

6. Conclusiones

En esta tesis se ha construido un nuevo reconocedor de voz de niños utilizando SONIC. Además, se ha mejorado el nivel de reconocimiento del experimento base, aplicando métodos de adaptación y normalización.

Las limitaciones de este trabajo dependieron de la disponibilidad y calidad del corpus de voces de niños (problemas con las transcripciones o errores en la grabación) [5], de acuerdo a este último punto se decidió el porcentaje de locutores a utilizar para el entrenamiento y evaluación de los experimentos. Debido a la complejidad de algunas de las técnicas para mejorar reconocimiento de voz sólo se utilizaron algunas para nuestros experimentos.

Para estos experimentos, podemos decir que implementando enfoques como agrupamiento de clases nos puede llevar a ciertas mejoras en el desempeño. Se ha investigado y corroborado también el uso de la normalización de la longitud del tracto vocal (VTLN) en niños, obteniendo un porcentaje de error por palabra de 42.6% y una reducción del error de 8.76% respecto al experimento base. Para el caso de adaptación se han investigado dos métodos, el primero utiliza la estimación de las matrices para maximizar la probabilidad de los modelos transformados generando los datos de adaptación. Y una técnica de adaptación de los modelos ocultos de Markov usando regresiones lineales basadas en la estimación de la probabilidad máxima *a posteriori*.

En los experimentos de las regresiones lineales de probabilidad máxima (MLLR) se han obtenido mejores resultados que en la evaluación de los modelos base y los modelos

normalizados con VTLN. El mejor porcentaje de este experimento fue de 41.2% y la reducción del error relativo con respecto a los modelos base en este experimento fue de 10.68%.

Para los experimentos desarrollados con la técnica de regresiones lineales de probabilidad máxima *a posteriori* (MAPLR), los resultados han sido los mejores en el desempeño general de los modelos desarrollados para esta tesis. La idea principal de este enfoque es restringir los valores de los parámetros de transformación usando una distribución previa, obteniendo así un porcentaje de error por palabra de 31.9%, mostrando una reducción del error relativo de 22.94% respecto a los experimentos base.

Las principales contribuciones de esta tesis son:

- El desarrollo de un reconocedor de voz de niños utilizando SONIC.
- Reducción de los porcentajes de error de palabra en cada uno de los experimentos con respecto a cada uno de los experimentos base, utilizando técnicas de adaptación y normalización para un reconocimiento robusto de voz usando SONIC.

6.1 Discusión

Existen muchas técnicas empleadas en cada uno de los módulos de la arquitectura de un sistema de reconocimiento de voz. Algunas de ellas tienen un costo computacional alto pero son redituables al momento de aplicarlas en el mundo real. Un mejor entendimiento de los problemas encontrados en los corpus y la aplicación de las técnicas exactas para disminuir estos errores nos llevarán a mejores resultados de evaluación. A

pesar de que hoy en día ningún reconocedor puede estar al nivel humano, la capacidad de cómputo puede ser de gran ayuda para que la interacción humano-computadora sea natural. En el caso particular de México, es importante analizar el contexto socio-cultural de la población objetivo al construir un reconocedor de voz, ya que la riqueza lingüística de nuestro país, además de los problemas fonéticos y fonológicos de los locutores hacen que el desempeño del reconocimiento de voz disminuya.

6.2 Trabajo a Futuro

El análisis previo de un corpus es esencial para aplicar técnicas de mejoramiento en la evaluación de los mismos. Se propone desarrollar un estudio del corpus de voz de niños donde reflejen los tipos de variabilidad y problemas lingüísticos existentes para desarrollar y aplicar técnicas actuales que disminuyan estas diferencias.

Existen otras técnicas que pueden ser aplicadas al reconocimiento de voz de niños y que mejorarían el desempeño del reconocedor. Los resultados en otras investigaciones pueden servir como guía para emplear esas técnicas, diferentes a las ya descritas en esta tesis.

Actualmente el reconocimiento de voz en diferentes ambientes acústicos ha incrementado su desempeño debido a estas técnicas, pero uno de los mayores problemas para mejorar la calidad de los reconocedores es la falta de datos, esto es, recolectar corpus que contengan variabilidades ambientales y entre locutores. Algunas de estos corpus pueden ser, grabaciones provenientes de celular con voces de niños o adultos de habla hispana (México).