

Introducción

Reconocimiento fonético

Existen varias formas para diseñar un sistema de reconocimiento de voz, y una de ellas es el enfoque basado en los fonemas. Este método está basado en la teoría acústica fonética, la cual postula que existen unidades fonéticas, finitas y distintivas, que se caracterizan por un conjunto de propiedades manifestadas en la señal de voz o espectro, a lo largo del tiempo [Rabiner93].

Características del reconocimiento fonético

Se puede considerar un sistema de reconocimiento de voz basado en fonemas como una caja negra que toma como entrada la señal de voz y produce como salida una secuencia de fonemas. La meta a largo plazo del reconocimiento de voz basado en fonemas es que este tipo de reconocedores sean independientes del hablante, del vocabulario y puedan ser usados para diferentes tareas en varios lenguajes [Lamel93].

El reconocimiento de voz a nivel de fonemas se encuentra muy relacionado con reconocimiento a nivel palabras. Estn relacionados en el sentido que un reconocedor de palabras utiliza un reconocedor de fonemas como base. Sin embargo, éstos reconocedores difieren en la estructura del lenguaje que se quiere reconocer, palabras o fonemas. Un reconocedor fonético tiene la función de reconocer pero sin la influencia de restricciones léxicas y sintácticas. La gran mayoría de los sistemas modernos de reconocimiento de voz está basada en el reconocimiento fonético.

Planteamiento del problema

Existen muchas razones por las cuales es importante evaluar el reconocimiento fonético. Primero, las demandas de independencia de vocabulario y del hablante requieren de un examen basado en fonemas como unidades. Si los modelos fonéticos son mejores, el sistema entero tendrá un mejor desempeño en cualquier aplicación y con cualquier vocabulario. Si solo consideramos el desempeño de reconocimiento a nivel palabras, probablemente no se vean problemas que si se verían a nivel fonético, y además no se puede extrapolar los resultados a otros dominios.

Para el caso del idioma español, y más específico, el español hablado en México, no existe una investigación previa que nos pueda decir, “el desempeño del reconocedor fonético de tal laboratorio tiene una precisión de X porciento”. De ahí surge este proyecto, que tiene entre sus finalidades la evaluación del nivel de desempeño de reconocimiento, la cual puede servir como punto de partida para futuras investigaciones en reconocimiento fonético del lenguaje español.

Es muy aventurado emitir un juicio acerca del desempeño de un reconocedor sin correr experimentos controlados sobre grandes cantidades de locutores; es decir, no se sabe con certeza si el sistema realmente está cumpliendo conforme a lo esperado. Por otro lado, si se tiene conocimiento del porcentaje de eficiencia de un segundo reconocedor probado con un corpus específico, es posible hacer una comparación con los resultados obtenidos por el primero en base al mismo corpus. Estas técnicas se usan mucho en otros lenguajes para evaluar el desempeño de reconocedores.

Objetivo general

La finalidad de esta tesis es desarrollar y evaluar un reconocedor fonético para el español hablado en México.

Objetivos específicos

- Definir el corpus con el que se va a trabajar, el cual debe estar balanceado fonéticamente.
- Construir y entrenar el reconocedor.
- Evaluar el desempeño del reconocedor sobre cada avance logrado.
- Mejorar el reconocedor utilizando técnicas dependientes del contexto.
- Formular conclusiones acerca del desempeño obtenido por el reconocedor fonético en base a un corpus.
- Publicar el corpus y los detalles de los experimentos para que otros grupos educativos puedan comparar sus resultados.

Esta tesis se ha organizado en cinco capítulos. Los temas de los capítulos se describen a continuación.

En el capítulo I se da una breve introducción a los sistemas de reconocimiento de voz, los antecedentes, su arquitectura y principales características. Por ultimo, se mencionan algunas de las principales aplicaciones.

El capítulo II trata sobre el reconocimiento de voz usando redes neuronales en el CSLU *Toolkit*. El corpus de voz con el que se trabajó en esta tesis se explica en el

capítulo III, donde además se describen las condiciones de grabación, el contenido y archivos que lo conforman.

El capítulo IV corresponde a la descripción del desarrollo del reconocedor. En esta sección se describe la metodología seguida para obtener como resultado un reconocedor fonético. Este reconocedor es el caso base hecho en base a modelos independientes del contexto.

En el capítulo V se muestran los detalles de las mejoras realizadas al reconocedor presentado en el capítulo IV. Algunas de estas mejoras fueron usar modelos dependientes del contexto e incremento de las duraciones mínimas de cada fonema. También se presenta la prueba final realizada sobre un conjunto de *stories* y los resultados obtenidos.

Finalmente se comentan los resultados, el estado actual, las perspectivas y limitaciones del proyecto desarrollado.