

# Software decodificador de texto a voz para ayuda a personas con capacidades diferentes

Luis Javier Mona Peña, Joaquin Antonio Alvarado Bustos,  
Octavio Mendez Hernandez

Instituto Tecnológico de Saltillo, Departamento de Ingeniería Eléctrica-Electrónica,  
Saltillo, Coahuila, México

lmona@isicnc.net,jalvarado@its.mx,omenh@hotmail.com

**Resumen.** El desarrollo de la tecnología históricamente ha ido de la mano de la resolución de problemas o necesidades que contribuyen al bienestar de la Humanidad. Cuando la tecnología se aplica al mejoramiento de la calidad de vida de personas con alguna discapacidad física y/o mental es mucho más evidente el propósito de la misma. El presente artículo se enfoca al uso del software LabVIEW (*Laboratory Virtual Instrumentation Engineering Workbench*) en el desarrollo de una herramienta capaz de auxiliar a personas con capacidades diferentes, en específico débiles visuales y/o mudos, aclarando que también es útil al resto de la población. Dicha herramienta es un convertidor de Texto a Voz (TTS) con la capacidad de leer archivos de texto en código ASCII (.txt), imágenes con extensión .jpg y texto introducido directamente desde la misma interfaz gráfica de la aplicación. Además es capaz de convertir el texto interpretado en documentos codificados en Braille y desplegarlos en un archivo del procesador de texto más común que es Microsoft Word. Como punto final, la aplicación es capaz de guardar la lectura computarizada en archivos de audio tipo WAV (*WAVE form audio file format*) y así crear audiolibros de las fuentes de texto ya descritas.

**Palabras clave:** ASCII, Braille, JPG, LabVIEW, TTS, WAV.

## 1. Introducción

La poca disponibilidad y accesibilidad de fuentes bibliográficas en Código Braille para personas con debilidad visual y el poco entendimiento de la mayor parte de la población del lenguaje a señas en el caso de personas con dificultad en el lenguaje hablado, merman su calidad de vida, y trae como consecuencias:

- Falta de interés en conocer la actualidad social.
- Problemas de interacción social.
- Rechazo desde/hacia el entorno.
- Sentimiento de frustración, etc.
- Poco o nulo desarrollo educativo profesional, etc.

Existen diferentes opciones tecnológicas para apoyar a las personas con discapacidad tales como los convertidores de Texto a Voz (TTS) los cuales son una

herramienta esencial en el apoyo de personas con capacidades diferentes. Existen varios tipos de los mismos, algunos de ellos requieren equipo especial tal como escáneres de alta velocidad, cámaras de alta resolución, teclados y sistemas de cómputo dedicados, los mismos que pueden llegar a tener costos prohibitivos para la mayoría de la población. Además de que su arquitectura es cerrada siendo muy complicado o imposible que pueda adaptarse a cambios personalizados requeridos por el usuario.

El reto principal era desarrollar una herramienta por software que no dependiera de hardware y software especial, es decir que pueda funcionar en cualquier computadora y utilizar cualquier cámara o escáner para la adquisición de la imagen, aunado a que fuera flexible al momento de poder escoger entre diferentes idiomas y voces para la lectura de los datos, además de convertir los documentos en texto convencional ASCII a documentos en Código Braille, dejando dichos documentos listos en un archivo en Microsoft Word para imprimirse y hacerlos disponibles para las personas que requieren leer en dicho código.

Para mayor comodidad del usuario, la aplicación es capaz de guardar los textos leídos en archivos de audio, en específico en formato WAV. Esto abre la cantidad de literatura disponible en formato Braille además de crear una audioteca digital.

## **2. Metodología**

La implementación de la aplicación se puede realizar por medio de diversas herramientas, tales como Sistemas de visión con el uso de algoritmos de extracción morfológica, aprendizaje de patrones y su posterior entrenamiento, Algoritmos matemáticos para Reconocimiento Óptico de Caracteres (OCR) y por medio de la utilización de Referencias ActiveX.

En este caso se optó por el uso de la tecnología ActiveX con el propósito de reutilizar código ya compilado tanto del sistema operativo Windows (XP, Vista, 7, 8) y de Microsoft Office (2003, 2007, 2010), en específico el **Microsoft Office Document Imaging**.

El sistema operativo Windows trae la opción de Texto a Voz para lo cual utiliza la Referencia ActiveX **Microsoft Speech SDK 5.1**, para la obtención de la lectura computarizada de texto plano, la cual, sin embargo es incapaz de obtener texto desde imágenes. Microsoft Office por medio de la herramienta del Módulo Office Document Imaging **MODI.IDocument** permite la extracción de texto de imágenes para su posterior uso o edición. La aplicación por lo tanto hizo uso de dichas referencias ActiveX (**Microsoft Speech SDK 5.1** y **MODI.IDocument**) además del generador del módulo de Microsoft Office para LabVIEW 2013 para la integración de la misma.

## **3. Implementación de la aplicación**

La Secuencia de operación en programación, involucra el uso de lo que el sistema operativo tiene instalado para herramientas TTS, es decir, voces computarizadas y

dispositivos de audio principalmente. Por lo que lo primero es llamar a la referencia Microsoft Speech SDK 5.1 para saber cuáles voces y que idiomas hay disponibles en el Sistema Operativo. Dicha información se muestra en la interfaz de usuario en forma de lista.

Cabe aclarar que existen diferentes ejemplos en la web para el uso del ActiveX de Microsoft Speech SDK lo que facilita la implementación de la aplicación. La figura 1 muestra el código necesario para dicha acción.

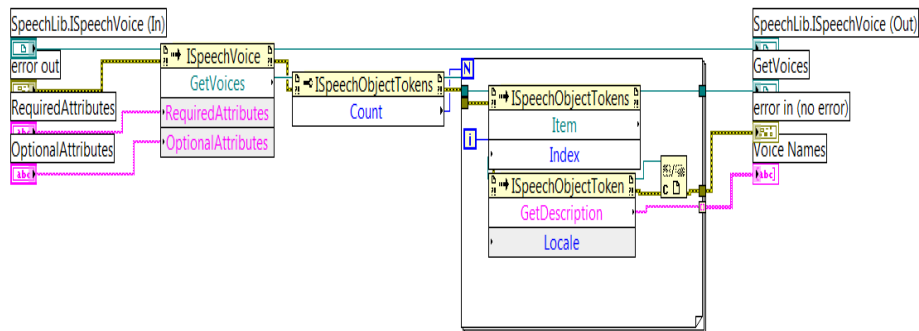


Fig. 1. Obtención de las voces del sistema.

Una vez inicializadas las voces del sistema se procede a obtener información del tipo de dispositivos de audio que están disponibles para escuchar la lectura. Esto se realiza haciendo la requisición directamente al sistema operativo por medio de la librería ya mencionada. El código de la figura 2 muestra el procedimiento realizado.

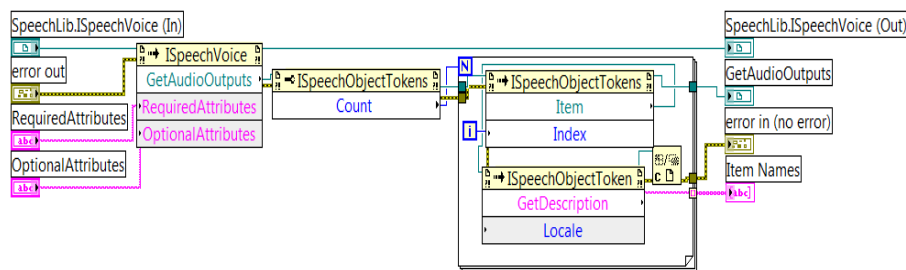


Fig. 2. Obtención de dispositivos de audio disponibles.

Una vez que se han determinado las voces y los dispositivos de audio con que cuenta la PC, se procede a la inicialización de los parámetros de velocidad, volumen y fuente de texto de la lectura. Esto se realiza con una estructura tipo case donde dependiendo de las acciones del usuario será el código ejecutado. Hay tres posibles opciones:

1. En el caso de que la fuente del texto sea un archivo .txt la interfaz despliega los elementos necesarios para que el usuario busque el archivo dentro de la estructura

del sistema operativo. Es importante señalar que esta versión del software solo acepta archivos de texto plano (.txt), por lo que archivos con otras extensiones (.pdf, Word, etc.) deben convertirse previamente a texto plano.

2. Si la opción seleccionada es por texto directamente en la interfaz, entonces el usuario tiene una caja de texto donde puede teclear todo el texto que requiera ser leído. No existe restricción en cuanto a la cantidad de caracteres.

3. La tercera opción es Texto a partir de una imagen. Para esta opción se utilizó la referencia ActiveX MODI.IDocument, la cual permite obtener diversa información sobre una imagen incluyendo el texto que está en ella, por lo que ya no es necesario realizar algoritmos OCR sino que solo utilizar las propiedades y métodos de dicho ActiveX. La figura 3 muestra parte del código necesario para obtener el texto en forma de oración.

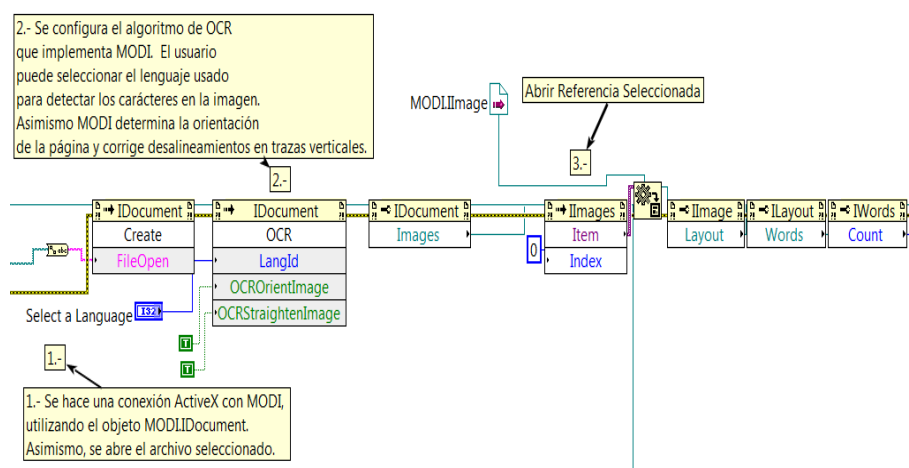


Fig. 3. Obtención de caracteres a partir de imagen.

Una vez inicializado el sistema se procede a realizar la acción de lectura. Dicha acción ejecuta código del ActiveX SpeechLib.ISpeechVoice el cual es parte del Microsoft Speech SDK. A dicho ActiveX es necesario pasar los datos previos de Voz, Volumen, Velocidad y Dispositivo de audio, dicho código se muestra en la figura 4.

El usuario escuchará en la lectura de acuerdo a la inicialización que haya configurado. Es importante recalcar que existen voces para lectura en español, inglés, ruso, etc. Estas voces son independientes de la aplicación y se instalan al sistema operativo para hacerlas disponibles. La conversión a Código Braille se llevó a cabo utilizando el Toolkit de Microsoft Office para LabVIEW el cual permite dar formato a un documento en Word, esto incluye la capacidad de cambiar el tipo de fuente que es la clave para dicha conversión. Únicamente se cambia el tipo de fuente a una tipo Braille, en este caso a la fuente tipo “XBraille”, con lo cual se obtiene cualquier documento codificado en Braille. El código que realiza dicha conversión se muestra en la figura 5.

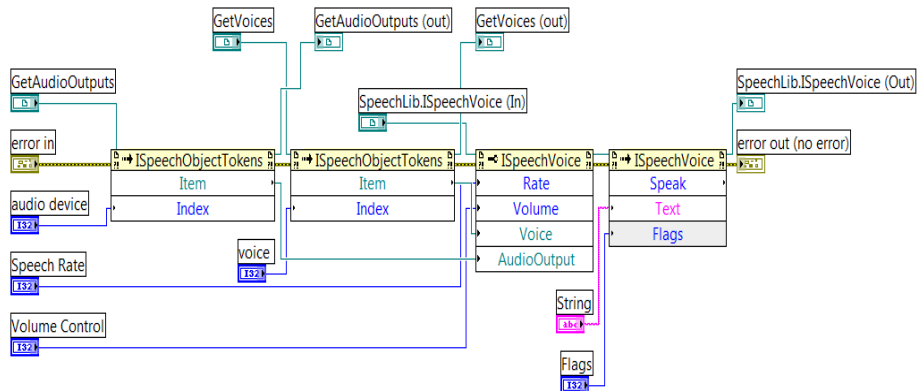


Fig. 4. Código para la acción de lectura.

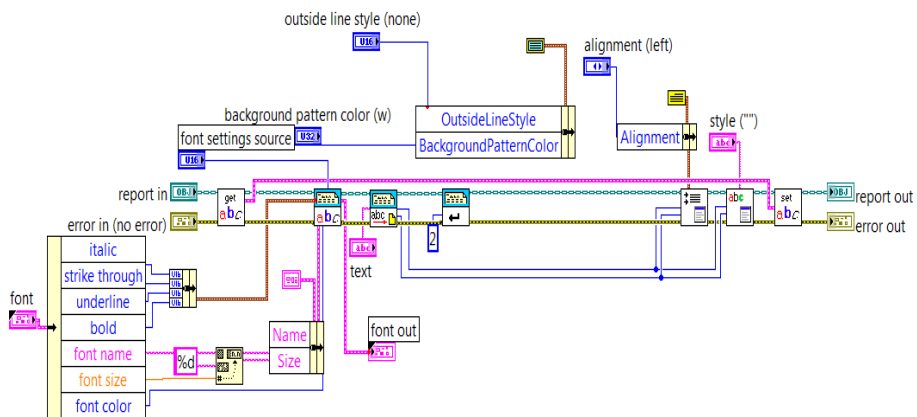


Fig. 5. Conversión a Braille.

Para la generación de los archivos de audio .wav se utilizaron las herramientas de LabVIEW de la paleta de Graphics and Sound y solo requiere una reconfiguración a la tarjeta de sonido para que permita grabar el audio que sale por las bocinas, muy parecido a utilizar el grabador de sonido de Windows.

Para el almacenamiento de los archivos en el disco duro, la aplicación crea una carpeta específica dentro del directorio raíz llamada Audio Texto a Voz (C:\Audio Texto a Voz\), donde se guardan los archivos de manera secuencial tomando el nombre del archivo de texto o de imágenes original. La aplicación verifica que no se sobrescriba el archivo y genera un número consecutivo que le anexa al nombre base en caso de que ya exista un archivo con el mismo nombre base, por lo que permite crear archivos secuenciales por ejemplo de los capítulos o temas de libros.

La interfaz de la aplicación con la que el usuario tiene interacción es parte fundamental ya que tiene las medidas necesarias para su rápida inicialización además

de impedir que se puedan cometer errores al no seguir el orden antes descrito. Dicha interfaz se muestra en la figura 6.

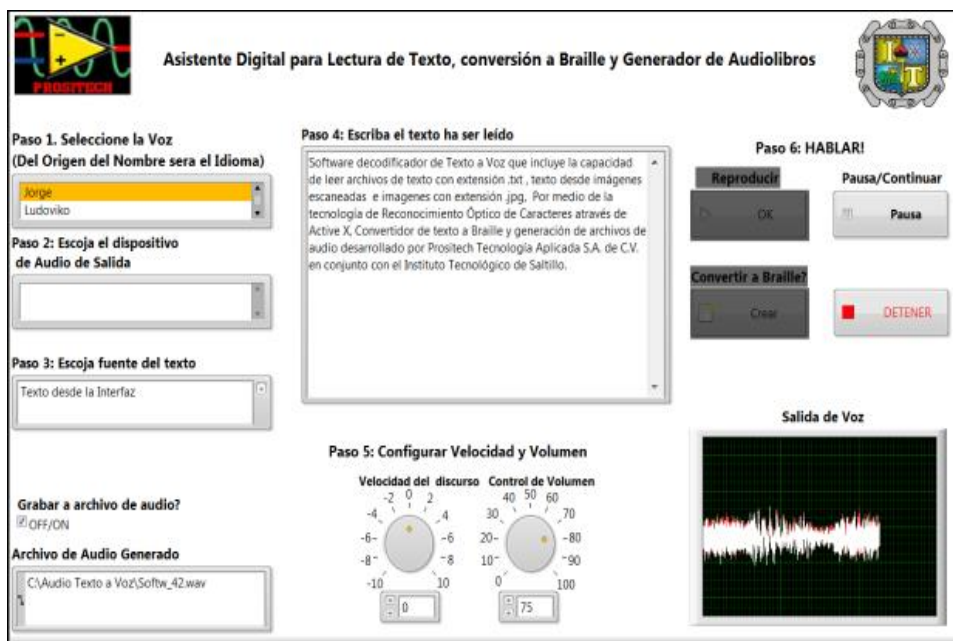


Fig. 6. Interfaz de Usuario

#### 4. Resultados

El resultado es una herramienta sumamente efectiva para leer documentos por medio de software, ya sea escaneados (imágenes) o directamente en texto, lo que permite que cualquier documento se convierta en un audiolibro con las ventajas que eso representa, en especial para las personas con debilidad visual, además es una aplicación poderosa para las personas con algún problema de habla ya que les permite comunicarse escribiendo en el teclado lo que quieren decir y la aplicación hablarlo por ellos, lo que permite que puedan comunicarse con personas que no entiendan lenguaje de señas que son la mayoría. La conversión a Código Braille amplía el acervo de documentos disponibles para personas con debilidad visual siendo prácticamente ilimitada la cantidad de información que tendrían disponible en dicho Código.

Esta herramienta informática ha sido probada con una muestra de diez personas con discapacidad visual y tres con discapacidad de lenguaje hablado. En todos los casos las personas han experimentado una mejora en su calidad de vida debido a que han podido comunicarse de una manera asertiva, a la vez de que han tenido acceso a grandes cantidades de información lo que les ha permitido una mayor integración con el resto de la población mejorando su calidad de vida. Asimismo, el software se

pondrá al alcance de La Asociación Mexicana de Enfermedades Visuales y Estudios de la Retina (**AMEVER**) en la ciudad de Saltillo, Coahuila para su uso generalizado y poder obtener información cualificada y cuantificada del impacto de esta herramienta en dicha población.

## **5. Conclusiones**

La implementación del desarrollo de Software para la solución de problemas específicos; en este caso, para ayuda a personas con problemas de visión o de habla ha resultado en una herramienta sumamente útil que colabora con el aumento en su calidad de vida.

Se desarrolló una herramienta de bajo costo que no requiere hardware especial para su funcionamiento, ya que funciona en cualquier PC bajo ambiente Windows y los textos pueden ser leídos desde diferentes fuentes incluyendo aquellos provenientes de cualquier escáner comercial. El impacto social que puede llegar a tener es muy alto ya que permite una comunicación más efectiva además de mayor interacción en el entorno de los usuarios.

## **Referencias**

1. National Instruments, "LabVIEW 2012 User Manual", February 2012 Edition.
2. National Instruments, "LabVIEW 2012 Connectivity User Manual", February 2012 Edition.
3. Mona Peña, Luis Javier, "Introducción a la Instrumentación Virtual", Manual de Curso, Instituto Tecnológico de Saltillo (2009)
4. Microsoft Corporation, "Microsoft Windows Software Development Kit (SDK)", Technical Notes.
5. Microsoft Corporation, "Microsoft Windows Microsoft Office Document Imaging (MODI) Technical Notes".