

# El Reconocimiento de Voz como alternativa de inclusión para discapacitados auditivos en un entorno educativo

Adriana Montoto, Eduardo Alvarez, Gabriel Chavira, Jorge Orozco, Gerardo Quiroga y Salvador W. Nava  
Facultad de Ingeniería “Arturo Narro Siller  
Universidad Autónoma de Tamaulipas  
Tampico, Tams.; México  
[amontoto, ccalvar, gchavira, jorozco, gquiroya, snava] @docentes.uat.edu.mx

**Abstract**— Estamos ante una sociedad donde el derecho a la igualdad es cada vez más proclamado por aquellos que sufren de discapacidad auditiva, la imparcialidad actúa incluso en el entorno educativo universitarios ya que por ley este tipo de estudiantes deben ser aceptados. No obstante, no existe todavía estándares, metodologías o herramientas capaces de compensar el hecho de no poder escuchar lo que ocurre en el entorno, por lo tanto, la dificultad en la interacción y en la transición de conocimientos difícilmente puede llevarse a cabo. La forma en que tradicionalmente se establece la comunicación entre profesor y estudiante no ayuda a la preparación de un alumno con discapacidad auditiva. De modo que si tratamos de convertir el discurso hablado a modo textual detonaría un cambio en la transición de conocimiento hacia los discapacitados auditivos, por lo tanto, procederemos a explorar las técnicas en el reconocimiento de la voz con la finalidad de realizar la evaluación en cada una de ellas y con esto determinar la más adecuada para incorporar en un desarrollo tecnológico que pueda trabajar como alternativa de inclusión.

**Palabras claves.** entorno educativo; discapacidad auditiva; reconocimiento de voz.

## I. INTRODUCCIÓN

El concepto de inclusión que frecuentemente es usado dentro de los sistemas de educación se enfoca principalmente en adaptar los estilos de aprendizaje a la manera de aprender y a las habilidades de los estudiantes, sin embargo una inclusión educativa abarca mucho más que eso, se constituye de un proceso vinculado a la transformación profunda de los sistemas educativos establecidos, de modo que en la educación formal (y no formal) se promueva y asegure la defensa de la dignidad e igualdad de todos los alumnos, como derechos inalienables, desde el reconocimiento y respeto de su diversidad [1]

Para lograr una inclusión educativa como tal, hace falta considerar diversos aspectos, entre ellos la adaptación del ambiente educativo con el fin de cubrir las necesidades que demandan los estudiantes y más aún si alguno de ellos cuenta con una discapacidad como es el caso de la auditiva la cual imposibilita a la persona en la percepción de sonido.

Las capacidades de comunicación en este tipo de estudiantes difieren mucho de las de un estudiante promedio, debido a que al carecer del sentido para escuchar su habilidad para hablar en algunas ocasiones también se ve afectada, por tal motivo el proceso enseñanza-aprendizaje se vuelve complicado.

Regularmente una persona que sufre discapacidad auditiva desde temprana edad cuenta con una formación que le permite comunicarse mediante movimientos en brazos, manos y dedos, a lo cual se le denomina como lenguaje de señas; pero que lamentablemente la sociedad no está capacitada para hablarlo ni para interpretarlo lo que da origen a una barrera en la comunicación y la falta de integración en diversos entornos.

Particularmente en un ambiente educativo donde los profesores, los materiales de clase y la infraestructura no cuentan con las herramientas adecuadas para hacer frente a este tipo de situación los estudiantes con discapacidad auditiva suelen quedar aislados en el proceso de enseñanza-aprendizaje ya que el habla es el

medio que predomina en la transmisión de conocimientos y en la interacción para relacionarse con los demás.

Por otro lado, hoy en día la tecnología aparece como una alternativa dedicada a mejorar la calidad de vida de las personas que cuentan con impedimentos físicos, brindándoles nuevas oportunidades en la interacción con los demás y con el entorno.

Por su parte la computación ubicua trabaja en este tipo de problemáticas integrando elementos informáticos que permitan el desarrollo de interfaces las cuales se adapten a los requerimientos y limitaciones que pueda presentar una persona, buscando siempre como objetivo compensar su condición al mismo tiempo que espera se vean implicado el menor esfuerzo durante el empleo de dichas herramientas tecnológicas.

Actualmente los avances tecnológicos permiten abordar las cuestiones implicadas dentro de la discapacidad auditiva desde diferentes perspectivas a través del uso de herramientas que incorporan potenciales técnicas en el reconocimiento de imágenes y de la voz adicionalmente estas requieren de un procesamiento avanzado [2]; sin embargo, el reconocimiento de la voz resulta una buena alternativa para abordar los inconvenientes que mencionamos anteriormente ya que hace posible la generación de información visual gracias a la conversión de voz a texto.

Por tal motivo es que basamos nuestra investigación en realizar una evaluación de las técnicas asociadas al reconocimiento de la voz con la finalidad de proponer la técnica más adecuada en el desarrollo de una aplicación que consiga la conversión del discurso del profesor a texto en apoyo a la inclusión de los estudiantes con discapacidad auditiva dentro del entorno educativo de las Universidades de la Zona Sur de Tamaulipas.

## II. RECONOCIMIENTO DE VOZ

El proceso de conversión de la voz ha dado paso al desarrollo de aportaciones tecnológicas verdaderamente relevantes en cuanto a traducción se refiere, este tipo de contribuciones se ven implementadas en apoyo a la inclusión de personas con discapacidades o limitaciones físicas, otorgando una nueva forma de interacción usuario-maquina.

El desarrollo de interfaces cada vez está más a favor de incluir el reconocimiento de voz para el manejo de aplicaciones comerciales que ponen en tendencia la manipulación de objetos y dispositivos que forman parte de nuestras actividades cotidianas.

Es importante mencionar que el reconocimiento de voz permite la traducción del discurso bajo dos modalidades: la conversión textual y la emisión de comandos de voz; derivando de ellas otras tareas que también se consideran sobresalientes para emplear sobre herramientas tecnológicas que tienen diferentes perspectivas en cuanto a su uso y que nos proveen de servicios amigables y de acceso para todos.

Las innovaciones tecnológicas están a la búsqueda de alcanzar una interacción más natural y siendo esta la verdadera contribución del reconocimiento de la voz actualmente, los portales con mayor popularidad en internet ya lo están incorporando, aprovechándose de los atributos que forman parte de esta herramienta.

La traducción gramatical de un idioma a otro es cada vez un proceso más sencillo, la interacción en la manipulación de aparatos electrónicos hasta la forma tradicional en la cual se opera un teléfono se ha vuelto una tarea más sencilla para aquellos que les es imposible hacerlo manualmente.

Google Translate es un ejemplo de un sistema fundamentado en una maquina traductora pero basado en texto que aplica técnicas de aprendizaje estadístico para construir el lenguaje y la traducción [3]. Por

su parte las plataformas Android, Microsoft y Apple operan dispositivos incorporando el reconocimiento de la voz para proveer determinados funcionamientos o búsquedas mediante la voz.

En referencia a cubrir en el ámbito de la discapacidad la Universidad de Santa Maria la cual está ubicada en Canadá lleva casi 20 años utilizando el reconocimiento de voz dentro del aula, con el objetivo de mejorar la experiencia de aprendizaje para los estudiantes que sufren de discapacidad auditiva, mediante la digitalización del discurso a texto el cual es desplegado sobre pantallas como una solución para las cátedras y las conferencias, utilizando lo anterior como alternativa en la toma de notas, beneficio que logra impactar en estudiantes que sufren de dificultades para realizar anotaciones [4].

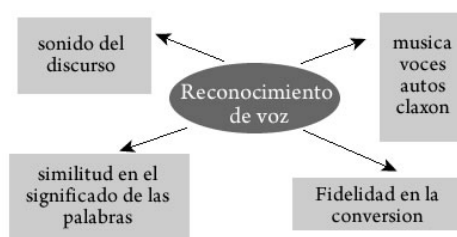
Sin lugar a duda los ejemplos anteriormente mencionados certifican que el reconocimiento de voz en cualquiera de sus modalidades tiene un rol importante en la inclusión de la población con limitaciones físicas, favoreciéndolos tanto en la manipulación de aplicaciones y elementos que forman parte del entorno pero sobre todo se establece como un nuevo paradigma en la interacción con los demás rompiendo barreras en la comunicación hasta el punto de dotarlos de una mayor independencia al sugerir soluciones tecnológicas transparentes acordes a los requerimientos sociales.

#### RECONOCIMIENTO DEL DISCURSO AUTOMÁTICO

Independientemente de la técnica a considerar para la generación del discurso automático es importante conocer que hay una serie de factores a contemplar para lograr el éxito esperado:

- 1.- La fonética en el discurso ya que en cada persona habla de un modo diferente
- 2.- Ruidos existentes en el ambiente
- 3.- Variantes en la conversión de las palabras, debido a la existencia de sinónimos
- 4.- La precisión en la conversión del discurso tomando en consideración los aspectos anteriormente mencionados

Los tratamientos de estos elementos conducirán a la conversión esperada, ya que estos constituyen parte fundamental en el funcionamiento de los sistemas que incorporan el reconocimiento de voz (véase Fig. 1).



*Fig. 1. Factores a considerar en la conversión de la voz*

No obstante, los elementos anteriormente mencionados suelen estar ya incorporados dentro de las maquinas traductoras que existen comercialmente en el mercado, en las plataformas de desarrollo o en su defecto sea necesario considerar su desarrollo o implementación por medio de modelos y algoritmos que pautaran los procesos a seguir para cada una de estas tareas a solucionar (véase Fig. 2); otro elemento que precisa el cumplimiento de un resultado de conversión optimo será contar siempre con los archivos indicados los cuales contienen el lenguaje con el idioma en el que de desea realizar la traducción

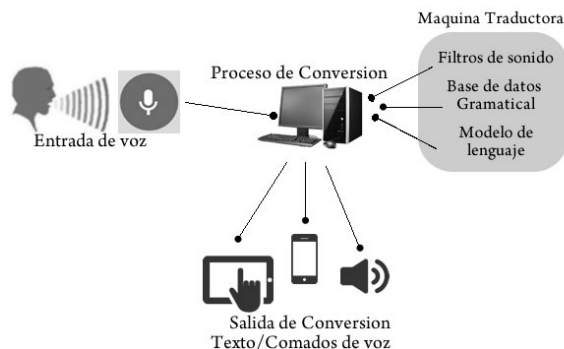


Fig. 2. Elementos que incorpora una máquina traductora.

Es importante saber que no solo considerar los factores que influyen en el proceso de conversión y la selección de la tecnología a aplicar en el desarrollo computacional podrán determinar el éxito en la conversión del discurso.

Sino que además es necesario tener en claro el tipo de solución se quiere ofrecer al usuario final, esperando siempre que esta sea la que mejor se adapte como solución a la problemática. Trabajar en los tipos de entradas y de salidas es otra tarea se suma relevancia ya que esto definirá el modelo ya sea el Léxico o el Acústico a implementar y junto a esto la decisión en cuanto a los dispositivos que soportaran la información (véase Fig 3).

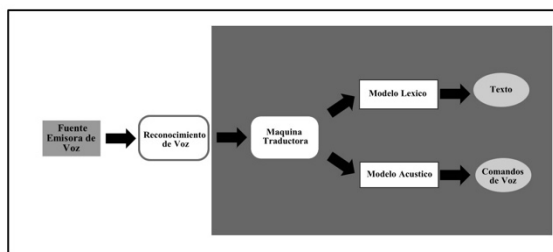


Fig. 3. Modelos a implementar de acuerdo al tipo de salida de información planificada.

### III. TÉCNICAS COMPUTACIONALES PARA EL RECONOCIMIENTO DE LA VOZ

Numerosas son las posibilidades en cuanto a la implementación de código de programación para el desarrollo de un sistema basado en el reconocimiento de la voz, pero al igual que en el resto de las tecnologías existen variantes las cuales indicaran ventajas y desventajas al incorporarlas como parte de nuestra codificación, no obstante el analizar cada una de ellas nos permitirá definir cuál de todas será la más conveniente no solo dados sus atributos sino por los conocimientos previos que demanda el utilizar cada una de ellas.

Queda claro que si bien estamos interesados en trabajar con el reconocimiento de voz como una herramienta que nos ayude en la inclusión de personas con discapacidad específicamente la auditiva, existen áreas que posiblemente hasta el día de hoy hayan sido inexploradas con anterioridad por nosotros mismos, sin embargo no es imposible gracias a los indicios que marcan los modelos y algoritmos que ya

han sido creados y probados por otros investigadores y que solo el trabajo y la dedicación a lograr nuestro objetivo serán lo que nos impulse a continuar nuestra propia investigación.

Además de examinar detalladamente otras obras cuyo propósito tal vez no es el mismo a nuestro trabajo pero si algo similar a lo que estamos buscando, por otro lado, los sistemas comerciales nos ofrecen un claro ejemplo de lo que se puede llegar a hacer bajo la incorporación de alguna de las técnicas en el reconocimiento de la voz puesto que sin lugar a duda más de una vez de manera consciente o inconsciente hemos manipulado alguna aplicación, objeto o dispositivo que tiene una interfaz que se opera a través de la voz o que simplemente nos devuelva subtítulos o texto a partir de ordenes o discursos que emitimos con la finalidad de otorgarnos un mayor confort.

Por lo que a continuación comenzaremos a detallar cada una de las técnicas para la implementación de un sistema reconocedor de la voz, acompañando cada una de estas técnicas de ejemplos que las incorporen y que dichos sistemas han servido para promover la inclusión de personas con discapacidad desde diferentes puntos de vista de cada una de las obras de investigación.

### MODELO OCULTO DE MÁRKOV (HMM)

Tras la transición que origino la aparición de la computación ubicua donde el uso de dispositivos computacionales no es solo a nivel de escritorio, sino que ahora están distribuidos en todas las áreas de nuestras vidas mediante conexiones de redes inalámbricas, se ha logrado una interacción desde cualquier lugar y momento [5]

Detrás de la aparición del concepto de computación ubicua fue necesaria la exploración de técnicas y modelos que dieran paso al surgimiento de interacciones más naturales, el uso del Modelo Oculto de Markov tiene cada vez mayor presencia en aplicaciones que impulsan la inclusión de personas con edad avanzada y con discapacidad destacando como componente verdaderamente esencial.

Un ejemplo es el desarrollo del sistema LSESpeak, un generador de español hablado en apoyo de personas sordas. Este sistema integra dos herramientas principales: un lenguaje de signos en el sistema de traducción de voz y un SMS (Servicio de Mensajes) en el sistema de traducción de voz.

La primera herramienta se compone de tres módulos: una interfaz visual (donde una persona sorda puede especificar una secuencia de signos), un traductor de idiomas (para generar la secuencia de palabras en español), y finalmente, un texto emocional a voz (TTS) para generar la conversión del español hablado.

La interfaz visual permite una secuencia señal que definirse utilizando varias utilidades. Los Conversor de TTS emocional se basan en modelos ocultos de Markov que permiten el género de voz, tipo de emoción, y la fuerza emocional para ser controlados [6]

La evaluación del modelo oculto de Markov también aparece como parte de la solución para asistir a personas con edad avanzada que presentan alguna emergencia dentro del hogar, la activación del sistema se realiza por medio del reconocimiento de voz, utilizando las señales de audio, el modo de trabajo de esta plataforma es por medio de la captura de llamadas de socorro por voz para seguido a esto comunicar a las unidades de emergencias.

Sin lugar a dudas el modelo oculto de Markov es aplicado para reconocer ciertos comportamientos mediante patrones para nuestro caso específicamente del habla veamos a continuación como es que actúa.

Queda claro que el habla se compone de un conjunto de palabras que están ordenadas de forma simultanea las cuales es posible almacenarlas dentro de una base de datos, paso seguido a esto es necesario un sistema que realice una decodificación de la voz y esto es posible gracias al análisis de cada vector que se tiene como entrada, posteriormente a esto se continuara con la debida extracción de los vectores, una vez

logrado esto en el proceso de la decodificación se obtendrá la mayor probabilidad de incidencia de cada una de las entradas lo que permitirá lograr una exactitud de los resultados [7].

En el grado de exactitud para la conversión de la voz es donde radica el éxito del modelo oculto de Markov y lo que lo ha posicionado como una de las técnicas más precisas a utilizar en diversas investigaciones si el objetivo principal es obtener los mejores resultados a partir del sonido del habla, por otro lado como ya lo habíamos mencionado antes la implementación de dicho modelo requiere de un mayor trabajo en cuanto a codificación se refiere, sin embargo si lo que un investigador está buscando es poder realizar la conversión de voz pero de una forma más sencilla definitivamente es preciso manipular otro tipo de técnica.

#### MOTORES DE RECONOCIMIENTO DE VOZ

Ante la necesidad de incorporar el reconocimiento de voz para el desarrollo de sistemas que permitan la inclusión de personas con discapacidad auditiva en el entorno educativo veamos una manera de realizarlo, pero de una manera automática sin implicar tanta complejidad en el tratamiento y síntesis del habla.

Esto es posible mediante los motores de reconocimiento de voz los cuales ya incorporan rutinas y subrutinas necesarias para alcanzar el propósito que planteamos dentro de nuestra investigación, simplemente optaremos por evaluar principalmente sobre qué tipo de plataforma deseamos poner en marcha la aplicación que estamos por desarrollar.

Actualmente Google, Apple y Microsoft nos ofrecen motores para reconocimiento de voz adaptables al propósito en el cual estamos interesados, por otro lado, la empresa Nuncce también nos da la oportunidad de usar su motor pero implicando un costo de por medio que es bastante considerable el cual tendríamos que someter a evaluación si es posible obtener de acuerdo al presupuesto con el que contemos.

#### SAPI

La empresa Microsoft siempre interesada en proveer de herramientas que den soporte al desarrollo de código de programación, crea un conjunto de interfaces capaces de facilitar las tareas de aquellos programadores que tienen inclinación por la implementación de aplicaciones para ejecutarse sobre su sistema operativo Windows.

Un ejemplo de ello es SAPI (Interfaz de programación de aplicaciones de voz) cuyo objetivo principal es producir de forma automática el reconocimiento y síntesis de la voz desde tu propia aplicación pudiendo acceder a ello desde diversos lenguajes de programación [8]

Lo que se podría verse como limitante por parte de la interfaz SAPI es que su servicio solo aplica para software dirigido a equipos de escritorio, por otro lado, lo que podríamos ver como un beneficio es que es de libre distribución y que logra una conversión del discurso a texto.

Un aspecto que adquiere relevancia son los requerimientos tecnológicos que demandan el desarrollo de interfaces orientadas a convertir la voz con ayuda de una API como SAPI, entre estos es necesario contar con un dispositivo que capture la voz, regularmente usando uno o varios micrófonos que estén conectados permanentemente a una computadora personal, la cual se encargara de procesar la conversión del discurso capturado a través del sistema, el cual integra un motor de reconocimiento de voz conocido como API (Interfaz de Programación de Aplicaciones), y por ultimo un dispositivo de salida que reúna las características que solicite el tipo de conversión resultante.

SAPI resulta un motor de conversión interesante para experimentar a través de él los distintos tipos de conversión del habla, es por eso que investigaciones alrededor del mundo están trabajando en ello en referencia el Departamento de Ciencias de la Computación en Bangladesh interesados en convertir la voz a texto trabajan sobre esta API con el objetivo de buscar una conversión lo bastante precisa para la lengua Bangla, encontrado en SAPI una herramienta lo suficientemente potencial para lograrlo.

Prueba de ello es el estudio experimental en el cual someten a SAPI mediante una conversión de discurso continuo a texto con la finalidad de plasmar el resultado en un documento el cual finalmente es producido como un artículo de periódico, esto muestra una aproximación del 78% de precisión final [9]. Lo que nos demuestra que solo es necesario trabajar un poco más en la parte gramatical para lograr la coincidencia planificada lo que aumentara el porcentaje de coincidencia y por tanto en nuestra búsqueda de soluciones a personas con discapacidad auditiva SAPI cumple con los elementos necesarios para ubicarla como una buena alternativa a nuestro propósito.

#### *RECONOCIMIENTO DE VOZ EN IOS*

Ante el desafío de estar en tendencia dentro del mercado para el desarrollo de software, Apple ofrece a los desarrolladores una API capaz de traducción de forma automática la voz de una persona pero que además está habilitada para interactuar por medio de la voz con los dispositivos propios de Apple incluyendo la administración de las aplicaciones incorporadas en el sistema operativo iOS; lo que ha permitido una mejora en la búsqueda de información mediante la asistencia de lo que conocemos como Siri.

Su API de reconocimiento de voz nos ofrece la capacidad de realizar la traducción de la voz transcribiéndola a texto, dicha traducción se realiza de acuerdo al idioma configurado en el sistema operativo, sus beneficios principales son el soporte de audio pregrabado, una alta precisión de traducción gracias a sus más de 50 idiomas y entre sus inconvenientes están la limitación de traducciones al día, la necesidad de activar la conexión a la red y la restricción en la entrada de audio permitiendo solo un minuto como máximo. [8]

El reconocimiento de voz para aplicaciones de iOS opta por una nueva experiencia en cuanto a manipulación de sus aplicaciones, sin embargo no hay que dejar de lado el hecho que para poder hacer uso de este tipo de interfaz es necesario activar la entrada de voz mediante el servicio correspondiente para otorgándole los permisos necesarios, en caso contrario permanecerá desactivado; además otro punto importante a considerar son los posibles errores en la conversión a texto ocasionados por las fallas en la conexión a red ya que esto imposibilita la conexión a los servidores de Apple los cuales gestionan la transferencia del texto en el idioma correspondiente [9].

Con el paso del tiempo los motores de voz han sufrido cambios con el fin de proveer un mejor soporte creando plataformas comerciales ya sea para desarrolladores expertos y para los que no lo son tanto, si bien existe una diferencia en el modo de implementarlo el objetivo sigue siendo el mismo que es lograr sintetizar la voz y poderla convertir a texto.

El funcionamiento de la API de Apple todavía muy lejos de catalogarse como la mejor interfaz en cuanto a voz se refiere, existen investigaciones que apuntan que existen mejores reconocedores de voz en el mercado puesto que el motor de Apple es carente de precisión debido a que no logra controlar la pronunciación o lo hace de una manera ineficiente lo que origina alteración en el texto [12]

#### *GOOGLE CLOUD SPEECH*

Google en una iniciativa por prevalecer como el sistema con mayor presencia entre los dispositivos portables otorga a los desarrolladores las directrices necesarias para implementar el reconocimiento de voz a través de su API en la producción de nuevas aplicaciones.

La API Speech de Google Cloud facilita la traducción del discurso a texto independientemente si la voz se está produciendo justo en el momento o si previamente se almacena en la nube, al incorporar esta API en nuestra propia codificación facilitara el proceso de codificación ya que esta API ya añade potenciales modelos de redes neuronales los cuales permiten la excelente captura de la voz evitando con ello ruidos externos lo que producirá una alta precisión en la conversión del habla; además su gran gama de

vocabulario hace de esta API una herramienta idónea para la conversión a texto ya que dispone de más de 80 idiomas para el reconocimiento y traducción de la voz [13]

Queda claro que en la producción de motores de voz podemos encontrar comercialmente una diversidad lo que marcará la pauta para decidir con que producto trabajar será aquel que más se apegue a nuestras necesidades, es por esta razón que existen múltiples artículos dirigidos a debatir cuál de ellos es el mejor.

Tal es el caso de la comparativa entre Sphinx que no es más que un reconocedor de código abierto y la API de Google esperando determinar cuál de ellos es mejor en cuanto a pruebas de rendimiento, se demuestra en una prueba para reconocer oraciones de un sistema de información para autobuses que Sphinx supera los ciclos de optimización del API de Google [14]

Independientemente a esto debemos saber que cada motor tiene su propósito específico, no hay que dejar de mencionar que si bien la API para reconocer la voz de Google es menos productiva también es la más óptima para contextos dentro de la web cosa que para otras interfaces pudiera parecer más complicada su adaptación.

#### *DRAGON NATURALLY SPEAKING*

Si nuestro objetivo primordial es producir una aplicación con la capacidad de mostrar el texto resultante del discurso continuo olvidándonos de los detalles en cuanto a reconocimiento de voz se refiere, la API de Dragon NaturallySpeaking es la solución óptima a nuestras necesidades.

La empresa Nuance líder en ventas ofrece por medio de sus productos la posibilidad de acceder a diversos recursos dirigidos al reconocimiento de la voz con un costo implícito dependiendo del alcance de cada una de ellas.

A diferencia de otras interfaces la capacidad que encontramos en Dragon NaturallySpeaking refiere a una mayor potencialidad esto es resultado de los años de trabajo en el tratamiento y síntesis de la voz por parte de Nuance, por otro lado, para poder trabajar con ella el presupuesto del desarrollador será la limitante para poder hacerlo.

La finalidad de Dragon NaturallySpeaking es aumentar la productividad dentro del sector empresarial al trabajar directamente desde una computadora personal por medio del dictado lo que representa una mayor rapidez que hacerlo del modo tradicional con teclado [15]

El reconocimiento de voz abarca una gran gama de perspectivas para abordar incluidos los aspectos de la salud donde los Departamentos de Radiología y Medicina en Hamilton Health Sciences los cuales están ubicados en Ontario Canadá ponen a prueba la eficiencia de Dragon Naturally Speaking para que los médicos que asisten en esta especialidad lo evaluaran y lo compararan con el proceso de dictado convencional [16]

#### IV. CONCLUSIONES

Ahondando en artículos de investigación tecnológica podemos encontrar las distintas maneras en la que se abordan problemáticas similares a la que planteamos, donde el interés principal radica en la búsqueda de soluciones tecnológicas que compensen la discapacidad auditiva dentro de instituciones de educación tradicional.

No obstante, la conversión de forma automática del discurso a modo textual se posiciona como una buena alternativa para poder lograr una formación profesional como la del resto de la población estudiantil, si bien en nuestra investigación hemos estudiado cada una de las posibles técnicas más populares y potenciales que existen actualmente queda en cada desarrollador optar por la que mejor convenga.



Desde nuestro punto de vista SAPI se posiciona como una técnica suficientemente potencial para cubrir las necesidades que se presentan en un entorno educativo el cual comúnmente es un ambiente cerrado y menos propenso a la gran mayoría de ruido que existen en el exterior.

El usar SAPI no implica de costos adicionales para su referencia además de esto el equipamiento que existe comúnmente ya se encuentra ubicado dentro de los salones de clases universitarios, recurrentemente en otras obras podemos observar que esta técnica es sometida a la experimentación lo que denota resultados aceptables y con el simple hecho de trabajar en el enriquecimiento de la gramática y el entrenamiento del hablante será los dos elementos a trabajar con mayor énfasis.

Otra de las ventajas es que SAPI ya incorpora la base de datos con vocabulario en español por lo que tampoco será necesario trabajar en ello, sin embargo la decisión final en la elección de la técnica para el manejo del reconocimiento de la voz dentro de una aplicación propia se basará en aspectos que parten de las habilidades y conocimientos de programación, el nivel de familiarización con la plataforma de desarrollo, los dispositivos tecnológicos disponibles y sobre todo el alcance económico para poder llevar en marcha el proyecto planeado.

De tal forma que una vez decidida la técnica esta debe de cumplir con los requerimientos básicos para cubrir las necesidades que demanda la presencia de un discapacitado auditivo dentro de un entorno educativo.

Si bien no existen estándares o técnicas formalizadas para el proceso de inclusión de un discapacitado auditivo en algunas de las instituciones universitarias de la zona sur de Tamaulipas la propuesta de utilizar el reconocimiento de la voz como una alternativa para la conversión del discurso a texto de manera automática resulta bastante atractivo a modo de inclusión ya que en nuestra investigación describimos las ventajas y limitaciones que brindan cada una de las técnicas el resto para lograr la construcción de una herramienta tecnológica potencial dependerá del grupo de investigadores interesados en llevarla a cabo como una solución a la problemática planteada.

Lejos de saber que por ley a ninguna persona con discapacidad le puede ser negado el derecho a la educación, actualmente no se cuenta con ningún sistema automatizado que permita la comunicación de un discapacitado auditivo con alumnos y profesores dentro de las universidades del sur de Tamaulipas sin embargo existen actualmente opciones tecnológicas significativas a vida de compensar la falta de capacidad para escuchar lo único que hace falta para lograrlo sería fomentar a través de las instituciones de educación un mayor compromiso social mediante financiamientos innovaciones tecnológicas para lograr una inclusión integral para todos

## REFERENCIAS

- [1] X. Etxebarria, «Ética y discapacidad,» *Revista Siglo Cero*, vol. 35, nº 2, pp. 210:68-79, 2014.
- [2] L. S. R. Rabiner, «Introduction to Digital Speech Processing,» *Foundations and Trends in Signal Processing*, p. 1–194 (2007), 2007.
- [3] R. P. F. C. a. F. L. T. Duarte, «Speech Recognition for Voice-Based Machine Translation,» *IEEE Software*, vol. 31, nº no. 1, pp. 26-31, 2014.
- [4] S. H. B. a. M. W. Keith Bain, «Speech recognition in university classrooms: liberated learning project.,» *n Proceedings of the fifth international ACM conference on Assistive technologies (Assets '02)*. ACM, New York, NY, USA, pp. 192-196., 2002.
- [5] M. T. West., «Ubiquitous computing,» *In Proceedings of the 39th annual ACM SIGUCCS conference on User services (SIGUCCS '11)*. ACM, New York, NY, USA, . , pp. 175-182, 2011.

- [6] R. B.-C. S. L. J. M. M. R. S.-S. Verónica López-Ludeña, «LSESpeak: A spoken language generator for Deaf people,» , *Expert Systems with Applications* , vol. 40, n° 4, pp. 1283-1295, 2013.
- [7] B. N. Tony Ayres, «Voice activated command and control with speech recognition over WiFi,» *Science of Computer Programming*, vol. 59, n° 1, pp. 109-126, 2006.
- [8] BIBLPROG, «Speech API,» 2005. [En línea]. Available: [https://biblprog.com/en/speech\\_api/](https://biblprog.com/en/speech_api/). [Último acceso: 17 08 2017].
- [9] M. A. H. A. P. K. D. a. M. M. H. R. S. Sultana, «Bangla Speech-to-Text conversion using SAPI,» *2012 International Conference on Computer and Communication Engineering (ICCCCE), Kuala Lumpur* , pp. 385-390., 2012.
- [10] ENLUME, «How to use iOS 10 Speech Recognition API to convert Voice to Text,» 2016. [En línea]. Available: <http://www.enlume.com/ios-10-speech-recognition-api-convert-voice-text/>. [Último acceso: 17 08 2017].
- [11] F. Speech, «APPLE DEVELOPER,» 2017. [En línea]. Available: <https://developer.apple.com/documentation/speech>. [Último acceso: 17 AGOSTO 2017].
- [12] A. Lampert, «Text-to-Speech Markup Languages,» 2004.
- [13] A. S. D. C. GOOGLE, «GOOGLE CLOUD PLATFORM,» 2017. [En línea]. Available: <https://cloud.google.com/speech/?hl=es>.
- [14] P. & S.-O. D. Lange, «Tuning Sphinx to Outperform Google's Speech Recognition API,» *Proc. of the ESSV 2014, Conference on Electronic Speech Signal Processing* , pp. 1-10, 2014.
- [15] NUANCE, «Software de reconocimiento de voz Dragon,» NUANCE , 2017 08 2017. [En línea]. Available: <http://www.nuance.es/dragon/index.htm>.
- [16] R. M. & J. I. H. Issenman, «Use of voice recognition software in an outpatient pediatric specialty practice.,» *Pediatrics*, vol. 114, n° 3, pp. 290-293., 2004.
- [17] C. Folke, «Respetar los límites del planeta y recuperar la conexión con la biosfera,» de *¿Es aún posible lograr la sostenibilidad? La situación del mundo 2013. Informe anual del Worldwatch Institute sobre la Sostenibilidad*, Madrid, FUHEM Ecosocial/Icaria Editorial, 2013, pp. 51-62.
- [18] S. S. R. B. G. F. F. P. Emanuele Principi, «An integrated system for voice command recognition and emergency detection based on audio signals,» *Expert Systems with Applications*, vol. 42, n° 13, pp. 5668-5683, 2015.