

Capítulo 3

Protocolo y grabaciones

3.1 Protocolo de grabación

El protocolo de grabación es una parte importante del reconocedor de voz, por que es un documento que ha sido balanceado fonéticamente con un programa el cual nos da una estadística de cuantas letras y tipos de conjugaciones hay dentro del documento. Este documento no sólo esta elaborado con palabras y frases, también tiene números y fechas, que se para que el locutor pueda realizar las grabaciones con ayuda de una aplicación elaborada con CSLU Toolkit.

El protocolo de grabación para este corpus esta dividido en 4 partes que son:

- **Nombres propios y comunes:** Estos nombres son de personas y objetos con raíces de origen español Mexicano para poder tener un mejor entrenamiento del reconocedor de voz. En el apéndice “A” se presenta la lista completa de estas palabras.
- **Números:** Es una lista de números elaborada aleatoriamente con un proceso que genera segmentos con 4 dígitos para que el locutor los pueda decir como quisiera: uno por uno o hacer su unión de los números por pares o como el quisiera, solamente respetando el orden de los números. En el apéndice “B” se presentan los números que se usaron para las grabaciones.
- **Fechas:** Es una lista de fechas elaborada por medio un programa que aleatoriamente generó el día, mes y año. En el apéndice “C” podemos ver las distintas fechas.

- **Oraciones:** Las oraciones fueron sacadas de un libro de texto, cuentos, e Internet. En el apéndice “D” podemos ver las frases que el locutor leyó para la grabación.

3.2 Grabación

La grabación de los archivos de sonido, es la segunda fase en la operación de este sistema, su finalidad es grabar dentro de un directorio el protocolo como archivos de sonido. Dicha fase se realiza usando un sistema de grabación hecho con el sistema CSLU Toolkit, como se muestra en la figura 3.1.

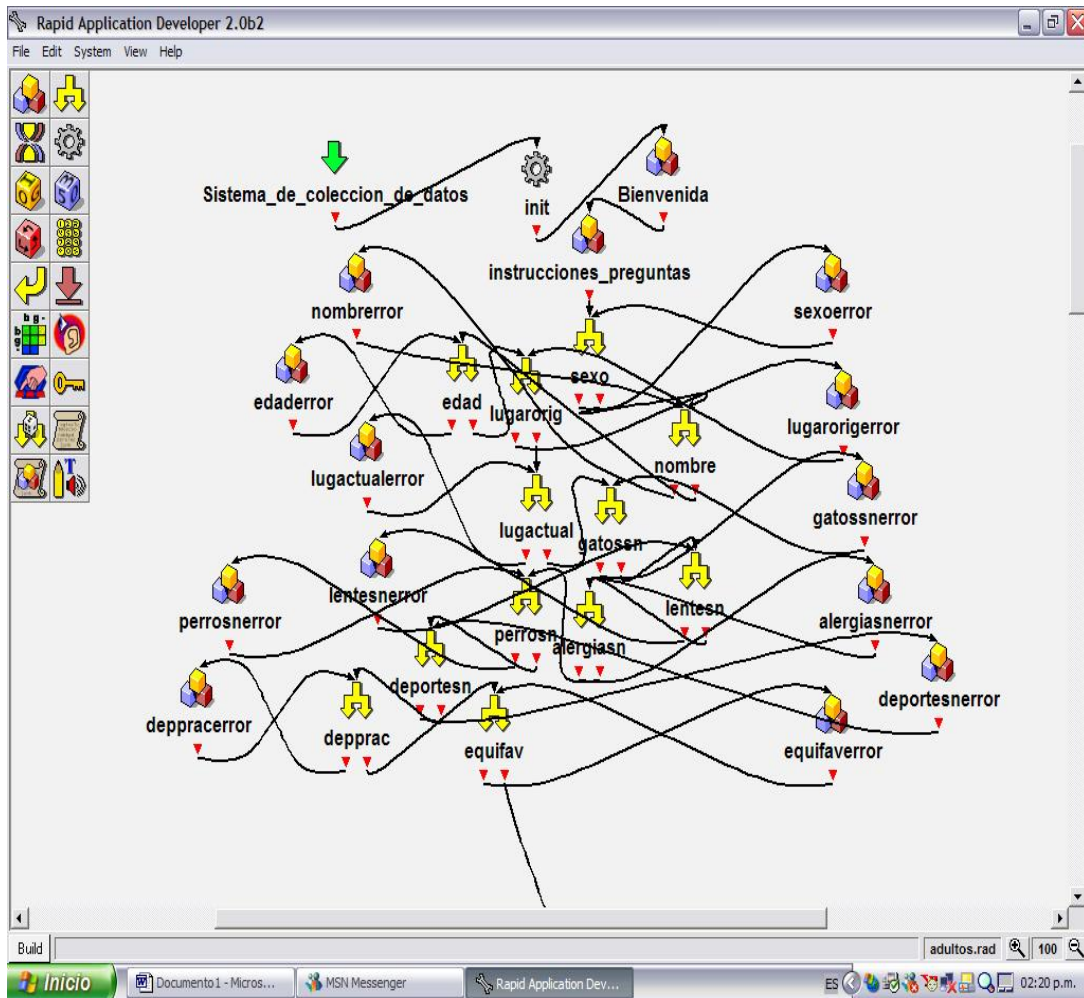


Figura 3.1: Aplicación para grabar.

EL CSLU Toolkit es un conjunto de herramientas que provee un ambiente poderoso y flexible para el desarrollo de sistemas de voz, esta diseñado para facilitar el desarrollo rápido de sistemas del lenguaje hablado para una variedad de aplicaciones así como también para proveer una plataforma para realizar investigación de tecnologías de voz [Serridge, 1998].

Las grabaciones fueron elaboradas en un lugar cerrado y con la menor cantidad de ruido que se pudo, pero no se buscó un lugar 100% aislado de ruido, ya que si el reconocedor va a ser usado en distintos lugares es necesario ver que tenga un poco de ruido para poder ser mejor el reconocimiento. Los componentes de las grabaciones fueron una computadora portátil, un micrófono de base y unos audífonos y el programa CSLU Toolkit.

En la computadora se elaboró una aplicación con CSLU Toolkit la cual graba las voces de nuestros distintos locutores, los audífonos fueron para que se los pusiera el locutor y pudiera escuchar las distintas instrucciones para poder grabar y no hiciera interferencia con el micrófono, el micrófono capturó la voz de nuestros distintos locutores. En la figura 3.1 se puede ver como se hicieron las grabaciones.










Figura 3.2: Locutor grabando el protocolo.



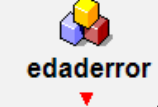

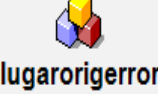

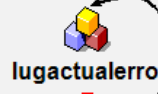

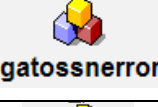



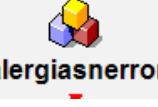

3.3 Proceso de grabación


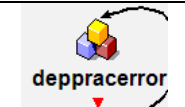





Las características de las personas que grabaron fueron las siguientes:

Personas mayores a los 18 años ya sean hombres y mujeres de distintas entidades de la republica, esto fue posible por que en los lugares en los que se hicieron las grabaciones habían personas de distintos lados de la republica los cuales tienen un acento característico de su lugar de origen y eso ayudo dentro del proceso del reconocimiento de voz para que fuera mas general nuestro entrenamiento.

Con este programa facilita mucho las grabaciones por que el mismo sistema muestra lo que el locutor en turno va a leer y también explica lo que tiene que decir el locutor siendo validado cada uno de los textos que el locutor dice. El archivo de audio va a ser grabado en un archivo extensión .wav el cual se guardó con el nombre del archivo de texto para después ver que lo que esta en el archivo de texto el locutor la grabo bien.

 Sistema_de_coleccion_de_datos	Inicio
 init	INIT es donde damos todas las variables y direcciones que necesitaremos para la grabación
 Bienvenida	Da un mensaje de bienvenida (Empezaremos a grabar tu voz)
 instrucciones_preguntas	Te da las instrucciones necesarias para poder comenzar la grabación (Por favor contesta lo siguiente, después de sonar la señal bip)
 SEXO	Pregunta si es hombre o mujer el locutor (¿Es hombre o mujer?)
 sexoerror	Si el locutor hablo muy fuerte o muy bajo la aplicación vuelve a pedir la información.
 nombre	Graba el nombre del locutor (¿Cuál es su nombre completo?)

 <p>nomberror</p>	<p>Si el locutor hablo muy fuerte o muy bajo la aplicación vuelva a pedir la información.</p>
 <p>edad</p>	<p>Graba la edad del locutor. (¿Qué edad tiene?)</p>
 <p>edaderror</p>	<p>Si el locutor hablo muy fuerte o muy bajo la aplicación vuelva a pedir la información.</p>
 <p>lugarorig</p>	<p>Graba el lugar de origen de la persona por el tipo de acento. (¿Cuál es su lugar de origen?)</p>
 <p>lugarorigerror</p>	<p>Si el locutor hablo muy fuerte o muy bajo la aplicación vuelva a pedir la información.</p>
 <p>lugactual</p>	<p>Graba el lugar donde ha vivido los últimos años. (¿Lugar donde ha vivido los últimos años?)</p>
 <p>lugactualerror</p>	<p>Si el locutor hablo muy fuerte o muy bajo la aplicación vuelva a pedir la información.</p>
 <p>gatossn</p>	<p>Graba un si o no (¿Le gustan los gatos si o no?)</p>
 <p>gatossnerror</p>	<p>Si el locutor hablo muy fuerte o muy bajo la aplicación vuelva a pedir la información.</p>
 <p>lentesn</p>	<p>Graba un si o no. (¿Usa lentes si o no?)</p>
 <p>lentesnerror</p>	<p>Si el locutor hablo muy fuerte o muy bajo la aplicación vuelva a pedir la información.</p>
 <p>alergiasn</p>	<p>Graba un si o no. (¿Tiene alguna alergia si o no?)</p>
 <p>alergiasnerror</p>	<p>Si el locutor hablo muy fuerte o muy bajo la aplicación vuelva a pedir la información.</p>
 <p>perrosn</p>	<p>Graba un si o no (¿Le gustan los perros si o no?)</p>

 <p>perrosnerror</p>	<p>Si el locutor hablo muy fuerte o muy bajo la aplicación vuelva a pedir la información.</p>
 <p>deportesn</p>	<p>Graba un si o no (¿Le gusta los deportes si o no?)</p>
 <p>deportesnerror</p>	<p>Si el locutor hablo muy fuerte o muy bajo la aplicación vuelva a pedir la información.</p>
 <p>deprac</p>	<p>Graba el nombre un deporte (¿Qué deporte practica usted?)</p>
 <p>depracerror</p>	<p>Si el locutor hablo muy fuerte o muy bajo la aplicación vuelva a pedir la información.</p>
 <p>equipfav</p>	<p>Graba el nombre de un equipo o jugador (¿Equipo o deportista favorito, que sea de México?)</p>
 <p>equipaverror</p>	<p>Si el locutor hablo muy fuerte o muy bajo la aplicación vuelva a pedir la información.</p>
 <p>instrucciones_letras</p>	<p>Da instrucciones. (Lee lo siguiente)</p>
 <p>letras</p>	<p>Nos indica el siguiente renglón del protocolo de grabación o si es todo lo que contenía el documento</p>
 <p>nombra_las_siguientes_letras</p>	<p>Se encarga de mostrar el contenido de cada uno de los renglones que contiene el protocolo</p>
 <p>grabar_letras</p>	<p>Graba lo que despliega la aplicación.</p>
 <p>actualiza</p>	<p>Es una decisión si es todo o continuamos grabando el protocolo</p>
 <p>instrucciones_numeros</p>	<p>Da instrucciones. (Lee lo números)</p>
 <p>numeros</p>	<p>Nos indica el siguiente renglón del protocolo de grabación o si es todo lo que contenía el documento</p>












 lee_los_numeros	Se encarga de mostrar el contenido de cada uno de los renglones que contiene el protocolo.
 grabar_numeros	Graba los números que leerá el locutor.
 actualiza	Es una decisión si es todo o continuamos grabando el protocolo
 instrucciones_palabra_frases	Se encarga de mostrar el contenido de cada uno de los renglones que contiene el protocolo
 state3	Nos indica el siguiente renglón del protocolo de grabación o si es todo lo que contenía el documento
 lee_las_frases	Se encarga de mostrar el contenido de cada uno de los renglones que contiene el protocolo
 grabar_frases	Graba las frases que despliega la aplicación.
 actualiza	Es una decisión si es todo o continuamos grabando el protocolo
 hablaexp	Graba un dialogo que el locutor quiera. (Habla un poco de lo que tú quieras.)
 lectura	Graba un texto que se le da al locutor. (Lee lo que se te pide.)
 adios	Fin de la grabación

Tabla 3.3: Elementos de la aplicación de grabación del corpus.

La interfaz que tiene y presenta en el sistema de grabación es muy amigable, por que se les da las instrucciones por medio de de la aplicación sin recurrir a las hojas de papel, ya que en las hojas es mas difícil por que se puede equivocar el locutor y seria una perdida de tiempo, como igual unas malas grabaciones.

3.4 Características de grabación.

Las grabaciones fueron realizadas en un lugar no completamente aislado del ruido ambiental, debido a que el reconocedor pretende ser de propósito general y no será utilizado en un lugar cerrado ni tampoco en condiciones muy ajenas al ruido ambiental sino que será usado en un algún lugar abierto con ruido.

Como se comento anteriormente las grabaciones que realizo cada uno de los locutores están compuestas del nombre de la persona, su sexo (masculino y femenino), de que lugar de la república son, esto para identificar el tipo de acento y pronunciación, preguntas de las cuales la respuesta era un si/no, si practicaban algún deporte y cual era, leían números que se generaron aleatoriamente, fechas compuestas de día mes y año, oraciones de libros de texto, cuentos, y al final se les dejaba que hablaran aproximadamente 40 a 50 segundos de lo que ellos quisieran esto para tener muestras de habla espontánea.

El numero de grabaciones que se tuvieron fueron 1822 a 16 Khz. con calidad micrófono, echas por 52 locutores que comprende 24 hombres 28 mujeres de distintas edades que van de 18 a 50 años.

Los problemas que se presentaron durante la grabación fue un exceso de ruido y murmullos lo cual no dejaba que avanzara la aplicación elaborada por CSLU Toolkit y que esta validaba la frecuencia de grabado si era muy alta los 16 Khz. solicitaba al locutor que repitiera la grabación o si estaba por debajo de la frecuencia igual solicitaba que se repitiera la grabación.

3.4 Características del protocolo de grabación.

El protocolo de grabación esta compuesto de una serie de preguntas, lectura de números, fechas, frases, lectura y habla espontánea.

Primero se le da una bienvenida al locutor y se le solicita que conteste la siguiente pregunta, y cada una de las grabaciones que hace el locutor es validada para que se hagan en una frecuencia de 16 Khz.

¿Es usted hombre o mujer? La respuesta que se recibe es hombre o mujer dependiendo el sexo del locutor.

Me podría decir su nombre, el locutor dice su nombre con esto para tener un poco de mayor vocabulario dentro de nuestro corpus.

¿Qué edad tiene usted? En esta pregunta el locutor contesta solo con el numero o también dice la palabra años esto queda abierto como quiera decirlo el locutor.

¿Cuál es el lugar de origen? La respuesta es el nombre de alguna ciudad o estado de la republica mexicana esto es para ver de donde es el locutor y la entonación con la cual habla.

¿Donde ha vivido los últimos años? La respuesta puede ser la misma que el lugar do origen y por lo mismo es para tener en cuenta si a cambiado su estilo de entonación.

¿Le gustan los gatos?, ¿Tiene alguna alergia?, ¿Le gustan los deportes? Cada una de estas preguntas fueron contestadas con una afirmación o una negación, esto se delimito así para que se tuviera un mejor control en las respuestas de los locutores.

Resumen

En este capitulo hablamos sobre el protocolo de grabación y como se hicieron las grabaciones, el protocolo de grabación esta compuesto de palabras como son nombres de personas, nombres de cosas, números aleatorios, fechas aleatorias, y frases escogidas de un libro de cuantos mexicanos.

Las grabaciones fueron echas con personas de distintas edades, sexo, regiones, esto para hacer un poco mas variado el tipo de acento de voz, y timbre de voz. Las personas fueron mayores de 18 años y con calidad micrófono por que aun no existe una aplicación para reconocer la voz adulta por medio de un micrófono del lenguaje español de México.

