

## **Aplicación multimedia para el entrenamiento en la certificación TOEFL mediante reconocimiento de voz**

Miguel Hernández Ramos, Rafael Lemuz López

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación,  
Puebla, México

miguellhdez@gmail.com, rlemuz@cs.buap.mx

**Resumen.** En este trabajo se presenta una aplicación para la práctica del inglés orientada a mejorar el desempeño de los alumnos en los exámenes de certificación TOEFL y motivar la práctica de hablar el idioma. Un aspecto esencial de la aplicación es que todas las respuestas a las preguntas de las diversas secciones son habladas. En el trabajo se describen, el diseño de la aplicación y las tecnologías de la información y comunicación utilizadas para su desarrollo, que van desde la base de datos, las APIs de reconocimiento y síntesis de voz, hasta los aspectos del diseño de la interacción humano computadora.

**Palabras clave:** TOEFL, entrenamiento, reconocimiento de voz, tecnologías Web.

### **Multimedia Application for Training in TOEFL Certification Using Speech Recognition**

**Abstract.** In this work we present an application to the practice of English geared to improving the performance of students in the certification exams TOEFL and motivate the practice of speak the language. An essential aspect of the application is that the answers to all the questions of the various sections are spoken. In the work described, the design of the application and the information and communication technologies used for its development, ranging from the database, the APIs speech recognition and synthesis, until those aspects of the design of the human computer interaction.

**Keywords:** TOEFL, training, speech recognition, Web technologies.

#### **1. Introducción**

El dominio del idioma inglés es importante porque permite mejorar la comunicación en una sociedad globalizada. También permite compartir experiencias y conocimientos

a través de los medios de comunicación tradicionales y emergentes basados en la Web. En especial en México es fundamental porque muchas de las relaciones laborales con los principales socios económicos son en inglés. Además, varias universidades actualmente tienen como requisito de titulación la acreditación de una lengua extranjera [1], y el idioma inglés, para muchos estudiantes, es la primera opción. Sin embargo, los métodos de enseñanza en general no incorporan las tecnologías de la información y la comunicación (TIC), y cuando lo hacen estas no están alineadas con los objetivos fundamentales de quienes toman un curso: aprender a hablar el idioma y acreditar los exámenes de certificación.

Lo anterior es la motivación fundamental de la aplicación multimedia que se propone, el objetivo es favorecer el desarrollo de ciertas habilidades como la percepción, la pronunciación, la habilidad auditiva y en particular se hace énfasis en que el alumno desarrolle su habilidad del habla.

En el 2015 el idioma inglés fue el principal idioma internacional, hoy en día se está convirtiendo en una habilidad básica para toda la fuerza laboral global [2]. Estudios recientes [3], muestran que Latinoamérica es una región con nivel bajo en cuanto al dominio de inglés, por primera vez, sólo tres países latinoamericanos quedaron en la banda del nivel más bajo. México se encuentra en la sexta posición de 14 países de América Latina y en la posición 40 de 70 países a nivel mundial.

Este trabajo analiza las soluciones existentes para el dominio del idioma inglés, en su mayoría aquellas que se sustentan de medios electrónicos para transmitir el conocimiento y determina en que puntos o partes del trabajo debe ponerse más dedicación para llenar los huecos que otros trabajos dejaron. La aplicación incorpora material de entrenamiento para la gramática, la pronunciación, la percepción y la lectura en cada uno de sus niveles para que el estudiante pueda desarrollar tareas mínimamente académicas.

## **2. Marco referencial**

El marco utilizado para este proyecto involucra temas de tecnologías web y APIs (Application Programming Interface) que pueden ser usadas en temas educativos.

### **2.1. Certificaciones de inglés**

Sirven para acreditar, de manera precisa, que la persona puede seguir un programa formativo en este idioma o que tiene las habilidades lingüísticas necesarias para desempeñar una función laboral completa o parcialmente en esta lengua. La elección de un certificado u otro depende del objetivo del interesado.

#### **2.4.1. TOEFL (Test of English as a Foreign Language)**

La Prueba del inglés como Lengua Extranjera fue creada por ETS, (Educational Testing Service), una organización con sede en los EU que se dedica al desarrollo de pruebas profesionales [6]. Es una prueba que evalúa el nivel de inglés americano de las personas que no son hablantes nativos de esta lengua. Evalúa la habilidad de una

persona al leer, escribir, hablar y comprender el inglés que se utiliza en la preparatoria o la universidad.

## **2.5. Tecnología de reconocimiento de voz**

El objetivo de la tecnología de reconocimiento de voz es hacer que la comunicación de los humanos con las máquinas principalmente sea más eficiente y fácil. Estas tecnologías de voz incluyen varias tecnologías: Speech analysis, Speech synthesis, Speech recognition y Speaker recognition.

## **2.6. Sintetizadores de voz**

También conocida como Síntesis del Habla, es la tarea de generar voz artificial partiendo de un texto. Un sistema Text To Speech (Texto a Conversación), también llamado TTS, es una aplicación que convierte un texto escrito a audio, permitiendo a un invidente escuchar un documento, un email o incluso llamar a un número de teléfono y poder escuchar una información que previamente alguien ha escrito.

## **2.7. Kits de desarrollo de software con reconocimiento de voz**

### **2.7.1. SAPI**

Microsoft ha desarrollado un set de herramientas para tareas de reconocimiento de voz desde las primeras versiones del sistema operativo. Este se incluye en la actualidad motores de reconocimiento y de síntesis de voz en varios idiomas que incluyen inglés, francés, chino, español, alemán, ruso, japonés entre otros. Desde la versión 5.0 de SAPI incluida en Windows 2000, se limita el acceso de la aplicación a los motores de reconocimiento y de síntesis de voz.

### **2.7.2. Java Speech**

La plataforma de desarrollo Java no tiene un motor de reconocimiento de voz, pero sí brinda al programador la facilidad para escribir aplicaciones de voz con un conjunto de clases e interfaces que modelan esta tarea. Este conjunto de clases e interfaces se las conoce como el JSAPI o Java Speech API. Estas clases fueron concebidas a manera de envoltura, ya que internamente en el entorno administrado de Java, estas clases no tienen implementación alguna. Es tarea del sistema operativo host a través de objetos nativos el proveer de la implementación de estas tareas de reconocimiento de voz. Java conecta a los objetos y a las interfaces de JSAPI con la implementación nativa a través de JNI14.

## **3. Soluciones existentes**

Se presentan sitios y/o aplicaciones web que al día de hoy son usadas por quienes están interesados en aprender o fortalecer sus conocimientos en el idioma inglés.

- Duolingo. Página Web y aplicación móvil basada en niveles que simulan un juego donde ganar puntos y avanzar de nivel es un reto, tiene contenido para casi cualquier situación de la vida diaria, pero carece de orientación para fines específicos.
- Interlingua. Aplicación móvil que presenta animaciones de tipo conversación. Está enfocada a la pronunciación.
- Saber inglés. Página de gran ayuda si se tiene conocimiento del idioma, enfocada principalmente a la gramática.
- Learn English. Página ideal para comenzar a aprender inglés. Lo mejor de este sitio es su sección de vocabulario ya que tiene todo tipo de palabras, conceptos y expresiones útiles para una comunicación básica en inglés.
- English Page. Está basada en el inglés americano, muchos cursos tradicionales la utilizan como apoyo adicional para sus estudiantes. Tiene cientos de ejercicios para cada una de las lecciones que ofrece.

#### 4. Metodología

La metodología aplicada para alcanzar el objetivo de este trabajo fue IDEAL que consta de las siguientes cinco fases (Bibiana D. Rossi, 2001):

1. Identificar la tarea mediante el plan de requisitos.
2. Desarrollo de prototipos. La concepción de la solución e implementación.
3. Ejecución de la construcción del sistema. Integración con otros sistemas, y la aceptación por el usuario del final.
4. Actuación para conseguir el mantenimiento perfectivo. Definir el mantenimiento del sistema global, de las bases de conocimientos.
5. Lograr la transferencia tecnológica mediante entrenamiento en sesiones de tutoría entre los diseñadores y los usuarios finales, (García Martínez, R y Britos P. 2004. Ingeniería de Sistemas Expertos. Editorial Nueva).

#### 5. Desarrollo

##### 5.1. Propuesta de solución

Como primera consideración para desarrollar la aplicación, de ahora en adelante *Angloparlante*, fue dar extras a las soluciones existentes. Las siguientes son características que la hacen diferente:

- Es una *aplicación para el entrenamiento* para un examen de certificación TOEFL, no enseña gramática ni vocabulario en inglés, es como su nombre lo indica aplicación para entrenar a las personas a fin de fortalecer los conocimientos que se tienen en el idioma inglés.
- Las respuestas a las evaluaciones para cada sesión de autoaprendizaje son habladas haciendo uso de un micrófono de computadora.
- Incluye actividades de la vida cotidiana del ámbito académico. Cuenta con 4 secciones: Reading, Listening, Writing y Speaking para las que tiene un apartado

de *learn* y otro de *practice*. La mayor parte de las actividades están dentro de Speaking debido a la naturaleza de la aplicación, el uso del reconocimiento de voz, y además porque se considera que es la sección dónde se ponen a prueba las otras 3 secciones.

- El aprendizaje puede ser guiado por un docente o de forma automática. Si se elige la forma guiada la respuesta es enviada vía mensaje, dentro de la misma aplicación, al docente y la calificación es asignada manualmente. En cambio, si se elige la forma automática, la aplicación recibe la entrada por voz, la convierte a texto y con el algoritmo del punto 5.2.3 se recibe la calificación, en caso de ser aprobatoria el alumno podrá realizar la siguiente actividad de evaluación, en otro caso podrá reintentar máximo 5 veces.
- Angloparlante conserva el último nivel alcanzado por el estudiante a fin de direccionarlo justo donde se quedó.
- A manera en que el estudiante avanza la dificultad es mayor, se usan palabras más difíciles de pronunciar, las respuestas son más extensas o cambian de contexto, pero siempre son frases de un entorno académico que es lo que evalúa TOEFL IBT.
- Los docentes intervienen únicamente cuando los alumnos eligen ser evaluados por ellos, y es en ese momento cuando orientan a los alumnos.
- Al realizar una evaluación, el estudiante, puede elegir entre ser evaluado por el docente o por la aplicación independientemente de cómo ha sido evaluado antes.
- Contiene tips para cada sección, estos son aleatorios y puede verse uno u otro cada vez que el estudiante ingresa a una sección.

## 5.2. Diseño e implementación del sistema

### 5.2.1. Modelo de casos de uso

El modelo de casos de uso describe a Angloparlante en términos de sus distintas formas de utilización, cada una de las cuales se conoce como un caso de uso. El estudiante y el docente son actores, es decir, entidades distintas a los usuarios, en el sentido de que éstos son las personas reales que utilizarán el sistema, mientras que los actores representan cierta función que una persona real realiza. En la terminología orientada a objetos se considera al actor una clase de usuario mientras que los usuarios se consideran como objetos o instancias de esa clase. Incluso, una misma persona puede aparecer como diversas instancias de diferentes actores.



Fig. 1. Diagrama de delimitación del sistema.

El diagrama de delimitación del sistema con nuestros actores se muestra en la figura 1, y tiene las características mencionadas.

Se puede notar que para interactuar con la aplicación ambos deben iniciar sesión y como consecuencia estar registrados. Actualmente la inscripción es manual debido a que no se cuenta con un administrador.

### 5.2.2. Diseño de la base de datos

La base de datos para Anglolarlante tiene la colección de datos necesaria para el correcto funcionamiento de la aplicación. Para realizarla, y una vez abstraído el problema, se determinó *el propósito de la base de datos* el cuál es disponer en todo momento de la información relacionada con la aplicación. El siguiente paso fue *determinar las tablas necesarias*, luego se *determinaron los campos necesarios para cada tabla*, el cuarto paso consistió en *determinar las relaciones* y finalmente se *perfeccionó el diseño* buscando errores usando datos de prueba para poder realizar consultas en modo consola y con esto determinar campos sobrantes o tablas innecesarias para disminuir la redundancia e inconsistencia de los datos.

La figura 3 muestra la estructura de la base de datos que se obtuvo al seguir los 5 pasos anteriores. En la figura se aprecian las tablas, los campos, la relación y la cardinalidad entre las tablas.

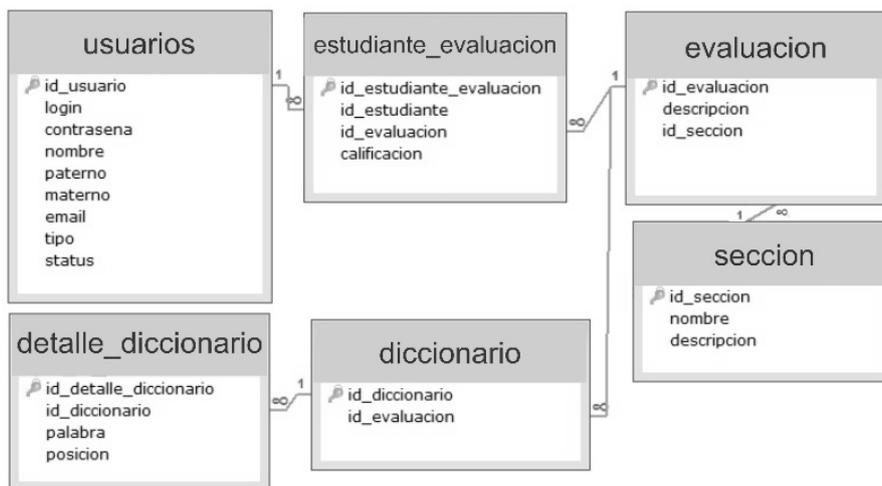


Fig. 3. Estructura gráfica de las tablas, la cardinalidad y relaciones de las tablas.

### 5.2.3. Técnicas y tecnologías empleadas

Respecto al reconocimiento de voz, se utilizó la modalidad de reconocimiento por medio de comandos debido a que disponemos de las palabras relacionadas a cada una de las respuestas, y además es una colección limitada dentro de la base de datos.

El reconocimiento de voz se hace con Web Speech API, una API diseñada por la comunidad de la W3C e impulsada por Google y openStream para la síntesis y el análisis del habla en aplicaciones Web [9].

En el diseño de la aplicación se consideraron aspectos de interacción humano computadora: un diseño sencillo y atractivo y responsivo para el usuario, con la capacidad de ocultar el menú lateral, fig. 4, para tener disponible la pantalla completa para las actividades.

El siguiente algoritmo realiza la evaluación automática, recibe como entrada dos arreglos, uno formado por las palabras reconocidas por Web Speech API y el otro con los datos asociados a la evaluación en curso, estos datos se recuperan previamente mediante una petición AJAX.

```
function evalua(arr_estudiante, arr_conocimiento){
  var arr_acertadas = [];
  var arr_errores = [];
  var resultado = [];
  for(var i=0;i<arr_conocimiento;i++){
    for(var j=0;j<arr_estudiante;j++){
      if(arr_conocimiento[i] == arr_estudiante[j]){
        arr_acertadas.push(arr_estudiante[j]);
      }
      else{
        arr_errores.push(arr_estudiante[j]);
      }
    }
  }
  resultado['pAciertos'] = arr_acertadas;
  resultado['pErrores'] = arr_errores;
  return resultado;
}
```

Posterior a la ejecución del algoritmo, escrito en JavaScript, lo que resta es mostrar el resultado al estudiante, Si el arreglo de errores está vacío y el arreglo de palabras acertadas coincide término a término con el arreglo de la petición la respuesta será correcta y su calificación aprobatoria, en caso contrario, si el 50% o más de las palabras de la petición no coinciden debe repetirse el ejercicio completo, en cambio si los errores son mínimos se le indica en que partes falló y tendrá la opción de corregir sólo eso. Respecto a los intervalos de calificación, si la respuesta llega en el primer intento la calificación es de 10, en el segundo intento (total o parcial) 9, para el tercer intento corresponde 8, para el cuarto 7, y para el quinto 6.

#### 5.2.4. Aplicación

La ventana principal de la aplicación tiene accesos directos a su último nivel alcanzado, sus calificaciones y un enlace para acceder a sus mensajes, esta ventana principal tiene un menú lateral con las secciones, cada una con su apartado de Learn y

Practice tal como se aprecia en la figura 4. Con la navegación lateral el estudiante puede ir a una u otra sección, y dentro de ella elegir si quiere aprender o practicar.

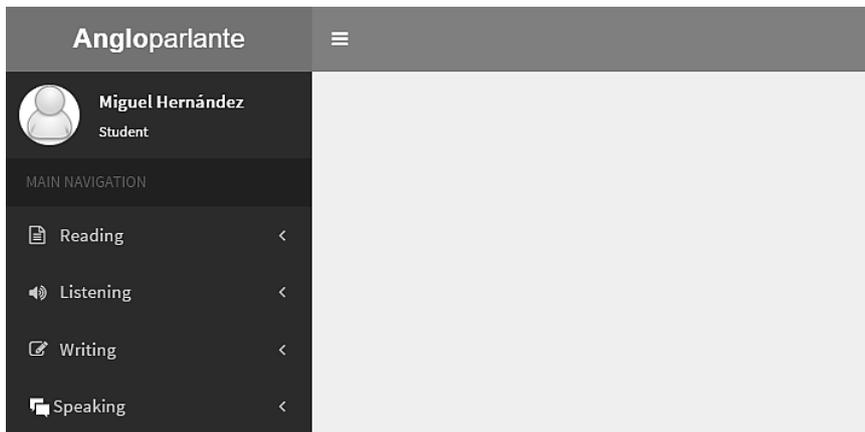


Fig. 4. Navegación lateral del estudiante.

La figura 5 muestra un ejemplo de evaluación. A la derecha está el texto que el usuario le dirá a Angloparlante. Para activar el reconocimiento el estudiante debe presionar el icono del micrófono y posteriormente comenzar a hablar.

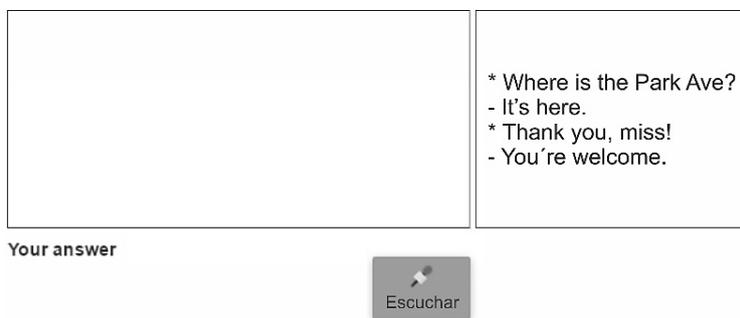


Fig. 5. Actividad de evaluación con reconocimiento de voz.

Para el caso del docente, sus actividades son de seguimiento al aprendizaje del estudiante como se aprecia en la figura 6.

Nombre	Lección superada	Regresar nivel	Activo	Suspender	Activar
Miguel Hernández Ramos	7	<	SI	×	✓
Violeta Elvia Mexicano Salas	0	<	SI	×	✓
Miguel H	0	<	SI	×	✓

Fig. 6. Ventana principal del docente.

## **6. Conclusiones**

En este trabajo presentamos un sistema de software para la práctica hablada del idioma inglés. En el trabajo se describe el diseño de la aplicación desde el diseño de la base de datos hasta el diseño de la interfaz gráfica para la interacción humano computadora. En un trabajo futuro se hará un estudio de usabilidad en un grupo piloto de estudiantes de nivel básico y medio para evaluar el sistema propuesto.

## **Referencias**

1. Criollo, R.: Titulación: Lengua Extranjera. Facultaddelenguas.com, available: <http://www.facultaddelenguas.com/caale/titulacion> (2016)
2. Ef.com.mx: EF Índice inglés Proficiency - inglés, Economía y Calidad de Vida. Available: <http://www.ef.com.mx/epi/insights/english-economics-and-quality-of-life> (2016)
3. Ef.com.mx: EF EPI - inglés e Innovación. Available: <http://www.ef.com.mx/epi/insights/english-and-innovation> (2016)
4. Castaño, C.: El rol del profesor en la transición de la enseñanza presencial al aprendizaje online (2003)
5. Marqués, P.: Multimedia educativo: clasificación, funciones, ventajas e inconvenientes (2004)
6. ETS Home. Ets.org, Available: <https://www.ets.org> (2016)
7. Cambridge Institute: Academia de inglés que imparte cursos de diferentes lenguas para que mejores tu nivel de idiomas - Cambridge Institute. CambridgeInstitute, Available: <http://www.cambridgeinstitute.net> (2016)
8. Salazar, J. A.: Reconocimiento de comandos de voz para control de las funciones multimedia dentro del hogar (2012)
9. Adorf, J.: Web Speech API. KTH Royal Institute of Technology (2013)