

1. Introducción

En el siglo XIX la tecnología sobre reconocimiento de voz tenía sus destellos iniciales, primero existieron dispositivos que intentaban producir voz imitando cada uno de los órganos vocales. Posteriormente con los avances de la electrónica y el desarrollo de tecnología en el área computacional, daban paso a una nueva era que permitiría aplicar ese conocimiento en el campo de reconocimiento de voz.

No fue sino hasta finales de 1960 cuando se tuvieron las bases para construir sistemas de reconocimiento de voz basados en nuevos procedimientos de aprendizaje como modelos ocultos de Markov [1,2] o redes neuronales. Los fundamentos de los conceptos anteriores están establecidos en el reconocimiento de patrones estadísticos y este a su vez en la teoría de Bayes. El resultado de la investigación en este campo la podemos ver desde el reconocimiento automático de manuscritos, ópticos, voz, hasta el diagnóstico médico a través de máquinas.

1.1 Definición del Problema

Durante la década pasada y principios de esta, los sistemas que utilizan reconocimiento de voz han tenido un mayor auge en sistemas de telefonía, educación, entretenimiento y militares. En cada uno de los campos mencionados han existido problemas y nuevos retos al utilizarlos en tiempo real. Algunos de estos problemas son: ruido (en el ambiente, señal), variabilidad temporal y espectral dependiente de la edad del

locutor, entre otros. La mayoría de la investigación se ha situado en sistemas que interactúan con adultos. Para enfrentar algunos de estos problemas se han investigado algoritmos o formas que minimicen el error en los reconocedores automáticos de voz y mejoren su desempeño [3].

Normalmente el reconocimiento automático de voz no había sido tan utilizado en los sistemas orientados a la educación y menos hacia niños debido a la falta de investigación en esta área. El laboratorio de tecnologías de voz TLATOA de la Universidad de las Américas-Puebla en conjunto con el CSLR* de la Universidad de Colorado en Boulder han trabajado en la versión en español de tutores inteligentes y libros interactivos para mejorar la lecto-escritura en escuelas públicas [4], los cuales utilizan tecnologías de reconocimiento de voz para niños.

Se requiere desarrollar un reconocedor de voz de niños que sea utilizado en este tipo de sistemas. Para lograrlo se debe escoger un conjunto de herramientas que nos permitan investigar y desarrollar reconocedores de voz utilizando técnicas para mejorar el desempeño en la tarea de reconocimiento de voz.

Actualmente no existe un reconocedor de voz de niños para el español hablado en México. Sistemas de lecto-escritura como los tutores y libros interactivos necesitan modelos de reconocimiento de voz para llevar a cabo algunas tareas como, dar seguimiento de la posición y ubicación de lectura de un niño, asistir detectando o leyendo errores, y para proveer información la cual puede ser usada para medir la fluidez de la lectura.

* *Center for Spoken Language Research*

1.2 Objetivos

1.2.1 Objetivos Generales

Los objetivos generales de esta investigación son:

- Desarrollar un reconocedor de voz que sea robusto.
- Presentar técnicas viables que ayuden a mejorar el reconocimiento de voz en niños en sistemas de lecto-escritura.

1.2.2 Objetivos Específicos

- Construcción de un reconocedor automático de voz de niños usando SONIC.
- Buscar y aplicar algunos métodos ya conocidos para mejorar el desempeño del reconocedor.
- Evaluación del reconocedor con los diferentes métodos investigados.
- Presentar resultados de cada uno de los experimentos con algún tipo de métrica estándar para el reconocimiento de voz.

1.3 Organización del Documento

El documento consiste en 6 capítulos. En el primer capítulo presentamos la definición del problema y los objetivos de esta tesis, además de listar los requerimientos.

En el capítulo 2 se da una breve explicación del problema de reconocimiento de voz que se

ve como un problema de reconocimiento de patrones, y la formulación básica para entender dicho problema, así como los conceptos que lo envuelven. En el inicio del capítulo se presentan las unidades básicas con las que trabaja un reconocedor y los problemas para su modelado. Las siguientes tres partes son conceptos matemáticos importantes, pues a partir de ellos se deriva lo que es reconocimiento de voz. En la última parte del capítulo se habla sobre cada uno de los elementos que componen a un reconocedor de voz y como es que se da todo el proceso de reconocimiento.

La evaluación de un reconocedor es un proceso importante, ya que de esta manera se puede verificar si los modelos obtenidos responden a variación ambiental y entre locutores. Para mejorar el desempeño de un reconocedor, existen técnicas que atacan estos problemas minimizando los errores que producen. En el capítulo 3 de esta tesis se plantean tres técnicas que ayudarán a mejorar dicho desempeño. En específico, se aplicará la normalización de la longitud del tracto vocal, regresiones lineales de probabilidad máxima y regresiones lineales de probabilidad máxima *a posteriori*.

En el capítulo 4 se describe el proceso para obtener los modelos acústicos y los modelos del lenguaje necesarios para evaluar el reconocedor, además de los detalles técnicos para el uso de nuestra herramienta de reconocimiento que es SONIC.

En el capítulo 5 se mencionan algunos parámetros de evaluación como la base de datos de voz, usada en cada experimento y sus detalles. Se presentan tablas indicando el desempeño de los modelos obtenidos y las ganancias al aplicar cada una de las técnicas de adaptación y normalización mencionadas en el capítulo 3.

Las conclusiones y trabajo a futuro se presentan en el último capítulo.