Diseño del corpus de voz.....	35
3.2.2.2	Preparación de las condiciones de grabación.....	36
3.2.2.3	Grabación.....	38
3.2.2.4	Etiquetado de los archivos de audio.....	38
3.2.2.5	Extracción de los fragmentos de audio.....	44
3.2.2.6	Base de datos de la aplicación.....	46
3.3	INTERFAZ GRÁFICA DE USUARIO (GUI).....	47
3.3.1	Panel de autenticación.....	47
3.3.2	Panel de registro de nuevos usuarios .....	48
3.3.3	Panel Principal CHSpeech.....	49
3.3.4	Recordatorio de contraseñas .....	51

3.3.5	Modificación de E-mail y Contraseña.....	52
3.3.6	Base de datos para el manejo de usuarios y frases predefinidas.....	53
3.4	MECANISMO DE TRANSMISIÓN.....	54
4	PRUEBAS Y RESULTADOS.....	58
4.1	COMPARACIÓN DE SEÑALES.....	58
4.2	PRUEBAS DE FUNCIONAMIENTO.....	60
4.2.1	Prueba al sintetizador.....	60
4.2.2	Prueba de ejecución en diferentes plataformas y medición de tiempos de respuesta.....	70
4.3	PRUEBAS DE ACEPTACIÓN CON USUARIOS.....	75
4.3.1	Formato de la prueba con Usuarios.....	75
4.3.2	Resultados de las Pruebas de Usuario.....	77
5	CONCLUSIONES.....	83
6	RECOMENDACIONES.....	86
7	REFERENCIAS.....	88

## LISTA DE FIGURAS

Figura 1.	Esquema de la herramienta software.....	4
Figura 2.	Esquema general del funcionamiento de un sintetizador de voz.....	8
Figura 3.	Modelo de Procesamiento de Medios Basados en el Tiempo.....	16
Figura 4.	Jerarquía de la clase Player.....	17
Figura 5.	Modelo de un Processor JMF.....	18
Figura 6.	Estados de un Player.....	20
Figura 7.	Clasificación Dialectal de Colombia.....	22
Figura 8.	Procesadores del Sintetizador.....	24
Figura 9.	Funcionamiento del Divisor de palabras.....	28
Figura 10.	Funcionamiento del Agrupador de Fonemas.....	28
Figura 11.	Fragmento del archivo grabaciones.properties.....	29
Figura 12.	Esquema de las actividades del Concatenador.....	31
Figura 13.	Gráfica de Flexibilidad Vs. Calidad.....	32
Figura 14.	Contenido del archivo chspeech.lst.....	39
Figura 15.	Fragmento del contenido del archivo chspeech.dat generado automáticamente por Diphone Studio.....	40
Figura 16.	Fragmento del contenido del archivo chspeech.dat con los valores reales	41
Figura 17.	Interfaz de Usuario Diphone Studio.....	42
Figura 18.	Inserción de etiquetas en Diphone Studio.....	43
Figura 19.	Fragmento del contenido final del archivo chspeech.dat.....	44
Figura 20.	Señal sin normalizar.....	45

Figura 21.	Señal normalizada.....	45
Figura 22.	Panel de Autenticación CHSpeech.....	48
Figura 23.	Panel de Registro de Nuevos Usuarios.....	49
Figura 24.	Panel Principal CHSpeech y sus componentes.....	50
Figura 25.	Menú cuenta.....	50
Figura 26.	Menú Ayuda.....	51
Figura 27.	Recordatorio de contraseñas.....	51
Figura 28.	Formulario Cambio E-mail.....	52
Figura 29.	Formulario Cambio de Contraseña.....	53
Figura 30.	Diseño de la base de datos para el manejo de usuarios y frases predefinidas.....	54
Figura 31.	Manos libres original.....	55
Figura 32.	Proceso de modificación del cable manos libres.....	56
Figura 33.	Cable manos libres modificado.....	57
Figura 34.	Conexión del cable manos libres al celular y al equipo de cómputo.....	57
Figura 35.	Señal de audio original.....	58
Figura 36.	Señal de audio obtenida del sintetizador.....	59
Figura 37.	Comparación de tiempos de respuesta de los tres equipos de prueba.....	73
Figura 38.	Resultados frases de prueba con usuarios.....	79
Figura 39.	Resultados opinión de los usuarios respecto a rapidez de la herramienta	81

## LISTA DE TABLAS

Tabla 1.	Condiciones de grabación corpus de voz.....	34
Tabla 2.	Fragmento de la tabla de difonemas.....	36
Tabla 3.	Condiciones de grabación corpus de difonemas.....	37
Tabla 4.	Resultados Prueba de funcionamiento del sintetizador.....	60
Tabla 5.	Características de los equipos utilizados en las pruebas.....	71
Tabla 6.	Tiempos de respuesta equipo 1 (Windows Vista).....	72
Tabla 7.	Tiempos de respuesta equipo 2 (Fedora 8).....	72
Tabla 8.	Tiempos de respuesta equipo 3 (Mac OS X).....	73
Tabla 9.	Frases utilizadas en la prueba de usuarios.....	78
Tabla 10.	Resultados inteligibilidad frases de prueba con usuarios.....	78

## LISTA DE ANEXOS

Anexo A.	Descripción de formatos y pronunciaciones.....	92
Anexo B.	Lista de abreviaturas.....	98
Anexo C.	Lista de símbolos.....	100
Anexo D.	Lista de extranjerismos.....	102
Anexo E.	Fonemas del español de Colombia.....	104
Anexo F.	Condiciones para asignación de fonemas.....	105
Anexo G.	Difonemas del español en Colombia.....	108
Anexo H.	Tabla de difonemas y palabras ejemplo.....	108
Anexo I.	Metodología de desarrollo Extreme Programming.....	114
Anexo J.	Manual de usuario.....	117
Anexo K.	Diagramas de clases.....	118
Anexo L.	Artículo pendiente de publicación: SINTESIS DE VOZ POR CONCATENACIÓN DE DIFONEMAS PARA EL ESPAÑOL DE COLOMBIA.....	123

## RESUMEN

**TÍTULO:** DISEÑO Y DESARROLLO DE UNA HERRAMIENTA SOFTWARE COMO SOPORTE A LA COMUNICACIÓN VÍA CELULAR MEDIANTE SÍNTESIS DE VOZ \*

**AUTORES:** CLAUDIA VICTORIA CORREA PUGLIESE \*\*  
HOOVER FABIÁN RUEDA CHACÓN \*\*

**PALABRAS CLAVE:** Síntesis de voz, corpus de voz, concatenación de difonemas, Español de Colombia, Java Media Framework.

Este proyecto presenta el diseño y desarrollo de una herramienta software que permite mantener conversaciones a través de teléfonos celulares en situaciones que limitan el uso directo de la voz. Esto es posible gracias a las tecnologías del habla, específicamente a la síntesis de voz.

La herramienta software desarrollada se compone de cuatro elementos principales que se presentan con detalle en el documento. El primero de ellos es un sintetizador de voz de dominio ilimitado que utiliza la técnica de concatenación de difonemas para el español de Colombia, el cual permite la producción de una onda sonora a partir de un texto de entrada. El segundo elemento es un corpus de voz para el mismo idioma que proporciona la voz al sintetizador. El manejo del audio dentro de la herramienta se realizó utilizando Java Media Framework, que permite el uso de archivos multimedia en aplicaciones desarrolladas en Java. La interfaz gráfica de usuario es el tercer componente de la herramienta, ésta permite que los usuarios accedan a las funcionalidades del sintetizador de manera sencilla. Finalmente, se implementó un mecanismo de transmisión que constituye el cuarto componente; la finalidad de éste mecanismo es transmitir la voz sintética desde un equipo de cómputo hacia un dispositivo celular para que sea reproducida en medio de una llamada.

---

\* Proyecto de grado

\*\* Facultad de Ingenierías Físico – Mecánicas. Escuela de Ingeniería de Sistemas e Informática.  
Director: Henry Argüello Fuentes.

## ABSTRACT

**TITLE:** DESIGN AND DEVELOPMENT OF A SOFTWARE TOOL AS SUPPORT FOR COMMUNICATIONS VIA MOBILE PHONES THROUGH SPEECH SYNTHESIS \*

**AUTORS:** CLAUDIA VICTORIA CORREA PUGLIESE \*\*  
HOOVER FABIÁN RUEDA CHACÓN \*\*

**KEYWORDS:** Speech synthesis, voice corpus, diphone concatenation, Colombian Spanish, Java Media Framework.

This project presents design and development of a software tool that allows conversations over mobile phones in situations that limit the direct use of voice. This is made possible by speech technology, specifically by speech synthesis.

The developed software tool consists of four major elements; the document presents those components in detail. The first element is the unlimited domain voice synthesizer; it uses the diphone concatenation technique for Colombian Spanish, which allows the production of a wave sound from an input text. The second element is the voice corpus for the same language that provides voice to the synthesizer. The audio management within the tool was made using Java Media Framework, which allows the use of multimedia in applications developed in Java. The graphic user interface (GUI) is the third component, it allows users simple access to the functionality of the synthesizer. Finally, a transmission mechanism was implemented as the fourth component; the purpose of this mechanism is to transmit the synthetic voice from a computer to a mobile phone; the voice will be reproduced during a call.

---

\* Research degree project.

\*\* Faculty of Physical-Mechanical Engineering. School Systems Engineering and Computer Science. Director: Henry Argüello Fuentes.

## INTRODUCCIÓN

La masificación del uso de la telefonía móvil ha provocado un cambio en las costumbres de los seres humanos y en la concepción de la comunicación a distancia. Las tecnologías del habla (síntesis y reconocimiento de voz) por su parte, han surgido como una alternativa de comunicación para personas con problemas en el uso directo de la voz, además de su aplicabilidad en la interacción hombre-máquina.

Cada una de estas tecnologías, goza de cierta popularidad entre usuarios e investigadores, y esto ha permitido que a su ritmo maduren hasta ubicarse en el estado en que se conocen actualmente. Por un lado la telefonía celular se encuentra muy bien ubicada en el mercado y la mayoría de personas cuentan al menos con un dispositivo móvil; por otro lado la síntesis de voz es objeto de estudio de diversos grupos de investigación en varios lugares alrededor del mundo, y de algunas compañías que comercializan con herramientas de este tipo, utilizadas principalmente en conmutadores y contestadores automáticos.

El estudio y la combinación de estas dos áreas del conocimiento han permitido a los autores del presente proyecto proponer una solución alternativa a los problemas de comunicación de aquellas personas con inconvenientes en la contestación de sus dispositivos celulares en situaciones que limitan el uso directo del habla.

La idea principal del proyecto es el desarrollo de una herramienta software que será utilizada en equipos de cómputo (computadores de escritorio o portátiles), que por medio del uso de la síntesis de voz en español de Colombia, permite a los usuarios contestar llamadas vía celular en las situaciones mencionadas anteriormente.

La herramienta software producto del presente proyecto se ha denominado CHSpeech 1.0, y tiene cuatro componentes principales. En primer lugar, CHSpeech 1.0 se vale, evidentemente, de un sintetizador de voz encargado de transformar los textos de entrada en sonidos que puedan ser reproducidos en la llamada; el segundo componente del proyecto corresponde al corpus de voz que utiliza el sintetizador para formar la voz que escuchará el interlocutor al otro lado del celular; el tercer elemento está constituido por un mecanismo hardware que permite la transmisión de la señal de audio desde el

computador hacia el celular; y por último se encuentra la interfaz gráfica de usuario desarrollada para facilitar la interacción del usuario con la herramienta.

Se espera que el desarrollo de este proyecto contribuya con la construcción de conocimiento en cuanto a síntesis de voz en español de Colombia, así como al mejoramiento de la producción de voces sintéticas en este idioma.

El presente libro contiene la documentación acerca del desarrollo de la herramienta software CHSpeech 1.0, esto incluye el trabajo realizado en cada uno de sus componentes. En la primera parte se presentan las especificaciones del proyecto, incluyendo la definición del problema, justificación, objetivo general y los objetivos específicos. El segundo capítulo contiene el marco teórico necesario para comprender el tema del proyecto, esto es lo concerniente a los conceptos básicos de síntesis de voz, las técnicas de síntesis, corpus de voz, JMF (Java Media Framework) utilizado para la reproducción de la voz, y documentación acerca del español en Colombia. El capítulo tres se enfoca en el desarrollo de los componentes de la herramienta, comenzando por el sintetizador, continuando con el corpus de voz y la interfaz gráfica de usuario y terminando con el mecanismo de transmisión. En el capítulo cuatro se presentan las pruebas realizadas a la herramienta y los resultados obtenidos. Finalmente se encuentran las conclusiones, recomendaciones y referencias utilizadas.

Los anexos presentan información de algunos detalles realizados en el desarrollo del sintetizador y del corpus, además de los diagramas de clases de la herramienta y el manual de usuario.

## **1. ESPECIFICACIONES DEL PROYECTO**

### **1.1 DEFINICIÓN DEL PROBLEMA**

La comunicación ha sido una de las principales necesidades de los seres humanos a través de la historia. En principio se utilizaron mecanismos como la pintura y representaciones por medio de signos y símbolos en las cavernas para transmitir mensajes o identificar tribus. Posteriormente con el desarrollo de la capacidad intelectual de los humanos se fueron originando los lenguajes y alfabetos [1], desde entonces ha sido posible referirse a comunicación oral y escrita.

Sin embargo, con el surgimiento de las grandes civilizaciones, se hizo necesaria la creación de mecanismos de comunicación a distancia, como los servicios postales [2], que hicieron posible que este tipo de comunicación evolucionara y se popularizara.

Adicionalmente, la invención del telégrafo sentó un gran avance dentro de la rama de la comunicación a distancia; pues el uso de la electricidad para transmitir mensajes letra a letra permitió posteriormente, la creación de un mecanismo más efectivo para la comunicación vocal: El teléfono[3].

Con el paso de los años, éste artefacto no fue solución suficiente a los requerimientos de comunicaciones, debido a que no era posible ubicar a las personas en cualquier momento y lugar, haciéndose necesaria la búsqueda de alternativas para resolver los problemas de portabilidad y disponibilidad; de esta manera surge el concepto de telefonía móvil, entendida como el tipo de comunicación en la cual emisor y/o receptor están en movimiento, limitando el uso de cables para establecer conexiones entre ellos [4][5].

Aunque la función principal de este tipo de comunicación es la transmisión de voz, actualmente existen diversos tipos de servicios como aplicaciones de audio, video, datos, etc.; esto ha sido posible gracias a que los dispositivos han ido aumentando sus capacidades de procesamiento, y a su vez, han ido incorporando otras tecnologías como conexión a redes, transmisión y almacenamiento de datos y sincronización con computadores.

A pesar de los avances que implica la utilización de las aplicaciones anteriormente mencionadas, existen aún ciertos defectos y desventajas en la tecnología celular; uno de los principales es que el hecho de tener un celular no garantiza el establecimiento de la comunicación, puesto que existen diversas situaciones que limitan el uso directo del habla. Por esta razón, el problema que tratará este trabajo de investigación es la imposibilidad de atender llamadas importantes en este tipo de situaciones, donde no es posible hablar normalmente, como reuniones y conferencias.

La solución presentada es una herramienta software que recurre a la utilización de síntesis de voz para apoyar la comunicación vía celular y contribuye con un mejor aprovechamiento de esta tecnología en las situaciones mencionadas anteriormente.

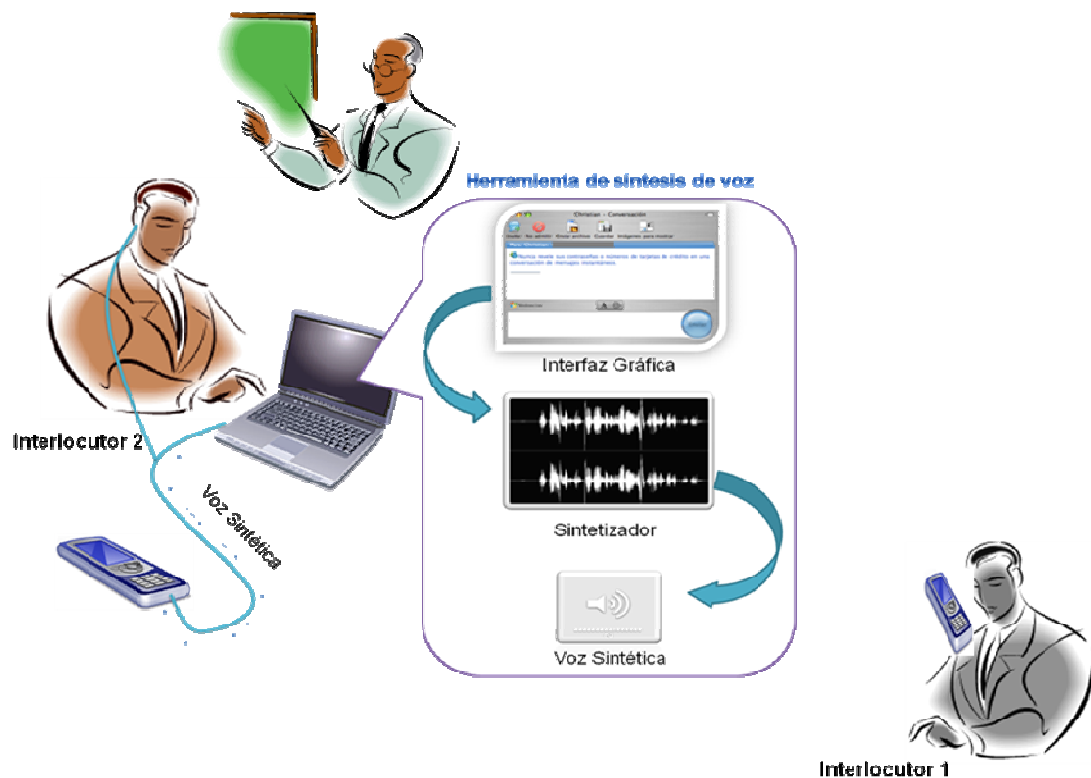


Figura 1. Esquema de la herramienta software

Fuente: Autores del Proyecto

La Figura 1 presenta el esquema general de la herramienta software propuesta. Cuenta con una interfaz gráfica sencilla que sirve para la introducción de los datos, que son posteriormente enviados al sintetizador para su procesamiento. El sintetizador

desarrollado en el proyecto cuenta con los algoritmos necesarios para la síntesis de voz en español de Colombia. Como resultado se obtiene la voz sintética, que es transmitida a través de un cable manos libres desde un equipo de cómputo hacia el dispositivo celular, para ser reproducido en medio de una llamada.

## **1.2 JUSTIFICACIÓN**

Los grandes cambios que se han dado en los últimos tiempos en el área de las comunicaciones, han revolucionado la forma de vida de las personas alrededor del mundo. Nuevas tecnologías, dispositivos y aplicaciones aparecen para facilitar las actividades cotidianas.

Actualmente, los teléfonos celulares son una de las formas de comunicación oral más utilizadas; el objetivo principal de esta tecnología: la movilidad, es tal vez una de las razones por la que resulta atractiva.

La posibilidad de estar en contacto permanentemente, en cualquier momento y en cualquier lugar, es de gran importancia en un mundo en el que las cosas cambian con tanta velocidad. Sin embargo, la disponibilidad no es siempre completa, existen ocasiones en que por un motivo u otro, no es posible atender una llamada celular, especialmente cuando se trata de personas que, debido a sus ocupaciones, están todo el tiempo atendiendo varias cosas a la vez. Es por esta razón, que se ha visto la necesidad de buscar una solución que brinde comodidad al momento de sostener una llamada al mismo tiempo que se atiende otro asunto como una reunión o conferencia.

Por otra parte, el proyecto tiene involucrada una tecnología que viene en ascenso y cada vez gana más popularidad, la síntesis de voz; por medio de ella es posible generar voz artificial a través de medios automáticos, siendo similar a la que produciría una persona al hablar[6].

La finalidad de este proyecto es proporcionar una herramienta que facilite la comunicación vía celular en ocasiones donde no sea posible el uso directo del habla, manteniendo vivos los conceptos de movilidad y disponibilidad; además servirá como base para futuros proyectos relacionados con tecnologías del habla, específicamente con síntesis de voz.

En la actualidad existen varios sintetizadores, algunos de ellos ofrecen voces en español de países como España, Venezuela y Argentina; sin embargo, no se tiene conocimiento acerca de desarrollos de síntesis de voz en Colombia, ni de aplicaciones de este tipo para solucionar problemas similares al presentado en este proyecto.

### **1.3 OBJETIVOS**

#### **1.3.1 Objetivo General**

Diseñar y desarrollar una herramienta software que sirva de soporte para la comunicación vía celular en situaciones que limiten el uso directo del habla utilizando síntesis de voz en español.

#### **1.3.2 Objetivos Específicos**

- Realizar el desarrollo de algoritmos que implementen las reglas para la síntesis de voz en español, basados en el análisis de la estructura y procesamiento de textos.
- Diseñar y desarrollar un prototipo software que sirva de interfaz entre el usuario y las funcionalidades ofrecidas por el sintetizador.
- Diseñar e implementar un mecanismo hardware para la transmisión de la voz sintetizada, desde un equipo de cómputo hacia el dispositivo celular.
- Verificar el desempeño de la herramienta desarrollada, en personas con inconvenientes en la contestación de sus dispositivos celulares debido a situaciones que limiten el uso directo del habla.

## 2. MARCO TEÓRICO

### 2.1 CONCEPTOS FUNDAMENTALES

- **Fonema:** Es la unidad lingüística mínima de una lengua, capaz de diferenciar significados, por ejemplo la palabra mar y par se diferencian por los fonemas /m/ y /p/. Sin embargo, los fonemas no son sonidos con entidad física, sino abstracciones mentales de los sonidos del habla. Para el español, se contabilizan 24 fonemas (5 vocales y 19 consonantes), aunque hay algunas variedades de la lengua que tienen una cantidad diferente de fonemas.
- **Difonema:** Es el segmento acústico que incluye la transición entre dos fonemas consecutivos, formado por la parte estacionaria del primero, la transición del primero al segundo, y la parte estacionaria del segundo. El español tiene aproximadamente 800 difonos [7], aunque ésta cantidad depende del número de fonemas de una variedad específica de la lengua.
- **Trifonema:** Es la unidad constituida por un fonema más la mitad del segmento precedente y la mitad del segmento siguiente. Su uso dota de más naturalidad a un sintetizador de voz, pero no todas las frases y palabras pueden formarse utilizando esta metodología, por lo que siempre suelen completarse las concatenaciones con fonemas y difonemas.
- **Semisílabas:** Es la unidad formada por la mitad de una sílaba, estableciendo el límite en el centro del núcleo silábico, que para el español es una vocal [7].
- **Sílabas:** Es cada una de las divisiones de una palabra. Se compone del núcleo silábico (la parte central de la sílaba y que tiene la mayor intensidad sonora), el ataque (precede el núcleo silábico) y la coda (sigue al núcleo silábico).
- **Prosodia:** Rama de la lingüística que se encarga del estudio de la pronunciación, acentuación y entonación. Se asocia físicamente con la frecuencia fundamental, duración e intensidad de la voz, constituyendo éstos, los parámetros prosódicos físicos.

- **Coarticulación:** Es el fenómeno de la fonética que se relaciona con la velocidad y la coordinación de los movimientos del tracto vocal. En cualquier representación acústica del habla puede observarse que los sonidos no se producen de manera aislada sino que se encadenan unos con otros, solapándose muchas veces las propiedades de un sonido con las de otro [8].

## 2.2 SÍNTESIS DE VOZ

La síntesis de voz es la producción artificial de habla humana. Su objetivo principal es la generación automática por medio de un sistema informático electrónico de una señal acústica que simule la voz humana, normalmente a partir de un texto de entrada [9]; sin embargo existen también sistemas que lo hacen a partir de la representación lingüística, es decir a partir de la representación gráfica de la pronunciación.

La Figura 2 brinda una idea general de los procesos que lleva a cabo un sintetizador de voz; se observa en ella, un prototipo de caja negra, con sus respectivas entrada y salida, en la cual se incluyen las funciones de análisis preliminar del texto y su respectivo análisis lingüístico, que incluye entre otras cosas, la duración, entonación y graduación; el resultado de éste análisis y procesamiento, es la señal acústica, siendo ésta la salida del sistema.

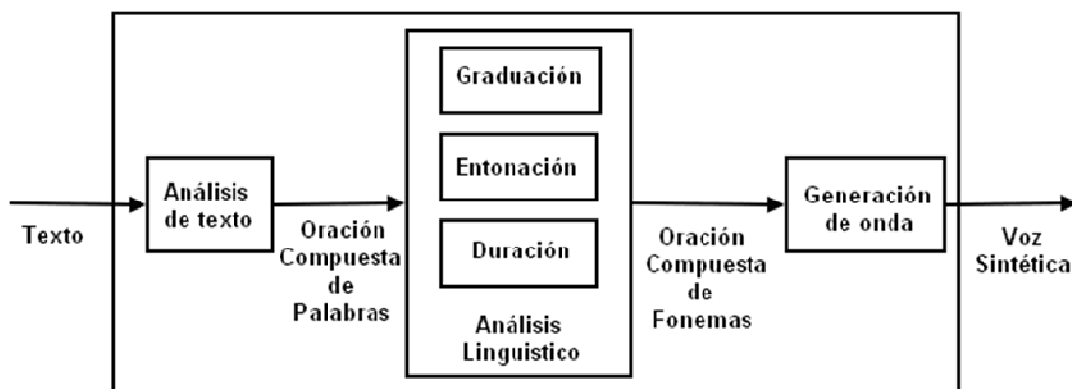


Figura 2. Esquema general del funcionamiento de un sintetizador de voz

Fuente: Wikipedia [10]

Más detalladamente, los pasos que realiza un sintetizador de voz son los siguientes (JSAPI, 1998, 27):

- **Análisis de la estructura:** Constituye el procesamiento del texto de entrada para determinar donde comienzan y terminan los párrafos, oraciones, y otras estructuras. La puntuación y el formato de los datos son usados en esta etapa.
- **Pre-procesamiento del texto:** En este paso se realiza un análisis del texto de entrada para identificar y procesar construcciones especiales del idioma. Por ejemplo: abreviaturas, acrónimos, fechas, horas, números, montos de dinero, cuentas de correo electrónico y otras formas.
- **Conversión Texto-Fonema:** En esta etapa se convierte cada palabra en fonemas, o la unidad de síntesis escogida. Un fonema es la unidad básica de sonido en un idioma.
- **Análisis prosódico:** Consiste en el procesamiento de la estructura de la oración, palabras y fonemas para determinar la prosodia adecuada de la oración. La prosodia es la parte de la gramática que enseña la correcta pronunciación y acentuación; incluye aspectos como la afinación (melodía), el compás (ritmo), las pausas, la rata de hablado, el énfasis en las palabras. Una correcta prosodia es importante para hacer correctos los sonidos del habla y el correcto significado de una oración.
- **Producción de la onda:** Los fonemas y la información prosódica son usados para producir la onda de audio para cada oración. Existen muchas formas en las que se puede producir el habla a partir de los fonemas y la información prosódica. Las más comunes son concatenación de trozos de voz humana grabada, y síntesis de formantes usando técnicas de procesamiento de señales basadas en conocimiento de cómo los fonemas suenan y cómo la prosodia afecta esos fonemas. Esta temática se explica con más detalle en la sección 2.2.2 Técnicas de síntesis.

Cada uno de los idiomas existentes contiene reglas fonéticas propias que indican cómo se debe realizar la pronunciación de las palabras. En el Español particularmente, la pronunciación coincide con la forma en que se escribe, sin embargo, estructuras como fechas, abreviaturas, cuentas de correo electrónico, números telefónicos, cuentas

bancarias, entre otras, tienen una manera particular de pronunciarse. Esta situación representa uno de los mayores inconvenientes para la síntesis de voz proveniente de texto escrito, por esta razón se debe realizar un análisis de la estructura del texto a sintetizar, además de un procesamiento del mismo donde por medio de algoritmos y aplicando las reglas propias del idioma español, se segmenta el texto en partes que sirvan de entrada al sintetizador, identificando separadores (espacios en blanco, signos de puntuación), acentos ortográficos (tildes), entre otros.

### **2.2.1 Parámetros de caracterización de los sistemas de síntesis de voz**

La calidad de los sistemas de síntesis de voz es medida desde varios puntos [7].

- La Voz
  - Naturalidad: Indica qué tanto se parece a una voz real, es decir que parezca una persona hablando, y no que por el contrario, se escuche una voz robótica.
  - Inteligibilidad: Indica la facilidad o dificultad con que se entiende.
- Versatilidad
  - Mensajes limitados: Tiene un rango limitado de palabras o frases, y en caso de ser probado el sistema con frases ajenas a dicho rango, no podrá responder de manera adecuada.
  - Mensajes ilimitados: El sistema está en capacidad de sintetizar gran cantidad de frases o palabras.
- Complejidad del procesamiento
  - Condicionantes lingüísticos: Se relaciona con la facilidad o dificultad de realizar el análisis de un idioma, y descomponerlo en unidades básicas para la realización de una herramienta de síntesis de voz.
  - Condicionantes tecnológicos: Hace referencia a las tecnologías necesarias para realizar las implementaciones: ¿existen?, ¿es posible acceder a ellas?

## **2.2.2 Técnicas de Síntesis**

Existen varias técnicas de síntesis, cada una de ellas busca el mejor equilibrio entre naturalidad e inteligibilidad. Sin embargo, algunas de ellas son mejores que otras en una u otra característica.

### **2.2.2.1 Síntesis por Concatenación**

Se basa en tener un conjunto de pequeños segmentos de voz tomados de un hablante, grabados previamente, los cuales se van concatenando para formar el discurso deseado [11]. Requiere un diccionario de unidades de síntesis. Esta técnica pretende mejorar la calidad de la voz sintetizada minimizando el ruido de codificación; aparte de tener una complejidad variable, genera una voz de alta calidad. Se debe tener especial atención en la unidad a escoger para la concatenación, ya que este es el parámetro clave de ésta técnica. Existen varias técnicas que pertenecen a este grupo y se diferencian principalmente en el tipo de unidad de síntesis escogida. Actualmente muchos sintetizadores utilizan este tipo de técnicas debido a que proporcionan buenos resultados con menores esfuerzos, en comparación con otras técnicas.

#### *2.2.2.1.1 Síntesis por selección de unidades*

Utiliza una base de datos de voz grabada (corpus de voz), obtenida a partir de la segmentación de una grabación en unidades como fonemas, sílabas, palabras y oraciones. De esta manera, al momento de la síntesis, el sistema hace uso de unidades de diferentes tamaños que se adapten a la frase que se desea procesar [12].

#### *2.2.2.1.2 Síntesis de Difonemas*

Conseguir habla sintetizada de calidad a base de concatenar sonidos aislados e intentando imitar el resultado acústico de los movimientos del tracto vocal en las transiciones entre sonidos es una operación muy difícil. Por ello, las unidades a partir de las cuales se construye un sistema de síntesis no suelen ser sonidos aislados, sino combinaciones de sonidos. Con el uso de difonemas se persigue que al concatenar las unidades la unión se produzca por las partes en las que existe una menor variación acústica (típicamente en el «centro» de un sonido) y no por aquellas en las que se encuentra la transición de un sonido a otro [8].

Esta técnica utiliza una base de datos que contiene un ejemplo de todos los difonemas de una lengua específica. La calidad de la voz es generalmente peor que la de la síntesis por selección de unidades, pero suena mejor que la síntesis por formantes [10].

#### *2.2.2.1.3 Síntesis específica para un dominio*

Utiliza la concatenación de palabras y frases grabadas, esto hace que el rango de mensajes que puede producir sea muy pequeño, lo que hace de esta técnica una buena opción en soluciones de propósito específico, pero una muy mala alternativa para soluciones de propósito general.

#### 2.2.2.2 Síntesis de Formantes

Esta técnica es también conocida como síntesis basada en reglas. A diferencia de la síntesis por concatenación, la síntesis de formantes no utiliza grabaciones de voz reales para construir una base de datos, sino que genera voces artificiales a partir de modelos acústicos, creando ondas sonoras mediante la variación de parámetros como los niveles de ruido, la frecuencia fundamental, amplitud y tiempo de apertura de la glotis en un periodo determinado [10].

La fuente de voz usada en los sintetizadores ha ido variando a través del tiempo. Los modelos más antiguos utilizaban trenes de impulsos o dientes de sierra. Más tarde se pasó a modelos matemáticos cada vez más complejos que permiten controlar los parámetros principales de la señal [12].

La principal ventaja de la síntesis de formantes sobre las otras técnicas es que no requiere una base de datos, por esta razón es de menor tamaño; sin embargo, las voces generadas suelen ser poco naturales.

#### 2.2.2.3 Síntesis Articulatoria

Se basa en los mecanismos naturales para la producción de la voz, utilizando parámetros como la posición de la lengua, el tamaño de la cavidad bucal, etc. Aplica señales armónicas a la señal sonora y establece una analogía entre parámetros relacionados con los órganos articulatorios, sus movimientos y parámetros. La ventaja de esta técnica es que proporciona voz sintética de alta calidad, pero su inconveniente es que los

parámetros son muy difíciles de obtener y controlarlos automáticamente; además de esto, no es muy comercial debido a que las aplicaciones desarrolladas con esta técnica son costosas y especializadas [12].

#### 2.2.2.4 Síntesis por derivación de técnicas de predicción lineal (LPC)

Trabaja con parámetros Lineal Predictive Coding (LPC) logrando sintetizadores de análisis-síntesis, para controlar la función de transferencia del filtro que simula el tracto vocal. LPC es uno de los métodos más útiles para codificar el habla con calidad en un índice binario. Proporciona estimaciones extremadamente exactas de los parámetros del habla, que resultan muy eficientes [12].

LPC asume que la señal del habla es producida por un zumbador en el extremo de un tubo. La glotis (espacio entre cuerdas vocales) produce el zumbido, que es caracterizado por su intensidad y frecuencia (pitch). La zona vocal (garganta y boca) forma el tubo, que es caracterizado por sus resonancias, que se llaman “formantes”. LPC analiza la señal del habla, estimando los formantes, quitando sus efectos y estimando la intensidad y la frecuencia del zumbido restante, este proceso recibe el nombre de “filtrado inverso” y la señal restante se denomina “residuo”.

LPC invierte el proceso para sintetizar la señal del habla de la siguiente manera: utiliza el residuo para crear la señal fuente, utiliza los formantes para crear un filtro (representa al tubo), y funciona la fuente a través del filtro, dando como resultado, el “habla”. El problema básico del sistema LPC, es determinar los formantes de la señal del habla; la solución utilizada básicamente es una ecuación diferencial, la cual expresa cada muestra de la señal como combinación lineal de muestras anteriores, a esta ecuación se le denomina “linear predictor”, de ahí el nombre de la técnica.

### 2.3 CORPUS DE VOZ

Se entiende por “corpus”, una recopilación de elementos lingüísticos seleccionados y ordenados de acuerdo con criterios explícitos, con la finalidad de ser usados para el estudio de una lengua o una determinada variedad lingüística. Según su tipología, el corpus básicamente se clasifica en dos: escritos y orales [13].

El desarrollo de este proyecto se centra en la clasificación oral del corpus; el cual se dedica al estudio fonológico, y está conformado por una recopilación de lenguaje hablado.

La estructura del lenguaje hablado está determinada por diferentes factores:

- La estructura fonológica, sintáctica y prosódica de lo que se dice.
- El entorno acústico y contexto en el que se produce (diferente entonación y ruido).
- Medio de transmisión empleado.

Se debe tener en cuenta que personas hablantes de la misma lengua varían en su hablar, no específicamente textualmente, sino dialéctica, prosódica, sintáctica y acústicamente, lo que conlleva a una infinidad de corpus de voz particulares. Esto está ligado a los patrones del habla, los cuales están influenciados por:

- El entorno (la calle, un estudio de grabación, un bar, un estadio, etc.).
- El contexto social de acuerdo al rol que desempeña la persona.
- El nivel social de los participantes.
- El estado emocional y físico.

Para poder diferenciar los efectos de estos factores de variabilidad en los diversos corpus, es necesario mantener documentación paralela referente a cada situación.

En el caso específico de la síntesis de voz basada en concatenación de unidades se requiere una base de datos de la cual se extraen dichas unidades para formar la voz sintética, esta base suele denominarse "corpus" y debe incluir el etiquetado de las unidades fonéticas. [14].

## **2.4 FREETS**

Actualmente existen en el mercado diversos productos para la síntesis de voz, la mayoría de ellos licenciados, sin embargo últimamente han surgido algunos proyectos de software libre de este tipo, en universidades y comunidades.

La principal limitación de estos productos son los idiomas que ofrecen, ya que la mayoría de éstos, tienen como lengua principal el inglés, aunque algunos ofrecen diferentes idiomas y variaciones de éstos.

FreeTTS es un sistema de síntesis de voz escrito completamente en el lenguaje de programación Java. Está basado en Flite: un pequeño motor de síntesis de voz en tiempo de ejecución desarrollado en la Universidad Carnegie Mellon. Flite es derivado de Festival, un sistema de síntesis de voz de la Universidad de Edimburgo y el Proyecto FestVox de la Universidad Carnegie Mellon [15].

Este sintetizador es open source y posee tres voces en inglés (dos voces masculinas basadas en difonemas, y una masculina para dominio limitado); además soporta importación de voces de otros sistemas de síntesis como FestVox<sup>1</sup> y MBROLA<sup>2</sup>, sin embargo, este soporte es sólo para voces en inglés, debido a que FreeTTS ha sido desarrollado para este idioma.

## **2.5 JAVA MEDIA FRAMEWORK (JMF)<sup>3</sup>**

Java Media Framework es un API (Application Programming Interface), para la incorporación de medios basados en el tiempo (time-based media) en aplicaciones Java y applets. Los medios basados en el tiempo son medios tales como audio, video, MIDI y animaciones que cambian respecto al tiempo. Inicialmente JMF (versión 1.0) fue creado para permitir a los desarrolladores de software Java incluir este tipo de medios en sus aplicaciones. Actualmente el JMF (versión 2.1) extiende el área de trabajo, proveyendo soporte para captura y almacenamiento de medios, controlando el tipo de procesamiento que es efectuado durante la reproducción y la personalización de procesos sobre flujos de medios (Captura – Procesamiento – Presentación)[16].

Toda información que presente cambios significativos respecto al tiempo, puede ser catalogada como un medio basado en el tiempo, como lo son clips de audio, secuencias

---

<sup>1</sup> Este proyecto es parte del trabajo del grupo de habla de la Universidad Carnegie Mellon, encaminado al avance del estado de la síntesis de voz. Mayor información en <http://festvox.org/>

<sup>2</sup> Proyecto iniciado en el Laboratorio TCTS de la Facultad Politécnica de Mons (Bélgica), cuyo objetivo es obtener un conjunto de sintetizadores de voz para tantos idiomas como sea posible. Mayor información en <http://tcts.fpms.ac.be/synthesis/mbrola.html>

<sup>3</sup> Disponible en: <http://java.sun.com/javase/technologies/desktop/media/jmf/>

MIDI, clips de video, entre otros. Estos medios pueden ser obtenidos de diversas fuentes como archivos locales o remotos, cámaras, micrófonos y difusiones en vivo.

A continuación se presenta un modelo que describe las características de estos medios y la manipulación que a estos se les puede aplicar:

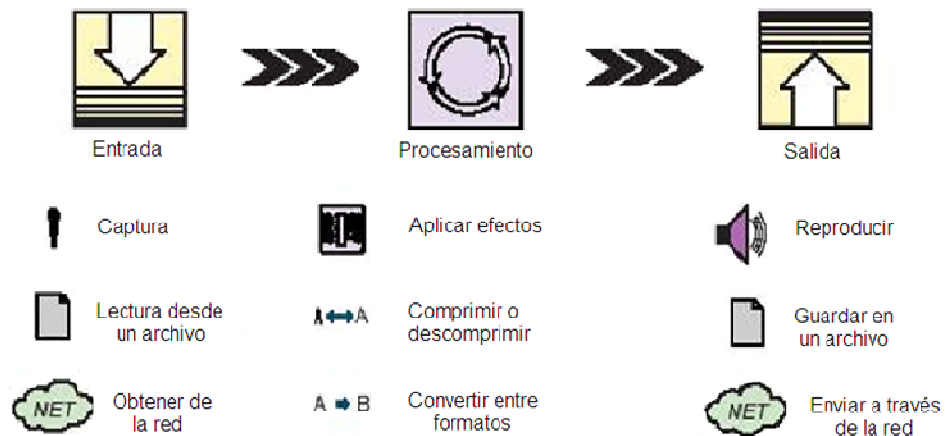


Figura 3. Modelo de Procesamiento de Medios Basados en el Tiempo

Fuente: [16]

Como muestra la Figura 3, los medios pueden ser capturados desde una fuente en vivo para procesarla o reproducirla, o puede ser adquirida de un archivo en forma local o remota (red). En varios casos, la información contenida en un medio necesita ser manipulada antes de ser presentada al usuario, ya sea, multiplexándola, filtrándola, comprimiéndola o convirtiéndola a otro tipo de medio. Por último la presentación o salida de medios debe o puede darse a través de dispositivos de salida como parlantes o monitores, o también almacenarse en un archivo localmente, o bien, ser transmitido a través de la red (difusión).

### 2.5.1 Captura de Medios (MediaLocator y DataSource)

Un dispositivo de captura multimedia puede actuar como una fuente de entrega de medios basados en el tiempo. Algunos dispositivos entregan múltiples flujos de datos que pueden ser separados mediante el procesamiento [16].

La entrada de datos en JMF se direcciona normalmente por un objeto *javax.media.MediaLocator*. Un *MediaLocator* se asemeja a una URL, indicando la ruta donde se encuentra el archivo o la ruta desde donde se va a capturar el medio.

Un ejemplo de *MediaLocator* para cargar un archivo de audio ya existente en una carpeta de la computadora sería: `C:/Proyecto/grabaciones/audio1.wav`

Una fuente de datos multimedia se modela mediante un objeto *DataSource*. De un *MediaLocator* se crea una instancia de *javax.media.protocol.DataSource* que representa la entrada de datos para cualquiera de los componentes de la JMF: *Players*, *Processors* o *DataSinks*. Este *DataSource* encapsula información sobre su ubicación, el protocolo para su transferencia al reproductor y el software a tratarlos.

### 2.5.2 Procesamiento de Medios (Processor)

Si se quiere tener un control total sobre el procesamiento de los medios (por ejemplo, controlar el tipo de códec que utiliza), o si desea mandar la información a otro destino distinto de la pantalla o altavoces entonces se debe utilizar la clase *Processor* (extiende de *javax.media.Player*). En la Figura 4, se muestra la jerarquía de la clase *Player*; allí se recalca la dependencia de la clase *Player* de la interfaz *Clock* y *Duration*, que indica que los medios a reproducir deben ser basados en el tiempo.

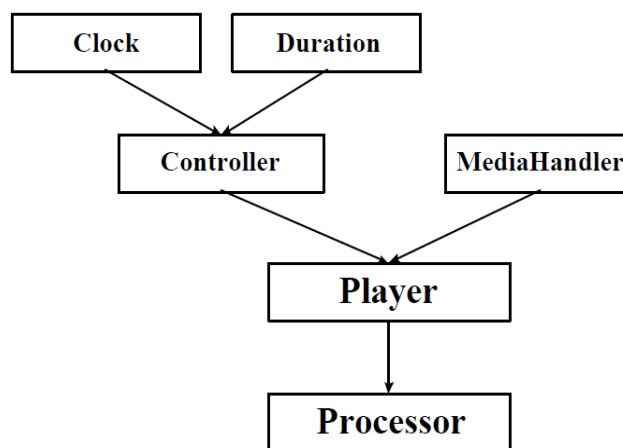


Figura 4. Jerarquía de la clase *Player*

Fuente: [17]

El Processor es un reproductor que toma una fuente de datos, realiza algunos procesos definidos por el usuario, y luego entrega la información de medio procesada. Las etapas por las que un medio atraviesa son la de: demultiplexación, transcoding (codificación o decodificación), multiplexación y renderización. (Ver Figura 5.)

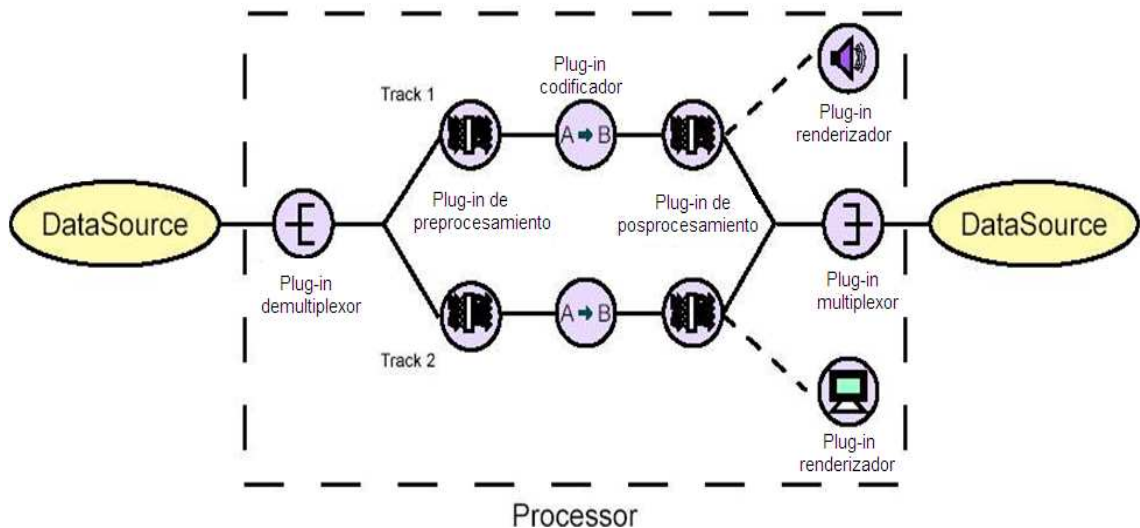


Figura 5. Modelo de un Processor JMF

Fuente: [18]

Para la creación de un Processor, existe el método `createManager()` de la clase `Manager`, el cual devuelve un `Processor`. Se puede crear un `Processor` desde cualquier tipo de medio ya sea `DataSource`, `MediaLocator` o `java.net.URL`, como sigue:

```
Processor p = Manager.createProcessor (Medio nombre_del_medio);
```

### 2.5.3 Reproducción de Medios (Player)

Para la presentación de medios, el JMF cuenta con los objetos `Player` los cuales se encuentran en el API por medio de `javax.media.Player`, extendiendo de la clase `Clock` (como se mostraba en la jerarquía del `Player` anteriormente), por eso se dice que son medios basados en el tiempo. La reproducción de los medios se puede controlar por medio de programación o bien a través de los componentes visuales de interacción con los controles (panel de control y control de ganancia) [18].

Si se tienen diferentes fuentes de medios se deberá utilizar un Player para la reproducción de cada una de ellas; pero si requiere una reproducción sincronizada de esas fuentes, entonces se puede hacer que un Player controle la reproducción de todos.

Para crear un reproductor también se hace uso de la clase Manager como se indicaba anteriormente para el caso de los procesadores. En este caso el parámetro pasado al método para la creación, también puede ser una *java.net.URL* o un *MediaLocator* o bien un *DataSource*. La instrucción para la creación es la siguiente:

```
Player player = Manager.createPlayer(Medio nombre_del_medio);
```

Al igual que el Processor, el Player debe pasar por unos estados anteriores antes de reproducir el medio, en este caso esos estados son: *Realizing* y *Prefetching*. Para pasar al primer estado (*Realizing*) se invoca al método *realize()*, aquí por ejemplo se realiza la tarea de averiguar el tamaño de la fuente de medios que se va a reproducir. Ahora teniendo el Player en el estado *Realizing* se debe pasar a *Prefetching*, para lograrlo se hace llamado al método *prefetch()*, en éste estado se inicia el llenado de los buffers de alimentación del decodificador. Los métodos *realize()* y *prefetch()* son asíncronos y retornan instantáneamente. Al cumplir cualquiera de los anteriores métodos, el player, lanza un evento *RealizeCompleteEvent* o *PrefetchCompleteEvent*.

Un Player en el estado *Prefetch* se encuentra listo para reproducir, para iniciar la presentación se necesita: Especificar el tiempo desde donde se quiere iniciar la reproducción del medio, invocando al método *setMediaTime()*, y llamar al método *start()* para iniciar la reproducción.

Para detener la reproducción de un medio se puede producir cualquier situación siguiente:

- Llamar al método *stop()* del Player.
- Se alcanza el final de la fuente reproducida.
- Llega al punto de stop de la reproducción, creado por medio del método *setStopTime()*.
- Cuando la tasa de recepción del medio es demasiado lenta para la reproducción del mismo a la tasa especificada.

Se debe tener especial atención en la liberación de recursos cuando se utiliza el JMF, debido a que se trabaja con archivos multimedia que generalmente son demasiado pesados para aplicaciones comunes. El JMF brinda el método *deallocate()* para que el Player libere el máximo número de recursos utilizados; éste método solo puede ser utilizado cuando el Player esté en estado *stop()*, de lo contrario se lanzará una excepción *ClockStartedError*. Si se llama a *deallocate()* una vez el Player a superado el estado *Realized*, éste volverá a ese estado; y si se encuentra en el estado *Realizing* el Player retorna al estado inicial o *Unrealized*

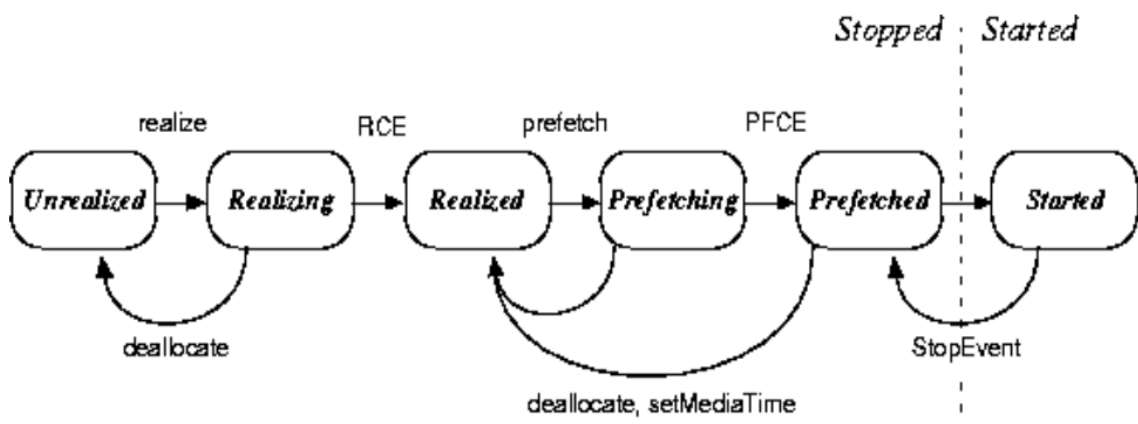


Figura 6. Estados de un Player

Fuente: [19]

#### 2.5.4 Almacenamiento de Medios (DataSink)

Esta opción contrasta con la de presentación de medios, ya que no es para reproducir, ni para actuar como un medio de difusión (papel que juega el Player). Aquí se trata de almacenar en un fichero el flujo de datos recibido de una fuente; para realizarlo, se necesita de un objeto de tipo *DataSink* para leer datos desde el objeto de salida de un Processor (*DataSource*) y enviarlos a un archivo.

Los pasos a realizar son [16]:

1. Recibir el *DataSource* de la salida de un Processor llamando al método *getDataOutput()*.

2. Construir un objeto *DataSink* por medio de un manager (`Manager.createDataSink`), pasándole como parámetros el *DataSource* obtenido del processor y la URL o *MediaLocator* donde se almacenará: `createDataSink(DataSource, MediaLocator)`.
3. Realizar el llamado el método `open()` del *DataSink*, para abrir el fichero donde se almacenará el flujo de datos.
4. Por último, invocar el método `start()` del *DataSink* para comenzar a escribir lo datos sobre el archivo de salida. El formato que tendrá dicho archivo es controlado por el processor del que se obtuvo el *DataSource*.

## 2.6 EL ESPAÑOL EN COLOMBIA

“Hasta dentro del rincón donde el castellano nació hay que rendirse a la evidencia de su falta de uniformidad”. (Vicente García de Diego, *El castellano como complejo...*, pág. 108).

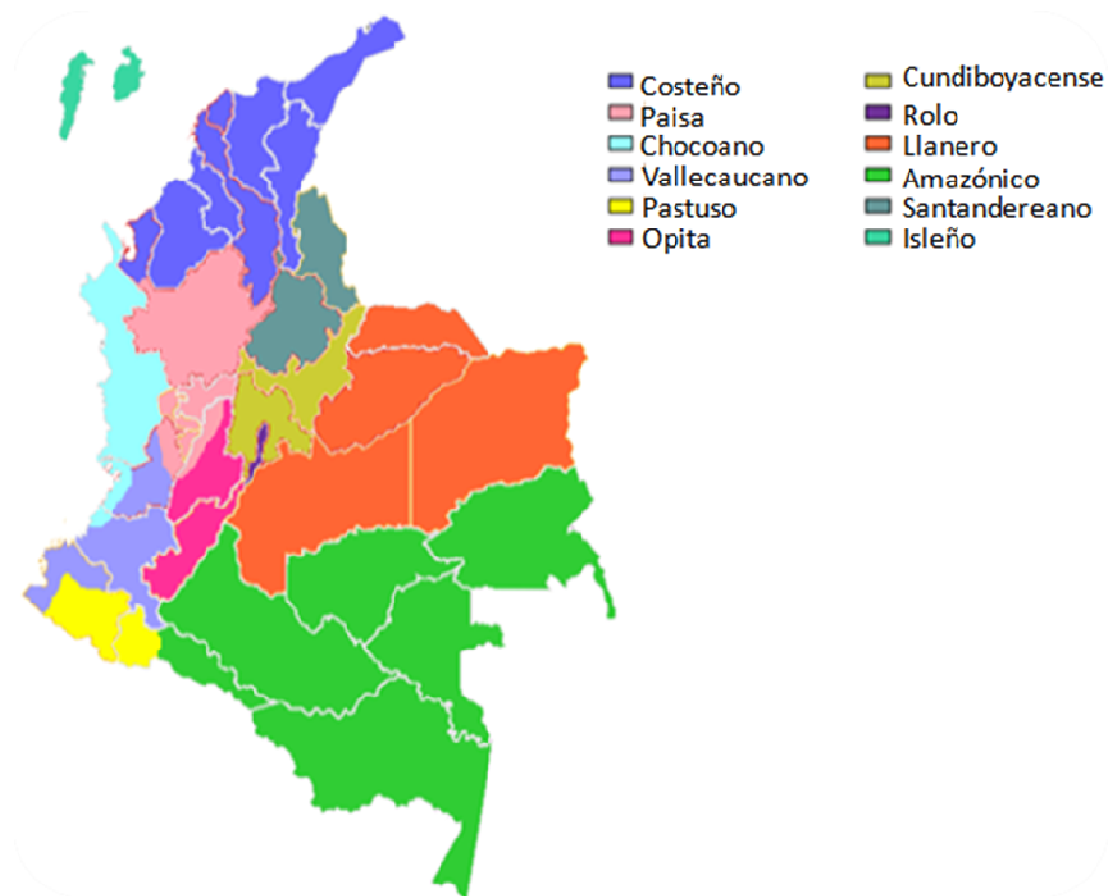
Notoria es la diferencia entre el castellano de un país como Colombia y otro como España; o si se toma un ejemplo más cercano, no es igual escuchar a un colombiano que a un argentino, aunque los dos hablen en esencia el mismo idioma. Respecto a esto José Joaquín Montes Giraldo expresa: “Una cosa sí parece quedar cada vez más clara en las discusiones de los últimos años: la esencial variedad o no unicidad del sistema que constituye una lengua real histórica” [20].

Aún dentro del mismo país, se distinguen diferencias en los hablantes; no sólo por la variedad de acentos que se presentan sino también por los modismos que se utilizan en cada una de las regiones. En este proyecto, específicamente para el sintetizador de voz, el interés está centrado en los acentos.

“El español de Colombia es una variedad de la lengua española compuesta por un conjunto de hablas regionales con rasgos específicos de pronunciación, de entonación y de vocabulario, principalmente, que permiten su clara diferenciación frente a otras modalidades nacionales”. [21].

De acuerdo a los rasgos que se definan para realizar una clasificación de los dialectos del país pueden generarse diferentes divisiones. La Figura 3 muestra una división dialectal de 12 regiones: costeño, paisa, chocoano, vallecaucano, pastuso, opita, cundiboyacense, rolo, llanero, santandereano, amazónico e isleño. Cada una de estas regiones se diferencia de las demás tanto por su acento específico, como por las variaciones en la pronunciación de algunos sonidos.

Por estas razones, la implementación de tecnologías del habla en Colombia requiere principalmente una contextualización en cuanto a la región en donde se aplicará. Específicamente para la síntesis de voz, es necesario un estudio de los sonidos utilizados en la región de aplicación, dependiendo de la técnica y de la unidad de síntesis escogida.



*Figura 7. Clasificación Dialectal de Colombia*

*Fuente: Wikipedia [22]*

### 3 HERRAMIENTA CHSPEECH

CHSpeech es la herramienta que se desprende como producto del presente proyecto. Está compuesta por un sintetizador de voz, un corpus o base de datos de voz, una interfaz gráfica que permite a los usuarios acceder más fácilmente a las funcionalidades del sintetizador, y un mecanismo de transmisión de datos desde un equipo de cómputo hacia el dispositivo celular. A continuación se presenta en detalle cada uno de estos componentes.

#### 3.1 SINTETIZADOR DE VOZ

El “corazón” de CHSpeech es un sintetizador de voz de dominio ilimitado, desarrollado completamente en el lenguaje de programación Java Standard Edition (JSE), utilizando el IDE Netbeans 6.5.1<sup>4</sup>. La técnica de síntesis utilizada se denomina concatenación de difonemas.

Las técnicas del habla, específicamente la síntesis de voz, requieren que se determine el idioma a utilizar de acuerdo al entorno de los usuarios finales; en el caso del presente proyecto, la herramienta ha sido ideada para ser utilizada en Colombia, por lo tanto, se hace necesario que el idioma manejado sea el español.

Existen en la actualidad varios motores de síntesis de voz para este idioma, sin embargo, cuando se habla de español, se hace referencia a la variante española (España), y es bien sabido que existen diferencias muy notorias, relacionadas con la pronunciación, entre dicha variante del idioma y la variante colombiana. Por esta razón, resultaría inadecuado utilizar un sintetizador para español de España.

Debido a que no se tiene conocimiento acerca de desarrollos de este tipo en Colombia, se optó por el desarrollo de un sintetizador y no la utilización de uno existente.

Como punto de partida para el desarrollo del sintetizador es necesario tener en cuenta las menciones hechas en el inciso 2.5. En Colombia se encuentran varios acentos diferentes correspondientes a las regiones o departamentos en que se encuentra dividido el país.

---

<sup>4</sup> Netbeans es un entorno de desarrollo libre, open-source, creado por Sun Microsystems. <http://www.netbeans.org/>

Por esta razón, la síntesis realizada intenta lograr el acento más neutro posible, de manera que la herramienta pueda ser utilizada en un amplio rango de lugares.

Es importante recalcar, que el sintetizador obtenido no incluye acentuación debida a signos de puntuación como: interrogantes (¿?) o admiración (¡!), pero si tiene en cuenta acentuación ortográfica representada por las tildes.

### 3.1.1 Procesadores

La función del sintetizador es convertir un texto de entrada en una salida de audio. Para lograr esto, es necesario que dicho texto pase por una serie de procesadores encargados de realizar diferentes tareas hasta llegar al objetivo. Se realiza un procesamiento en cadena, es decir que la salida de un procesador se convierte en la entrada del siguiente. (Ver Figura 4).

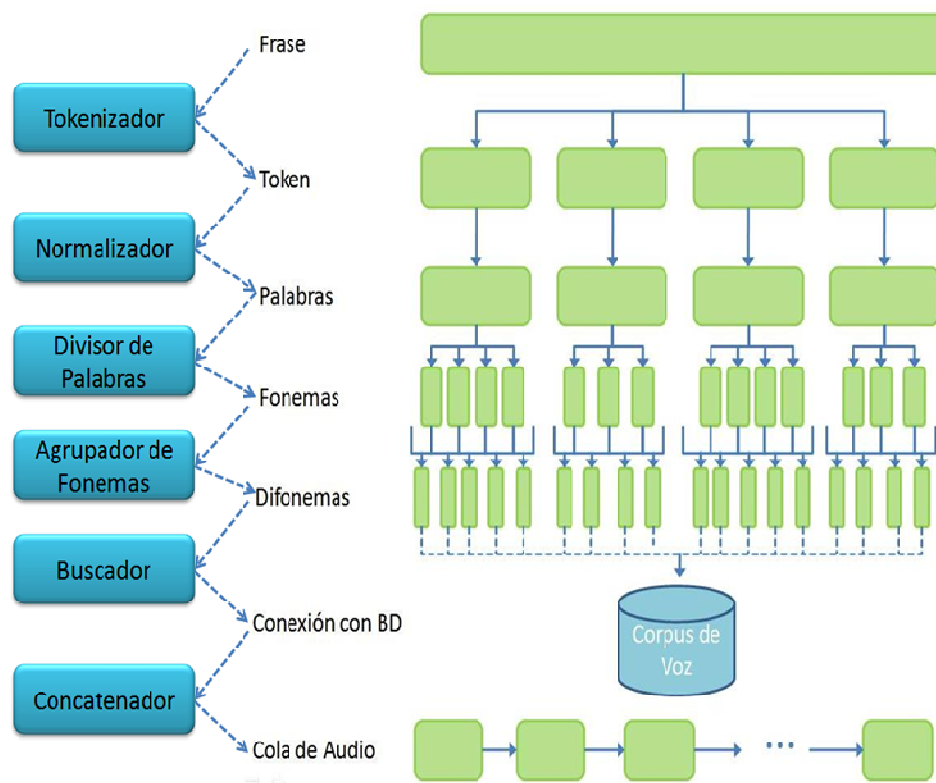


Figura 8. Procesadores del Sintetizador

Fuente: Autores del Proyecto

#### 3.1.1.1 Tokenizador

Es el procesador inicial; a éste ingresa una cadena de caracteres que se ha denominado frase, y la salida que proporciona es una lista de trozos o “tokens”. Su función es realizar particiones en el texto de entrada, de manera que quede dividido en trozos independientes. La forma de identificar el fin de un trozo y el comienzo de otro, es el espacio en blanco que existe entre ellos.

#### 3.1.1.2 Normalizador

La característica de dominio ilimitado del sintetizador implica que debe estar en condiciones de procesar cualquier entrada que realice un usuario, independientemente de la construcción que ésta presente. Debido a esto, se hace necesaria la implementación de un normalizador, que se encargue de identificar las diferentes estructuras ingresadas por el usuario y a su vez, establecer el modo en que se procesa cada una de ellas.

Para ello se definieron una serie de formatos que permiten la identificación de estructuras como fechas, horas, números telefónicos (celulares y fijos), números romanos, cantidades de dinero, direcciones, páginas web, correos electrónicos, fraccionarios, números ordinales, números decimales y enteros, o simplemente palabras.

Cada una de estas estructuras tiene una forma de pronunciación específica que intenta ser lo más intuitiva posible, y ha sido definida con base en las formas comunes en que lo hacen las personas cotidianamente, además de las guías del Diccionario Panhispánico de Dudas<sup>5</sup> de la Real Academia de la Lengua Española (RAE).

La explicación de los formatos y pronunciaciones se encuentra en el ANEXO A.

Además de identificar las formas anteriormente mencionadas, el normalizador también se encarga de procesar construcciones como abreviaturas, símbolos y algunos extranjerismos.

En el caso de las abreviaturas, debido a que es posible generar nuevas a partir de casi cualquier palabra, se ha definido una lista de las más utilizadas en el hablar cotidiano de

---

<sup>5</sup> Obra que se propone servir de instrumento eficaz para todas aquellas personas interesadas en mejorar su conocimiento y dominio de la lengua española. Ver <http://buscon.rae.es/dpdl/>

manera que al introducirse alguna de ellas en el sintetizador, éste pueda pronunciar la palabra completa en lugar de intentar leer la abreviatura. En caso de que ésta no se encuentre en la lista, el sintetizador intentará pronunciarla y como última opción la deletreará. El Diccionario Panhispánico de Dudas de la RAE contiene documentación para un correcto uso y creación de abreviaturas<sup>6</sup>.

Normalmente las abreviaturas se escriben con la primera letra mayúscula, y finalizadas en punto; para efectos de comodidad al escribir, el sintetizador aceptará una abreviatura escrita en minúsculas pero el punto debe incluirse.

El ANEXO B contiene la lista de abreviaturas implementadas.

El manejo de los símbolos es similar al de las abreviaturas, se ha definido una lista de símbolos para que en caso de que se inserte alguno, el sintetizador pueda identificarlo y pronunciar su nombre. Este caso es más sencillo que el de las abreviaturas, puesto que el número de símbolos es más reducido.

El ANEXO C presenta la lista de símbolos.

Finalmente, también se ha definido una lista de extranjerismos de uso común en el español, con base en una recopilación de la Academia Colombiana de la Lengua [23] y en una lista de las palabras más comunes del inglés disponible en internet [24]. Aunque no se trata de hacer que el sintetizador sea “multilingüe”, se vio la necesidad de implementar este caso puesto que existen palabras extranjeras de uso cotidiano sobre todo en áreas específicas y que no tienen una traducción literal para el español. Sin embargo, es necesario aclarar que la pronunciación de estas palabras no será la adecuada de acuerdo al idioma al cual pertenecen puesto que como se ha mencionado antes, cada idioma tiene sus sonidos (fonemas) propios y es apenas normal que muchos de ellos no existan en el español de Colombia.

En ocasiones la escritura de palabras en otros idiomas coincide con la escritura de palabras en español. Por ejemplo: “once” en inglés y el número “once” en español. En estos casos se omitió la inclusión de palabras extranjeras.

La lista de extranjerismos se puede ver en el ANEXO D.

---

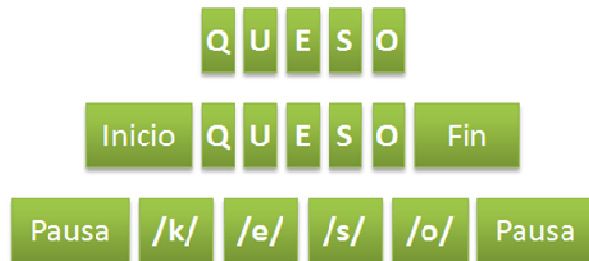
<sup>6</sup> Ver <http://buscon.rae.es/dpdl/SrvltGUIBusDPD?lema=abreviatura>

### 3.1.1.3 Divisor de palabras

La entrada a este procesador son las palabras obtenidas a través del normalizador. Se encarga, como su nombre lo dice, de dividir una palabra en sus fonemas correspondientes. La salida que proporciona este procesador es una lista de fonemas. El divisor de palabras realiza las siguientes tareas específicas:

- Recorrido de la palabra carácter por carácter: Se analiza cada una de las letras que conforman la palabra.
- Identificación del inicio y el fin de la palabra para introducir pausas.
- Identificación de la letra en posición “actual” y de sus vecinas (anterior y/o siguiente): En cada letra de la palabra se identifican las letras vecinas de manera que se pueda establecer el sonido que produce la letra en cuestión. Para la realización de este paso se deben tener en cuenta dos aspectos:
  - Previamente se identificaron los fonemas del español de Colombia, mediante el análisis a partir de las letras del alfabeto y los sonidos que se producen al pronunciar cada una de ellas; algunas letras presentan la misma pronunciación por lo que comparten el fonema. Esto haría que el número de fonemas fuese menor que el número de letras, pero es necesario contar también las vocales con tilde ya que el sonido que se produce es diferente al de las vocales sin tilde. En el desarrollo del presente proyecto se utilizó una nomenclatura de fácil comprensión para los autores y futuros lectores. Los fonemas para el español de Colombia se presentan en el ANEXO E.
  - Además de contar con los fonemas correspondientes al idioma, es necesario tener en cuenta las condiciones que se presentan para que una letra suene de una forma determinada. El sonido producido por la pronunciación de una letra en ocasiones depende de las letras que la rodean (vecinas). El ANEXO F muestra las condiciones que se deben dar para que una letra sea representada por cierto fonema.

- Asignación del fonema correspondiente: El fonema representa el sonido identificado en el paso anterior.



*Figura 9. Funcionamiento del Divisor de palabras*

*Fuente: Autores del Proyecto*

#### 3.1.1.4 Agrupador de Fonemas

A este procesador ingresa la lista de fonemas generada por el divisor de palabras. La función que desempeña el agrupador de fonemas es, como su nombre lo dice, agrupar los fonemas de una palabra, de manera que se obtenga la palabra compuesta por sus difonemas.

Siguiendo con el ejemplo de la Figura 9 se puede observar en la Figura 10 que cada uno de los fonemas se agrupa con sus vecinos, tanto con el anterior (izquierda) como con el siguiente (derecha). Los fonemas de comienzo y fin de palabra se agrupan con el fonema que se ha denominado pausa, encargado de introducir un pequeño silencio entre palabras.



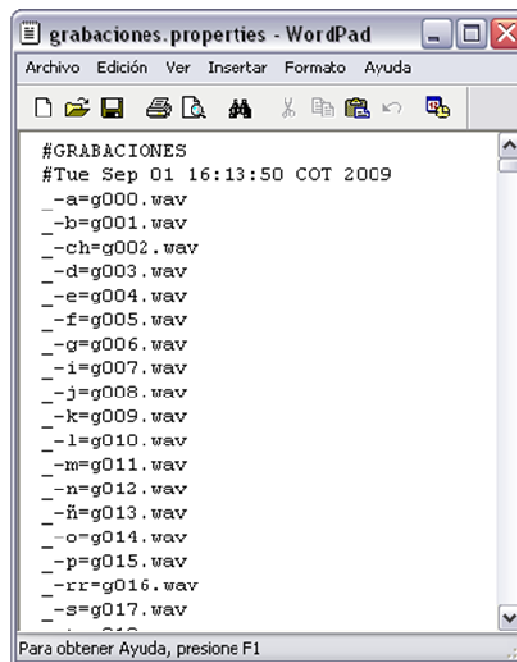
*Figura 10. Funcionamiento del Agrupador de Fonemas*

*Fuente: Autores del Proyecto*

### 3.1.1.5 Buscador

La entrada de este procesador es la lista de difonemas de las palabras obtenidas del proceso de normalización del texto. La función que cumple es realizar la conexión con la base de datos, en busca del archivo de audio correspondiente a cada uno de los difonemas de entrada.

Para el presente proyecto se utilizó un archivo .properties<sup>7</sup> que realiza el papel de la base de datos requerida. En este archivo se encuentra una serie de filas “clave-valor”; la clave corresponde a todos y cada uno de los difonemas existentes para el idioma en cuestión (lista de difonemas existentes, ver ANEXO G) y el valor corresponde al nombre del archivo de audio que contiene al difonema (clave). En la Figura 7 se muestra un fragmento del archivo .properties utilizado.



*Figura 11. Fragmento del archivo grabaciones.properties*

*Fuente: Autores del Proyecto*

<sup>7</sup> Archivo .properties es aquel que permite guardar propiedades de la aplicación como una serie de cadenas “clave-valor”; se utilizan para realizar referencias de localización de recursos.

### 3.1.1.6 Concatenador

Este es el último procesador del sintetizador; es el encargado de la generación de la señal que se reproducirá en el teléfono celular. A este procesador ingresa la lista de archivos correspondiente a los difonemas que se deben concatenar, para generar una única señal compacta que los contenga a todos y se pueda reproducir.

Para llevar a cabo su objetivo, este procesador utiliza el JMF (Java Media Framework) proporcionado por Sun Microsystems. Este framework permite que audio, video o cualquier otro tipo de archivo multimedia pueda ser añadido a aplicaciones y applets construidos con la tecnología Java. Entre sus principales funciones, se encuentran las de capturar, reproducir, transferir y codificar múltiples formatos multimedia que extienden la plataforma Java 2 Standard Edition (J2SE) para los desarrolladores multimedia, proporcionando un potente conjunto de herramientas para desarrollar aplicaciones escalables. Para este proyecto se utilizó la versión 2.1.1<sup>8</sup> del JMF, última actualización disponible a la fecha de realización del proyecto.

El Concatenador realiza las siguientes actividades principales:

- Carga los archivos de audio físicos a memoria RAM para poder procesarlos.
- Crea un procesador (JMF Processor) para cada uno de los archivos de audio cargados (implica creación de MediaLocator para cada uno).
- Extrae las pistas contenidas en cada uno de los archivos de audio a través de sus Processor, es decir, extracción de los DataSource de cada archivo de audio (el DataSource es la fuente principal de lo que va a sonar).
- Realiza la unión (concatenación) de todos los DataSource, convirtiendo la cola de audio en un único DataSource que será entregado al Player, el cual se encargará de su reproducción.

---

<sup>8</sup> JMF es un paquete opcional que debe descargarse y adjuntarse como librería al proyecto que desee utilizarlo. El paquete se acoge a la licencia de uso de Sun Microsystems; se encuentra disponible en: <http://java.sun.com/javase/technologies/desktop/media/jmf/2.1.1/download.html>.

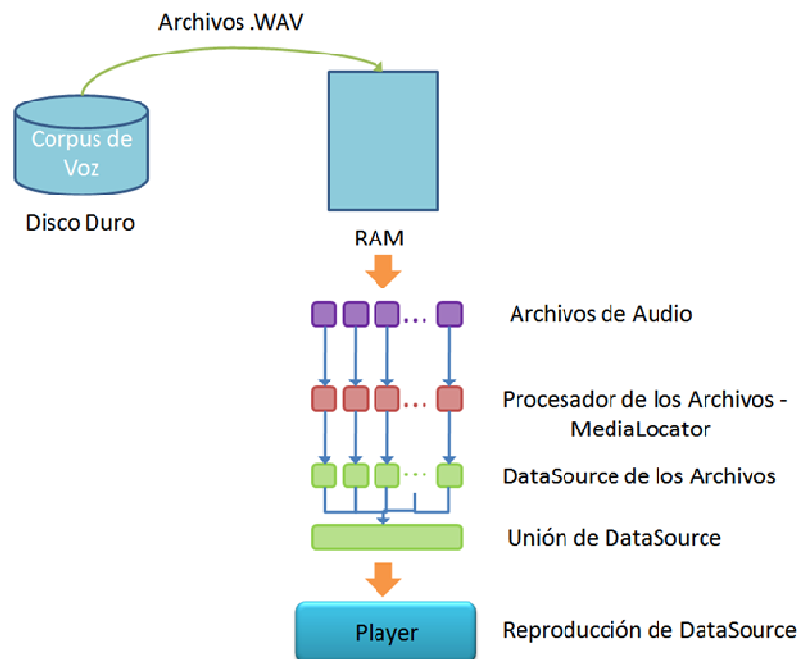


Figura 12. Esquema de las actividades del Concatenador

Fuente: Autores del Proyecto.

### 3.2 CORPUS DE VOZ

La herramienta CHSpeech requiere una base de datos de unidades de voz debido a que el sintetizador que fue desarrollado para ella, utiliza la técnica de concatenación de unidades. Específicamente se requiere una base de datos de difonemas, que es la unidad de síntesis seleccionada.

Las unidades más utilizadas son los fonemas, sílabas, semisílabas, difonemas y palabras. El hecho de escoger un tipo de unidad u otro, viene dado por un compromiso. En general, un idioma contiene entre 20 y 40 fonemas, así que si se utilizan fonemas como unidad de síntesis, se obtendría una base de datos de unidades de pequeñas dimensiones y de fácil manejo. Sin embargo, la calidad de la voz obtenida en este caso, no sería la deseada. Se puede captar intuitivamente que cuan mayor sean las unidades básicas de síntesis a utilizar, mayor será la calidad de la voz generada, pero si se utilizaran palabras o frases pregrabadas para sintetizar cualquier tipo de mensaje, las dimensiones de la base de

datos serían demasiado grandes y poco manejables (Ver Figura 13). Es así como se llega a un compromiso de “calidad – memoria” que conlleva a trabajar con difonemas.

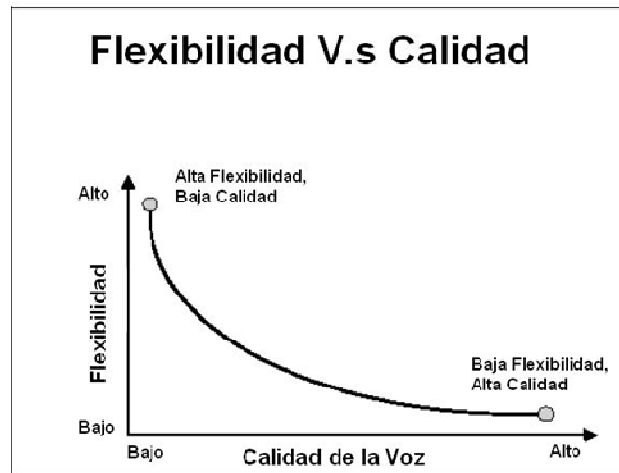


Figura 13. Gráfica de Flexibilidad Vs. Calidad

Fuente: [28]

Los difonemas son unidades de voz que contienen dos fonemas concatenados (vocal-consonante, consonante-vocal, vocal-vocal o consonante-consonante). Realmente lo que se va a utilizar al momento de sintetizar, es la parte de los difonemas comprendida entre el centro del primer fonema, pasando por la transición existente entre ambos, hasta la mitad del segundo fonema. El contener las transiciones de los difonemas pregrabadas es un buen indicio, hablando en términos de naturalidad e inteligibilidad.

Castillo en 1999 definió corpus de voz como “una colección de grabaciones de voz, las cuales están accesibles en forma de lectura dentro de una computadora y las cuales vienen con anotaciones y con documentación suficiente para permitir el reuso de los datos” [25].

Una colección de grabaciones de voz, como lo llama Castillo, no es más que una base de datos que permite la organización de esos fragmentos de audio de manera que puedan ser extraídos y utilizados en cualquier momento; eso es lo que precisamente se necesita para que el sintetizador pueda “hablar”: la voz.

La voz del sintetizador se obtiene extrayendo cada una de las grabaciones que conforman la base de datos. Estos fragmentos son las unidades escogidas para el sintetizador, es decir, si éste concatena difonemas entonces el corpus de voz estará formado por los difonemas del idioma para el cual se desarrolló. Las palabras de dicho idioma se formarán a partir de los difonemas extraídos de la base de datos, esta tarea de búsqueda es realizada por uno de los procesadores del sintetizador, explicado en el inciso 3.1.1.5.

Debido a que cada difonema es extraído de una grabación diferente, el conjunto de dichas grabaciones que proporcionan la voz al sintetizador debe cumplir ciertas características en uno de los intentos para que suene un poco más natural:

- Deben ser realizadas por un mismo locutor
- Mantener el tono y volumen al hablar
- Buena pronunciación y vocalización
- Neutralizar al máximo el acento de manera que el sintetizador pueda ser utilizado en varios lugares del país

El proceso de implementación del corpus de voz se realizó en dos etapas, la primera de ellas consta de un corpus de frases realizado a manera de prueba, y en la segunda se llevó a cabo la implementación del corpus de difonemas, que finalmente utiliza la herramienta CHSpeech.

### **3.2.1 Corpus de Frases**

A manera de prueba, con el fin de obtener mayor conocimiento acerca de las condiciones de grabación necesarias, se implementó un corpus de voz cuya unidad son frases. Se grabaron alrededor de 50 frases cortas, utilizadas comúnmente en conversaciones telefónicas.

Durante el diseño y la implementación de este corpus de voz se realizaron las siguientes actividades:

- Definición del lugar de grabación: Se escogió como escenario de grabación el apartamento de uno de los autores, cuidando que no hubiera ruido de personas, televisores, radios, etc. dentro de este.

- Determinación de la lista de frases a grabar: Como se mencionó anteriormente se escogieron alrededor de 50 frases cortas que se utilizan con frecuencia en una llamada telefónica como: aló, buenos días, buenas noches, hasta luego, ¿cómo está?, entre otras.
- Selección del formato de grabación: Partiendo de los formatos de audio que Java maneja, se definió que las grabaciones tendrían la extensión “.WAV”.
- Realización de las grabaciones: Con la voz de uno de los autores del proyecto, se procedió a grabar las frases seleccionadas.

La Tabla 1 resume las condiciones de grabación para esta fase.

<b>Lugar</b>	Sala de apartamento
<b>Micrófono</b>	Multimedia común
<b>Formato de Grabación</b>	WAV

*Tabla 1. Condiciones de grabación Corpus de Frases*

Finalizada esta etapa de grabación, se determinó que las condiciones dadas no eran suficientes para obtener audio de buena calidad, ya que en primera instancia, el lugar no era el adecuado puesto que no tenía la acústica necesaria y presentaba altos niveles de ruido del exterior. El micrófono utilizado no era de gran calidad, esto se veía reflejado en que cada grabación presentaba volumen diferente a pesar de que aparentemente, el micrófono se mantenía a la misma distancia del hablante y en la misma posición.

### **3.2.2 Corpus de Difonemas**

En esta segunda fase se realizó la implementación del corpus de voz que finalmente utiliza el sintetizador, cuya unidad son difonemas. En esta fase se realizaron diferentes actividades: diseño del corpus de voz, definición de las características de grabación, realización de las grabaciones, etiquetado de los archivos de audio, extracción de los fragmentos correspondientes a los difonemas y por último la implementación de la base de datos de la aplicación para que el sintetizador acceda a dichos archivos. Cada una de estas actividades se explica en detalle a continuación.

### 3.2.2.1 Diseño del corpus de voz

Inicialmente fue necesaria la identificación de los difonemas propios del español de Colombia, teniendo como punto de partida el listado de fonemas para el mismo idioma que se puede ver en el ANEXO E. Normalmente se piensa que el número de difonemas es igual al número de fonemas elevado al cuadrado (*difonemas = fonemas<sup>2</sup>*) puesto que los difonemas se producen al agrupar los fonemas en parejas (todos con todos), sin embargo se debe tener en cuenta que hay algunas combinaciones que no se presentan.

Para la identificación de los difonemas del español de Colombia se realizaron dos tablas, la primera de ellas se ha denominado tabla de difonemas y hace una representación de todas las posibles combinaciones de los fonemas especificando aquellas que se presentan y aquellas que no. La segunda tabla confirma la existencia de un determinado difonema puesto que para cada uno de ellos presenta una palabra ejemplo que lo contiene. Se utilizó Word Navigator<sup>9</sup> para la búsqueda de las palabras ejemplo y el diccionario de la Real Academia de la Lengua Española<sup>10</sup> para corroborar que la palabra encontrada efectivamente existe en el idioma español.

En muchos casos, existen varias palabras que contienen un difonema específico, por esta razón se escogieron, en la medida de lo posible, las palabras que contienen el difonema requerido en la mitad, no en los extremos. Sin embargo se presentaron excepciones en los casos que así lo requerían como difonemas de comienzo y fin de palabra. Los difonemas que se encuentran en la mitad de la palabra de donde se extraerán, son pronunciados de manera más natural que los que se encuentran en los extremos; es por eso que se hizo énfasis en encontrar palabras con el difonema requerido en el centro.

La Tabla 2 muestra una parte de la tabla de difonemas. El primer fonema del difonema está representado por la primera columna de la tabla y el segundo fonema por la primera fila. En cada celda aparece el nombre del difonema formado, y donde aparece únicamente un guión, significa que el difonema no existe. Así por ejemplo, el difonema ch-ch no existe, mientras que el difonema formado por los fonemas d y a se representa d-a.

---

<sup>9</sup> Página web que permite buscar palabras que contengan una serie de letras específicas. <http://wordnavigator.com/es/contains/>

<sup>10</sup> El diccionario puede consultarse en <http://www.rae.es/rae.html>

	pau	a	b	ch	d	e
pau	_ _	_ -a	_ -b	_ -ch	_ -d	_ -e
a	a _	a-a	a-b	a-ch	a-d	a-e
b	b _	b-a	b-b	b-ch	b-d	b-e
ch	-	ch-a	ch-b	-	-	ch-e
d	d _	d-a	d-b	-	-	d-e
e	e _	e-a	e-b	e-ch	e-d	e-e

Tabla 2. Fragmento de la Tabla de Difonemas

La tabla de difonemas completa, muestra los 590 difonemas encontrados para el español de Colombia. Puede verse en el ANEXO G.

El ANEXO H contiene la tabla con la palabra ejemplo correspondiente a cada difonema.

Los elementos del corpus de voz diseñado para este proyecto serán entonces un total de 590 grabaciones correspondientes a igual cantidad de difonemas. Expresado de otra manera el número de difonemas total existentes en el idioma Español de Colombia se define como:

$$difonemas_{totales} = fonemas^2 - difonemas_{inexistentes} = 590 \quad (1)$$

### 3.2.2.2 Preparación de las condiciones de grabación

Teniendo los difonemas y las palabras ejemplo, el siguiente paso en el desarrollo del corpus antes de la grabación es la definición de las condiciones para realizar este proceso.

Con base en la experiencia adquirida en la grabación del corpus de frases, se decidió cambiar radicalmente dichas condiciones. Principalmente fue necesaria la utilización de equipo técnico especializado, así como un lugar lo suficientemente aislado de ruido exterior. Además, se tuvieron en cuenta características más específicas en cuanto a la calidad del sonido como la frecuencia de muestreo<sup>11</sup>, resolución<sup>12</sup> y número de canales<sup>13</sup>.

<sup>11</sup> Número de muestras por unidad de tiempo que se toman de una señal continua para producir una señal discreta, durante el proceso necesario para convertirla de analógica a digital.

Es necesario recalcar que a pesar de tener en cuenta estos aspectos técnicos del sonido, la calidad de éste es buena mas no óptima, puesto que además de dichas características hay que tener en cuenta que el sintetizador debe manejar estas señales de audio, por lo tanto éstas deben limitarse a las características que soporte el lenguaje de programación utilizado, en este caso Java.

Así las cosas, en primer lugar se realizaron las gestiones necesarias para tener acceso a las cabinas de las emisoras de la UIS ya que éstas cuentan con el equipo técnico especializado para el manejo de voces, además están aisladas del ruido exterior. Seguidamente se definieron los aspectos técnicos mencionados anteriormente: frecuencia de muestreo, resolución y número de canales. La frecuencia de muestreo se estableció en 16 KHz debido a que un menor valor como 8KHz no proporciona calidad suficiente, y uno mayor como 32 KHz se torna muy lento de procesar, lo que no favorece a la herramienta CHSpeech. La resolución de 16 bits se definió en este valor para contribuir a la generación de audio de buena calidad y procesamiento rápido. Por otra parte, el número de canales es 1 (monofónico) puesto que sólo se reproducirá un sonido a la vez que es el de la voz, entonces no hay necesidad de utilizar más canales. Por último, el formato de grabación se mantuvo en .WAV, debido a su popularidad, calidad y a la facilidad que presenta su utilización en Java y en las herramientas de etiquetado.

La Tabla 3 resume las condiciones de grabación utilizadas en esta fase.

<b>Lugar</b>	Cabinas de radio – Emisoras UIS
<b>Frecuencia de Muestreo</b>	16 KHz
<b>Resolución</b>	16 bits
<b>Número de Canales</b>	1 (mono)
<b>Formato de Grabación</b>	WAV

*Tabla 3. Condiciones de grabación Corpus de Difonemas*

---

<sup>12</sup> Cantidad de bits utilizados para hacer el muestreo. Un muestreo con una resolución de 16 bits se aproximará más a la señal analógica original que uno de 8 bits puesto que el primero contendrá las frecuencias más altas.

<sup>13</sup> Especifica la cantidad de sonidos diferentes que pueden reproducirse simultánea e independientemente.

### 3.2.2.3 Grabación

El proceso de grabación del corpus de voz se realizó siguiendo las condiciones previamente descritas. En este corpus, a diferencia del corpus de frases, la voz encargada de realizar las grabaciones no es de uno de los autores sino de una persona con experiencia en el manejo de la voz, con buen tono, volumen, pronunciación, y que en general, contribuye notablemente en la mejora de la calidad del sonido respecto a la experiencia anterior.

Para la realización de las grabaciones se utilizó la frase portadora: “Él dijo \_\_\_\_\_. Yo sé que él dijo \_\_\_\_\_.”[26]. Insertando en cada espacio la palabra ejemplo correspondiente a cada difonema. El uso de la frase portadora permite al locutor mantener la naturalidad de la voz con mayor facilidad que cuando se leen palabras sueltas. También al pronunciar dos veces cada palabra dentro la frase es posible seleccionar la que tenga mejor pronunciación y así obtener una mejor representación del difonema.

Después de la grabación de las frases, se escogió entre las dos opciones de cada palabra ejemplo, teniendo como criterio de selección la claridad en la pronunciación, con el fin de extraer la palabra de la frase y así contar con información más precisa para conformar la base de datos. De esta manera se obtuvo el total de 590 archivos de audio, correspondientes a igual cantidad de palabras ejemplo.

### 3.2.2.4 Etiquetado de los archivos de audio

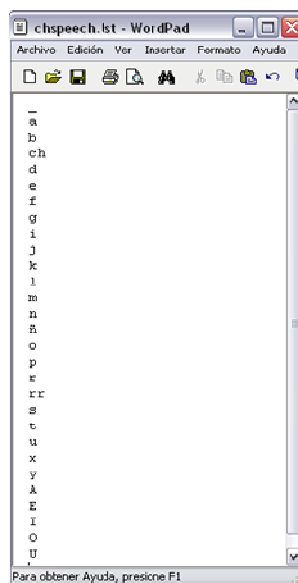
El etiquetado de un corpus de voz permite delimitar las fronteras de una unidad específica presente en las grabaciones. Las etiquetas se utilizan para marcar el inicio y el fin de estas unidades. Generalmente son archivos con información asociada a las grabaciones del corpus. Por medio de las etiquetas es posible extraer fragmentos de los archivos de sonido para utilizarlos en la creación de palabras y frases al unirlos con otros fragmentos en la síntesis por concatenación de unidades. [14]

La definición anterior proporciona una idea bastante clara acerca del concepto de etiquetado de un corpus. El proceso realizado en el desarrollo del presente proyecto involucra el etiquetado del corpus para identificar las unidades fonéticas utilizadas por el

sintetizador: los difonemas. Las etiquetas utilizadas indican el comienzo y el fin del difonema, y la transición entre los fonemas que lo conforman.

El etiquetado de las grabaciones se realizó utilizando el software Diphone Studio 1.3 desarrollado por Fluency<sup>14</sup>. Diphone Studio<sup>15</sup> es una herramienta para el desarrollo y mantenimiento de un conjunto de difonemas para su uso en síntesis de voz. [27]. Se escogió este software por su facilidad de uso, además porque utiliza como unidad fundamental los difonemas y finalmente porque es de uso libre para fines no comerciales.

El primer paso del etiquetado con Diphone Studio es incluir en un archivo de texto, el listado de fonemas del idioma en el formato que reconozca el sintetizador que utilizará el corpus de voz; el archivo debe guardarse con extensión “.lst”. La Figura 14 muestra el archivo *chspeech.lst*, que contiene los fonemas identificados en el presente proyecto.



*Figura 14. Contenido del archivo chspeech.lst*

*Fuente: Autores del Proyecto*

---

<sup>14</sup> Empresa que desarrolla software de conversión de texto a voz en holandés desde 1997. <http://www.fluency.nl>

<sup>15</sup> Disponible para descarga en <http://www.fluency.nl/dstudio/dstudio.htm>

Con base en el contenido del archivo “.lst”, Diphone Studio genera todos los difonemas posibles, combinando todos los fonemas entre sí; la cantidad de difonemas encontrados será el cuadrado de la cantidad de fonemas que contenga el archivo “.lst”.

Los difonemas generados se organizan a manera de tabla en un archivo con extensión “.dat” que es creado automáticamente por la herramienta. Cada línea del archivo corresponde a un difonema, además se dispone de cuatro columnas para agregar datos correspondientes a cada uno de ellos: las dos primeras indican el primer y el segundo fonema, respectivamente; la tercera columna indica el nombre del archivo de audio que lo contiene, y la cuarta indica la palabra ejemplo que contiene la grabación.

El software asigna inicialmente valores por defecto a la tercera y cuarta columna. En el caso de los nombres de los archivos de audio, Diphone Studio utiliza la nomenclatura “wn.wav” donde n varía dependiendo del número de difonemas que existan en el archivo. La cuarta columna aparece automáticamente llena con el término <UTTERANCE> puesto que no se han introducido en el archivo las palabras ejemplo.

Esta descripción puede verse más claramente en la Figura 15 que presenta un fragmento del archivo *chspeech.dat*, generado automáticamente por la herramienta Diphone Studio.

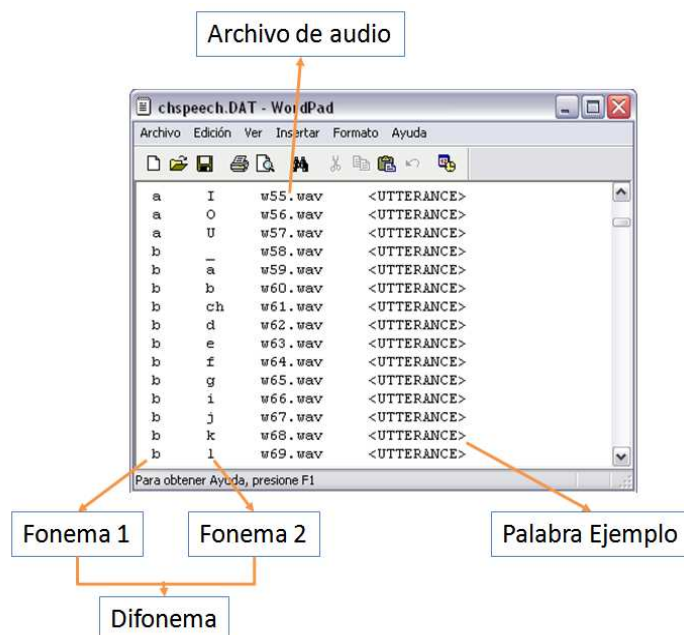
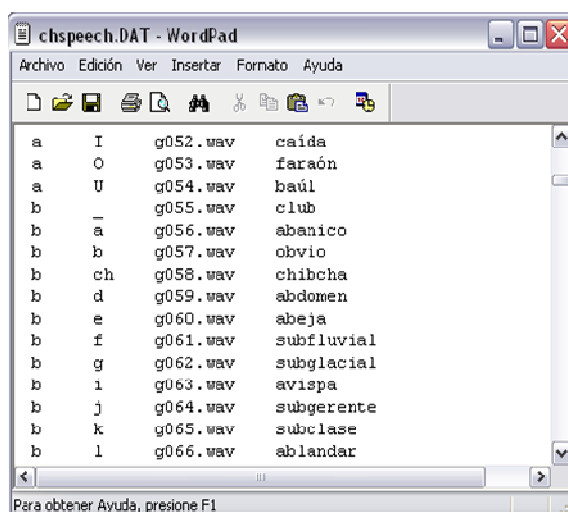


Figura 15. Fragmento del archivo *chspeech.dat* generado por Diphone Studio

Fuente: Autores del Proyecto

Antes de comenzar el etiquetado propiamente dicho, es necesario hacer una revisión del archivo “.dat” y eliminar de él los difonemas que no existan en el idioma en que se está trabajando; en el caso del presente proyecto, se editó el archivo de manera que los difonemas contenidos en él, coincidieran con el listado de difonemas identificados para el español de Colombia que se muestra en el ANEXO H.

También se editó el contenido de la tercera y cuarta columna. Esto implicó el cambio de los nombres de los archivos de audio, para hacerlos corresponder con los nombres de las grabaciones hechas previamente. Además del cambio de la palabra <UTTERANCE> por las palabras ejemplo de cada difonema. En la Figura 16 se puede ver el mismo fragmento del archivo que se presentó en la Figura 15, con la diferencia de que en esta nueva ilustración se aprecian los cambios realizados.



*Figura 16. Fragmento del contenido del archivo chspeech.dat con los valores reales*

*Fuente: Autores del Proyecto*

Una vez terminadas las modificaciones del archivo, es posible empezar el etiquetado. Como se mencionó en el comienzo de la sección, se utilizaron las tres etiquetas que permite insertar Diphone Studio que son: comienzo del difonema, parte media del difonema que separa los dos fonemas (donde termina el primero y empieza el segundo), y por último el fin del difonema.

Durante el proceso de etiquetado, el archivo “.dat” se seguirá modificando a medida que se insertan las etiquetas, esto significa que además de las cuatro columnas existentes, habrá tres columnas más, una para cada una de las etiquetas utilizadas.

La interfaz de usuario de Diphone Studio es bastante intuitiva y fácil de utilizar. Tiene cuatro componentes básicos. El primero y el más visible es el espacio de trabajo donde se muestra la onda producida por la grabación, sobre la que se ubicarán las etiquetas. El segundo componente es la barra de menú por medio de la cual se accede a las opciones para recorrer el archivo “.dat” y realizar el etiquetado. El tercer componente es la barra inferior que hace las veces de barra de estado pues muestra en diferentes recuadros información relevante acerca de la grabación como el nombre de ésta, la palabra y el difonema que contiene y una vez se han insertado etiquetas muestra la duración de cada fonema del difonema. El cuarto y último componente son dos barras de desplazamiento, la barra vertical permite hacer zoom en la onda para facilitar el etiquetado, y la barra horizontal permite movilizarse a través de la onda. La figura 17 muestra estos componentes detalladamente.

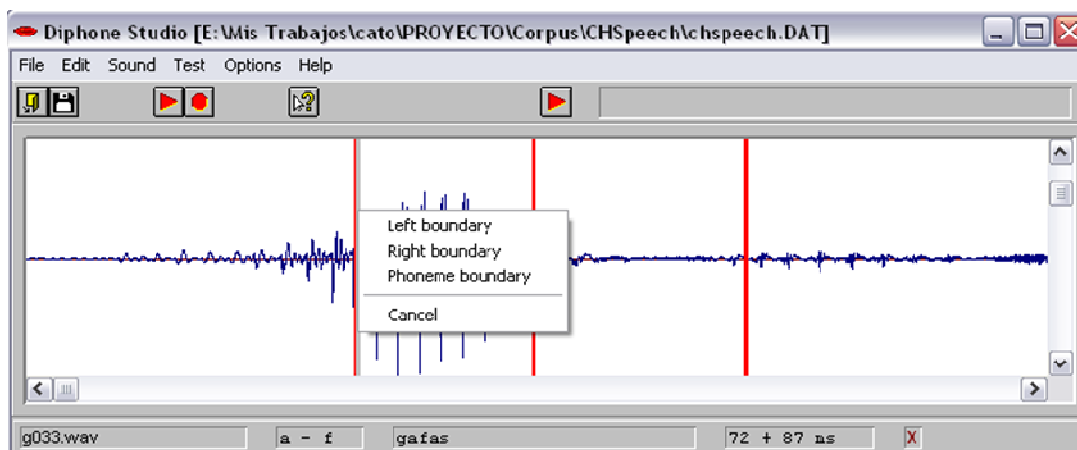


Figura 17. Interfaz de Usuario Diphone Studio

Fuente: Adaptación de los autores de la interfaz gráfica de Diphone Studio

Para comenzar con el etiquetado es necesario cargar el archivo “.dat” con la información de las grabaciones, palabras y difonemas, esto se hace a través del menú File – Open. Diphone Studio se ubica en el primer difonema del archivo, pero también es posible recorrer dicho archivo difonema por difonema a través del menú Edit - Next o Previous, o escoger directamente el difonema al cual se quiere ir mediante el menú Edit – Goto Diphone y tecleando el nombre. Cuando se ha ubicado el difonema que se va a trabajar, las etiquetas se insertan mediante un click y seleccionando la etiqueta que se quiere insertar en ese punto.

La Figura 18 muestra gráficamente el menú que es desplegado al hacer clic en una parte de la onda para escoger el tipo de etiqueta que se desea insertar.



*Figura 18. Inserción de etiquetas en Diphone Studio*

*Fuente: Diphone Studio*

Una vez se han insertado las tres etiquetas en un difonema, se hacen los mismos pasos con todos los demás hasta terminar el listado de difonemas del archivo “.dat”. Cuando esta labor se ha cumplido, se puede decir que se ha terminado el proceso de etiquetado. El archivo final contendrá en sus tres últimas columnas, el número de la “muestra” en que se ha insertado cada etiqueta. Un fragmento del contenido final del archivo chspeech.dat se muestra en la Figura 19.

a	l	g052.wav	caída	2155	3456	5392
a	o	g053.wav	taraón	4070	5400	7196
a	u	g054.wav	baúl	3088	4648	6414
b	-	g055.wav	club	3513	5379	6715
b	a	g056.wav	abanico	2493	3548	4359
b	b	g057.wav	obvio	2775	3412	3941
b	ch	g058.wav	chibcha	3625	4825	6982
b	d	g059.wav	abdomen	2647	4094	5696
b	e	g060.wav	abeja	4144	5181	6562
b	f	g061.wav	subfluvial	3461	6230	9366
b	g	g062.wav	subglacial	4578	6072	8380
b	i	g063.wav	avispa	2912	3789	4935
b	j	g064.wav	subgerente	3514	4365	6027
b	k	g065.wav	subclase	5075	6571	9108
b	l	g066.wav	ablandar	2315	3388	4345

Figura 19. Fragmento del contenido final del archivo chspeech.dat

Fuente: Autores del Proyecto

### 3.2.2.5 Extracción de los fragmentos de audio

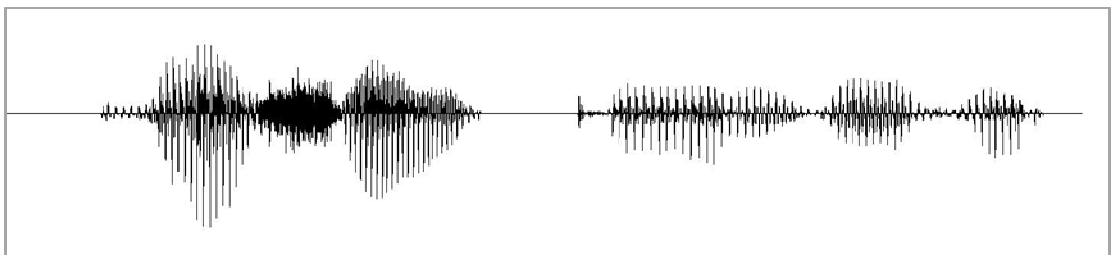
Teniendo terminado completamente el etiquetado del corpus, se continúa con la extracción de cada uno de los difonemas, desde cada uno de los archivos correspondientes. Para esto, se utilizó el archivo *chspeech.dat* que se obtuvo como resultado del etiquetado con el software Diphone Studio. De allí se extrajeron las columnas que indican las fronteras del difonema (inicio y fin), siendo éstas los números 5 y 7 respectivamente; con estas fronteras se procedió a extraer de cada archivo el difonema especificado.

El objetivo de ésta extracción, es construir la base de datos de difonemas correspondientes al idioma español de Colombia. Sin embargo, una vez se tuvieron las grabaciones de los difonemas por separado, se observó que algunas de ellas sonaban más bajo que otras debido a que en el momento de la grabación el locutor disminuyó la intensidad de la voz al pronunciar la palabra que contenía el difonema en cuestión; esta diferencia en las grabaciones se vería reflejada posteriormente en el sintetizador pues al concatenar los difonemas, las transición de uno a otro sería más notoria, afectando la inteligibilidad y la naturalidad del mismo. Por esta razón, se decidió aplicar un filtro de normalización a las grabaciones.

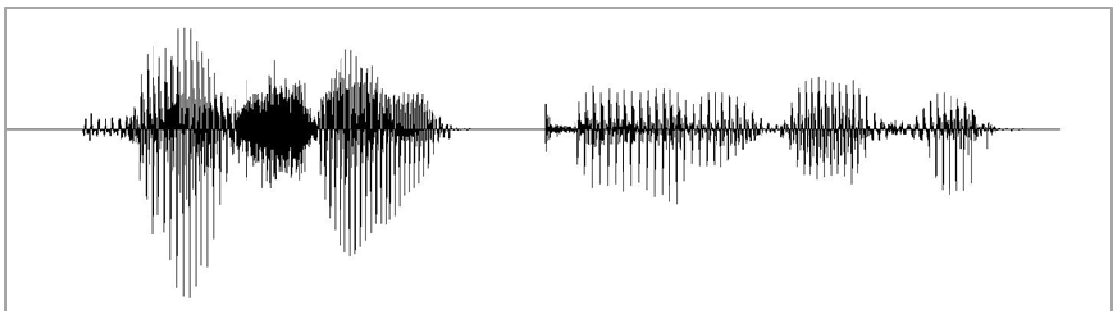
La normalización de la señal consiste en aumentar (o reducir) la amplitud de ésta a un punto deseado, exactamente se trata de llevar el pico de amplitud más alto de la señal justo por debajo del nivel de distorsión (antes de ruido); este proceso en general aumenta el nivel de sonido de la señal a un punto específico, aumentando en la misma proporción las partes de poco y mucho volumen (sin sobrepasar la barrera de la distorsión).

Para el presente proyecto se utilizó el proceso de normalización con el fin de obtener un mismo nivel de audio para todas las grabaciones realizadas, aumentando el nivel a las que presentaban menor calidad y manteniendo (o aumentando en menor proporción) las que en el momento de grabación obtuvieron mejor calidad.

La Figura 20 muestra la señal de audio original de la frase “Yo sé que él dijo” utilizada en el proceso de grabación del corpus antes de ser normalizada, y la Figura 21 presenta la señal de la misma frase después del proceso de normalización.



*Figura 20. Señal sin normalizar*  
*Fuente: Autores del Proyecto*



*Figura 21. Señal normalizada*  
*Fuente: Autores del Proyecto*

### 3.2.2.6. Base de datos de la aplicación

La herramienta CHSpeech cuenta con una base de datos embebida que se compone de dos elementos: el primero es una carpeta en la raíz del proyecto llamada “*difonemas*” que contiene todos los archivos .WAV (590 archivos) correspondientes a los difonemas; y el segundo es un archivo .properties (sección 3.1.1.5 Buscador) que contiene las referencias a los archivos de esta carpeta.

Un archivo .properties puede considerarse como una *tabla referencial* de una base de datos, puesto que contiene únicamente 2 campos, adaptándose a las necesidades del proyecto. Este tipo de archivo es útil cuando se trata de *tablas referenciales*, debido a que no soporta más de 2 campos, y la velocidad de búsqueda es relativamente igual a la de una base de datos, comparando tiempo de conexión y búsqueda de una base de datos, sólo con tiempo de búsqueda en un .properties.

El funcionamiento de la base de datos se explica a continuación:

- El archivo *grabaciones.properties* contiene 590 filas (una por cada difonema), cada una compuesta por dos cadenas llamadas “*clave=valor*”.
- La cadena “*clave*” del archivo *grabaciones.properties* hace referencia al nombre del difonema y la cadena “*valor*” hace referencia al nombre del archivo de audio que le contiene.
- La búsqueda de los archivos en la base de datos se realiza por medio del nombre del difonema solicitado.
- Al ser solicitado un difonema al archivo *grabaciones.properties*, éste realiza la búsqueda del mismo, asignando el nombre del difonema al campo “*clave*” y buscando el valor correspondiente.
- Al ser encontrado el difonema, el archivo *grabaciones.properties* retorna la cadena correspondiente al “*valor*”, que para el presente proyecto, corresponde al nombre del archivo de audio que contiene al difonema.

### **3.3 INTERFAZ GRÁFICA DE USUARIO (GUI)**

La interfaz gráfica de usuario (GUI) de CHSpeech permite que los usuarios accedan a las funcionalidades de la herramienta de una manera sencilla. El desarrollo de este componente del proyecto se realizó en el lenguaje de programación Java Standard Edition, utilizando el IDE Netbeans 6.5.1 y el Swing Application Framework<sup>16</sup>.

La GUI de CHSpeech está compuesta por tres pantallas principales: un panel de autenticación que permite el ingreso de los usuarios, un panel de registro en el que nuevos usuarios pueden crear sus cuentas para acceder a la herramienta, y un panel principal en el que se desarrolla la actividad esencial de CHSpeech. Además existen paneles de recordatorio de contraseñas, modificación de datos y ayuda.

A continuación se presenta en detalle cada uno de estos componentes.

El manual de usuario de CHSpeech explica detalladamente los procedimientos para acceder a las funciones de cada uno de los componentes de la GUI. Esta información puede verse en el ANEXO J.

#### **3.3.1 Panel de autenticación**

El panel de autenticación cumple tres funciones fundamentales:

1. Permitir el ingreso de usuarios con su nombre de usuario y contraseña
2. Permitir el registro de nuevos usuarios
3. Permitir recordatorio de contraseñas

Esta pantalla fue diseñada con el fin de contribuir con la seguridad de la aplicación, y fomentar la confianza de los usuarios en la herramienta. La utilización de la síntesis de voz en la telefonía móvil, no está exenta de casos de suplantación, más aún cuando se ofrece una única voz, como es el caso del presente proyecto. La idea de utilizar sesiones de usuario disminuye la posibilidad de que se presenten este tipo de casos, pensando en futuras aplicaciones, donde se puedan establecer parámetros para definir características de voz similares a las del usuario.

---

<sup>16</sup> Es un conjunto pequeño de clases que simplifica el desarrollo de aplicaciones de escritorio. Mayor información en <https://appframework.dev.java.net/>

La Figura 22 presenta el panel de autenticación CHSpeech.

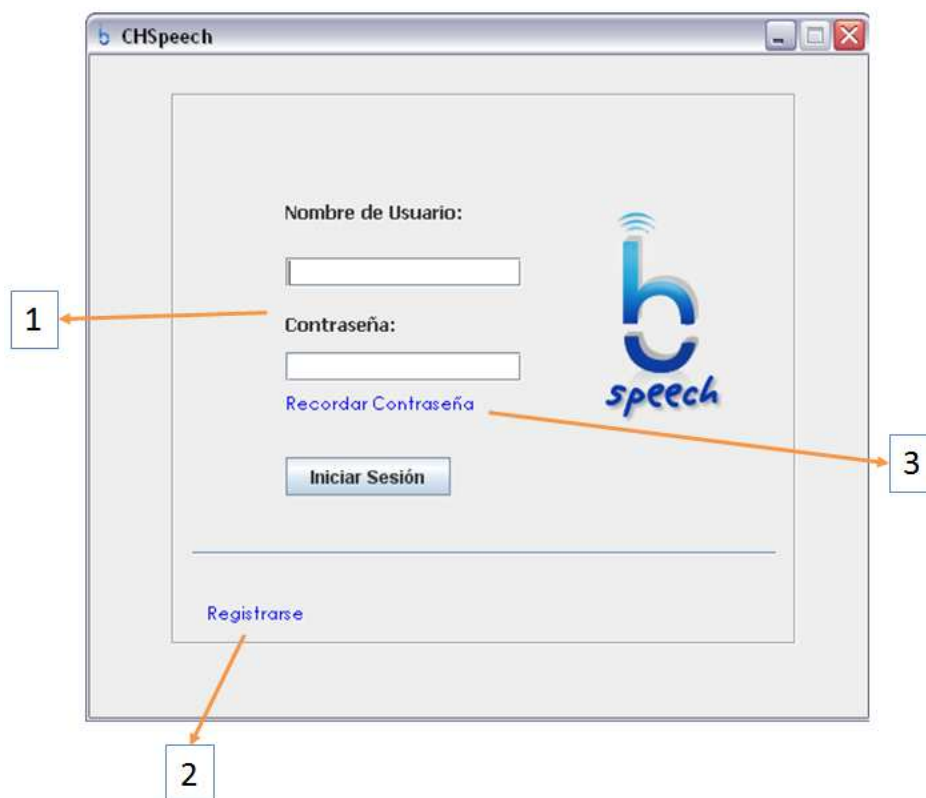


Figura 22. Panel de Autenticación CHSpeech

Fuente: Autores del Proyecto

### 3.3.2 Panel de registro de nuevos usuarios

La función que cumple el panel de registro es permitir que nuevos usuarios creen sus cuentas para acceder a las funcionalidades de CHSpeech. La herramienta maneja una base de datos local embebida que mantiene los datos de los usuarios.

La Figura 23 muestra el formulario de registro que debe diligenciarse.

CHSpeech

**REGISTRO DE USUARIOS**

Nombre de Usuario:

Contraseña: (Mínimo 6 caracteres)

Repetir Contraseña:

Correo Electrónico:

El correo electrónico será usado únicamente en caso de olvido de contraseñas.

*Figura 23. Panel de Registro de Nuevos Usuarios*

*Fuente: Autores del Proyecto*

### **3.3.3 Panel Principal CHSpeech**

El panel principal de CHSpeech es donde realmente se accede a la funcionalidad de síntesis de voz para la que fue diseñada la herramienta. Esta pantalla se muestra al usuario una vez ha ingresado con un nombre de usuario y contraseña válidos.

Tres secciones componen el panel:

1. Barra de menús: Ofrece opciones de configuración y ayuda.
2. Panel de frases predefinidas: permite un acceso rápido a frases que comúnmente se utilizan en conversaciones telefónicas evitando así la escritura de éstas, pues a través de un doble clic la frase seleccionada se muestra como si el usuario la hubiese digitado; además permite que cada usuario adicione o elimine sus propias frases mediante el uso de los respectivos botones ubicados en la parte inferior.
3. Área de trabajo: Este último componente alberga dos grandes áreas de texto, la primera es el historial que muestra las frases sintetizadas anteriormente en la sesión actual, y la segunda es donde el usuario ingresa el texto que se va a procesar.

En la Figura 24 se muestra el panel principal de CHSpeech.

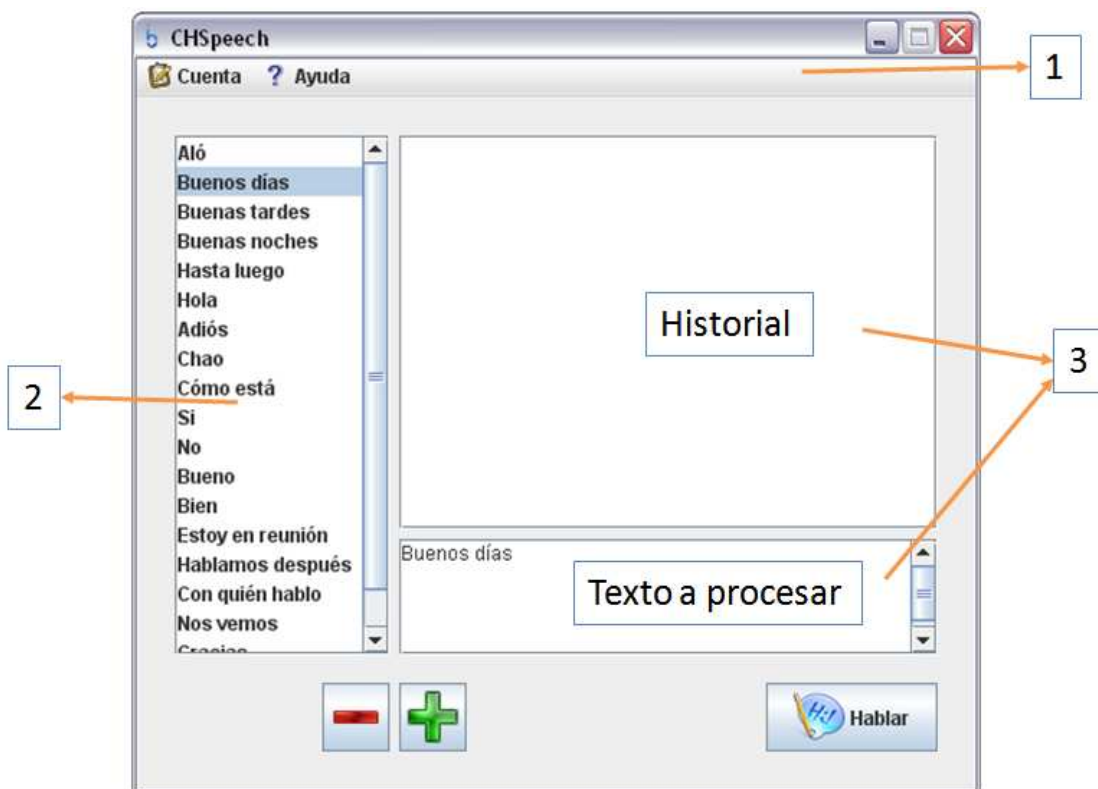


Figura 24. Panel Principal CHSpeech y sus componentes

Fuente: Autores del Proyecto

En el menú Cuenta del panel principal, es posible acceder a opciones de modificación de datos como contraseña o E-mail; además de eliminar su cuenta, si así lo desea, o cerrar sesión. La Figura 25 presenta las opciones de este menú.

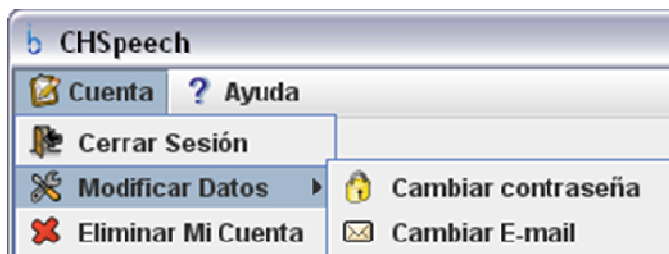
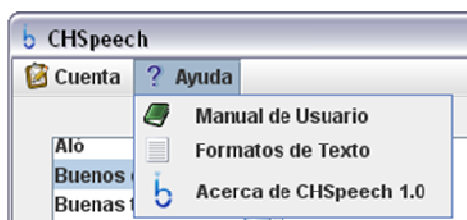


Figura 25. Menú cuenta

Fuente: Autores del Proyecto

Además del menú Cuenta, se encuentra el menú Ayuda donde el usuario puede encontrar información acerca de la herramienta, acceder al manual de usuario o visualizar los formatos predefinidos para CHSpeech que se presentan en el ANEXO A. La Figura 26 muestra el contenido del menú Ayuda.



*Figura 26. Menú Ayuda*  
*Fuente: Autores del Proyecto*

### 3.3.4 Recordatorio de contraseñas

Cuando se manejan sesiones de usuarios es muy probable que alguno de ellos olvide su contraseña de ingreso. Por esta razón, es necesario implementar un mecanismo de recordatorio de contraseñas.

Desde el panel de autenticación de CHSpeech es posible acceder al formulario para recordar contraseñas. La Figura 27 muestra esta pantalla.




*Figura 27. Recordatorio de contraseñas*  
*Fuente: Autores del Proyecto*

Para recordar una contraseña, CHSpeech utiliza el correo electrónico del usuario que está almacenado en la base de datos; la contraseña actual de dicho usuario es enviada a este E-mail, recomendándole que la cambie al ingresar nuevamente a la herramienta. De esta manera, se evita que una persona diferente al dueño de la cuenta pueda acceder a la contraseña de la misma, puesto que para conocerla debe saber también la contraseña del E-mail asociado al usuario.

### 3.3.5 Modificación de E-mail y Contraseña

En la descripción del panel principal se mencionó que el menú Cuenta presenta la opción de modificar datos como E-mail y contraseña de un usuario. Las figuras 28 y 29 presentan el formulario para modificar el correo electrónico y contraseña respectivamente.

El cambio de correo electrónico requiere que el usuario ingrese la contraseña de su cuenta CHSpeech; esto evita que una persona diferente al dueño de la cuenta modifique los datos registrados en la base de datos.



The image shows a screenshot of a web application window titled "CHSpeech". The main content area is titled "Cambio de E-mail". It contains the following fields and elements:

- Nombre de Usuario:** A text input field containing the value "cato0385".
- Contraseña Actual:** A password input field, currently empty.
- Nuevo E-mail:** A text input field, currently empty.
- A warning message: "Asegúrese de ingresar una cuenta de correo válida."
- Two buttons at the bottom: "Guardar Cambios" and "Cancelar".

*Figura 28. Formulario Cambio E-mail*

*Fuente: Autores del Proyecto*

CHSpeech

**Cambio de Contraseña**

Nombre de Usuario:  
cabo0386

Contraseña Actual:

Nueva Contraseña:

Repetir Nueva Contraseña:

*Figura 29. Formulario Cambio de Contraseña*

*Fuente: Autores del Proyecto*

Al igual que la modificación del correo electrónico, el cambio de contraseña requiere que el usuario ingrese su contraseña actual, y su nueva contraseña con su repetición.

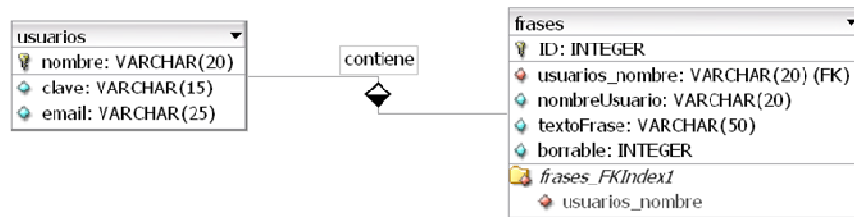
### **3.3.6 Base de datos para el manejo de usuarios y frases predefinidas**

El manejo de los usuarios y las frases predefinidas en CHSpeech se realiza mediante una pequeña base de datos embebida en la aplicación. Para este fin se utilizó Apache Derby<sup>17</sup> como motor de base de datos puesto que es open source, está implementado completamente en Java, y permite su uso en aplicaciones desarrolladas con esta tecnología.

La base de datos diseñada para este fin consta de dos tablas, una de ellas registra los datos pertenecientes a los usuarios como nombre, contraseña y correo electrónico. La segunda tabla contiene las frases predefinidas agregadas por los usuarios. De esta manera un usuario sólo puede ver y eliminar las frases que él mismo ha agregado y no las de los demás usuarios en el mismo equipo. El diseño de la base de datos implementada puede verse en la Figura 30.

---

<sup>17</sup> Más información acerca de Apache Derby en <http://db.apache.org/derby/>



*Figura 30. Diseño de la base de datos para el manejo de usuarios y frases predefinidas*  
*Fuente: Autores del Proyecto*

### 3.4 MECANISMO DE TRANSMISIÓN

El presente proyecto está orientado a la utilización de síntesis de voz como soporte en la comunicación vía celular. Esto implica la implementación de un mecanismo de transmisión que permita que la señal auditiva producida por la herramienta de síntesis sea usada en una llamada realizada a través de este medio.

Existen varias maneras de transmitir señales (audio, imagen, video, etc.) de un computador a un celular. Básicamente se distinguen dos categorías: medios cableados y medios inalámbricos. Entre los medios cableados se distingue el uso del cable de datos suministrado con el dispositivo móvil. Por otra parte, entre los medios inalámbricos se destaca el uso del infrarrojo, y el conocido Bluetooth.

La utilización de un medio u otro depende principalmente de las características de los equipos involucrados (celular-computador o celular-celular), y además del objetivo que se quiere lograr al realizar una transmisión. En este caso los aspectos que se deben tener en cuenta son, en primera instancia que la transmisión de la señal se realizará desde un equipo de cómputo hacia un teléfono celular, y además, que el objetivo que se persigue es que la señal de audio transmitida se reproduzca durante una llamada en el celular.

El objetivo implica acceder al micrófono del dispositivo móvil, esta condición descarta de entrada el uso del cable de datos (USB) suministrado con el dispositivo. De otro lado se encuentran los medios inalámbricos mencionados anteriormente (infrarrojo y Bluetooth); el uso del infrarrojo como medio de transmisión en este proyecto se descarta debido a que

exige que los dispositivos se encuentren alineados para que la transmisión se realice con éxito, además la mayoría de computadores no tienen incorporada esta tecnología, lo que implicaría un costo adicional.

Con la tecnología Bluetooth ocurre algo similar. Muchos de los dispositivos celulares que existen en el mercado actualmente, sobre todo los de gama baja, no cuentan con soporte para esta tecnología, al igual que varias referencias de equipos de cómputo. El uso de Bluetooth como medio de transmisión en el presente proyecto, limitaría el rango de usuarios potenciales de la herramienta. Por esta razón, se descarta esta tecnología como medio de transmisión.

En vista de que las tecnologías existentes no satisfacen las condiciones y objetivos necesarios para su implementación en este proyecto, se optó por diseñar un mecanismo diferente para la transmisión de las señales de audio del computador al dispositivo celular. Para esto se tomó como elemento base el cable manos libres. Este cable es suministrado con la mayoría de celulares, y tiene acceso tanto al micrófono como al auricular del dispositivo. Estas características se convierten en puntos a favor para su utilización.

Sin embargo, el cable manos libres como tal, con sus características de fábrica no permite conseguir el objetivo propuesto, por lo tanto fue necesario realizar modificaciones. Un cable manos libres normalmente se compone de un audífono que es conectado al auricular del celular, un micrófono que se conecta al micrófono del dispositivo, y naturalmente, dos cables que por comodidad, casi siempre van unidos, para conducir las señales entre los dispositivos. La Figura 31 muestra un cable manos libres común, en su estado original.

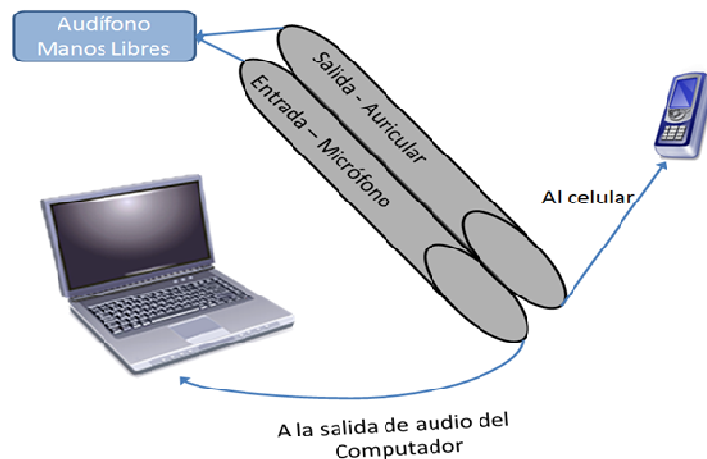


*Figura 31. Manos libres original*

*Fuente: Autores del Proyecto*

El proceso de modificación implementado consiste en tomar el cable del micrófono del manos libres y conectarlo a la salida de audio del computador, de manera que al producirse la voz sintética, sea transportada hacia el micrófono del celular.

La Figura 32 muestra un esquema general del procedimiento de modificación.



*Figura 32. Proceso de modificación del cable manos libres*

*Fuente: Autores del Proyecto*

Finalizado este proceso, se obtuvo el cable manos libres modificado que permite transmitir las señales de audio producidas por el sintetizador al dispositivo celular, en medio de una llamada. La Figura 33 presenta el resultado obtenido con el cable manos libres real.



*Figura 33. Cable manos libres modificado*

*Fuente: Autores del Proyecto*

Al conectar directamente la salida de audio del computador a la entrada del micrófono del celular, se transmiten todas las señales de audio generadas por el computador; es decir, que al utilizar este mecanismo de transmisión se reproducirán, además de la voz sintética en medio de una llamada, otros sonidos que esté reproduciendo el computador en ese momento como: música, o sonidos provenientes de otros programas en ejecución.



*Figura 34. Conexión del cable manos libres al celular y al equipo de cómputo*  
*Fuente: Autores del Proyecto*

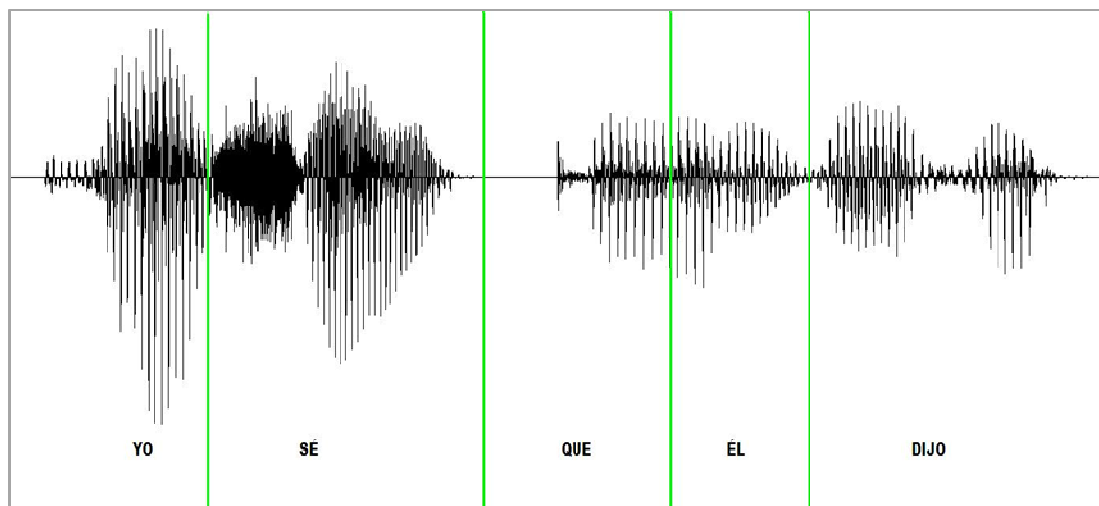
## 4 PRUEBAS Y RESULTADOS

### 4.1 COMPARACIÓN DE SEÑALES

El problema de la naturalidad es muy común en la mayoría de los sintetizadores de voz existentes actualmente. Muchos estudios se enfocan en encontrar una solución a este problema pues los métodos actuales no brindan los elementos necesarios para lograr voces de excelente calidad que puedan confundirse con una voz real.

A continuación se presenta la comparación de la representación gráfica de dos señales: la primera es la voz del locutor pronunciando una frase durante la realización de las grabaciones para la creación del corpus de voz y la segunda es la señal obtenida como salida del sintetizador de voz CHSpeech. Ambas señales corresponden a la frase “Yo sé que él dijo”.

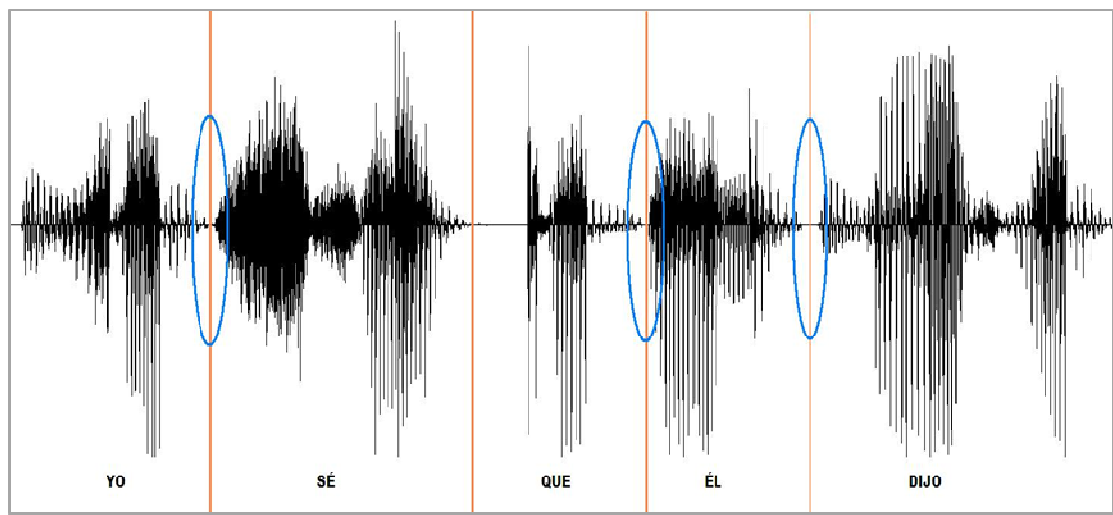
La Figura 35 hace referencia a la señal de audio correspondiente a la voz real, grabada para la creación del corpus de voz, por lo tanto se considera como una señal completamente natural. Esta se encuentra dividida por franjas verticales verdes que ayudan a la identificación de inicio y final de una palabra (transición entre palabras).



*Figura 35. Señal de audio original*

*Fuente: Autores del Proyecto*

La Figura 36 muestra la señal de audio correspondiente a la salida del sintetizador CHSpeech para la misma frase. Esta señal se encuentra al igual que la anterior, separada por franjas, esta vez naranjas que marcan la separación entre palabras. Además, incluye 3 elipses azules que ubican los lugares donde se hace notorio el problema de naturalidad en la voz sintética.



*Figura 36. Señal de audio obtenida del sintetizador*

*Fuente: Autores del Proyecto*

Al analizar las figuras 35 y 36 es posible notar que no coinciden perfectamente, pero la forma de la señal obtenida por el sintetizador es muy similar a la señal original. La señal de la voz sintética presenta picos mayores que la señal original debido al proceso de normalización aplicado.

Se resaltan con las elipses azules, los espacios en blanco o caídas de señal que indican la coarticulación entre una palabra y otra; éstas marcas identifican el problema de la naturalidad (tonalidad robótica) presente en la voz resultante de la herramienta software CHSpeech.

## 4.2 PRUEBAS DE FUNCIONAMIENTO

### 4.2.1 Prueba al sintetizador

<b>Objetivo</b>	Comprobar que la salida proporcionada por el sintetizador sea correcta respecto al texto ingresado, teniendo en cuenta los formatos establecidos.
<b>Actividad a realizar</b>	Introducir diferentes textos en español que contengan tanto palabras como números y elementos contemplados en los formatos establecidos (fechas, horas, páginas web, teléfonos, etc.).

Se tomaron 5 ejemplos para cada uno de los formatos definidos previamente (16 formatos), obteniendo un total de 80 frases de prueba. El resultado obtenido en esta prueba es satisfactorio debido a que el sintetizador pronunció correctamente las 80 frases introducidas, procesando de manera correcta cada uno de los formatos establecidos.

La Tabla 4 presenta los textos de prueba introducidos en la herramienta y la salida proporcionada por ésta.

	<b>ENTRADA</b>	<b>SALIDA</b>
<b>PALABRAS</b>	hOLa BuEnOs DiaZ	Hola buenos días
	Dr. Correa, la Sra. Ana María no se encuentra en el momento	Doctor Correa, la señora Ana María no se encuentra en el momento
	Vamos a comer lasagna al restaurante de la esq. Bolívar.	Vamos a comer lasaña al restaurante de la esquina bolívar.
	El shampoo anti-caspa es bueno para toda la flia., contenido mín. 35 ml.	El champú anti caspa es bueno para toda la familia, contenido mínimo treinta y cinco mililitros.
	Como instalar el paquete JRE	Como instalar el paquete jota ere e

	para ejecutar aplicaciones en Java	para ejecutar aplicaciones en java.
<b>FECHAS</b>	<b>ENTRADA</b>	<b>SALIDA</b>
	Mi fecha de nacimiento es 18/09/1986	Mi fecha de nacimiento es dieciocho de septiembre de mil novecientos ochenta y seis
	El 20/07/1810 se declara la independencia de Colombia	El veinte de julio de mil ochocientos diez se declara la independencia de Colombia.
	Último día de entrega de proyectos en la escuela es el 23/10/08	Último día de entrega de proyectos en la escuela es el veintitrés de octubre del cero ocho
	Bucaramanga 02/01/2008, empieza el proceso de búsqueda	Bucaramanga dos de enero de dos mil ocho, empieza el proceso de búsqueda
	Albert Einstein, nació en ULM el 14/03/1879 y murió en Princeton el 18/04/1955	Albert Einstein, nació el catorce de marzo de mil ochocientos setenta y nueve y murió en Princeton el dieciocho de abril de mil novecientos cincuenta y cinco.
<b>HORAS</b>	<b>ENTRADA</b>	<b>SALIDA</b>
	En este momento son las 3:18 p.m.	En este momento son las tres y dieciocho pe eme.
	Juan Pablo Montoya compite hoy en la NASCAR a las 02:15.	Juan Pablo Montoya compite hoy en la NASCAR a las dos y quince.
	Las clases este semestre	Las clases este semestre comienzan

	comienzan a las 6:00 y terminan a las 10:00:00	a las seis en punto y terminan a las diez en punto
	El tiempo de Bolt en los olímpicos fue de 00:11:01 en la maratón individual	El tiempo de Bolt en los olímpicos fue de cero y once con un segundo en la maratón individual
	Nos vemos en la cafetería a las 0:20 o no, mejor nos vemos en biblioteca a las 22:00:20	Nos vemos en la cafetería a las cero y 20 o no, mejor nos vemos en biblioteca a las veintidós con veinte segundos.
<b>TELEFONOS FIJOS</b>	<b>ENTRADA</b>	<b>SALIDA</b>
	El teléfono del grupo de investigación es 6-344000 ext. 2676 o 2476	El teléfono de la universidad es seis treinta y cuatro cuarenta cero cero extensión dos mil seiscientos setenta y seis o dos mil cuatrocientos setenta y seis.
	Llame al teléfono 6-123456 y gánese fabulosos premios	Llame al teléfono seis doce treinta y cuatro cincuenta y seis y gánese fabulosos premios
	Si vas a llamar a Bucaramanga marca el siguiente numero: (097)6-310109	Si vas a llamar a Bucaramanga marca el siguiente numero: cero noventa y siete seis treinta y uno cero uno cero nueve
	¿El último teléfono del directorio corresponde al número 9-999999?	El último teléfono del directorio corresponde al número nueve noventa y nueve noventa y nueve noventa y nueve.
	Si quiere comunicarse con	Si quiere comunicarse con riesgos

	riesgos profesionales marque al (051)9-251463 en Bogotá.	profesionales marque al cero cincuenta y uno nueve veinticinco catorce sesenta y tres en Bogotá.
<b>TELEFONOS CELULARES</b>	<b>ENTRADA</b>	<b>SALIDA</b>
	Diseño gráfico creArte, comuníquese con nosotros al: 300-5012359.	Diseño grafico create comuníquese con nosotros al: trescientos quinientos uno veintitrés cincuenta y nueve.
	Movistar la marca líder de telefonía móvil en Colombia, llama ahora al 315-0154196	Movistar la marca líder de telefonía móvil en Colombia, llama ahora al trescientos quince cero uno cinco cuarenta y uno noventa y seis
	El teléfono avantel de mi papá es 350-1122301	El teléfono avantel de mi papá es trescientos cincuenta ciento doce veintitrés cero uno
	Claudia tiene la tarea llámela al 311-2551285 y pídasela prestada	Claudia tiene la tarea llámela al trescientos once doscientos cincuenta y cinco doce ochenta y cinco y pídasela prestada
	El rango de números telefónicos de tigo varia entre 300-0000000 y 301-9999999	El rango de números telefónicos de tigo varia entre trescientos cero cero cero cero cero y trescientos uno novecientos noventa y nueve noventa y nueve noventa y nueve
<b>NUMEROS ROMANOS</b>	<b>ENTRADA</b>	<b>SALIDA</b>
	El príncipe Juan Carlos nació en	El príncipe Juan Carlos nació en el

	el s. XVI	siglo dieciséis.
	La biblia en el capitulo V versículo XXIII dice que dormir es pecado.	La biblia en el capitulo cinco versículo veintitrés dice que dormir es pecado
	El presente año romano es el MMIX	El presente año romano es el dos mil nueve
	En la sección III del libro pág. CVI se encuentra publicado un buen artículo.	En la sección tres del libro página ciento seis se encuentra publicado un buen artículo.
	El programa humorístico el chavo del ocho fue creado en MCMLXXXVI	El programa humorístico el chavo del ocho fue creado en mil novecientos ochenta y seis
	<b>ENTRADA</b>	<b>SALIDA</b>
<b>DINERO</b>	Una computadora portátil cuesta entre \$1500000 y \$3000000	Una computadora portátil cuesta entre un millón quinientos mil pesos y tres millones de pesos.
	La tasa representativa del mercado se ubicó en \$1889.85	La tasa representativa del mercado se ubicó en mil ochocientos ochenta y nueve pesos con ochenta y cinco centavos
	Debido a la crisis económica mundial, el activo de la empresa se ubicó en -\$101548000	Debido a la crisis económica mundial, el activo de la empresa se ubicó en menos ciento un millones quinientos cuarenta y ocho mil
	Las ganancias que reportó el sector financiero ascendieron a	Las ganancias que reportó el sector financiero ascendieron a los

	los \$951632547823316.20	novecientos cincuenta y un billones seiscientos treinta y dos mil quinientos cuarenta y siete millones ochocientos veintitrés mil trescientos dieciséis pesos con veinte centavos
	Me regalaron \$500000 para la matricula de la universidad	Me regalaron quinientos mil pesos para la matricula de la universidad
<b>DIRECCIONES</b>	<b>ENTRADA</b>	<b>SALIDA</b>
	Diag. 15 Cl. 32 bl. G apto. 301	Diagonal quince calle treinta y dos bloque je apartamento trescientos uno
	El viejo chiflas está ubicado en la kra. 58 # 78 - 44 Medellín	El viejo chiflas está ubicado en la carrera cincuenta y ocho numero setenta y ocho guion cuarenta y cuatro Medellín
	Av. El dorado # 34 A - 25 Edificio 13 apto. 1001 Bogotá	Avenida El dorado numero treinta y cuatro a guion veinticinco edificio trece apartamento mil uno Bogotá.
	Hubo un accidente en la Cra. 34 # 5C - 37 informo la policía	Hubo un accidente en la carrera treinta y cuatro número cinco ce guión treinta y siete informó la policía
	Cra. 53 Clle. 98 Esq. Centro comercial Buenavista	Carrera cincuenta y tres calle noventa y ocho esquina centro comercial Buenavista

	<b>ENTRADA</b>	<b>SALIDA</b>
<b>PAGINAS WEB</b>	<a href="http://cormoran.uis.edu.co/eisi/eisi.jsp?IdServicio=S25">http://cormoran.uis.edu.co/eisi/eisi.jsp?IdServicio=S25</a>	Hache te te pe dos puntos barra barra ce o ere eme o ere a ene punto uis punto edu punto co barra eisi barra eisi punto jota ese pe cierra interrogación idservicio igual ese dos cinco
	<a href="http://www.rae.es/rae.html">http://www.rae.es/rae.html</a>	Hache te te pe dos puntos barra barra doble u doble u doble u punto rae punto es barra rae punto hache te eme ele
	<a href="http://java.sun.com/javase/technologies/desktop/media/jmf/2.1.1/download.html">http://java.sun.com/javase/technologies/desktop/media/jmf/2.1.1/download.html</a>	Hache te te pe dos puntos barra barra jota a ve a punto sun punto com barra javase barra te e ce hache ene o ele o ge i e ese barra desktop barra media barra jota eme efe barra dos punto uno punto uno barra download punto hache te eme ele
	<a href="https://www.uis.edu.co/webUIS/es/index.jsp">https://www.uis.edu.co/webUIS/es/index.jsp</a>	Hache te te pe ese dos puntos barra barra doble u doble u doble u punto uis punto edu punto co barra ueb uis barra es barra index punto jota ese pe
	<a href="http://www.banrep.gov.co">www.banrep.gov.co</a>	Doble u doble u doble u punto banrep punto gov punto co

	<b>ENTRADA</b>	<b>SALIDA</b>
<b>CORREOS ELECTRÓNICOS</b>	claudia.correa@correo.uis.edu.co	Claudia punto correa arroba correo punto uis punto edu punto co
	hfnacho18@hotmail.com	Hache efe ene a ce hache o uno ocho arroba jotmeil punto com
	correo_prueba@yahoo.com	Correo guión de piso prueba arroba yaju punto com
	admincorreo@correo.uis.edu.co	Admin correo arroba correo punto uis punto edu punto co
	chspeech.uis@gmail.com	Ce hache ese pe e e ce hache punto uis arroba yimeil punto com
	<b>ENTRADA</b>	<b>SALIDA</b>
<b>NUMEROS (En general)</b>	La colonia de hormigas tenía un total de 485867473834 individuos	La colonia de hormigas tenía un total de cuatrocientos ochenta y cinco mil ochocientos sesenta y siete millones cuatrocientos setenta y tres mil ochocientos treinta y cuatro individuos
	3500000 personas vieron el programa	Tres millones quinientos mil personas vieron el programa
	-1436574832195837	Menos mil cuatrocientos treinta y seis billones quinientos setenta y cuatro mil ochocientos treinta y dos millones ciento noventa y cinco mil ochocientos treinta y siete

	El trabajo tiene 49 páginas	El trabajo tiene cuarenta y nueve páginas
	847362759483716473	Ochocientos cuarenta y siete mil trescientos sesenta y dos billones setecientos cincuenta y nueve mil cuatrocientos ochenta y tres millones setecientos dieciséis mil cuatrocientos setenta y tres
<b>NUMEROS FRACCIONARIOS</b>	<b>ENTRADA</b>	<b>SALIDA</b>
	43/67	Cuarenta y tres sesenta y sieteavos
	-3/4	Menos tres cuartos
	7/9	Siete novenos
	1024/2320	Mil veinticuatro dos mil trescientos veinteavos
	95/136	Noventa y cinco ciento treinta y seisavos
<b>NUMEROS ORDINALES</b>	<b>ENTRADA</b>	<b>SALIDA</b>
	136 <sup>a</sup>	Centésima trigésima sexta
	54 <sup>o</sup>	Quincuagésimo cuarto
	409 <sup>no</sup>	Cuadrigentésimo noveno
	672 <sup>da</sup>	Sexcentésima septuagésima segunda
	5 <sup>to</sup>	Quinto

	<b>ENTRADA</b>	<b>SALIDA</b>
<b>NUMEROS DECIMALES</b>	37.543326171	Treinta y siete punto cincuenta y cuatro treinta y tres veintiséis diecisiete uno
	+654.98674	Más seiscientos cincuenta y cuatro punto noventa y ocho sesenta y siete cuatro
	-132543.4684	Menos ciento treinta y dos mil quinientos cuarenta y tres punto cuarenta y seis ochenta y cuatro
	-987.764e+137	Menos novecientos ochenta y siete punto setenta y seis cuatro e a la más ciento treinta y siete
	4.87E-9	Cuatro punto ochenta y siete e a la menos nueve
	<b>ENTRADA</b>	<b>SALIDA</b>
<b>TEMPERATURAS</b>	El aire acondicionado está a 20 °C	El aire acondicionado está a veinte grados centígrados
	Pre caliente el horno a 180 °F antes de hacer la mezcla	Pre caliente el horno a ciento ochenta grados Fahrenheit antes de hacer la mezcla
	La temperatura máxima de la ciudad en el día de hoy será de 35 °C	La temperatura máxima de la ciudad en el día de hoy será de treinta y cinco grados centígrados
	El punto de ebullición del agua	El punto de ebullición del agua es

	es 100 °C o 212 °F	cien grados centígrados o doscientos doce grados Fahrenheit
	El niño tuvo fiebre de 39.3 °C	El niño tuvo fiebre de treinta y nueve punto tres grados centígrados
<b>SIMBOLOS</b>	<b>ENTRADA</b>	<b>SALIDA</b>
	;	Punto y coma
	~	Virgulilla
	>	Mayor que
	%	Por ciento
	–	Guión de piso

Tabla 4. Resultados Prueba de funcionamiento del sintetizador

#### 4.2.2 Prueba de ejecución en diferentes plataformas y medición de tiempos de respuesta

<b>Objetivo</b>	<p>Verificar el desempeño de la herramienta en diferentes sistemas operativos y arquitecturas.</p> <p>Realizar una medición de los tiempos que tarda el sintetizador en proporcionar una salida a un texto ingresado. Es decir, el tiempo de respuesta comprende el lapso entre el envío de la orden de hablar al sintetizador y el comienzo de la reproducción del sonido.</p>
<b>Actividades a realizar</b>	<p>Ejecutar la herramienta en diferentes sistemas operativos, verificando que lo haga correctamente.</p> <p>Ejecutar la herramienta con textos de diferentes longitudes y medir los tiempos de respuesta para cada uno de ellos teniendo en cuenta las condiciones anteriores para evaluar la influencia de las características de software en el desempeño de la aplicación.</p>

En esta prueba se ejecutó la herramienta en tres equipos diferentes. Las características de éstos se presentan en la Tabla 5.

<b>EQUIPO 1</b>	<b>SOFTWARE</b>	
	Sistema Operativo	Microsoft Windows Vista Home Basic Service Pack 2
	JRE	JRE 6 Update 15
	<b>HARDWARE</b>	
	Memoria RAM	2 GB
	Procesador	AMD Athlon 64 X2 Dual Core Processor 5600+ 2.9 GHz
<b>EQUIPO 2</b>	<b>SOFTWARE</b>	
	Sistema Operativo	Fedora 8
	JRE	JRE 6 Update 16
	<b>HARDWARE</b>	
	Memoria RAM	2 GB
	Procesador	AMD Athlon 64 X2 Dual Core Processor 5600+ 2.9 GHz
<b>EQUIPO 3</b>	<b>SOFTWARE</b>	
	Sistema Operativo	Mac OS X 10.5.8
	JRE	Java for Mac OS X 10.5 Update 5 (Versión 1.6.0_15)
	<b>HARDWARE</b>	
	Memoria RAM	2.5 GB
	Procesador	Intel Core 2 Duo 2.16 GHz

*Tabla 5. Características de los equipos utilizados*

En cada uno de los equipos se probó el sintetizador con conjuntos de diez frases de 5, 10, 15, 20, 25, 30, 35, 40, 45 y 50 palabras cada uno. Las Tablas 6, 7 y 8 presentan los resultados obtenidos para los equipos 1, 2 y 3, respectivamente.

	Número de palabras									
	5	10	15	20	25	30	35	40	45	50
Tiempos de respuesta (ms)	287	460	513	536	633	540	618	767	769	701
	255	453	504	538	584	512	681	642	681	694
	221	463	529	361	581	650	578	627	721	710
	401	432	472	484	496	451	702	709	710	786
	235	383	469	535	539	629	670	711	588	653
	372	627	460	608	463	665	703	632	709	591
	222	459	470	459	408	546	616	706	712	781
	362	533	460	401	578	709	653	557	696	804
	273	408	379	471	537	439	591	593	705	690
	282	476	570	440	484	605	633	703	660	657
	<b>Prom.</b>	<b>291</b>	<b>469,4</b>	<b>482,6</b>	<b>483,3</b>	<b>530,3</b>	<b>574,6</b>	<b>644,5</b>	<b>664,7</b>	<b>695,1</b>

Tabla 6. Tiempos de respuesta Equipo 1 (Windows Vista)

	Número de palabras									
	5	10	15	20	25	30	35	40	45	50
Tiempos de respuesta (ms)	267	536	604	731	819	740	864	653	725	733
	292	590	528	678	772	741	983	725	693	647
	390	526	555	689	844	797	867	729	798	765
	223	714	524	785	815	739	726	659	790	752
	337	605	677	783	778	861	704	779	697	820
	531	690	805	871	834	883	840	798	865	818
	297	733	699	655	683	879	735	855	877	816
	425	600	705	823	818	850	758	730	891	793
	270	530	693	814	819	793	907	739	842	725
	325	636	668	736	903	827	789	667	876	778
	<b>Prom.</b>	<b>335,7</b>	<b>616</b>	<b>645,8</b>	<b>756,5</b>	<b>808,5</b>	<b>811</b>	<b>817,3</b>	<b>738,4</b>	<b>805,4</b>

Tabla 7. Tiempos de respuesta Equipo 2 (Fedora 8)

	Número de palabras									
	5	10	15	20	25	30	35	40	45	50
Tiempos de respuesta (ms)	354	592	677	669	576	566	587	667	648	727
	409	608	669	565	744	573	598	673	700	572
	346	612	597	697	617	589	621	742	684	677
	329	621	687	547	606	595	618	637	747	571
	326	717	662	669	488	618	609	685	619	647
	290	629	630	698	513	791	588	656	604	635
	274	704	623	518	604	625	617	766	693	628
	394	687	639	621	701	623	670	647	572	693
	402	595	572	627	698	591	691	652	556	712
	343	607	564	721	677	683	798	605	609	604
	<b>Prom.</b>	<b>346,7</b>	<b>637,2</b>	<b>632</b>	<b>633,2</b>	<b>622,4</b>	<b>625,4</b>	<b>639,7</b>	<b>673</b>	<b>643,2</b>

Tabla 8. Tiempos de respuesta Equipo 3 (Mac OS X)

La Figura 37 muestra una comparación de los tiempos promedio obtenidos durante esta prueba.

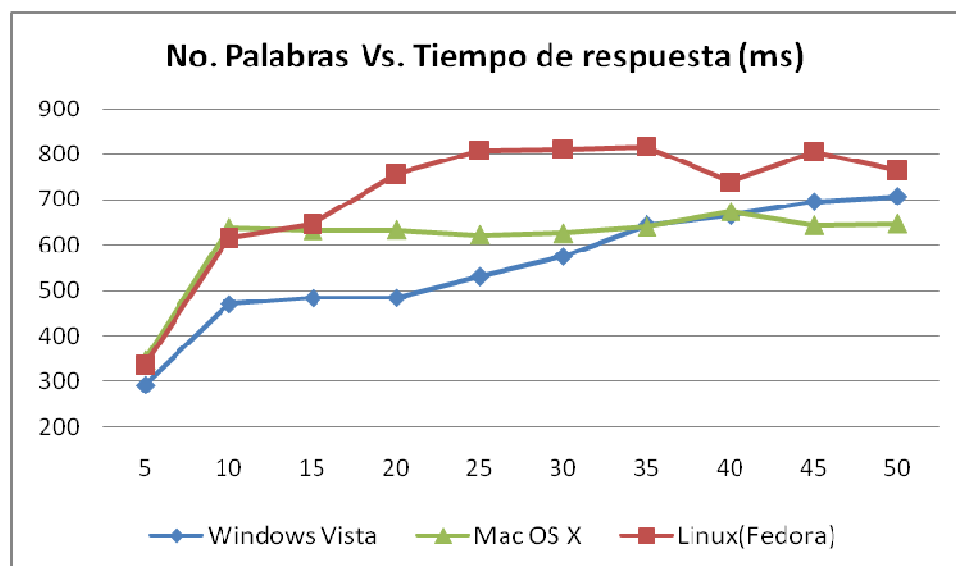


Figura 37. Comparación de tiempos de respuesta de los tres equipos de prueba

Los resultados anteriores muestran que los menores tiempos de respuesta promedio son los obtenidos para el sistema operativo Mac OS X, pues en el peor de los casos tardó casi 7 décimas de segundo en procesar el texto. Por otra parte, el equipo con Windows Vista proporcionó los segundos mejores resultados, en el peor caso tardó 7 décimas de segundo, un poco más que Mac OS X. Por último, el equipo con Fedora 8 presentó los mayores tiempos de respuesta, la mayoría de las veces tardó más de 0.7 segundos en procesar los textos, aunque nunca superando los 0.9 segundos.

La Figura 37 permite notar la diferencia entre los tiempos de respuesta de Fedora y los otros sistemas operativos, aunque se trate de centésimas de segundo. Sin embargo, las pruebas efectuadas en Windows y Mac permiten establecer que los tiempos de respuesta para estos dos sistemas son similares, y el desempeño de la herramienta no se ve afectado en este aspecto por el sistema operativo.

Por otra parte, se puede establecer que los tiempos de respuesta de la aplicación no dependen exclusivamente del tamaño del texto que se va a procesar, debido a que el algoritmo desarrollado procesa por bloques de palabras, esto permite que los tiempos de respuesta sean muy similares para diferentes longitudes de texto.

### 4.3 PRUEBAS DE ACEPTACIÓN CON USUARIOS

<b>Objetivo</b>	Conocer la opinión de usuarios finales respecto a la herramienta evaluando aspectos como la inteligibilidad de la voz, el desempeño del sintetizador, entre otros.	
<b>Actividad a realizar</b>	Usuarios finales pondrán a prueba la aplicación por un tiempo determinado, a la vez diligenciarán un formulario a manera de encuesta para evaluar varios aspectos sobre la herramienta.	
<b>Aspectos a evaluar</b>	<b>ASPECTO</b>	<b>OBJETIVO</b>
	<b>Inteligibilidad de la voz</b>	Evaluar qué tanto se entiende la voz producida por el sintetizador.
	<b>Sintetizador</b>	Evaluar desde el punto de vista del usuario, si la salida proporcionada por el sintetizador es acorde con el texto ingresado.
	<b>Tiempos de respuesta</b>	Conocer la opinión del usuario respecto al tiempo que tarda la aplicación en proporcionar una respuesta.
	<b>Facilidad de uso</b>	Evaluar la comodidad de los usuarios al utilizar la herramienta, así como la suficiencia de la información proporcionada por las ayudas de la misma.
	<b>Mecanismo de transmisión</b>	Realizar una estimación del desempeño del mecanismo de transmisión diseñado con base en la opinión de los usuarios de la aplicación.

#### 4.3.1 Formato de la prueba con Usuarios

La prueba se realizó con veinte (20) usuarios. El formato utilizado se presenta a continuación.

### **Inteligibilidad de la Voz**

*Pregunta 1.* Se probará la herramienta con 10 frases desconocidas para usted, después de escuchar cada una, escriba lo que entendió.

Frase 1: \_\_\_\_\_

Frase 2: \_\_\_\_\_

Frase 3: \_\_\_\_\_

Frase 4: \_\_\_\_\_

Frase 5: \_\_\_\_\_

Frase 6: \_\_\_\_\_

Frase 7: \_\_\_\_\_

Frase 8: \_\_\_\_\_

Frase 9: \_\_\_\_\_

Frase 10: \_\_\_\_\_

### **Sintetizador**

*Pregunta 2.* Ponga a prueba el sintetizador con 10 frases elegidas por usted, y responda las siguientes preguntas.

<b>Nro. de frase</b>	<b>Nro. de palabras contenidas en la frase</b>	<b>Nro. de palabras pronunciadas erróneamente</b>
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

### **Tiempos de respuesta**

*Pregunta 3.* Teniendo en cuenta el tiempo que se tarda en pronunciar las frases, usted considera que la herramienta es:

Rápida \_\_\_\_      Aceptable \_\_\_\_      Lenta \_\_\_\_

### **Facilidad de Uso**

*Pregunta 4.* Usted considera que la herramienta es:

Fácil de usar \_\_\_\_      Difícil de usar \_\_\_\_

*Pregunta 5.* Las ayudas escritas proporcionadas por la herramienta, son suficientes para un buen manejo de esta:

Excelentes \_\_\_\_      Aceptables \_\_\_\_      Malas \_\_\_\_

### **Mecanismo de Transmisión**

*Pregunta 6.* Al utilizar el cable manos libres modificado usted:

Escucha la voz perfectamente \_\_\_\_      Escucha ruidos además de la voz \_\_\_\_

Escucha ruidos pero no la voz \_\_\_\_      No escucha nada \_\_\_\_

#### **4.3.2 Resultados de las Pruebas de Usuario**

Los resultados de la prueba, clasificados por pregunta se presentan a continuación

- En la primera pregunta los usuarios escucharon diez frases para poner a prueba la inteligibilidad del sintetizador, se utilizaron 181 palabras; las palabras que se tuvieron en cuenta fueron las que se producen al leer las frases de prueba de la forma en que las pronuncia el sintetizador. La Tabla 9 presenta las frases utilizadas.

# Frase	Frase	# Palabras
1	Bienvenido, gracias por participar en la fase de pruebas de nuestro proyecto	12
2	Último día de entrega de proyectos en la escuela es el 23/10/08	17
3	Las clases este semestre comienzan a las 6:00 y terminan a las 10:00:00	17
4	Claudia tiene la tarea llámela al 311-2551285 y pídasela prestada	18
5	La tasa representativa del mercado se ubicó en \$1889.85	19
6	Diríjase a la Av. El dorado # 34 - 25 Edificio 13 apto. 1001 Bogotá	18
7	La página de la Real Academia Española es <a href="http://www.rae.es/rae.html">http://www.rae.es/rae.html</a>	30
8	Espero el 1er trabajo en mi correo electrónico: correo_prueba@yahoo.com	17
9	1829000 personas vieron el concierto de Juanes, 1/3 de la población de Cuba	18
10	El punto de ebullición del agua es 100 °C o 212 °F	15

*Tabla 9. Frases utilizadas en la prueba de usuarios*

Los resultados de esta prueba se dividieron en 4 categorías de acuerdo a los porcentajes de palabras acertadas por los usuarios: la primera categoría abarca desde el 90 hasta el 92.5% de palabras acertadas, la segunda hasta el 95% de las palabras, la tercera hasta el 97.5% y la última hasta el 100%. La Tabla 10 presenta detalladamente los resultados obtenidos.

	<b>Categoría 1</b> <b>97.51%-100%</b>	<b>Categoría 2</b> <b>95.1%-97.5%</b>	<b>Categoría 3</b> <b>92.51%-95%</b>	<b>Categoría 4</b> <b>90%-92.5%</b>
<b>Palabras Correctas</b>	176-181	172-176	167-172	163-167
<b># Usuarios</b>	17	2	0	1

*Tabla 10. Resultados inteligibilidad frases de prueba con usuarios*

La mayoría de los usuarios, 17 personas que corresponden al 85%, acertaron más del 97% de las palabras; dos usuarios que corresponden al 10% del total, acertaron entre 172 y 176 palabras es decir entre el 95% y el 97.5%; ningún usuario se ubicó en la tercera categoría; mientras que sólo un usuario, correspondiente al 5% acertó menos del 92.5% de las palabras, correspondiente a 164 palabras. Con base en esto, es posible afirmar que los resultados obtenidos son buenos ya que sólo se presentó un caso de la cuarta categoría, donde el mínimo de palabras acertadas por un usuario fue del 90.6% del total de palabras evaluadas. Además la mayoría de los usuarios, el 95% acertó más del 95% de las palabras de prueba.

La Figura 38 muestra una gráfica estadística con los resultados obtenidos para esta pregunta.

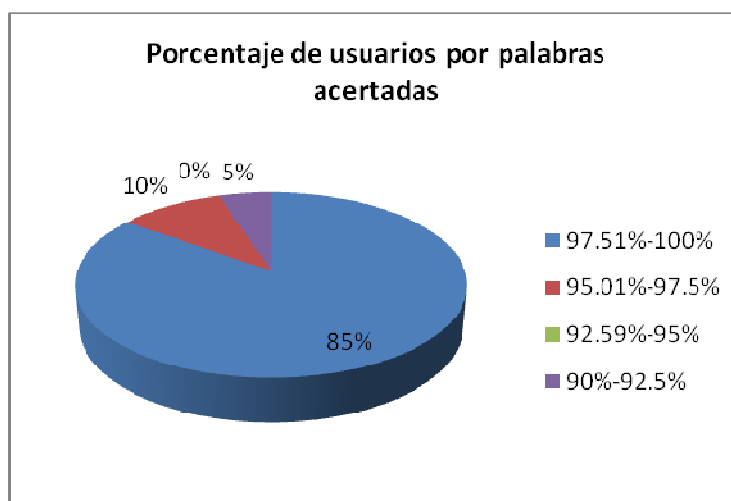


Figura 38. Resultados frases de prueba con usuarios

Con base en estos datos se calculó un porcentaje de inteligibilidad del sintetizador desarrollado, teniendo en cuenta que para los 20 usuarios se tuvo un total de 3620 palabras, correspondientes al 100% de inteligibilidad.

$$\%Intelig = \frac{((4*181)+180+(6*179)+178+177+(4*176)+175+173+164)*100}{3620} \quad (2)$$

$$\%Inteligibilidad = 98.04\%$$

De esta manera se obtuvo para el sintetizador desarrollado en este proyecto, un porcentaje de inteligibilidad del 98.04%. Esto muestra que aunque la voz desarrollada suena robótica, es de buena calidad respecto a la inteligibilidad que proporciona, que permite que las frases sintetizadas a través de la herramienta se pronuncien adecuadamente y se entiendan.

Durante el análisis de los resultados de esta pregunta, se obtuvo que varios de los usuarios fallaron en la identificación de la página web de la frase número 7, esto se debe a que estas personas no están muy familiarizadas con este tipo de construcciones del idioma y por lo tanto se les dificulta entender la pronunciación de las mismas. Por el contrario, otros usuarios no tuvieron dificultad al entender la pronunciación de la página web utilizada como prueba. Se puede decir entonces, que un inconveniente que se presentó en esta parte, fue la utilización de esta página web que resultó compleja para algunos de los usuarios, en su lugar, se pudo haber utilizado una dirección más sencilla de entender para cualquier tipo de persona.

Adicional a esto, se percibió, que varias personas encuestadas fallaron en algunas palabras debido a la falta de retentiva, aunque entendían el significado de las frases, al escribirlas olvidaban algunas palabras o en su lugar pusieron sinónimos, lo que hace entender que la idea general de la frase fue transmitida pero el inconveniente estuvo en la transcripción por parte de los usuarios.

- En la pregunta número dos, los usuarios probaron el sintetizador con un total de cuatro mil ciento cincuenta y tres (4153) palabras, de las cuales treinta y seis (36) fueron pronunciadas incorrectamente desde el punto de vista de los usuarios. Varias palabras identificadas como pronunciadas incorrectamente correspondieron a correos electrónicos; aunque el sintetizador está en capacidad de procesarlos, existen construcciones que no identifica como pronunciables, por lo tanto las deletrea y este fue el inconveniente que se encontró.

Estos resultados sirven como identificador de las fallas de los procesadores del sintetizador para un futuro mejoramiento del mismo.

Con base en los datos presentados anteriormente, se estableció un porcentaje de pronunciación correcta del sintetizador así:

$$\% \text{ Pronunciación Correcta} = 100 - \frac{36}{4153} * 100 = 99.13\% \quad (3)$$

Este porcentaje permite establecer que el sintetizador está en capacidad de pronunciar correctamente la gran mayoría de las palabras del español de Colombia cuando se tienen en cuenta los formatos predefinidos.

- La tercera pregunta tenía como finalidad conocer la opinión de los usuarios respecto al tiempo que tarda la aplicación en proporcionar una respuesta, es decir, respecto al tiempo que demora en comenzar a reproducir la voz a partir del momento en que se le da la orden de procesar un texto.

En este aspecto las opiniones de los usuarios estuvieron divididas, pues la mitad de ellos opinó que la herramienta es rápida, y la otra mitad dijo que es aceptable el tiempo de espera. Ningún usuario respondió que la aplicación es lenta. La Figura 39 presenta un gráfico con los resultados de esta pregunta.



*Figura 39. Resultados opinión de los usuarios respecto a rapidez de la herramienta*

Estos resultados muestran que respecto a la velocidad de procesamiento de la herramienta, todos los usuarios estuvieron conformes, pues ninguno tuvo una opinión negativa al respecto.

- En la cuarta pregunta se hizo referencia a la facilidad de uso de la herramienta, el 100% de los usuarios coincidió en que es fácil de utilizar. Este aspecto es de gran

importancia pues una herramienta complicada disminuye la aceptación de los usuarios, y por lo tanto no es una buena solución a una problemática específica.

- Respecto a las ayudas escritas proporcionadas (manual de usuario y ayuda de la herramienta) el 90% de los usuarios que participaron en las pruebas, es decir 18 de ellos, afirman que son excelentes, mientras que sólo dos de ellos, es decir el 10%, piensa que son aceptables. Ninguno de los usuarios encuestados piensa que las ayudas son malas. Este aspecto evaluado es bastante importante, puesto que al momento de usar una herramienta el usuario no cuenta con el apoyo de los desarrolladores, y un buen manual de usuario es la única guía, aparte de la intuición, con la que cuentan los usuarios.
- La última pregunta hace referencia al mecanismo de transmisión diseñado en el proyecto. Los resultados obtenidos en este aspecto son totalmente favorables, pues el 100% de los usuarios encuestados escuchó la voz perfectamente en medio de una llamada a través del manos libres modificado.

En general, los resultados obtenidos por medio de esta prueba son favorables, pues arrojaron altos porcentajes de inteligibilidad y pronunciación correcta, además de una actitud positiva y de aprobación por parte de los usuarios.

## 5 CONCLUSIONES

- Este proyecto permitió la integración de dos tecnologías de gran importancia en la actualidad, dándole una aplicación diferente a la conversión de texto en voz. El uso de la síntesis de voz en la telefonía celular a través de herramientas como CHSpeech permite que los usuarios puedan atender sus llamadas a pesar de los inconvenientes que puedan presentarse en el uso directo del habla.
- El estudio de las técnicas de síntesis de voz y de los sonidos del idioma en el país permitió obtener como producto de este trabajo un primer prototipo de sintetizador de voz de dominio ilimitado basado en la técnica de concatenación de difonemas para el español de Colombia, que puede ser utilizado en diversas aplicaciones y a su vez, sirve de punto de partida para el avance en el estudio de este tema en la región.
- El desarrollo de este proyecto permitió obtener resultados en el campo de la síntesis de voz, adicionales al sintetizador. El estudio realizado sobre las reglas tanto a nivel oral como escrito del idioma español en su variante colombiana permitió el desarrollo de una base de datos (*corpus de voz*) de difonemas, para ser utilizada en la técnica de síntesis por concatenación de unidades, como origen de la voz sintética de acento neutro. Este estudio incluye las tablas de identificación de los difonemas que se presentan en el idioma español de Colombia, además de las palabras ejemplo que los representan, y la documentación necesaria para su uso, implementación y mejoramiento. Adicional a esto, se obtuvo una voz sintética de género masculino con calidad aceptable y acento neutro, que puede utilizarse en varias regiones del país.
- El mecanismo implementado para la transmisión de la voz en medio de una llamada arrojó resultados positivos, ya que la señal enviada desde la computadora al dispositivo celular no sufrió pérdidas de ninguna clase, así como tampoco se presentó ruido; por el contrario la mayoría de los usuarios que participaron en la fase de pruebas de la

herramienta afirmaron que la voz se escuchaba mejor a través del cable manos libres que directamente desde la computadora.

- En todo idioma existen formatos formales de escritura y otros que son regionalistas o personales, converger a un solo punto de formatos es complejo, y hacer que la herramienta CHSpeech se ajuste al gusto y escritura de todos los usuarios es aún más complicado porque depende de una visión subjetiva; la construcción de los formatos predefinidos que utiliza la herramienta CHSpeech encargados de clasificar los textos de entrada para su correcta interpretación y procesamiento permitió dar solución a este inconveniente. De esta manera se abarca un grupo numeroso de construcciones del idioma que permiten que el sintetizador desarrollado pueda calificarse como un sintetizador de dominio ilimitado.
- A través de las pruebas realizadas al sintetizador pudo notarse que las señales de audio obtenidas son notablemente parecidas a las señales de audio producidas por un hablante natural; sin embargo aún existen secciones en las que pueden notarse las fallas en la coarticulación entre palabras, lo que disminuye la calidad de la voz respecto a su naturalidad, por esta razón se tiene en cuenta como un aspecto susceptible a mejoras.
- Los tiempos de respuesta obtenidos en la fase de pruebas dan una buena idea de la rapidez de procesamiento de la herramienta CHSpeech, ya que en el peor de los casos analizados éste no fue mayor a los 0.9 segundos, ni mayor a 0.7 segundos en el mejor; la herramienta software CHSpeech fue realizada con el objetivo de ser utilizada en medio de una llamada en la telefonía celular, lo que implica realizar el procesamiento de síntesis de voz en tiempo real; un tiempo de respuesta entre 0.7 y 0.9 segundos resulta bastante aceptable a la hora de su uso.

- El uso de hilos en la herramienta CHSpeech evita que ésta colapse en momentos de alta exigencia por parte del usuario. Esta técnica mejora el desempeño de la herramienta en el procesamiento de textos extensos sin que el usuario note prolongaciones en los tiempos de respuesta.
- La programación en el lenguaje Java trae consigo todos los beneficios brindados por esta tecnología, entre ellos, la independencia de plataforma de ejecución permitiendo ejecutar la herramienta software CHSpeech sobre cualquier sistema operativo; el paradigma orientado a objetos que facilita la programación y permite una excelente organización de codificación, además de permitir documentación implícita para un mejor y fácil entendimiento del código desarrollado; su licenciamiento open source, además de su robustez y documentación existente hacen de Java un lenguaje de programación muy poderoso, que permitió el desarrollo de una herramienta rápida y eficaz.
- El uso de Java Media Framework (JMF) proporcionado por Sun Microsystems fue vital para la realización del sintetizador porque proporcionó las herramientas necesarias para el manejo de los archivos de audio, facilitando la edición, concatenación y reproducción de la voz de forma óptima, rápida y sin sacrificar demasiados recursos computacionales.

## 6 RECOMENDACIONES

- La herramienta desarrollada sirve como punto de partida para la implementación de características que mejoren la aplicación de la síntesis de voz en la telefonía celular; tal es el caso de la creación de perfiles personales que permitan captar parámetros específicos de las voces de los usuarios de manera que la voz sintética que se reproduzca sea similar a la de éstos. Con esto se haría un aporte a la seguridad de la aplicación en cuanto a la suplantación.
- Uno de los grandes inconvenientes de la síntesis de voz es alcanzar altos niveles de naturalidad en las voces. Es recomendable mejorar la calidad de la voz obtenida en este trabajo, en cuanto a naturalidad se refiere, de manera que se escuche menos robótica y sea más similar a la voz de un ser humano.
- El estudio de los difonemas del español de Colombia realizado en este proyecto no tuvo en cuenta los diferentes sonidos que puede producir una misma letra de acuerdo a la posición que ocupa dentro de una palabra. Se recomienda que futuros trabajos en el área de síntesis de voz en el país, realicen un estudio más exhaustivo teniendo en cuenta éste y otros aspectos lingüísticos; para esto es necesaria la participación de personal capacitado en asuntos de la lengua que aporten desde su conocimiento al desarrollo de un corpus de voz mucho más completo.
- La transmisión de la voz sintética en este trabajo se realizó por medio de un cable manos libres que fue modificado para cumplir esta función. Pero existen otros posibles mecanismos de transmisión que pueden utilizarse para tal fin, ya que los dispositivos móviles cada día incorporan mejores tecnologías. Un posible candidato es Bluetooth, su uso significaría sacrificar un rango de celulares que no lo tienen incorporado, pero para ellos existe la opción implementada en el presente proyecto.

- La animación de acuerdo a las emociones y la acentuación por signos de puntuación (interrogación y admiración) es una tendencia que se está utilizando en los sintetizadores de voz alrededor del mundo; sin embargo, son los de licencias privadas, en su mayoría, los que cuentan con esta opción. Esta es una característica que resulta muy llamativa para los usuarios, y teniendo como base el sintetizador desarrollado en este proyecto, es posible adicionarla puesto que gran parte del trabajo está hecho.

## 7 REFERENCIAS

- [1] Historia de la comunicación. Disponible en <http://www.rppnet.com.ar/hiscomunicacion.htm> [Fecha de consulta: Noviembre de 2008]
- [2] Correo: La importancia del correo a través del tiempo. El correo en la antigüedad. Disponible en <http://www.adpostal.gov.co/secs/corporativa/historia1.htm> [Fecha de consulta: Noviembre de 2008]
- [3] Bahamonde M., Ángel. Otero C., Luis Enrique. El teléfono. El nacimiento de un nuevo medio de comunicación, 1877-1936. En: Las comunicaciones en la construcción del Estado contemporáneo en España. 1993. Disponible en <http://www.ucm.es/info/hcontemp/leoc/telefono.htm> [Fecha de consulta: Noviembre de 2008]
- [4] Jiménez, José J. Evolución e historia de la telefonía celular. Febrero 2004. Disponible en <http://www.revistaciencias.com/publicaciones/EpZyEkkuVlwxnCdrUU.php> [Fecha de consulta: Noviembre de 2008]
- [5] Sánchez O, José Antonio. Comunicaciones móviles. Agosto 2003. Disponible en <http://www.ilustrados.com/publicaciones/EpVAZVuFFuWmFmcpHE.php> [Fecha de consulta: Noviembre de 2008]
- [6] Tutorial de fonética – Síntesis de habla. Universidad de los Andes. Mérida, Venezuela. Disponible en [http://www.ceidis.ula.ve/cursos/humanidades/fonetica/tutorial\\_de\\_linguistica/sintes.html](http://www.ceidis.ula.ve/cursos/humanidades/fonetica/tutorial_de_linguistica/sintes.html) [Fecha de consulta: Noviembre de 2008]
- [7] LLISTERRI, J. “La síntesis del habla”. Presentación I Jornada de Tecnologías del habla Universidad de Sevilla – Universidad de Granada. Sevilla. 7 Noviembre 2000. Disponible en: [http://liceu.uab.es/~joaquim/speech\\_technology/tecnol\\_parla/synthesis/synthesis\\_general/sintesi\\_general.pdf](http://liceu.uab.es/~joaquim/speech_technology/tecnol_parla/synthesis/synthesis_general/sintesi_general.pdf) [Fecha de consulta: Octubre de 2008]

- [8] LLISTERRI, J. La conversión de texto en habla. Quark, Ciencia, medicina, comunicación y cultura 21 (2001): 79-89. Disponible en: <http://www.prbb.org/quark/21/021079.htm> [Fecha de consulta: Julio de 2009]
- [9] Características del software sintetizador de voz Loquendo TTS. Disponible en <http://www.loquendo.com/en/technology/TTS.htm> [Fecha de consulta: Noviembre de 2008]
- [10] Definición Síntesis de voz (Speech Synthesis), Disponible en [http://en.wikipedia.org/wiki/Speech\\_synthesis](http://en.wikipedia.org/wiki/Speech_synthesis) [Fecha de consulta: Octubre de 2008]
- [11] Guzmán Arreola, M. A. 2004. Sintetizador de voz para la enseñanza de la lectura a niños mexicanos. Tesis Licenciatura. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas Puebla. Enero. Capítulo 2: Sintetizadores y corpus de voz. Disponible en [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lis/guzman\\_a\\_ma/capitulo\\_2.html](http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/guzman_a_ma/capitulo_2.html) [Fecha de consulta: Octubre de 2008]
- [12] Guzmán Arreola, M. A. 2004. Sintetizador de voz para la enseñanza de la lectura a niños mexicanos. Tesis Licenciatura. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas Puebla. Enero. Capítulo 5: Síntesis de voz utilizando selección de unidades. Disponible en [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lis/guzman\\_a\\_ma/capitulo5.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/guzman_a_ma/capitulo5.pdf) [Fecha de consulta: Octubre de 2008].
- [13] Memorias Doctorado en Ingeniería Lingüística aplicada al Procesamiento de Documentos. Raquel Martínez Unanue. Disponible en <http://www.escet.urjc.es/~rmartine/IL-sesion2.pdf> [Fecha de consulta: Noviembre de 2008].
- [14] Diseño e Implementación de un Sistema de Síntesis de Voz. K. Palacio, J. Auquilla, E. Calle. Facultad de Ingenierías – Carrera de Ingeniería Electrónica, Universidad Politécnica Salesiana. Disponible en: [http://www.rte.espol.edu.ec/archivos/Revista\\_2008/15-299Final.pdf](http://www.rte.espol.edu.ec/archivos/Revista_2008/15-299Final.pdf) [Fecha de consulta: Agosto de 2009]

- [15] Descripción del proyecto open-source de síntesis de voz en inglés, FreeTTS. Disponible en: <http://freetts.sourceforge.net/docs/index.php> [Fecha de consulta: Noviembre de 2008]
- [16] Proyecto semestral: “Java Media Framework”, Cristian Nievas, Sergio Catalán. Departamento de Electrónica. Universidad Técnica Federico Santa María. Chile. Disponible en: <http://profesores.elo.utfsm.cl/~agv/elo330/2s03/projects/JavaMediaFramework/descargas/Informe%20JMF.pdf> [Fecha de consulta: Mayo de 2009]
- [17] Tratamiento multimedia en Java con JMF. Disponible en: <http://cprades.eresmas.com/Tecnica/programarJMF.html> [Fecha de consulta: Mayo de 2009]
- [18] Introducción al JMF. Servidores de Información Multimedia. 2º Ingeniero Técnico de Telecomunicación –Imagen y Sonido. Departamento de Ingeniería telemática, Universidad Carlos III Madrid. España. Disponible en: [http://www.it.uc3m.es/~labsimitis/sesiones/teoria/01\\_IntroJMF.pdf](http://www.it.uc3m.es/~labsimitis/sesiones/teoria/01_IntroJMF.pdf) [Fecha de consulta: Marzo de 2009]
- [19] Java Media Framework, Programmers Guide, Overview. Disponible en: <http://java.sun.com/javase/technologies/desktop/media/jmf/1.0/guide/JavaMediaFramework.fm1.html> [Fecha de consulta: Marzo de 2009]
- [20] Otros estudios sobre el español de Colombia. Publicaciones del Instituto Caro y Cuervo. Santafé de Bogotá. 2000. Pág. 31. [Fecha de consulta: Marzo de 2009]
- [21] El Español de Colombia. María Bernarda Espejo – Investigadora Titular Instituto Caro y Cuervo. Disponible en: <http://www.universia.net.co/vigia-del-idioma/articulos-especiales/el-espanol-de-colombia.html> [Fecha de consulta: Abril de 2009]
- [22] El idioma español en Colombia – Dialectos del Español en Colombia. Disponible en: [http://es.wikipedia.org/wiki/Idioma\\_espa%C3%B1ol\\_en\\_Colombia](http://es.wikipedia.org/wiki/Idioma_espa%C3%B1ol_en_Colombia) [Fecha de consulta: Febrero de 2009]

- [23] Extranjerismos en el diccionario de la lengua española. Cleóbulo Sabogal Cárdenas. Disponible en: <http://www.universia.net.co/vigia-del-idioma/view-document/documento-127.html> [Fecha de consulta: Mayo de 2009]
- [24] Las 500 palabras más comunes utilizadas en el lenguaje inglés. Disponible en: <http://www.world-english.org/english500.htm> [Fecha de consulta: Julio de 2009]
- [25] Castillo Condado, O 1999. Evaluación de un reconocedor fonético para el español hablado en México. Tesis Licenciatura. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería de Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas, Puebla, México. (pág. 31). Disponible en: [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lis/castillo\\_c\\_o/capitulo3.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/castillo_c_o/capitulo3.pdf) [Fecha de consulta: Noviembre de 2008]
- [26] Síntesis de voz en el dialecto venezolano por medio de la concatenación de difonos. \*M. Rodríguez, \*\*E. Mora, \*Departamento de Electrónica y Comunicaciones, Facultad de Ingeniería, \*\*Departamento de Lingüística, Facultad de Humanidades, Universidad de los Andes, Mérida, Venezuela. Disponible en: <http://www.saber.ula.ve/bitstream/123456789/16449/1/conversor-texto.pdf> [Fecha de consulta: Febrero de 2009]
- [27] Documentación del software para el etiquetado de difonemas, Diphone Studio. Disponible en: <http://www.fluency.nl/dstudio/dstudio.htm> [Fecha de consulta: Diciembre de 2009]
- [28] Síntesis en Español Mexicano con el Método de Selección de Unidades de Longitud Variable. Leonardo Flores, Alcira Vargas, Alejandra Olivier, Ingrid Kirschning, Ofelia Cervantes. TLATOA-CENTIA. Depto. Ingeniería en Sistemas Computacionales. Universidad de las Américas, Puebla, México. Disponible en: [http://ict.udlap.mx/people/ingrid/ingrid/articulo\\_33.pdf](http://ict.udlap.mx/people/ingrid/ingrid/articulo_33.pdf) [Fecha de consulta: Noviembre de 2009]

## ANEXO A.

### DESCRIPCIÓN DE FORMATOS Y PRONUNCIACIONES

- Fecha: Se acepta como valores de fecha la estructura conformada por tres campos numéricos (día, mes y año), separados por una barra (slash). El campo destinado para indicar el año acepta números tanto de 4 dígitos como de 2.

dd / mm / aaaa      dd / mm / aa

<b>Entrada al Normalizador</b>	<b>11/08/2009</b>
<b>Salida del Normalizador</b>	Once de agosto de dos mil nueve

- Hora: Para indicar horas se aceptan dos formatos. El primero de ellos incluye tres campos numéricos (horas, minutos y segundos), separados por dos puntos “:”. El segundo formato acepta únicamente los campos de horas y minutos, de igual forma separados por dos puntos. En ambos casos es posible, cuando se trata de horas de un solo dígito, omitir el cero. Por ejemplo, no es necesario 05:34, es suficiente poner 5:34.

hh : mm : ss      hh : mm

<b>Entrada al Normalizador</b>	<b>8:43:25</b>
<b>Salida del Normalizador</b>	Ocho y cuarenta y tres con veinticinco segundos

- Números Telefónicos: En esta categoría se ubican dos tipos de números telefónicos: fijos y celulares.
  1. Teléfonos Fijos: En Colombia los números telefónicos fijos constan de 7 dígitos. Se definieron dos formatos diferentes para este tipo de números. El primero acepta 7 dígitos divididos en un grupo de un dígito y otro de 6, separados por un guión. El segundo formato incluye indicativo de larga distancia nacional entre paréntesis, y el resto del número dividido de igual forma que el primer formato. El trío de dígitos que conforma el indicativo está restringido de manera que el primero de ellos debe



tampoco hay caracteres entre el signo y los números. El segundo formato es básicamente igual que el primero, la diferencia es que se le agrega la posibilidad de ingresar dos dígitos decimales precedidos por un punto. Sólo se aceptan 2 decimales debido a que éstos representan los centavos, y 100 centavos son 1 peso, entonces no tiene sentido expresar pesos en centavos. El rango de 18 dígitos permite expresar cantidades hasta “novecientos noventa y nueve mil novecientos noventa y nueve billones novecientos noventa y nueve mil novecientos noventa y nueve millones novecientos noventa y nueve mil novecientos noventa y nueve pesos”.



<b>Entrada al Normalizador</b>	<b>\$5639000</b>
<b>Salida del Normalizador</b>	Cinco millones seiscientos treinta y nueve mil pesos

- Direcciones: Existen diferentes formas de expresar direcciones, cada una de esas formas puede ser muy diferente respecto a las demás. Por esta razón no se definió un formato específico para este tipo de construcción; sin embargo la restricción que se estableció es que los componentes deben ir separados por un espacio en blanco. Por ejemplo:



En el primer campo se aceptan palabras como: calle, carrera, diagonal, avenida, y sus respectivas abreviaturas: Clle. , Cll. , Cra. , Kra. , Diag. , Av. , Avda. , entre otras. Ver Listado de abreviaturas en el ANEXO B.

<b>Entrada al Normalizador</b>	<b>Cra. 143 # 64 – 19</b>
<b>Salida del Normalizador</b>	Carrera ciento cuarenta y tres número sesenta y cuatro diecinueve

- Páginas Web: Una de las características principales de las direcciones web es que no tienen espacios. Además de esto, siempre empiezan por “www.” o algún protocolo

como “http” seguido de “://”. También es posible que estas dos construcciones nombradas anteriormente se combinen. El formato definido para páginas web permite las dos opciones, es decir, que se introduzca únicamente el protocolo (http://, ftp://, https://) o en su lugar “www.”, y la otra opción es que se realice una combinación de los dos. Enseguida debe ir el nombre de la página, que debe estar compuesto por una serie de caracteres alfanuméricos seguida de un punto. Por último se aceptan grupos compuestos por cualquier conjunto caracteres excepto espacios en blanco, y que se separan entre sí por un punto.



<b>Entrada al Normalizador</b>	<b>http://www.uis.edu.co</b>
<b>Salida del Normalizador</b>	Hache te te pe dos puntos barra barra doble u doble u doble u punto uis punto edu punto co

- Correos Electrónicos: El formato de correo electrónico es casi siempre el mismo. Un nombre de usuario, el símbolo “@”, el nombre del servidor, y el dominio. En el nombre de usuario es posible combinar caracteres alfanuméricos con los símbolos de punto “.”, guión “-” y guión de piso “\_”, pero éstos no pueden ir seguidos; además después de un símbolo de estos es necesario que haya por lo menos un carácter alfanumérico. En cuanto al servidor de correo, normalmente son utilizados nombres que contienen caracteres alfanuméricos. El dominio es un grupo de caracteres alfabéticos; es posible encontrar uno o más dominios asociados a una dirección de correo, pero éstos van separados por un punto. Por ejemplo: nombre.apellido@correo.uis.edu.co



<b>Entrada al Normalizador</b>	<b>pedro.perez@correo.uis.edu.co</b>
<b>Salida del Normalizador</b>	Pedro punto perez arroba correo punto uis punto edu punto co

- **Números en general:** El sintetizador brinda soporte para lectura de números hasta de 18 dígitos (999999999999999999). Para la identificación de números enteros en general, es necesario que éstos no estén separados por ningún símbolo, es decir, que únicamente sean ingresados los números. Por ejemplo:

Cuatro mil doscientos treinta y seis

4236 ✓

4.236 ✗

<b>Entrada al Normalizador</b>	<b>4236</b>
<b>Salida del Normalizador</b>	Cuatro mil doscientos treinta y seis

- **Números Fraccionarios:** El formato de este tipo de números está compuesto por un número para el numerador, una barra (slash), y un número para el denominador. En el caso de este último, es necesario que sea diferente de cero y de uno.



<b>Entrada al Normalizador</b>	<b>1/2</b>
<b>Salida del Normalizador</b>	Un medio

- **Números Ordinales:** Un número ordinal puede ser expresado de dos maneras, la primera es agregando después del número la terminación correspondiente a la pronunciación del ordinal del último dígito; en la gráfica se especifica la terminación adecuada para cada número. La segunda manera es agregando simplemente el símbolo “º” o “ª”. El sintetizador acepta cualquier a de los dos formatos, restringiendo el uso de números ordinales hasta 999, es decir que acepta hasta tres dígitos.

3 Dígitos

ro , er , ra  
 do , da  
 to , ta  
 no , na  
 vo , va  
 mo , ma  
 º , ª

1,3→ro , er , ra  
 2→do , da  
 4,5,6→to , ta  
 7→mo , ma  
 8→vo , va  
 9→no , na  
 10→mo , ma

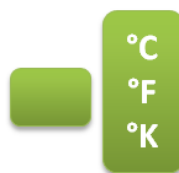
<b>Entrada al Normalizador</b>	27°
<b>Salida del Normalizador</b>	Vigesimoséptimo

- **Números Decimales:** El sintetizador acepta dos formatos para números decimales; en los dos es posible ingresar el signo. En el primer formato acepta hasta 18 dígitos en la parte entera, enseguida debe ir un punto, y en la parte decimal es posible ingresar cuantos dígitos quiera. El segundo formato es básicamente igual que el primero pero con la posibilidad de agregar exponentes con signo, dicho exponente puede ser de hasta 18 dígitos.



<b>Entrada al Normalizador</b>	-37.52
<b>Salida del Normalizador</b>	Menos treinta y siete punto cincuenta y dos

- **Temperatura:** El formato para indicar valores de temperatura es muy sencillo, simplemente se ingresa un número seguido del símbolo “°”, y por último la letra correspondiente a la unidad (Centígrados, Kelvin o Fahrenheit).



<b>Entrada al Normalizador</b>	37°C
<b>Salida del Normalizador</b>	Treinta y siete grados centígrados

## ANEXO B.

### LISTA DE ABREVIATURAS

La selección de abreviaturas se realizó con base en el listado de abreviaturas del Diccionario Panhispánico de Dudas de la RAE, dando mayor prioridad a las de mayor uso en Colombia.

Abreviatura	Lectura	Abreviatura	Lectura
<b>Sr.</b>	Señor	<b>Pág.</b>	Página
<b>Sra.</b>	Señora	<b>Pbro.</b>	Presbítero
<b>Srta.</b>	Señores	<b>P.D.</b>	Postdata
<b>Sres.</b>	Señorita	<b>Ppal.</b>	Principal
<b>Reg.</b>	Registro	<b>Priv.</b>	Privado
<b>abrev.</b>	Abreviatura	<b>Q.E.P.D.</b>	Que en paz descanse
<b>adj.</b>	Adjetivo	<b>S.S.</b>	Su Santidad
<b>Admón.</b>	Administración	<b>Sto.</b>	Santo
<b>a.C.</b>	Antes de Cristo	<b>Sta.</b>	Santa
<b>aprox.</b>	Aproximadamente	<b>Ud.</b>	Usted
<b>Art.</b>	Artículo	<b>Uds.</b>	Ustedes
<b>Atte.</b>	Atentamente	<b>V.B.</b>	Visto Bueno
<b>Av.</b>	Avenida	<b>Vol.</b>	Volumen
<b>Avda.</b>	Avenida	<b>Cll.</b>	Calle
<b>Cap.</b>	Capítulo	<b>Clle.</b>	Calle
<b>C.C.</b>	Cédula de Ciudadanía	<b>Cra.</b>	Carrera
<b>Cód.</b>	Código	<b>Kra.</b>	Carrera
<b>Cta.</b>	Cuenta	<b>Diag.</b>	Diagonal
<b>Cte.</b>	Corriente	<b>US\$</b>	Dólares
<b>c/u</b>	Cada uno	<b>Acept.</b>	Aceptación
<b>d.C</b>	Después de Cristo	<b>Alfz.</b>	Alférez
<b>Doc.</b>	Documento	<b>Almte.</b>	Almirante
<b>Dpto.</b>	Departamento	<b>a.m.</b>	A eme
<b>Dr.</b>	Doctor	<b>Apdo.</b>	Apartado
<b>Dra.</b>	Doctora	<b>Apto.</b>	Apartamento
<b>Ej.</b>	Ejemplo	<b>Arq.</b>	Arquitecto
<b>Esq.</b>	Esquina	<b>Arz.</b>	Arzobispo
<b>etc.</b>	Etcétera	<b>A.T.</b>	Antiguo Testamento
<b>Flia.</b>	Familia	<b>Atto.</b>	Atento
<b>Fig.</b>	Figura	<b>Autop.</b>	Autopista
<b>Form.</b>	Formulario	<b>Bco.</b>	Banco
<b>Gral.</b>	General	<b>Bibl.</b>	Biblioteca
<b>Hno.</b>	Hermano	<b>Brig.</b>	Brigada
<b>Hna.</b>	Hermana	<b>Bs.As.</b>	Buenos Aires
<b>Ing.</b>	Ingeniero	<b>Bl.</b>	Bloque
<b>Kg.</b>	Kilogramo	<b>Bmo.</b>	Beatísimo
<b>Km/h</b>	Kilómetros por hora	<b>c/c</b>	Cuenta Corriente
<b>Lic.</b>	Licenciado	<b>Cdad.</b>	Ciudad

<b>Min.</b>	Minutos	c.e.	Correo Electrónico
<b>mm.</b>	Milímetros	Cent.	Centavo
<b>Mín.</b>	Mínimo	Cént.	Céntimo
<b>Máx.</b>	Máximo	Cgo.	Cargo
<b>C.I.</b>	Cédula de Identidad	Izqdo.	Izquierdo
<b>Cía.</b>	Compañía	JJ.OO.	Juegos Olímpicos
<b>Cmdt.</b>	Comandante	K.O.	Knock Out
<b>Cmte.</b>	Comandante	Ltd.	Limitado
<b>Cnel.</b>	Coronel	Ltda.	Limitada
<b>Coord.</b>	Coordinador	Mons.	Monseñor
<b>C.P.</b>	Código Postal	Nro.	Número
<b>Dcho.</b>	Derecho	Núm.	Número
<b>Depto.</b>	Departamento	Párr.	Párrafo
<b>D.F.</b>	Distrito Federal	Pdte.	Presidente
<b>Dicc.</b>	Diccionario	Pg.	Página
<b>Dir.</b>	Dirección	Prof.	Profesor
<b>Dña.</b>	Doña	Pról.	Prólogo
<b>Dto.</b>	Descuento	Prov.	Provincia
<b>Dupdo.</b>	Duplicado	Rdo.	Reverendo
<b>Ed.</b>	Edición	Rep.	República
<b>Edit.</b>	Editorial	R.P.M.	Revoluciones por minuto
<b>EE.UU.</b>	Estados Unidos	RR.HH.	Recursos Humanos
<b>E.S.M.</b>	En sus manos	Rte.	Remitente
<b>Exc.</b>	Excelente	S.	Siglo
<b>Fasc.</b>	Fascículo	S.A.	Sociedad Anónima
<b>FF.AA.</b>	Fuerzas Armadas	Sgto.	Sargento
<b>Impto.</b>	Impuesto	Sig.	Siguiente
<b>Incl.</b>	Inclusive	S/N	Sin Número
<b>Inst.</b>	Instituto	Soc.	Sociedad
<b>Izq.</b>	Izquierda	Tel.	Teléfono
<b>Izqdo.</b>	Izquierdo	Tít.	Título
<b>JJ.OO.</b>	Juegos Olímpicos	Trad.	Traducción
<b>K.O.</b>	Knock Out	Tte.	Teniente
<b>Ltd.</b>	Limitado	Univ.	Universidad
<b>Ltda.</b>	Limitada	Vda.	Viuda
<b>Mons.</b>	Monseñor	Vdo.	Viudo
<b>Incl.</b>	Inclusive	V.O.	Versión Original
<b>Inst.</b>	Instituto	V°B°	Visto Bueno
<b>Izq.</b>	Izquierda	Vs.	Versus

**ANEXO C.**  
**LISTA DE SIMBOLOS**

Símbolo	Lectura	Símbolo	Lectura
.	Punto	¶	Fin de párrafo
#	Número	1	Uno
\$	Pesos	§	Sección
%	Porciento	¥	Yenes
/	Barra	«	Apertura
\	Contra barra	»	Cierre
-	Guión	-	Raya alta
—	Guión de Piso	us\$	dólares
*	Asterisco	°c	Grados centígrados
@	Arroba	°F	Grados Fahrenheit
€	Euros	°k	Grados kelvin
£	Libras Esterlinas	mm <sup>2</sup>	Milímetros cuadrados
&	Ampersand	mm <sup>3</sup>	Milímetros cúbicos
^	Circunflejo	cm <sup>2</sup>	Centímetros cuadrados
~	Virgulilla	cm <sup>3</sup>	Centímetros cúbicos
°	Grados	dm <sup>2</sup>	Decímetros cuadrados
®	Marca Registrada	dm <sup>3</sup>	Decímetros cúbicos
™	Marca comercial	m <sup>2</sup>	Metros cuadrados
¡	Abre admiración	m <sup>3</sup>	Metros cúbicos
¿	Abre interrogación	km <sup>2</sup>	Kilómetros cuadrados
“	Comillas dobles	km <sup>3</sup>	Kilómetros cúbicos
‘	Comilla simple	Å	Angstrom
˘	Tilde	α	Alfa
ˆ	Acento	β	Beta
(	Abre paréntesis	γ	Gamma
{	Abre llave	δ	Delta
[	Abre corchete	ε	Épsilon
!	Cierra admiración	ζ	Zeta
?	Cierra interrogación	η	Eta
)	Cierra paréntesis	θ	Teta
}	Cierra llave	ι	Iota
]	Cierra corchete	κ	Kappa
:	Dos puntos	λ	Lambda
,	Coma	μ	Mi
;	Punto y coma	ν	Ni
	Ó	ξ	Xi
¨	Diéresis	ο	Omicron

¢	centavos	π	Pi
©	Derechos de autor	ρ	Ro
¬	Negación lógica	σ	Sigma
	Barra partida	∞	Infinito
±	Más o menos	√	Raíz cuadrada
τ	Tau	<sup>3</sup> √	Raíz cubica
υ	Ípsilon	≈	Aproximadamente
φ	Fi	Δ	Incremento
χ	Chi	∈	Pertenece a
ψ	Psi	∉	No pertenece a
ω	omega	⊆	Subconjunto de
×	Multiplicado por	⊂	Contenido en
÷	Dividido en	∩	Intersección
∀	Para todo	∪	Unión
∃	Existe	∫	integral
>	Mayor que	∫∫	Integral doble
<	Menor que	∫∫∫	Integral triple
≤	Menor o igual	≡	Equivalente
≥	Mayor o igual	∇	Gradiente
=	Igual	∂	Derivada parcial de
+	Mas	∏	Productoria
∅	Vacio	≠	Diferente de
∑	Sumatoria	ç	Cedilla

ANEXO D.

LISTA DE EXTRANJERISMOS

<b>able</b>	<b>bye</b>	<b>earth</b>	<b>good</b>	<b>key</b>
<b>about</b>	caddy	east	graffiti	keylogger
<b>above</b>	call	eat	great	kind
<b>add</b>	came	email	green	kleenex
<b>affaire</b>	care	enough	groggy	knockout
<b>again</b>	carry	even	ground	know
<b>against</b>	cash	example	group	lady
<b>age</b>	cassette	express	hacker	language
<b>ago</b>	catcher	eye	had	large
<b>air</b>	cause	face	half	lasagna
<b>all</b>	champagne	facebook	hall	late
<b>also</b>	change	fall	hamburguer	lead
<b>always</b>	checking	family	hand	learn
<b>answer</b>	class	father	handicap	leave
<b>any</b>	clear	feedback	happen	letter
<b>appear</b>	click	feel	hard	life
<b>are</b>	close	feeling	hardware	light
<b>baby</b>	cognac	feet	has	like
<b>baguette</b>	cold	few	have	line
<b>basketball</b>	collage	field	head	listen
<b>be</b>	common	find	hear	little
<b>beam</b>	contain	fine	hello	live
<b>beauty</b>	cool	fire	help	lobby
<b>been</b>	copyright	first	her	look
<b>before</b>	could	five	here	love
<b>began</b>	country	flash	high	lycra
<b>begin</b>	course	flashback	him	machine
<b>behind</b>	cricket	fly	hippie	madame
<b>beige</b>	croissant	follow	hobby	made
<b>better</b>	cross	food	hockey	magazine
<b>between</b>	cry	foot	home	mail
<b>bird</b>	day	football	hooligan	main
<b>blue</b>	delicatessen	force	horse	make
<b>boat</b>	differ	found	hot	malware
<b>book</b>	diskette	free	hour	manager
<b>boom</b>	do	full	house	many
<b>both</b>	does	game	how	may
<b>boulevard</b>	done	gay	hundred	mean
<b>boutique</b>	door	gentleman	interview	measure

<b>brother</b>	down	get	jazz	meeting
<b>buffet</b>	draw	girl	jersey	messenger
<b>build</b>	dribbling	give	jet	might
<b>bulldozer</b>	drive	glamour	jockey	mile
<b>but</b>	dry	go	just	miss
<b>by</b>	each	god	keep	money
<b>moon</b>	perhaps	science	study	two
<b>more</b>	picture	scooter	such	under
<b>mother</b>	piece	see	sure	unit
<b>mouse</b>	piercing	seem	surface	until
<b>mousse</b>	place	seller	sweater	up
<b>move</b>	plain	sentence	table	us
<b>mozzarella</b>	plane	shampoo	tail	use
<b>much</b>	planning	shape	take	vedette
<b>music</b>	play	she	talk	voice
<b>my</b>	possible	short	teach	vowel
<b>name</b>	pound	should	teacher	wait
<b>near</b>	power	side	tell	water
<b>need</b>	press	since	that	way
<b>new</b>	pull	size	the	we
<b>night</b>	question	sky	their	week
<b>noun</b>	race	skype	them	weekend
<b>now</b>	racing	sleep	then	well
<b>number</b>	rain	small	there	were
<b>nylon</b>	rally	soccer	these	what
<b>object</b>	reach	software	they	wheel
<b>off</b>	read	some	this	where
<b>one</b>	ready	soon	those	while
<b>operator</b>	remember	sound	thought	white
<b>other</b>	right	spaghetti	three	who
<b>our</b>	roaming	speaker	time	why
<b>out</b>	room	spell	to	will
<b>output</b>	root	sprint	together	women
<b>own</b>	rootkit	spyware	too	would
<b>paddle</b>	round	staff	took	write
<b>page</b>	run	state	tour	year
<b>panzarotti</b>	said	stay	town	you
<b>paper</b>	same	stood	travelling	your
<b>pass</b>	saw	store	tree	zoom
<b>pattern</b>	say	street	true	
<b>pedigree</b>	scanner	stress	try	
<b>people</b>	school	striptease	twitter	

**ANEXO E.**  
**FONEMAS DEL ESPAÑOL EN COLOMBIA**

Fonema	Letras que lo producen
/a/	a
/b/	b , v
/ch/	Combinación c-h
/d/	d
/e/	e
/f/	f
/g/	g*
/i/	i
/j/	j , g*
/k/	c* , k , q
/l/	l
/m/	m
/ñ/	ñ
/n/	n
/o/	o
/p/	p
/r/	r*
/rr/	r* , Combinaciones con r*
/s/	s , z , c*
/t/	t
/u/	u* , ü
/x/	x , Combinación de c*
/y/	y , ll
/A/	á
/E/	é
/I/	í
/O/	ó
/U/	ú

\*El fonema se presenta bajo ciertas condiciones. Ver ANEXO F.

**ANEXO F.**  
**CONDICIONES PARA ASIGNACIÓN DE FONEMAS**

Letra Anterior	Letra Actual	Letra Siguiente	Fonema (Actual)
NI	a	NI	/a/
NI	b	NI	/b/
NI	c	e,i,é,í	/s/
		c	/x/
		h	/ch/
		Otra letra	/k/
NI	d	NI	/d/
NI	e	NI	/e/
NI	f	NI	/f/
NI	g	e,i,é,í	/j/
		Otra letra	/g/
NI	h	NI	Es muda y el fonema /ch/ ya se estudió en la letra c
NI	i	NI	/i/
NI	j	NI	/j/
NI	k	NI	/k/
NI	l	l	/y/
		Otra letra	//
NI	m	NI	/m/
NI	n	NI	/n/
NI	ñ	NI	/ñ/
NI	o	NI	/o/
NI	p	NI	/p/
NI	q	NI	/k/

<b>NI</b>	<b>r</b>	<b>r</b>	<b>/rr/</b>
<b>l,n,s</b>		<b>NI</b>	<b>/rr/</b>
<b>Inicio de palabra</b>		<b>NI</b>	<b>/rr/</b>
<b>NI</b>		<b>Fin de palabra</b>	<b>/r/</b>
<b>Otra letra</b>		<b>Otra letra</b>	<b>/r/</b>
<b>NI</b>	<b>s</b>	<b>h</b>	<b>/ch/</b>
		<b>Otra letra</b>	<b>/s/</b>
<b>NI</b>	<b>t</b>	<b>NI</b>	<b>/t/</b>
<b>g,q</b>	<b>u</b>	<b>e,i,é,í</b>	<b>No suena la u</b>
<b>Otra letra</b>		<b>Otra letra</b>	<b>/u/</b>
<b>NI</b>	<b>v</b>	<b>NI</b>	<b>/b/</b>
<b>NI</b>	<b>w</b>	<b>NI</b>	<b>/u/</b>
<b>NI</b>	<b>x</b>	<b>NI</b>	<b>/x/</b>
<b>NI</b>	<b>y</b>	<b>Ninguna</b>	<b>/i/</b>
		<b>Alguna</b>	<b>/y/</b>
<b>NI</b>	<b>z</b>	<b>NI</b>	<b>/s/</b>
<b>NI</b>	<b>á</b>	<b>NI</b>	<b>/A/</b>
<b>NI</b>	<b>é</b>	<b>NI</b>	<b>/E/</b>
<b>NI</b>	<b>í</b>	<b>NI</b>	<b>/I/</b>
<b>NI</b>	<b>ó</b>	<b>NI</b>	<b>/O/</b>
<b>NI</b>	<b>ú</b>	<b>NI</b>	<b>/U/</b>
<b>NI</b>	<b>ü</b>	<b>NI</b>	<b>/u/</b>

NI: No Importa

## ANEXO G.

### DIFONEMAS DEL ESPAÑOL EN COLOMBIA

	pau	a	b	ch	d	e	f	g	i	j	k	l	m	n	ñ	o	p	r	rr	s	t	u	x	y	A	E	I	O	U
pau	_ _	_a	_b	_ch	_d	_e	_f	_g	_i	_j	_k	_l	_m	_n	_ñ	_o	_p	-	_rr	_s	_t	_u	_x	_y	_A	_E	_I	_O	_U
a	a_	a-a	a-b	a-ch	a-d	a-e	a-f	a-g	a-i	a-j	a-k	a-l	a-m	a-n	a-ñ	a-o	a-p	a-r	a-rr	a-s	a-t	a-u	a-x	a-y	-	a-E	a-I	a-O	a-U
b	b_	b-a	b-b	b-ch	b-d	b-e	b-f	b-g	b-i	b-j	b-k	b-l	b-m	b-n	-	b-o	b-p	b-r	-	b-s	b-t	b-u	-	b-y	b-A	b-E	b-I	b-O	b-U
ch	-	ch-a	ch-b	-	-	ch-e	-	-	ch-i	-	-	-	-	-	-	ch-o	-	-	-	-	-	ch-u	-	-	ch-A	ch-E	ch-I	ch-O	ch-U
d	d_	d-a	d-b	-	-	d-e	-	d-g	d-i	d-j	d-k	d-l	d-m	d-n	-	d-o	d-p	d-r	-	d-s	-	d-u	-	d-y	d-A	d-E	d-I	d-O	d-U
e	e_	e-a	e-b	e-ch	e-d	e-e	e-f	e-g	e-i	e-j	e-k	e-l	e-m	e-n	e-ñ	e-o	e-p	e-r	e-rr	e-s	e-t	e-u	e-x	e-y	e-A	-	e-I	e-O	e-U
f	f_	f-a	-	-	-	f-e	-	f-g	f-i	-	-	f-l	-	f-n	-	f-o	-	f-r	-	-	f-t	f-u	-	-	f-A	f-E	f-I	f-O	f-U
g	g_	g-a	g-b	-	g-d	g-e	-	-	g-i	-	-	g-l	g-m	g-n	-	g-o	-	g-r	-	g-s	-	g-u	-	-	g-A	g-E	g-I	g-O	g-U
i	i_	i-a	i-b	i-ch	i-d	i-e	i-f	i-g	-	i-j	i-k	i-l	i-m	i-n	i-ñ	i-o	i-p	i-r	i-rr	i-s	i-t	i-u	i-x	i-y	i-A	i-E	-	i-O	i-U
j	j_	j-a	-	-	-	j-e	-	-	j-i	-	-	-	-	-	-	j-o	-	-	-	-	-	j-u	-	-	j-A	j-E	j-I	j-O	j-U
k	k_	k-a	k-b	-	k-d	k-e	-	-	k-i	-	-	k-l	k-m	k-n	-	k-o	-	k-r	-	k-s	k-t	k-u	-	-	k-A	k-E	k-I	k-O	k-U
l	l_	l-a	l-b	l-ch	l-d	l-e	l-f	l-g	l-i	l-j	l-k	-	l-m	l-n	-	l-o	l-p	-	l-rr	l-s	l-t	l-u	-	-	l-A	l-E	l-I	l-O	l-U
m	m_	m-a	m-b	-	-	m-e	-	-	m-i	-	m-k	m-l	-	m-n	-	m-o	m-p	-	-	-	m-t	m-u	-	-	m-A	m-E	m-I	m-O	m-U
n	n_	n-a	n-b	n-ch	n-d	n-e	n-f	n-g	n-i	n-j	n-k	n-l	n-m	n-n	-	n-o	n-p	-	n-rr	n-s	n-t	n-u	-	n-y	n-A	n-E	n-I	n-O	n-U
ñ	-	ñ-a	-	-	-	ñ-e	-	-	ñ-i	-	-	-	-	-	-	ñ-o	-	-	-	-	-	ñ-u	-	-	ñ-A	ñ-E	ñ-I	ñ-O	-
o	o_	o-a	o-b	o-ch	o-d	o-e	o-f	o-g	o-i	o-j	o-k	o-l	o-m	o-n	o-ñ	o-o	o-p	o-r	o-rr	o-s	o-t	o-u	o-x	o-y	o-A	o-E	o-I	o-O	-
p	p_	p-a	-	-	p-d	p-e	-	-	p-i	-	-	p-l	-	p-n	-	p-o	-	p-r	-	p-s	p-t	p-u	-	-	p-A	p-E	p-I	p-O	p-U
r	r_	r-a	r-b	r-ch	r-d	r-e	r-f	r-g	r-i	r-j	r-k	r-l	r-m	r-n	-	r-o	r-p	-	-	r-s	r-t	r-u	r-x	r-y	r-A	r-E	r-I	r-O	r-U
rr	-	rr-a	-	-	-	rr-e	-	-	rr-i	-	-	-	-	-	-	rr-o	-	-	-	-	-	rr-u	-	-	rr-A	rr-E	rr-I	rr-O	-
s	s_	s-a	s-b	s-ch	s-d	s-e	s-f	s-g	s-i	s-j	s-k	s-l	s-m	s-n	s-ñ	s-o	s-p	-	s-rr	s-s	s-t	s-u	-	s-y	s-A	s-E	s-I	s-O	s-U
t	t_	t-a	t-b	-	t-d	t-e	t-f	-	t-i	t-j	-	t-l	t-m	t-n	-	t-o	t-p	t-r	-	t-s	-	t-u	-	-	t-A	t-E	t-I	t-O	t-U
u	u_	u-a	u-b	u-ch	u-d	u-e	u-f	u-g	u-i	u-j	u-k	u-l	u-m	u-n	u-ñ	u-o	u-p	u-r	u-rr	u-s	u-t	u-u	u-x	u-y	u-A	u-E	u-I	-	-
x	x_	x-a	x-b	-	-	x-e	x-f	-	x-i	-	x-k	-	-	-	-	x-o	x-p	-	-	-	x-t	x-u	-	-	x-A	x-E	x-I	x-O	-
y	-	y-a	-	-	-	y-e	-	-	y-i	-	-	-	-	-	-	y-o	-	-	-	-	-	y-u	-	-	y-A	y-E	y-I	y-O	y-U
A	A_	-	A-b	A-ch	A-d	-	A-f	A-g	-	A-j	A-k	A-l	A-m	A-n	A-ñ	-	A-p	A-r	A-rr	A-s	A-t	A-u	A-x	A-y	-	-	-	-	-
E	E_	E-a	E-b	-	E-d	-	E-f	E-g	E-i	E-j	E-k	E-l	E-m	E-n	E-ñ	E-o	E-p	E-r	E-rr	E-s	E-t	E-u	E-x	E-y	-	-	-	-	-
I	I_	I-a	I-b	-	I-d	I-e	I-f	I-g	-	I-j	I-k	I-l	I-m	I-n	-	I-o	I-p	I-r	I-rr	I-s	I-t	-	I-x	I-y	-	-	-	-	-
O	O_	-	O-b	O-ch	O-d	O-e	O-f	O-g	-	O-j	O-k	O-l	O-m	O-n	-	-	O-p	O-r	O-rr	O-s	O-t	-	O-x	-	-	-	-	-	-
U	U_	U-a	U-b	U-ch	U-d	U-e	U-f	U-g	-	U-j	U-k	U-l	U-m	U-n	U-ñ	U-o	U-p	U-r	-	U-s	U-t	-	-	U-y	-	-	-	-	-

Nota: El primer fonema del difonema está indicado por la primera columna, y el segundo por la primera fila. Las mayúsculas indican las vocales con tilde.

ANEXO H.

TABLA DE DIFONEMAS Y PALABRAS EJEMPLO

No.	Difonema	Palabra
1	__-a	amarillo
2	__-b	banco
3	__-ch	choza
4	__-d	dinero
5	__-e	enano
6	__-f	falta
7	__-g	gato
8	__-i	indio
9	__-j	jaula
10	__-k	casa
11	__-l	luna
12	__-m	mujer
13	__-n	nadie
14	__-ñ	ñame
15	__-o	oruga
16	__-p	padre
17	__-rr	rico
18	__-s	saco
19	__-t	tapa
20	__-u	uva
21	__-x	xenón
22	__-y	yuca
23	__-A	ábaco
24	__-E	hélice
25	__-I	índigo
26	__-O	óleo
27	__-U	último
28	a-__	poca
29	a-a	aarón
30	a-b	cabra
31	a-ch	racha

32	a-d	cada
33	a-e	traer
34	a-f	gafas
35	a-g	traga
36	a-i	aire
37	a-j	caja
38	a-k	vaca
39	a-l	calar
40	a-m	trampa
41	a-n	cantar
42	a-ñ	caña
43	a-o	caos
44	a-p	capa
45	a-r	cara
46	a-rr	carrera
47	a-s	pasar
48	a-t	acatar
49	a-u	pausa
50	a-x	laxante
51	a-y	callar
52	a-E	aéreo
53	a-I	caída
54	a-O	faraón
55	a-U	baúl
56	b-__	club
57	b-a	abanico
58	b-b	obvio
59	b-ch	chibcha
60	b-d	abdomen
61	b-e	abeja
62	b-f	subfluvial
63	b-g	subglacial

64	b-i	avispa
65	b-j	subgerente
66	b-k	subclase
67	b-l	ablandar
68	b-m	submarino
69	b-n	subnormal
70	b-o	abogar
71	b-p	subproducto
72	b-r	abrir
73	b-s	subsidio
74	b-t	obtuso
75	b-u	abusar
76	b-y	subyace
77	b-A	acrobático
78	b-E	alfabético
79	b-I	herbívoro
80	b-O	albóndiga
81	b-U	albúmina
82	ch-a	cachaco
83	ch-b	pechblenda
84	ch-e	cochera
85	ch-i	achicar
86	ch-o	macho
87	ch-u	cachucha
88	ch-A	cháchara
89	ch-E	cliché
90	ch-I	muchísimo
91	ch-O	colchón
92	ch-U	manchú
93	d-__	maldad
94	d-a	unidad
95	d-b	advertir

96	d-e	arder
97	d-g	édgar
98	d-i	odio
99	d-j	adjunto
100	d-k	adquirir
101	d-l	dejadle
102	d-m	admitir
103	d-n	adnato
104	d-o	dedo
105	d-p	adpostal
106	d-r	cedro
107	d-s	adscrito
108	d-u	adular
109	d-y	adyacente
110	d-A	didáctico
111	d-E	académico
112	d-l	audífono
113	d-O	paradójico
114	d-U	adúltero
115	e- <u>  </u>	elefante
116	e-a	cefalea
117	e-b	cebo
118	e-ch	acechar
119	e-d	pedal
120	e-e	proveer
121	e-f	jefe
122	e-g	elegante
123	e-i	ley
124	e-j	ceja
125	e-k	pecar
126	e-l	pelea
127	e-m	femenina
128	e-n	lente

129	e-ñ	leña
130	e-o	buceo
131	e-p	arepa
132	e-r	acero
133	e-rr	aterrar
134	e-s	festejo
135	e-t	acetato
136	e-u	deuda
137	e-x	anexa
138	e-y	muelle
139	e-A	oceánico
140	e-l	cafeína
141	e-O	ideólogo
142	e-U	transeúnte
143	f- <u>  </u>	surf
144	f-a	alfabeto
145	f-e	afecta
146	f-g	afgano
147	f-i	afilar
148	f-l	chiflar
149	f-n	hafnio
150	f-o	enfoque
151	f-r	cifra
152	f-t	naftalina
153	f-u	afuera
154	f-A	enfático
155	f-E	profético
156	f-l	difícil
157	f-O	polifónico
158	f-U	sulfúrico
159	g- <u>  </u>	zigzag
160	g-a	abogado
161	g-b	rugby

162	g-d	magdalena
163	g-e	juguete
164	g-i	alguien
165	g-l	manglar
166	g-m	segmento
167	g-n	magnesio
168	g-o	algoritmo
169	g-r	agrario
170	g-s	tungsteno
171	g-u	agudo
172	g-A	orgánico
173	g-E	burgués
174	g-l	sanguíneo
175	g-O	antagónico
176	g-U	ningún
177	i- <u>  </u>	penalti
178	i-a	alergia
179	i-b	arriba
180	i-ch	chicha
181	i-d	cocido
182	i-e	pierna
183	i-f	arrecife
184	i-g	amigo
185	i-j	sufijo
186	i-k	abanico
187	i-l	afilar
188	i-m	animar
189	i-n	adivino
190	i-ñ	cariño
191	i-o	ansioso
192	i-p	diploma
193	i-r	admirar
194	i-rr	virrey

195	i-s	abismo
196	i-t	apetito
197	i-u	ciudad
198	i-x	asfixia
199	i-y	cepillo
200	i-A	asiático
201	i-E	higiénico
202	i-O	adición
203	i-U	veintiún
204	j- <u>  </u>	reloj
205	j-a	aflojar
206	j-e	ajedrez
207	j-i	bajista
208	j-o	abejorro
209	j-u	bejuco
210	j-A	bajá
211	j-E	ejército
212	j-l	cojín
213	j-O	callejón
214	j-U	júpiter
215	k- <u>  </u>	bistec
216	k-a	bocadillo
217	k-b	bricbarca
218	k-d	anécdota
219	k-e	aquello
220	k-i	alquiler
221	k-l	folclor
222	k-m	dracma
223	k-n	acné
224	k-o	acoger
225	k-r	concreto
226	k-s	fucsia
227	k-t	afecto

228	k-u	oscuro
229	k-A	mecánica
230	k-E	caquéctico
231	k-l	botiquín
232	k-O	ecónomo
233	k-U	catecúmeno
234	l- <u>  </u>	ojal
235	l-a	mulata
236	l-b	silbar
237	l-ch	acolchado
238	l-d	alcalde
239	l-e	colero
240	l-f	golfo
241	l-g	colgar
242	l-i	policía
243	l-j	álgebra
244	l-k	alcance
245	l-m	colmo
246	l-n	vulnerar
247	l-o	alocado
248	l-p	pulpo
249	l-rr	alrededor
250	l-s	pulso
251	l-t	cultura
252	l-u	absoluto
253	l-A	equilátero
254	l-E	milésima
255	l-l	curvilíneo
256	l-O	analógico
257	l-U	telúrico
258	m- <u>  </u>	álbum
259	m-a	plasmado
260	m-b	combinar

261	m-e	cometer
262	m-i	amigo
263	m-k	cumquibus
264	m-l	kremlin
265	m-n	himno
266	m-o	gomoso
267	m-p	campo
268	m-t	tamtan
269	m-u	almuerzo
270	m-A	cinemática
271	m-E	simétrica
272	m-l	mamífero
273	m-O	armónico
274	m-U	submúltiplo
275	n- <u>  </u>	dicen
276	n-a	empanada
277	n-b	duunvirato
278	n-ch	rancho
279	n-d	candela
280	n-e	moneda
281	n-f	confiar
282	n-g	mango
283	n-i	anidar
284	n-j	canjear
285	n-k	banco
286	n-l	enlace
287	n-m	conmemorar
288	n-n	innovar
289	n-o	conocer
290	n-p	bienplaciente
291	n-rr	honrado
292	n-s	consejo
293	n-t	cantar

294	n-u	denuncia
295	n-y	conllevar
296	n-A	dinámica
297	n-E	magnético
298	n-l	magnífico
299	n-O	económico
300	n-U	cornúpeta
301	ñ-a	caña
302	ñ-e	añejo
303	ñ-i	ceñido
304	ñ-o	añorar
305	ñ-u	ñu
306	ñ-A	gañán
307	ñ-E	montañés
308	ñ-l	compañía
309	ñ-O	cañón
310	o- _	mosco
311	o-a	atoar
312	o-b	alcoba
313	o-ch	coche
314	o-d	modal
315	o-e	poema
316	o-f	atrofiar
317	o-g	ahogado
318	o-i	ovoide
319	o-j	flojo
320	o-k	loco
321	o-l	color
322	o-m	coma
323	o-n	sondeo
324	o-ñ	coñac
325	o-o	cooperar
326	o-p	copa

327	o-r	corona
328	o-rr	correo
329	o-s	rosa
330	o-t	cota
331	o-u	bou
332	o-x	boxeo
333	o-y	pollo
334	o-A	coágulo
335	o-E	poético
336	o-l	heroína
337	o-O	zoólogo
338	p- _	clip
339	p-a	opaco
340	p-d	capdal
341	p-e	imperio
342	p-i	alpina
343	p-l	soplar
344	p-n	apnea
345	p-o	reposo
346	p-r	sorpresa
347	p-s	elipse
348	p-t	reptil
349	p-u	opuesto
350	p-A	antipático
351	p-E	apéndice
352	p-l	empírico
353	p-O	propósito
354	p-U	crepúsculo
355	r- _	nadar
356	r-a	veterano
357	r-b	curva
358	r-ch	archivo
359	r-d	sorda

360	r-e	arepa
361	r-f	perfecto
362	r-g	cargar
363	r-i	victoria
364	r-j	tarjeta
365	r-k	barco
366	r-l	burla
367	r-m	normal
368	r-n	carne
369	r-o	coronel
370	r-p	cuerpo
371	r-s	farsa
372	r-t	norte
373	r-u	oruga
374	r-x	marxista
375	r-y	superyó
376	r-A	temporáneo
377	r-E	interés
378	r-l	perímetro
379	r-O	pirómano
380	r-U	forúnculo
381	rr-a	borrador
382	rr-e	corredor
383	rr-i	aburrido
384	rr-o	barrote
385	rr-u	carrusel
386	rr-A	hemorrágico
387	rr-E	herrén
388	rr-l	terrícola
389	rr-O	erróneo
390	s- _	hojas
391	s-a	pasado
392	s-b	esbelto

393	s-ch	deschapa
394	s-d	desde
395	s-e	caseta
396	s-f	asfalto
397	s-g	sesgo
398	s-i	racimo
399	s-j	desjuntar
400	s-k	pascua
401	s-l	muslo
402	s-m	chisme
403	s-n	limosna
404	s-ñ	ñisñil
405	s-o	tesoro
406	s-p	caspa
407	s-rr	israel
408	s-s	piscina
409	s-t	castillo
410	s-u	censura
411	s-y	disyunto
412	s-A	bursátil
413	s-E	arsénico
414	s-l	basílica
415	s-O	insólito
416	s-U	súbito
417	t- <u>  </u>	internet
418	t-a	dictado
419	t-b	futbolista
420	t-d	postdata
421	t-e	ratero
422	t-f	postfijo
423	t-i	objetivo
424	t-j	roentgen
425	t-l	atleta

426	t-m	logaritmo
427	t-n	etnia
428	t-o	notorio
429	t-p	postpalatal
430	t-r	maestro
431	t-s	etcétera
432	t-u	postura
433	t-A	tentáculo
434	t-E	sintética
435	t-l	cartílago
436	t-O	teutónico
437	t-U	litúrgico
438	u- <u>  </u>	tribu
439	u-a	actual
440	u-b	curuba
441	u-ch	lucha
442	u-d	fraude
443	u-e	escuela
444	u-f	estufa
445	u-g	tortuga
446	u-i	pingüino
447	u-j	flujo
448	u-k	glucosa
449	u-l	ambulancia
450	u-m	volumen
451	u-n	asunto
452	u-ñ	cuñado
453	u-o	suntuoso
454	u-p	supuesto
455	u-r	apurado
456	u-rr	susurro
457	u-s	museo
458	u-t	ruta

459	u-u	duunvirato
460	u-x	succión
461	u-y	maullar
462	u-A	acuático
463	u-E	huérfano
464	u-l	lingüística
465	x- <u>  </u>	tórax
466	x-a	anexa
467	x-b	exvoto
468	x-e	boxeo
469	x-f	exfoliar
470	x-i	asfixia
471	x-k	excusa
472	x-o	exonerar
473	x-p	inexperto
474	x-t	extremo
475	x-u	sexual
476	x-A	hexágono
477	x-E	exégesis
478	x-l	taxímetro
479	x-O	exótico
480	y-a	pillada
481	y-e	ballena
482	y-i	gallina
483	y-o	bellota
484	y-u	velludo
485	y-A	maracuyá
486	y-E	payés
487	y-l	jollín
488	y-O	pabellón
489	y-U	mayúscula
490	A- <u>  </u>	mamá
491	A-b	guanábana

492	A-ch	cháchara
493	A-d	esporádica
494	A-f	metáfora
495	A-g	decágono
496	A-j	pájaro
497	A-k	carácter
498	A-l	decálogo
499	A-m	cerámica
500	A-n	triángulo
501	A-ñ	cañamo
502	A-p	galápagos
503	A-r	antártica
504	A-rr	párrafo
505	A-s	cetáceo
506	A-t	acuático
507	A-u	hidráulico
508	A-x	máximo
509	A-y	náyade
510	E- <u>  </u>	corsé
511	E-a	océano
512	E-b	débil
513	E-d	inedito
514	E-f	encéfalo
515	E-g	orégano
516	E-i	béisbol
517	E-j	estratégico
518	E-k	anécdota
519	E-l	evangélico
520	E-m	académico
521	E-n	congénito
522	E-ñ	péñola
523	E-o	alvéolo
524	E-p	epiléptico

525	E-r	américa
526	E-rr	acérrimo
527	E-s	vigésimo
528	E-t	cinética
529	E-u	terapéutica
530	E-x	léxico
531	E-y	pléyade
532	I- <u>  </u>	guaraní
533	I-a	anarquía
534	I-b	carnívoro
535	I-d	paracaídas
536	I-e	hematíe
537	I-f	pacífico
538	I-g	polígono
539	I-j	indígena
540	I-k	artículo
541	I-l	bisílaba
542	I-m	centímetro
543	I-n	rectilíneo
544	I-o	etíope
545	I-p	atípico
546	I-r	empírico
547	I-rr	pírrico
548	I-s	ateísmo
549	I-t	capítulo
550	I-x	píxide
551	I-y	trailla
552	O- <u>  </u>	nacionalizó
553	O-b	discóbolo
554	O-ch	abróchate
555	O-d	melódico
556	O-e	bóer
557	O-f	sarcófago

558	O-g	camarógrafo
559	O-j	nitrógeno
560	O-k	binóculo
561	O-l	enólogo
562	O-m	termómetro
563	O-n	antónimo
564	O-p	acrópolis
565	O-r	calórico
566	O-rr	prórroga
567	O-s	depósito
568	O-t	caótico
569	O-x	monóxido
570	U- <u>  </u>	champú
571	U-a	cacatúa
572	U-b	república
573	U-ch	feúcho
574	U-d	impúdico
575	U-e	lúes
576	U-f	búfalo
577	U-g	lúgubre
578	U-j	brújula
579	U-k	azúcar
580	U-l	penúltimo
581	U-m	catecúmeno
582	U-n	pedúnculo
583	U-ñ	cortaúñas
584	U-o	flúor
585	U-p	estúpido
586	U-r	quirúrgico
587	U-s	crepúsculo
588	U-t	inútil
589	U-y	aúllo
590	- <u>  </u>	“pausa”

## ANEXO I.

### METODOLOGÍA DE DESARROLLO EXTREME PROGRAMMING

Para la realización del presente proyecto, se utilizó la metodología ligera de desarrollo de software, llamada eXtreme Programming o XP, creada por Kent Beck, quien escribió el primer libro sobre la materia *Extreme Programming Explained: Embrace Change (1999)*. Esta metodología es considerada la más destacada entre los procesos ágiles de desarrollo de software, debido a que realiza más énfasis en la adaptabilidad que en la previsibilidad (cambios en los requisitos sobre la marcha). XP opta por la capacidad de adaptación a los cambios en cualquier instante del ciclo de vida del proyecto, que en definir todos los requisitos al comienzo y después invertir más de lo previsto en el control de los cambios que puedan suscitarse.

La eXtreme Programming se basa en tres principios fundamentales: Simplicidad, comunicación y realimentación.

#### 1. Simplicidad:

La simplicidad, se refiere a la facilidad de entendimiento del código programado, implicando organización y documentación. En el presente proyecto se intentó tener la mayor simplicidad posible, creando paquetes e incluyendo en cada uno de ellos las clases correspondientes con nombres que reflejen lo que realmente esa clase o ese paquete incluyen, con el fin de tener una organización de código; además se incluyó la correspondiente documentación, tanto de las clases como de los métodos, objetos y subrutinas que se desarrollaron, reflejando así mayor comodidad y entendimiento de cada uno de los procesos. Para este lograr buena simplicidad, se tuvo especial cuidado en la escogencia de los nombres utilizados, haciendo énfasis en que reflejaran o insinuaran la función que llevarían a cabo, teniendo presente que los nombres largos, no decrementan la eficiencia del código, ni retrasan el tiempo de desarrollo, debido a las actuales herramientas de autocompletado y refactorización.

#### 2. Comunicación:

La eXtreme Programming hace especial énfasis en la comunicación que siempre debe existir entre el equipo de desarrollo, lo que se denomina *Programación en Pares*. Esta programación implica y se enfoca en grupos de programación de 2 integrantes trabajando

en una misma estación de trabajo, donde cada miembro realiza acompañamiento a la actividad que esté realizando el otro; aquí se aplica la teoría “*cuatro ojos ven más que dos*”, el código será de mayor calidad y tendrá menos fallos. Éste principio de la XP, se realizó a lo largo del proyecto, pero especialmente en la fase de diseño y desarrollo de la herramienta software y en la fase de pruebas. La aplicación de éste principio ayudo no solo a la creación de una buena herramienta software, sino además impulsó la colaboración mutua, el acompañamiento en cuanto comprensión y aclaración de conceptos, el compartir pensamientos, ideas, logrando una realimentación mutua, y primordialmente el aprender a trabajar en equipo.

### 3. Realimentación:

La realimentación hace referencia tanto al hallazgo de errores que se conocen en tiempo real, debido a los cortos ciclos que se realizan y a la integración del usuario con el grupo de desarrollo, como a la reutilización de código. En este proyecto, el usuario y el programador eran la misma persona, entonces se tenían presente siempre las necesidades del usuario final de la aplicación; esto permitía realizar ciclos pequeños (1 a 2 semanas) en los cuales se realizaba uno de los requisitos del usuario y se probaba inmediatamente, para verificar su correcto funcionamiento, permitiendo añadir al software final funcionalidades pequeñas pero importantes a medida que se desarrollaban, en vez de realizar un gran bloque de código y después probar, presentándose errores comunes, pero difíciles de hallar debido al volumen del mismo.

## Herramientas Utilizadas para el Desarrollo de la Metodología

### 1. NETBEANS IDE<sup>18</sup>

En el presente proyecto se utilizó el entorno de desarrollo integrado de software NetBeans, herramienta que permite el desarrollo de aplicaciones en lenguaje Java para cualquiera de las arquitecturas software para las que fue diseñado: JEE (Java Enterprise Edition), JSE (Java Standard Edition) y JME (Java Micro Edition). En el caso particular de este proyecto (desarrollo de una herramienta JSE para síntesis de voz en Español), se utilizó la versión 6.5.1 de NetBeans.

Entre las características especiales que brindó éste IDE se encuentran:

---

<sup>18</sup> Sitio Web: <http://www.netbeans.org>

- Facilidad en la creación del diseño de la aplicación orientado a objetos, utilizando, paquetes, clases, métodos y objetos implementados por los procesadores del sintetizador, en la creación del corpus de voz y en la interfaz gráfica de usuario, utilizada en la herramienta CHSpeech.
- Permite el manejo (conexión, búsqueda, inserción y actualización) de registros de la base de datos embebida de DERBY, para el manejo y autenticación de los usuarios de la aplicación.
- Permite el acceso directo al repositorio de control de versiones SUBVERSION (utilizando un plugin), para manejo de errores y actualización de código fuente.
- Cuenta con la función PROFILE que habilita al desarrollador un mecanismo de monitoreo de la aplicación, permitiendo observar el consumo de CPU, de memoria, la cantidad de subprocesos desencadenados (hilos) y el tiempo de ejecución de cada uno de los métodos, entre otras.
- NetBeans es *multiplataforma*, lo que permitió desarrollar y probar la herramienta en diferentes sistemas operativos, además de habilitar un mundo amplio de ambientes para realizar su ejecución.

## 2. SUBVERSION

Es un sistema para el control de versiones que fue utilizado primordialmente para el almacenamiento continuo de los cambios realizados sobre el software desarrollado (versiones), teniendo un respaldo de seguridad del proyecto en un repositorio de internet, y teniendo la posibilidad de actualizar los cambios realizados por los programadores desde diferentes estaciones de trabajo sobre el mismo proyecto, permitiendo tener siempre actualizado el código fuente.

SubVersion es un proyecto Open Source que se integra fácilmente al IDE NetBeans a través de un modulo (plug-in) que se instala en el equipo, facilitando su utilización.

## **ANEXO J.**

### **MANUAL DE USUARIO**

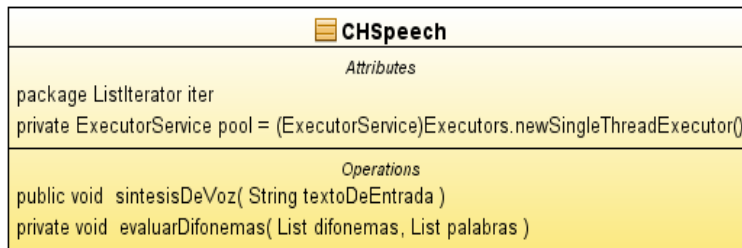
El manual de usuario de CHSpeech 1.0 puede verse en el CD.

## ANEXO K.

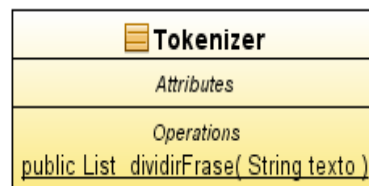
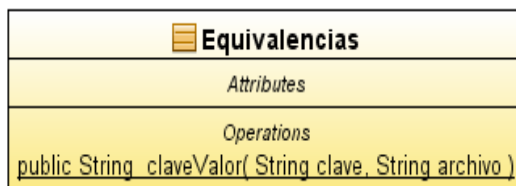
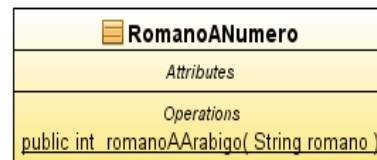
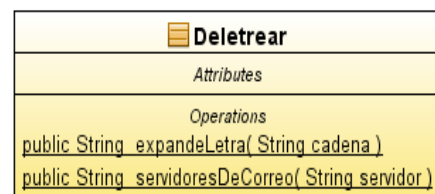
### DIAGRAMAS DE CLASES

#### 1. Sintetizador

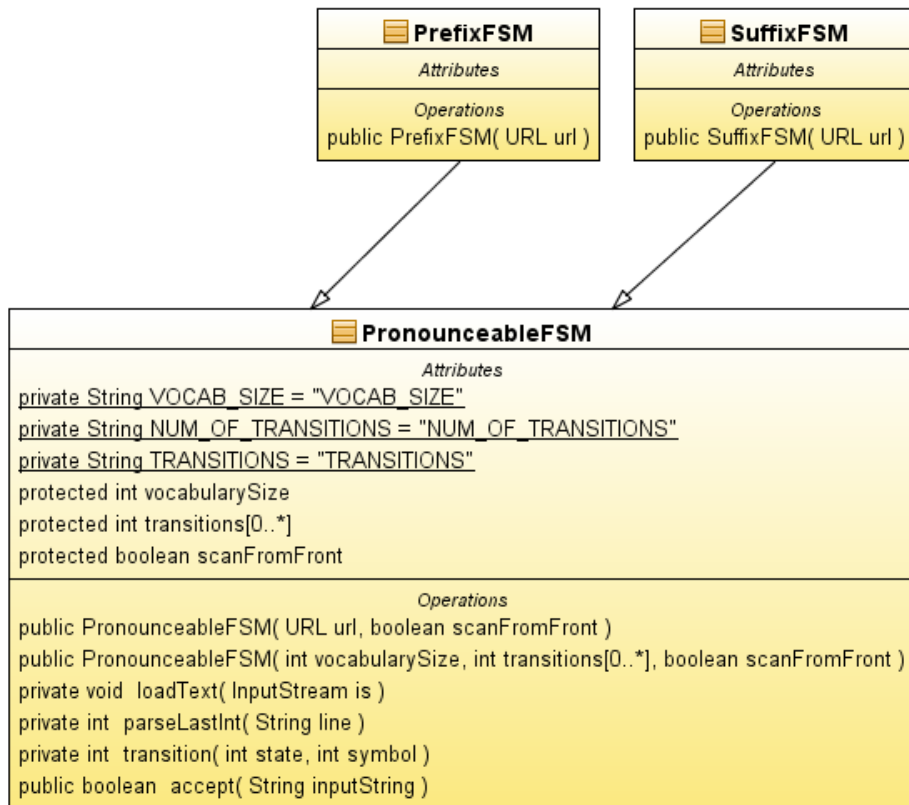
- Paquete app



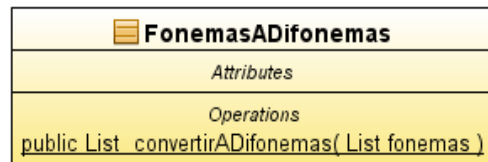
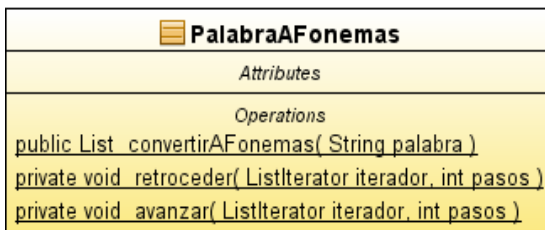
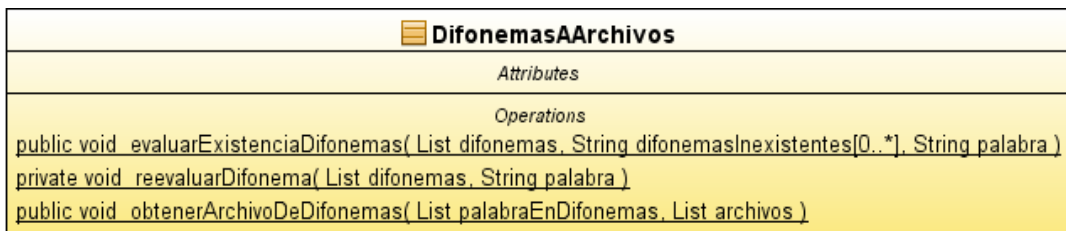
- Paquete normalizador



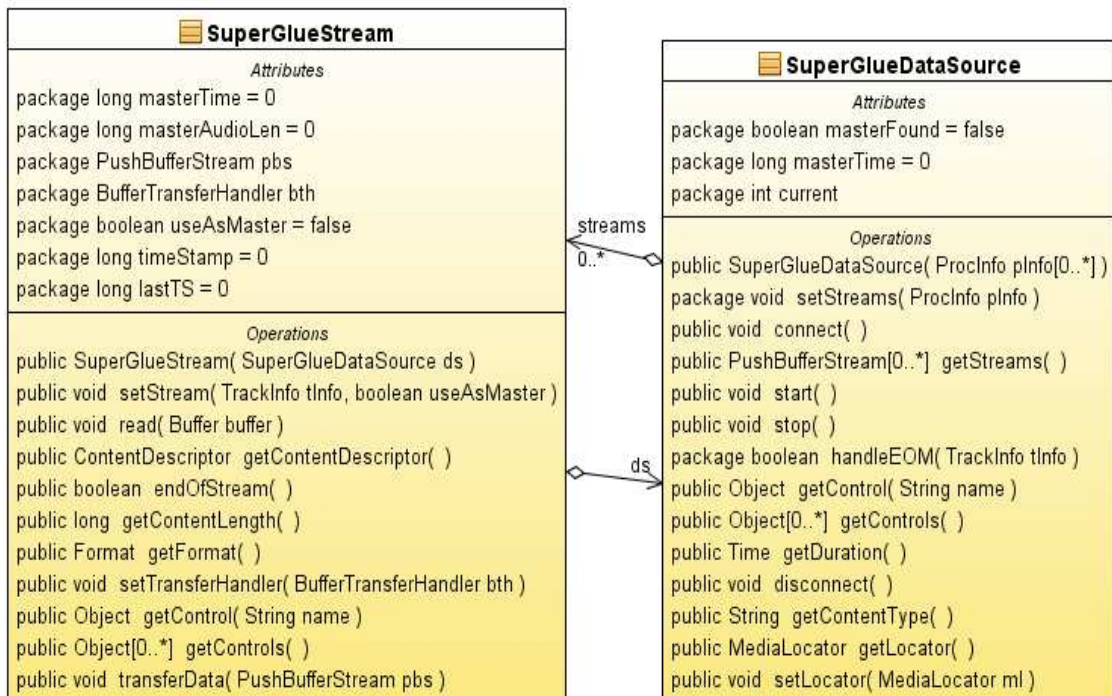
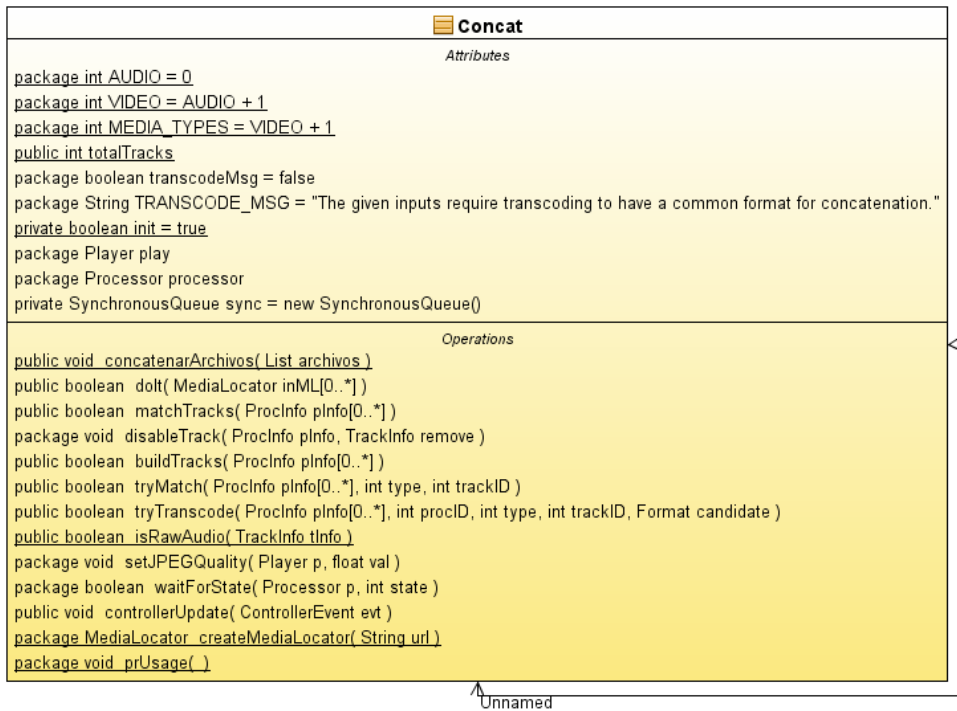




- **Paquete sintetizador**



## 2. Corpus de voz



### 3. GUI

- **Paquete interfaces**

Estos diagramas se omiten puesto que hacen referencia al manejo de eventos en la interfaz gráfica de usuario a través de Swing.

- **Paquete manejoDatos**



## ANEXO L.

### SINTESIS DE VOZ POR CONCATENACIÓN DE DIFONEMAS PARA EL ESPAÑOL DE COLOMBIA

**CLAUDIA VICTORIA CORREA PUGLIESE**  
10º Nivel Ingeniería de Sistemas  
Miembro del Grupo de Investigación en  
Ingeniería Biomédica GII B  
Universidad Industrial de Santander  
COLOMBIA  
[cato0386@hotmail.com](mailto:cato0386@hotmail.com)

**HOOVER FABIAN RUEDA CHACÓN**  
10º Nivel Ingeniería de Sistemas  
Miembro del Grupo de Investigación en  
Ingeniería Biomédica GII B  
Universidad Industrial de Santander  
COLOMBIA  
[hfnacho18@hotmail.com](mailto:hfnacho18@hotmail.com)

**MPE. HENRY ARGUELLO FUENTES**  
Profesor Titular Escuela de Ingeniería de Sistemas e Informática  
Director del Grupo de Investigación en Ingeniería Biomédica GII B  
Universidad Industrial de Santander  
COLOMBIA  
[henarfu@uis.edu.co](mailto:henarfu@uis.edu.co)

#### RESUMEN

Las tecnologías del habla, tanto el reconocimiento como la síntesis de voz, han ganado popularidad en los últimos tiempos. Actualmente son utilizadas en una amplia variedad de aplicaciones y lugares alrededor del mundo; es así como es posible escuchar voces sintéticas en diferentes idiomas (inglés, italiano, español, portugués, entre otros). Sin embargo, aún dentro de un mismo idioma existen variaciones en cuanto a la pronunciación de las palabras, por esta razón se han desarrollado voces sintéticas para diferentes lugares en los que se habla la misma lengua, como es el caso del español en España, Argentina y México.

Este artículo presenta el desarrollo y evaluación de un sintetizador de voz de dominio ilimitado utilizando la técnica de concatenación de difonemas para el Español de Colombia, el desarrollo incluye desde la concepción del corpus de voz hasta los procesadores implementados, finalizando con los resultados obtenidos.

**PALABRAS CLAVE:** Síntesis de voz, Concatenación de difonemas, Corpus de voz, Español de Colombia.

#### ABSTRACT

Speech technologies, both recognition and synthesis, have gained popularity in recent times. They are currently used in a wide variety of applications and locations around the world, and it is possible to hear synthetic voices in different languages (English, Italian, Spanish, Portuguese, among others). However, even within the same language there are variations in the pronunciation of the words, for this reason there have been developed synthetic voices for different places where people speak the same language, as is the case of the Spanish language in Spain, Argentina and Mexico.

This article presents the development and evaluation of an unlimited domain voice synthesizer using the diphone concatenation technique for Colombian Spanish, the development includes from the determination of the voice corpus until the implemented processors, ending with the obtained results.

**KEY WORDS:** Speech synthesis, Diphone concatenation, Voice corpus, Colombian Spanish.