Instituto Tecnológico de Estudios Superiores de Monterrey

Campus Querétaro

Diseño de sistemas en chip

4.4.3 Síntesis de voz

Profesora: Eloina Rodriguez González

Marcos Eduardo García Ortiz A01276213
Jesús David Talamantes Morales A01706335

Santiago de Querétaro, Querétaro, 29 de mayo de 2021

Resumen de Fundamentación técnica de la síntesis de voz

Antecedentes

En la historia antes del nacimiento de la electrónica digital hubo diversos inventores que trataban de imitar la voz humana como por ejemplo Silvestre II (d. 1003 AD), Alberto Magno (1198–1280) y Roger Bacon (1214–1294), en 1779 el científico Christian Kratzenstein,construyó modelos del tracto vocal humano que podían reproducir los sonidos de las cinco vocales. En la década de los treinta, los laboratorios Bell desarrollaron el vocoder, el cual automáticamente analizaba el habla a través de su nota fundamental y resonancias. De su trabajo con el vocoder, Homer Dudley desarrolló un sintetizador operado por un teclado llamado The Voder, el cual fue exhibido en la New York World's Fair de 1939.

Los primeros sistemas de computadora basados en la síntesis de voz fueron creados en los cincuenta. El primer sistema general de inglés de texto-habla fue desarrollado por Noriko Umeda et al. en 1968 en Laboratorio Electrotécnico en Japón.

En 1961, el físico John Larry Kelly, Jr y su colega Louis Gerstman usaron una computadora IBM 704 para sintetizar la voz, un evento importante en la historia de los laboratorios Bell. El sintetizador de voz de Kelly (vocoder) reprodujo la canción "Daisy Bell" con el acompañamiento musical de Max Mathews.

Dispositivos móviles electrónicos incluyendo síntesis de voz comenzaron a aparecer en los setentas. Uno de los primeros fue la calculadora para ciegos Speech+ de Telesensory Systems Inc. (TSI) en 1976. El primer juego electrónico multijugador en usar la síntesis de voz fue "Milton" de Milton Bradley Company

Elementos de la síntesis

Síntesis de difonos

La síntesis de difonos usa una base de datos de voz mínima que contiene todos los difonos (transiciones entre sonidos) que ocurren en el lenguaje. El número de difonos depende de la fonotáctica del lenguaje: por ejemplo, en el idioma español existen alrededor de 800 difonos y en el alemán 2500. En la síntesis de difonos, solo un ejemplo de cada difono es almacenado en la base de datos de voces. En el tiempo de ejecución, la prosodia objetivos de una oración es superpuesta en estas unidades mínimas a través de técnicas de procesamiento digital de señal como la codificación predictiva lineal o o técnicas más recientes como la codificación del tono en el dominio de la fuente empleado la transformada de coseno discreta.32

La síntesis de difonos sufre de glitches sonidos de la síntesis concatenativa y el sonido de naturaleza robótica de la síntesis de formantes y tiene pocas ventajas sobre cualquier otro acercamiento más que su tamaño. Su uso en aplicaciones comerciales ha disminuido, aunque sigue siendo investigada debido a su número de aplicaciones en software gratuito.

Síntesis de dominio específico

Concatena palabras y frases pregrabadas para crear enunciados completos. Es usada en aplicaciones donde la variedad de los textos del sistemas está limitada a una salida de audio en un dominio particular, como los anuncios en un calendario de tránsito o reportes del clima,La tecnológica es muy simple de implementar y ha sido empleada de manera comercial por varios años en dispositivos como calculadoras o relojes parlantes. El nivel de naturalidad de estos sistemas puede ser muy alto debido a que la variedad los tipos de oraciones está limitada y logran estar muy cerca de la prosodia y entonación de las grabaciones originales.

Síntesis Concatenativa

La síntesis concatenativa es la más eficiente en sistemas de síntesis al día de hoy. En la síntesis concatenativa se pueden modificar más detalladamente las unidades mínimas de lenguaje logrando una mayor naturalidad cuando éstos se producen.

Como consecuencia de lo anterior, la inteligibilidad y entonación de una voz artificial de síntesis concatenativa superan a aquellas logradas con síntesis articulatoria o con síntesis de formantes.

Uso para eliminar barreras

La síntesis de voz ha sido una de las herramientas vitales de tecnologías de apoyo y su aplicación en esta área es significativa y de gran uso. Permite que las barreras ambientales sean removidas para personas con diferentes discapacidades. La aplicación con mayor uso han sido los lectores de pantalla para personas con discapacidades visuales, pero los sistemas de texto-voz ahora son comúnmente usados por personas con dislexia y otras dificultades para la lectura, así como para los niños. También son frecuentemente empleados para ayudar a aquellos con discapacidades comunicativas usualmente a través de una voz de ayuda.

Fig. 1 Imagen demostrativa de apps

Diseño de la solución diagrama de flujo (Generación de archivos en consola)

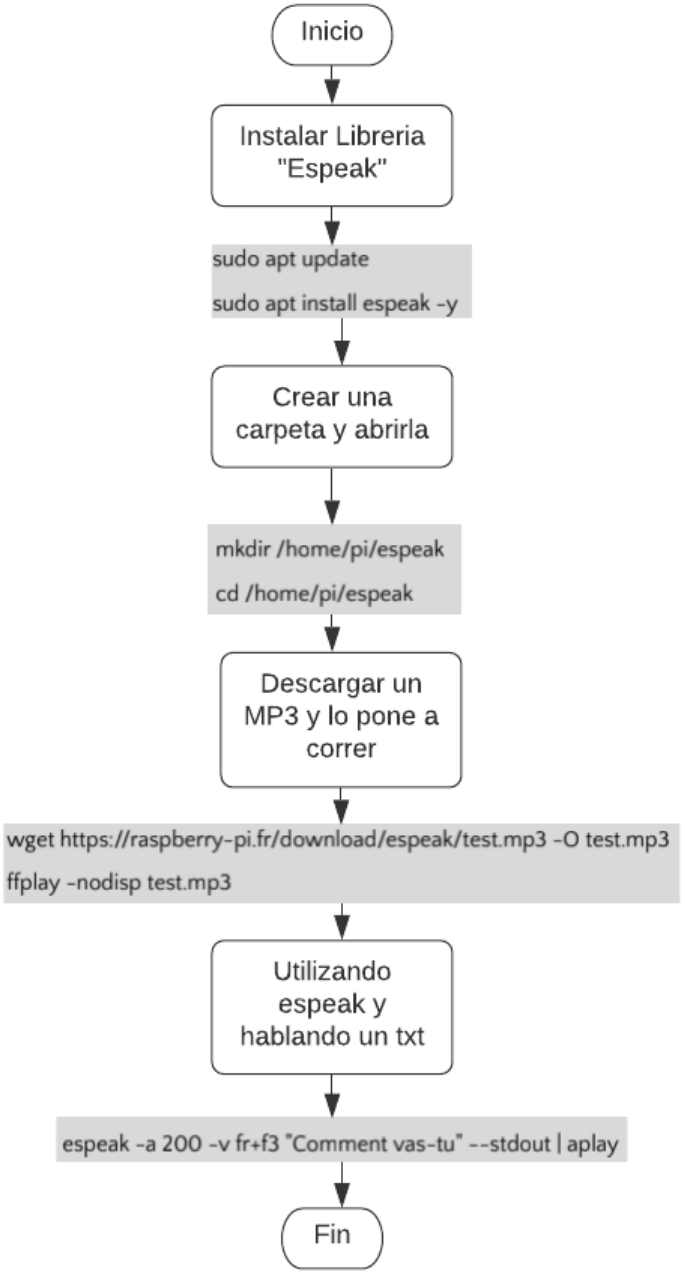

Fig. 2 Diagrama de flujo de uso de espeak

Resolver el reto:
Diseño de la solución del Reto (diagrama de flujo)

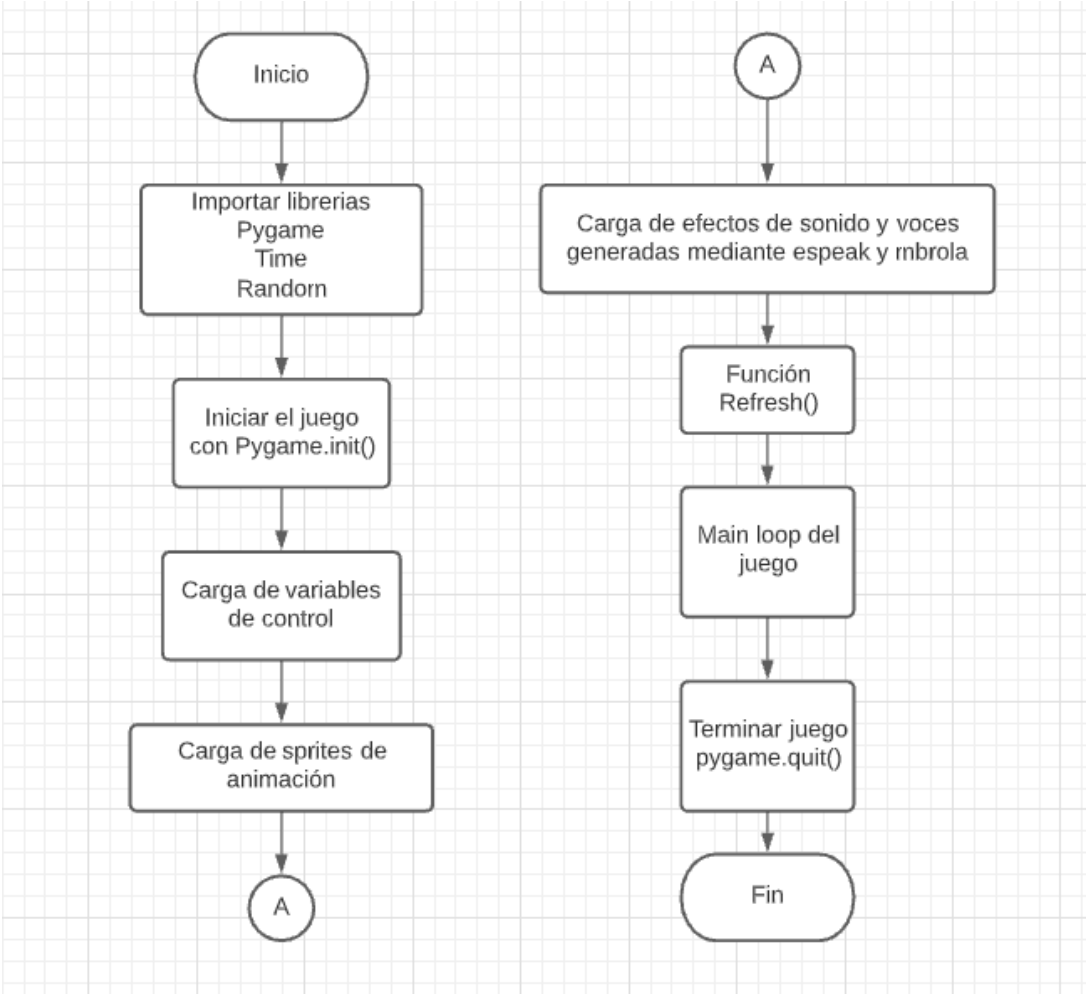

Fig. 3 Diagrama de flujo de Funcionamiento del Juego

Programa creado en el IDE de Thonny

Debido a la complejidad del programa y también a que no queremos que sea tan extenso con imágenes, decidimos crear un Github el cual puede visitar para mejor comprensión del programa. En este caso ya teníamos implementado las mejorar para esta actividad

https://github.com/JesusTalMor/Actividad_4.4.2

Explicación detallada

En esta práctica continuamos implementando mejoras a nuestro pequeño videojuego de Naruto donde mediante las funciones dadas por espeak mbrola en raspberry Pi al ejecutar en consola el comando espeak -a 200 -v mb-es1 -s 150 "Mensaje" -w

prueba.wav es posible generar un archivo .wav con el texto que exista entre comillas además de poder elegir el idioma según el API de la liberia donde “es” para obtener la síntesis en castellano y mx en español con acento mexicano

Con ayuda de esta herramienta generamos los mensajes que dirá nuestro personaje al presionar la tecla S que será “Mi nombre es Naruto Uzumaki y voy a ser hokage” en español y en inglés junto con dos frases extra que contrastan pues son las dichas por la actriz de doblaje que lo dobla en japón notando como la síntesis de voz suena de manera particular.

Se incluyó además la librería random pues las frases fueron incluidas en un arreglo para que al presionar la tecla la frase generada sea aleatoria pudiendo alternar entre las cuatro variantes programadas.

Ejecución

https://youtu.be/HRDXqEXkRyw

Conclusiones individuales

Marcos Eduardo García Ortiz A01276312

En esta práctica se lograron realizar mejoras al código de videojuego de Naruto donde ahora se incrementaron funcionalidades de habla por síntesis de voz permitiendo al presionar una tecla que el personaje genere diferentes diálogos característicos del personaje logrando una implementación más compleja todo ello mediante entender la manera en que se realiza la síntesis de voz a texto mediante las librerías de espeak mbrola las cuales cuentan con múltiples comandos que generan voz en distintos idiomas.

Jesús David Talamantes Morales A01706335

Con este último de ejercicio logramos integrar en nuestro proyecto de las voces generadas por medio de un ordenador para poder darle aún más vida a nuestro personaje, aunque la voz generada no se parezca en nada a la voz del personaje que estamos utilizando, el tener un voz generada por computadora abre bastante el panorama para nuestra implementación del reto y la implementación de proyectos más pequeños como un chat bot.

Contribución o impacto de esta actividad en la solución de la situación problema (relacionar entregas anteriores con el avance de esta actividad)

En el desarrollo de esta actividad hicimos uso de las librerías espeak y mbrola disponibles en raspberry pi, las cuales fueron instaladas con la ayuda de la consola y posteriormente mediante comandos y especificaciones fue posible generar diferentes síntesis de voz que fueron capaces de transformar texto a voz

Esta función como ya se mencionó anteriormente hace posible activar las opciones de accesibilidad para personas con discapacidad visual donde los dispositivos sean capaces de indicarle al usuario que elementos se están presionando algo que resulta de vital importancia al momento de diseñar un sistema de infoentretenimiento pues le da a toda persona la posibilidad de interactuar con este sistema

Referencias

Martí Roca, J. | SITUACION ACTUAL DE LA SINTESIS DE VOZ [Ebook]. Barcelona: Departamento de Acústica de la Escuela Universitaria de Telecomunicaciones. Retrieved from https://www.ub.edu/journalofexperimentalphonetics/pdf-articles/EFE-IV-JMarti-Sintesis_voz.pdf

Síntesis de Voz | EcuRed (2013). Consultado el 29 de Mayo de 2021, de https://www.ecured.cu/S%C3%ADntesis_de_voz