

Capítulo 2.

Planteamiento del problema

Para la mayoría de las personas el habla forma parte integral de sus vidas. En el contexto de sistemas computacionales, el poder realizar sistemas que permitan a una persona comunicarse con una máquina utilizando el lenguaje del usuario representa el reto final hacia el entendimiento del proceso de producción y percepción en la comunicación [Cole et al., 99].

2.1 Antecedentes

Durante las últimas dos décadas, la práctica de ejercitar con las habilidades del lenguaje hablado ha recibido una importante atención entre educadores. Los planes de estudios del idioma extranjero se centran en habilidades productivas con especial énfasis en la capacidad comunicativa. La capacidad de los estudiantes de ajustarse a la interacción conversacional en un lenguaje específico se considera la meta principal en el aprendizaje de un segundo idioma. Este cambio del énfasis ha generado una necesidad creciente de proveer los materiales de instrucción que proporcionen una oportunidad para practicar el lenguaje de forma controlada fuera de un salón de clase.

En la Universidad de las Américas, como en muchas otras instituciones, los estudiantes extranjeros pasan sus veranos en cursos para aprender el español, un gran porcentaje de estos estudiantes son de los Estados Unidos, la creciente demanda de cursos de español requiere de otras forma de aprendizaje para estos estudiantes.

Hasta el momento existe software para enseñanza de español, pero no existe una herramienta que pueda corregir la pronunciación y le muestre al estudiante donde tuvo el error. Como parte de éste proyecto, el interés se ha enfocado en la verificación de la pronunciación correcta de una palabra o frase ya que el aprendizaje de un idioma no solamente es el poder escribirlo sino también hablarlo. El entrenamiento de una pronunciación correcta es importante. En el área de aprendizaje de un idioma se considera útil e incluso necesario que el estudiante lo practique hablando.

2.2 El problema de corrección de pronunciación

El aprendizaje de un segundo idioma requiere practicar la

pronunciación, por tanto es útil un sistema que pueda evaluar dicho rubro. Sin embargo, es difícil determinar a que se refiere una pronunciación correcta debido a la naturaleza del habla, que varía debido a razones fonológicas y de producción. En el caso de Reconocimiento de Voz, 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 [Korkmazskiy et al., 98]. 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.

Una aplicación útil y notable de reconocimiento de voz y tecnología de procesamiento ha sido demostrada por un número de investigadores y laboratorios comerciales en el área de entrenamiento de la pronunciación. Tutores para la pronunciación interactiva invitan al estudiante a repetir las palabras y frases de forma hablada o leer en voz alta enunciados en el idioma a aprender con el propósito de practicar el sonido y la entonación del lenguaje. La clave para enseñar pronunciación satisfactoriamente es corregir dando retroalimentación, mas específicamente, un tipo de retroalimentación que no haga que el estudiante pierda la atención e interés sobre lo que esta haciendo . Un gran número de sistemas experimentales han implementado puntaje automático de pronunciación para evaluar las producciones del locutor en términos de fluidez, calidad en los segmentos (fonemas) y rasgos supra-segmentales(entonación). El puntaje generado automáticamente puede entonces ser usado como una base para proveer otras formas de retroalimentación correctiva [Ehsani & Knodt, 99].

2.3 Parámetros en la practica de la pronunciación

De los principios que usualmente se describen como contribuyentes para la práctica de pronunciación [Kenworthy, 87], se listan cinco que son aplicables para la práctica automática del lenguaje:

Se deben producir grandes cantidades de enunciados.

Se deben recibir, pertinentemente, correcciones.

Se deben escuchar muchos modelos diferentes de la pronunciación.

La prosodia (amplitud, duración y tono) debe ser enfatizada.

Experimentar en situaciones fáciles en el aprendizaje del lenguaje.

Los primeros cuatro principios se refieren al ambiente "externo" del aprendizaje de idiomas mientras que el quinto principio se dirige a los principiantes, los ambientes "internos", es decir, sus actitudes y capacidades para el aprendizaje de idiomas [Celce & Goodwin, 91].

Mientras que los estudiantes son jóvenes principiantes del lenguaje, producen los nuevos sonidos fácilmente, conforme la edad aumenta, la facilidad para percibirlos y producirlos disminuye. Los estudiantes adultos, necesitan sentirse seguros de sí mismos y ser motivados para producir nuevos sonidos sin sentirse inhibidos. Los estudiantes que tienen alguna enfermedad o discapacidad tienen un riesgo más alto de realizar o decir mal las cosas, y quizá abandonen el aprendizaje de idiomas.

2.3.1 Producción de grandes cantidades de enunciados.

Es importante que el estudiante pueda repetir diferentes enunciados propios del lenguaje que desea aprender. La producción activa de la voz en tales ambientes prepara a principiantes para participar con eficacia en conversaciones en diferentes situaciones. Sin embargo, un curso particular con los instructores humanos es generalmente demasiado costoso y no práctico.

En realidad, la mayoría de los estudiantes toman clases en las cuales tienen que compartir a su profesor, ya que éste debe poner atención a más de un alumno. Esto reduce grandemente la cantidad de tiempo que cada estudiante pasa en producir la voz en el idioma extranjero.

2.3.2 Retroalimentación y correcciones

Los profesores de idiomas marcan una pronunciación incorrecta en el momento en que esta fue hecha, ya que deben intervenir de inmediato para evitar que el estudiante se cree malos hábitos en la pronunciación y que se sienta satisfecho con los resultados que va obteniendo.

Celce Murcia y Goodwin señalan que muchos profesores utilizan la técnica "escucha e imita" que implica la presentación de pares mínimos tales como el las palabras en Inglés *tear* y *tore*. Su investigación sugiere que la repetición de sonidos es una manera eficaz para que los estudiantes aprendan pronunciar los sonidos en un segundo idioma. Por ejemplo, si un estudiante norteamericano quiere comenzar a aprender francés y oye el sonido / y / en la palabra *tu*, él o ella lo asociará más probablemente al sonido / u / de la palabra *toot* en inglés, puesto que / y / todavía no se ha enseñado como fonema separado para el estudiante [Celce & Goodwin, 91].

Las técnicas de la enseñanza han seguido el principio de que se deben escuchar los sonidos no nativos, categorizarlos, y producirlos, los estudiantes deben tener la instrucción específica de cómo articularlos. Se sabe que los estudiantes pueden experimentar físicamente la articulación del sonido y poder producirlo antes de que puedan oírlo como elemento separado en la lengua objetivo. Investigaciones recientes muestran que este método es muy efectivo [Akhane-Yamada, et al., 96].

2.3.3 Diferentes modelos de la pronunciación

Es importante que el estudiante escuche muchas veces a muchas personas nativas del lenguaje que desea aprender. Esta es la técnica conocida como "escucha e imita". Para hacer más sencilla la labor, se recomienda obtener materiales educativos, tales como cintas de audio y de vídeo. Aunque puede resultar muy poco costoso.

2.3.4 Prosodia

Cuando los estudiantes comienzan a aprender un nuevo lenguaje, se dedican generalmente a aprender los sonidos que no se encuentran en su lengua materna. La entonación mantiene la consistencia en un mensaje. Indica qué palabras son importantes, quita ambigüedades de alguna parte de un enunciado, y realza el significado, estilo y emoción.

2.3.5 Experimentar en diferentes situaciones

Cuando los estudiantes tienen que producir los sonidos que no existen en su lengua materna, tienden a perder la confianza en sí mismos [Laroy, 95]. Consecuentemente, pueden dejar de practicar la pronunciación del segundo idioma confiando solamente en los sonidos de su lengua materna. Según Laroy, la confianza del estudiante puede ser aumentada al corregir solamente cuando es necesario, reforzando la buena pronunciación, y evitando una retroalimentación negativa.

Un sistema automático se puede utilizar como complemento al profesor humano en la práctica de la pronunciación. En este escenario el profesor debe proporcionar una atmósfera positiva para el aprendizaje, que explique las diferencias entre el lenguaje que se desea aprender y el lenguaje materno, mientras que la computadora asume el control de los aspectos de la práctica de la pronunciación.

2.4 Enseñanza de lenguajes

Cotton afirma que el aprendizaje es un proceso de adquisición de un nuevo conocimiento y habilidad. Para que este proceso pueda ser calificado como aprendizaje, en lugar de una simple retención pasajera, debe implicar una retención del conocimiento o de la habilidad en cuestión que permita su manifestación en un tiempo futuro. El aprendizaje puede definirse de un modo más formal "como un cambio relativamente permanente en el comportamiento o en el posible comportamiento, fruto de la experiencia" [Cotton, 89].

Los avances tecnológicos han levantado las expectativas y las demandas han puesto a las computadoras como herramienta potencial para el aprendizaje. Los educadores y los investigadores de

la enseñanza de un segundo idioma están exigiendo sistemas inteligentes, que deben poder diagnosticar los problemas de un estudiante con la pronunciación, la sintaxis, o el uso, y después decidir inteligentemente entre un rango de opciones (por ejemplo: repetir, parafrasear, corregir, o dirigir al estudiante explicaciones a fondo) [Warschauer, 96].

Las estrategias de enseñanza del lenguaje han sido identificadas como factores significativos en SLA (*Second Language Acquisition*). Una filosofía de aprendizaje de cómo el lenguaje se aprende puede ser determinante en el logro de los objetivos, lo cual influye en el nivel de alcances logrados.

2.5 Trabajos relacionados a la enseñanza de un segundo idioma

En recientes años, varios laboratorios han comenzado a desarrollar aplicaciones CALL con tecnología de voz. Los resultados muestran sistemas prototipos con interacción de la voz para la enseñanza de pronunciación, lectura y la práctica de las habilidades en conversaciones [Ehsani & Knodt, 99].

Un número de técnicas han sido sugeridas para el reconocimiento y puntuación automática de la voz extranjera [Bernstein, 97; Franco, et al., 97; Kim et al., 97; Witt & Young, 97]. En términos generales, el procedimiento consiste en construir modelos basados en las "pronunciaciones nativas" y después llevar a cabo un análisis sobre las pronunciaciones hechas por los locutores extranjeros en base a los modelos nativos. Esto requiere que los modelos sean entrenados con datos de voz nativos y no nativos, y luego utilizar un conjunto de los algoritmos para medir las variables acústicas que han probado ser útiles en distinguir voz nativa de la extranjera. Estas variables incluyen tiempo de espera de la respuesta, la duración del segmento, pausas entre palabras (en frases), probabilidad espectral, y la frecuencia fundamental (F0). Los puntajes de la máquina se calculan de la estadística derivada de comparar los valores extranjeros para estas variables a los modelos nativos.

En un paso final, la máquina genera la puntuación de la pronunciación la cual es validada comparando estos puntos con el juicio humano. La exactitud de la puntuación aumenta con la duración de la pronunciación que se evaluará. El Instituto de Investigación de Stanford (SRI) ha demostrado un 0.44 de acierto entre las puntuaciones generadas por la computadora y las puntuaciones de humanos a nivel de teléfono. A nivel de enunciados, los aciertos máquina-humano fueron de 0.58, y en a nivel de locutor fue de 0.72 para un total de 50 pronunciaciones por locutor [Franco et al., 1997; Kim et al., 1997].

Otros han utilizado sistemas expertos sobre los errores sistemáticos de las pronunciaciones hechas por aprendices adultos de un segundo

idioma para diagnosticar y corregir tales errores. Un sistema es el proyecto SPELL, de la Comunidad Europea, para la revisión y mejora automatizada de la pronunciación del idioma extranjero [Hiller, et al., 94]. Este sistema utiliza tecnologías avanzadas procesamiento y reconocimiento de voz para evaluar errores en la pronunciación de los estudiantes de Inglés (franceses o italianos) y para proporcionar una retroalimentación correctiva de inmediato.

Un sistema para enseñar la pronunciación de vocales largas japonesas, el *mora nasal* y los *mora obstruents*, fue construido recientemente en la universidad de Tokio. Este sistema permite a los estudiantes practicar las diferencias de los fonemas en japonés, que se conocen como desafíos especiales para los principiantes de este segundo idioma. Invita a estudiantes pronunciar pares mínimos (ej. vocales cortas y largas) y da una retroalimentación inmediata en base a la duración del segmento. De acuerdo con los datos limitados, el sistema parece absolutamente eficaz en esta tarea determinada.

Otro tipo de retroalimentación para la pronunciación incluye la visualización gráfica de una cara, del tracto vocal, la información del espectro, y de las formas de onda de la voz. Los experimentos han mostrado que una representación visual del locutor mejora no solamente exactitud de la identificación de la palabra [Bernstein & Christian, 96], sino también ritmo y sincronización del habla [Markham & Nagano, 97]. En la figura 2.1, se muestra un ejemplo de representaciones visuales del locutor.

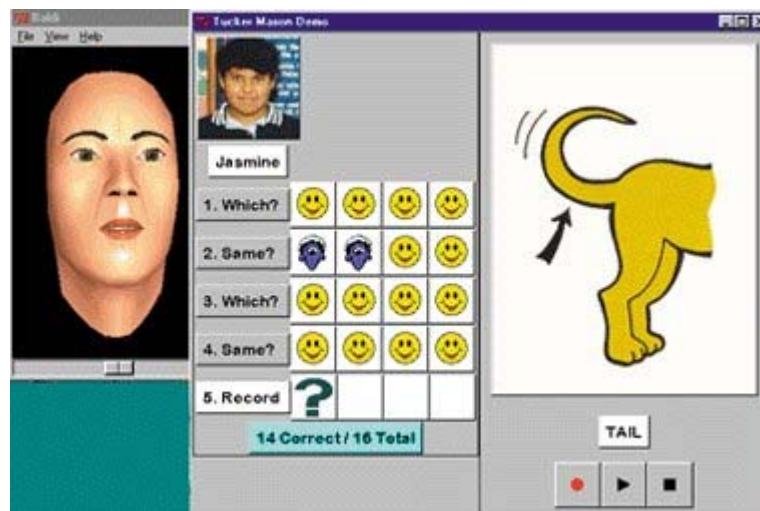


Figura 2.1. Ejemplo de representaciones visuales del locutor.

La práctica intensiva en la lectura en voz alta ayuda a estudiantes a establecer la asociación convencional entre los sonidos y su forma escrita, una habilidad que requiere años de práctica en niños jóvenes y estudiantes de idiomas con escritura no-fonética, tal como japonés o chino [Ehsani & Knodt, 99]. En la figura 2.2 se muestran varias formas alternativas de representar la frase *He was shot in the back*.

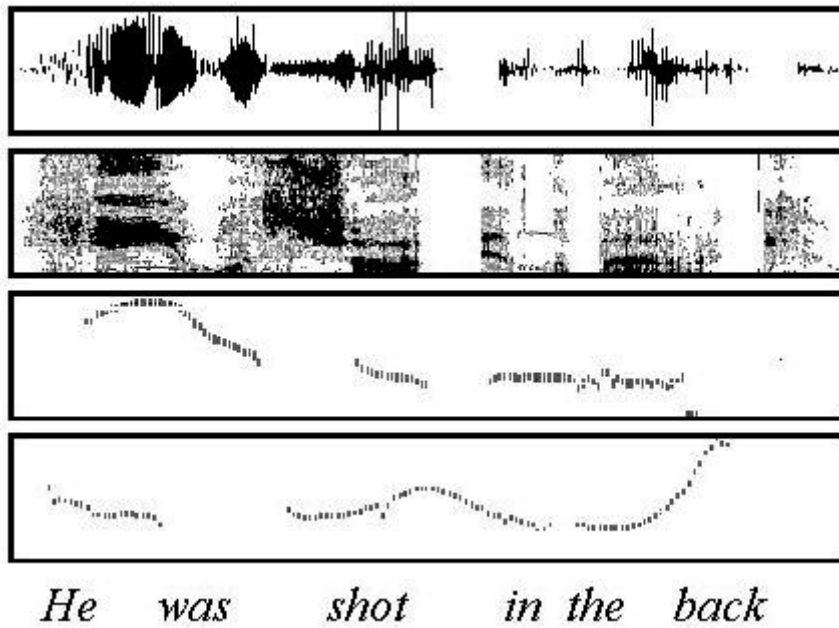


Figura 2.2. Formas alternativas de representar la frase *He was shot in the back*.

Otro sistema es FLUENCY (figura 2.3), desarrollado en la Universidad de Carnegie Mellon (CMU). El proyecto se ha desarrollado como una solución para los usuarios que participan activamente en múltiples situaciones a elegir. El sistema simplemente mapea ejemplares de los sonidos de acuerdo a la señal de entrada. Es posible predecir lo que es estudiante dirá para satisfacer las necesidades del reconocedor, mientras se da la libertad de construir pronunciaciones hechas por el mismo.

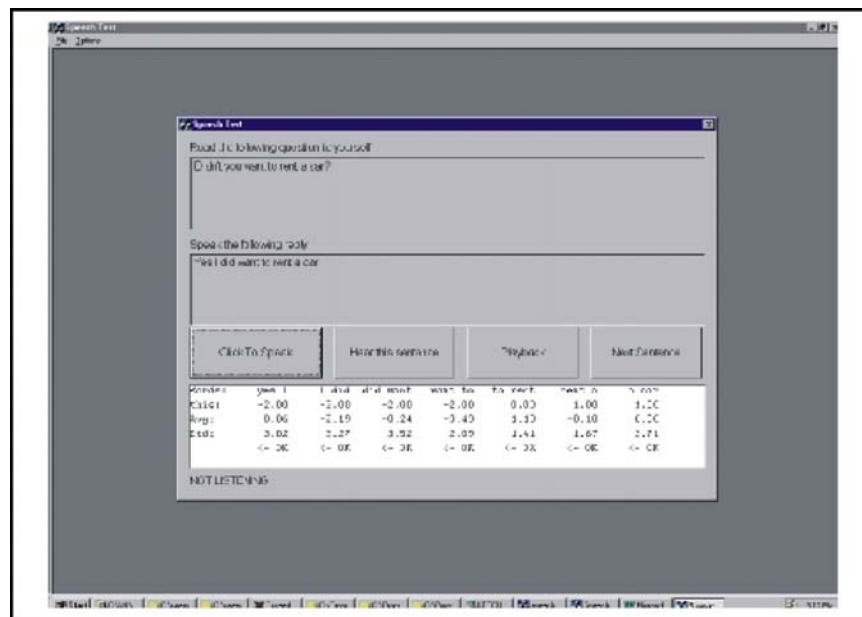


Figura 2.3. Sistema FLUENCY.

Hasta ahora hemos visto algunos de los enfoques que se han tomado en diferentes laboratorios que han comenzado a desarrollar aplicaciones CALL. 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.

2.6 La herramienta de verificación de pronunciación

Haciendo un análisis sobre lo que han comenzado a hacer varios laboratorios en el desarrollo de aplicaciones CALL, podemos darnos cuenta de que no existe una herramienta tal como la que se plantea en este proyecto.

Hasta ahora, lo que proponen es dar un puntaje sobre la pronunciación del locutor, el cual se obtiene en base a modelos nativos de la pronunciación, con los que se compara a los extranjeros, se calcula un puntaje y dicha puntuación se valida comparando estos puntos con el juicio humano.

Tomando como base algunas de estas ideas, la herramienta de verificación de pronunciación, desarrollada para este proyecto, comparará las pronunciaciones hechas por los estudiantes extranjeros contra un modelo basado en pronunciaciones nativas.

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.

La retroalimentación para el estudiante no se dará como un puntaje sobre la pronunciación, si no que se le indica, en forma textual, que fue lo que pronunció mal. También, como apoyo al aprendizaje, se le mostrará al estudiante una imagen de la palabra o frase que esta practicando y el sonido de la misma.

El desempeño de la herramienta al verificar la pronunciación de un estudiante esta validada comparando los resultados del sistema con los resultados dados por el juicio humano.

Por lo tanto podemos definir a este proyecto como una aplicación enfocada al apoyo de estudiantes norteamericanos que desean practicar su pronunciación. Esta herramienta les permite verificar su pronunciación de forma interactiva, rápida y eficiente por medio de una interfaz gráfica amigable diseñada precisamente para estudiantes.

El aprendizaje ocurre a través de la interacción con palabras, imágenes y sonidos. El sistema desarrollado está basado en algunas de las técnicas establecidas por el desarrollo de ambientes de aprendizaje para un idioma utilizando la computadora.

Nuestro objetivo es entonces desarrollar una herramienta de verificación de pronunciación para un ambiente de aprendizaje de Español Mexicano, que permita que estudiantes extranjeros con idioma inglés estadounidense puedan aprender el idioma pronunciando palabras o frases comunes del español usado en México y sirva además como base para futuras investigaciones.

Para lograr este objetivo se necesitó realizar varias actividades conjuntas como la revisión del código del CSLU Toolkit para realizar modificaciones, ya que no presentaba la capacidad de manejo/diseño de una herramienta del tipo requerido para éste proyecto, definición del vocabulario y el tipo de interacción del estudiante con la computadora, diseñar una base de datos de voz adecuada y grabarla. Las grabaciones fueron hechas por mexicanos y norteamericanos. Se desarrolló un método para hacer el análisis de la pronunciación del usuario mediante la aplicación de un algoritmo basado en la salida de una red neuronal. Luego se desarrolló un prototipo accesible localmente basado en:

Metodología aceptada para el aprendizaje de un segundo idioma utilizando una computadora.

Vocabulario restringido, pero enfocado a la aplicación.

Reconocimiento a nivel de fonemas.

Este enfoque usa un reconocedor de voz de propósito general desarrollado en Tlatoa.

Conclusiones

En este capítulo se describieron los antecedentes para el problema de verificación de pronunciación, se identificó en qué consiste el problema y se explicaron algunos parámetros a seguir. Finalmente se analizaron algunos trabajos relacionados. En el siguiente capítulo se describen las metodologías y herramientas utilizadas para la solución del problema.

[índice](#) [resumen](#) [introducción](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [A](#) [B](#) [C](#) [D](#) [referencias](#)

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.
Derechos Reservados © 1999, Universidad de las Américas-Puebla.