Capítulo 6.

Conclusiones

Las aplicaciones CALL con reconocimiento de voz han comenzando a despertar el interés de muchas personas. Diferentes laboratorios han desarrollado aplicaciones de este tipo, los resultados muestran sistemas prototipos con interacción de voz para la enseñanza de pronunciación, lectura y la práctica de las habilidades en conversaciones. En la mayoría el procedimiento consiste en construir modelos nativos de la pronunciación y después comparar los resultados obtenidos por los extranjeros contra los de los modelos nativos, luego se calcula un puntaje en base a la estadística derivada de comparar los valores extranjeros para estas variables a los modelos nativos. Y finalmente, la puntuación generada es validada comparando estos puntos con el juicio humano. Los modelos son entrenados con datos de voz nativos y no nativos. Muchos utilizan representaciones gráficas y dan retroalimentación al locutor en base al puntaje obtenido en su pronunciación. Hasta ahora este tipo de sistemas se han enfocado a el idioma Inglés, Alemán, Japonés y Holandés [Akhane-Yamada, et al., 96; Bernstein & Christian, 96; Cucchiarini et al., 99; Eskezani, 99; Franco et al., 97; Hiller et al., 94; Kim, et al., 97; Korkmazskiy, et al., 98].

El objetivo de este proyecto era desarrollar una herramienta de verificación de pronunciación para un ambiente de aprendizaje de Español Mexicano, que permitiera que estudiantes extranjeros con idioma inglés estadounidense pudieran aprender el idioma pronunciando palabras o frases comunes del español usado en México y sirviera además como base para futuras investigaciones.

El conjunto de fonemas usados en el Español Mexicano es prácticamente un subconjunto de los fonemas usados en Inglés, por lo tanto, un reconocedor entrenado para el Español Mexicano no podrá detectar los fonemas fuera de éste conjunto. Entonces, un punto importante es el de entrenar el modelo con datos de voz nativos y no nativos, ya que de esta manera podremos identificar los errores que el estudiante tenga, lo que se busca es poder generalizar a partir de las múltiples pronunciaciones del mismo fonema. Mientras más pronunciaciones alternativas se incorporen en un sistema, el reconocimiento de la voz podrá mejorar potencialmente en cuanto a su exactitud. Por lo tanto, se considera práctica recolectar un corpus extenso que represente las pronunciaciones de los fonemas en un idioma determinado, para entrenar un reconocedor.

En un primer experimento se utilizó una red neuronal entrenada para el reconocimiento de voz del Español Mexicano. La gramática permitía cualquier combinación de letras. El problema que surgió fue que el

reconocedor tenía que ser muy exacto, preferiblemente dependiente del locutor y utilizado en un ambiente libre de ruido, evitando, de igual forma, el ruido de respiración en el micrófono, de otro modo se insertaba una gran cantidad de basura entre los fonemas reconocidos.

El restringir la gramática sólo para las palabras necesarias para la aplicación puede causar que el sistema rechace las palabras mal pronunciadas o el vocabulario tiene que incluir una lista extensiva con muchas palabras y todas sus posibles pronunciaciones. Esto incrementa el espacio de búsqueda en gran medida, causando así un incremento en el porcentaje de error, especialmente un alto número de errores de eliminación de fonemas.

Esto nos llevó a crear un vocabulario de letras, no palabras, y aquí cada letra incluye todas las posibles pronunciaciones que el estudiante, tanto en español como en inglés. Aun así la red neuronal puede confundirse fácilmente e insertar una gran cantidad de fonemas adicionales o basura. Usando algunas ideas de la técnica *de forced aligment*, el proceso de verificación genera únicamente una gramática específica para la palabra o frase a ser analizada, la cual se supone deberá pronunciar el estudiante

El desempeño del reconocedor es bueno, pero tiene muchos problemas en reconocer algunos fonemas con problemas como la 'r', la 'o' y la 'z'. Esto se debe a que no existieron muestras suficientes en el proceso de entrenamiento. Muchas veces el reconocedor marca errores que el juicio humano no vio. Como se mencionó, el humano no se percata de los errores de pronunciación al hablar, incluso, dichos errores forman parte del habla.

Aportaciones

Se demostró que una red neuronal entrenada con voces de mexicanos y norteamericanos es capaz de reconocer los errores de las pronunciaciones.

Se puede concluir que el método ofrece una forma de verificar y detectar los errores en la pronunciación de palabras o frases.

El reconocedor se encuentra restringido al propósito del reconocimiento, a la gramática y a un vocabulario predefinido, aunque basándonos en la idea del *forced aligment* se pueden eliminar errores de inserciones.

Perspectivas

El vocabulario actual del prototipo es pequeño y puede ser adecuado para las necesidades del estudiante que desea aprender a hablar el Español de México.

El nivel de reconocimiento del reconocedor puede mejorarse entrenado éste con un número mayor de muestras, especialmente aquellos fonemas donde existe problemas de reconocimiento como en el caso de los fonemas 'ow','Z' y 'R'.

De cualquier forma, esta es la primera parte del desarrollo de herramientas con Reconocimiento de Voz para un Ambiente de Aprendizaje del Español Mexicano.

Aguas García, N. 1999. Verificación de Pronunciación Basada en Tecnología de Reconocimiento de Voz para un Ambiente de Aprendizaje. Tesis Licenciatura. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas-Puebla. Diciembre.