

## **Conclusión**

El corpus que se desarrolló en este proyecto fue balanceado por un programa hecho en java, el cual compara una lista de sílabas (ver Apéndice B) con el protocolo de grabación, éste consta de frases y palabras obtenidas por medio de libros de la SEP, chistes, refranes, revistas, pensamientos, cuentos, diccionarios y páginas de internet (ver Apéndices C y D). Generando la cantidad de mil cuatrocientas oraciones efectivas para contener toda la conjugación de los sonidos. Se grabó en siete sesiones para poder tener una mejor calidad de voz, debido a que el protocolo de grabación es demasiado amplio.

Se etiquetó el corpus a nivel texto, palabras y fonemas de forma manual, para tener un mejor trabajo con el sistema de selección de unidades (véase capítulo 5). Este sistema fue desarrollado por Leonardo Flores en su tesis de licenciatura de la UDLA [Flores, 2001] el cual trabaja con el corpus Fraga, y fue elaborado con documentos como, revistas, periódicos y artículos políticos.

Para este trabajo se utilizó dicho sistema con el fin de reproducir el corpus Gama que se generó con palabras que no reproduce el corpus de Fraga, con el fin de mejorar la voz artificial.

Para la realización de este proyecto se comenzó con la búsqueda de información sobre la comunicación, que es la base del lenguaje, de tal manera se va tomando parte del conocimiento requerido para empaparse sobre este tema. También fue necesario saber sobre los sintetizadores existentes y algunos trabajos de investigación.

El protocolo de grabación se fue generando poco a poco, con el fin de abarcar todos los posibles sonidos del lenguaje español mexicano. Se buscó lograr un equilibrio

fonético y silábico con los documentos que integran el mismo. Véase los apéndices C y D que contienen todo el protocolo que fue grabado.

La grabación fué la segunda parte y la más importante del proyecto, ya que es la materia prima de la reproducción del sonido, es necesario una voz clara del locutor y paciencia para leer los documentos, así como también equilibrar el estado de ánimo del mismo para lograr la calidad necesaria y que no exista demasiada variación en la calidad de las grabaciones. Debido a que el protocolo es demasiado largo se realizaron pruebas y se llegó a la conclusión que una persona, que no es un locutor profesional, no varía su tono de voz al leer aproximadamente doscientas cincuenta frases o grabaciones por lo que el total de grabaciones para este proyecto fueron siete, y una para los errores de pronunciación o distracción.

El etiquetado fue la tercera parte más importante del proyecto debido a que se tiene que hacer minuciosamente para poder alinear los archivos de sonido (.wav) con los archivos de texto (.txt), palabra (.wrđ) y fonema (.phn). De esta manera, el sintetizador por concatenación ejecuta la voz clara y uniforme por medio del método de Unit Selection.

El sistema de unit selection realizado en el año 2001 se utilizó con el fin de poder comparar el corpus Fraga, con el corpus Gama realizado en éste trabajo, logrando mejoras que se mencionan a continuación:

- Las pruebas del sintetizador demuestran mejoras en la voz, como la claridad del timbre y mayor volumen en comparación con el corpus Fraga. Esto es debido a

que este corpus está diseñado para cualquiera de los contextos fonéticos del español. Fue creado grabando documentos de revistas y periódicos sobre finanzas y política, por lo que no concatena lo que un niño puede entender.

- El corpus Gama permite reproducir palabras que el corpus Fraga no puede, por ejemplo palabras en diminutivo o algunas pronunciaciones que solo se utilizan en el lenguaje español hablado en México.
- Otra de las mejoras y la más importante de todas, que ahora se cuenta con dos corpus, uno con voz de hombre y otro con voz de mujer, ambos grabados con el mismo protocolo que contiene una amplia gama de palabras y sonidos los cuales se pueden utilizar con fines generales para el uso de cualquier proyecto que se desee con una voz artificial.

- Alguna de las diferencias significativas entre el corpus Gama y el corpus Fraga son:

<b>Gama</b>	<b>Fraga</b>
<ul style="list-style-type: none"> <li>• Dos voces, femenino y masculino.</li> </ul>	<ul style="list-style-type: none"> <li>• Una voz masculina</li> </ul>
<ul style="list-style-type: none"> <li>• Tonalidad alta en la reproducción de voz.</li> </ul>	<ul style="list-style-type: none"> <li>• Tonalidad media en la reproducción de voz.</li> </ul>
<ul style="list-style-type: none"> <li>• 1400 grabaciones en archivos .wav, con las características de 256 kbps y 16 kHz</li> </ul>	<ul style="list-style-type: none"> <li>• 800 grabaciones en archivos .wav, con las características de 128 kbps y 8 kHz</li> </ul>
<ul style="list-style-type: none"> <li>• Amplia gama de grabaciones con sonidos utilizados en México con palabras como Tzintzuntzan, xocohitzli, etc</li> </ul>	<ul style="list-style-type: none"> <li>• Grabaciones de sonidos para adultos omitiendo sonidos para niños y mexicanismos como Agravios, Progresistas, etc.</li> </ul>
<ul style="list-style-type: none"> <li>• Corpus con fines generales</li> </ul>	<ul style="list-style-type: none"> <li>• Corpus con fines específicos</li> </ul>

Para este trabajo se tuvieron las siguientes limitantes:

1. Costos altos para contratar un locutor profesional y una cabina de grabación.
2. Falta de disponibilidad de tiempo de los locutores para hacer las grabaciones.

### 3. Falta de tiempo para lograr la etiquetación completa de los dos locutores.

Con este trabajo se logra un avance más para el estudio de la voz artificial en el área del lenguaje español hablado en México. Gracias a esta investigación se pueden realizar otros trabajos a futuro para mejorar la voz:

- Realizar la etiquetación del corpus Gama con un grupo de trabajo, para lograr una mejor calidad en su alineación de las fronteras y así tener el corpus en su totalidad.
- La búsqueda de otro algoritmo de concatenación de unidades para lograr una mejor y clara pronunciación del locutor y quitarle la robotización a la voz que pueda generar.
- Mejorar el tiempo de respuesta del sintetizador a lo que el usuario desea reproducir en texto a voz.
- Generar interfases para trabajar con la lectura de mail.
- Conmutadores y páginas electrónicas en servicio telefónico como cine, teatro, eventos artísticos y deportivos, etc.
- En el área académica, involucrar a personas en temas interesantes y culturales por medio del habla.
- La lectura y escritura en zonas rurales como apoyo a los profesores para comenzar con una cultura tecnológica, ya que en la actualidad todos tenemos derecho a conocerla.
- Sistemas para personas invidentes que les permitan interactuar mediante la tecnología utilizando el lenguaje braille. Lo más importante es que una persona con una discapacidad podrá hacer uso de las tecnologías

computacionales en cualquier lugar usando una computadora común y corriente.

En la actualidad ya existen programas para personas invidentes pero tienen un costo alto y no todas las personas con esta discapacidad son capaces de tener el conocimiento, ni la facilidad para poder adquirir un sistema de esta índole, por lo cual, es necesario trabajar un poco para la sociedad y no quitarle a las personas de bajos recursos la facilidad de poder crecer con todo el conocimiento y la tecnología que día con día avanza.

El presente trabajo logra grandes avances en el campo tecnológico de la voz artificial. Esto desboca una explosión de problemas los cuales pueden ser de utilidad para la sociedad ya que se considera su fin el desarrollo de tecnologías.