

SISTEMA DE EVALUACIÓN SUBJETIVA VÍA INTERNET PARA APLICACIONES DE TECNOLOGÍAS DEL HABLA

Iratxe Martin	Eva Navas	Inmaculada Hernáez	Jon Sánchez
Dpto. Electrónica y Telecomunicaciones	Dpto. Electrónica y Telecomunicaciones	Dpto. Electrónica y Telecomunicaciones	Dpto. Electrónica y Telecomunicaciones
Universidad del País Vasco			
e-mail : iratxem@bips.bi.ehu.es	e-mail : eva@bips.bi.ehu.es	e-mail : inma@bips.bi.ehu.es	e-mail : ion@bips.bi.ehu.es

Abstract-This paper describes a tool for evaluating the quality of natural and synthetic speech signals using subjective techniques. The designed evaluation system allows evaluating all of the aspects that determine their quality, like intelligibility, naturalness and expression of emotions. The system is flexible and totally configurable. The subjective tests created by this tool consist of WEB forms that are easily accessible by the evaluators. This system has been successfully used to evaluate the quality of an emotional database.

I. INTRODUCCIÓN

Debido al progresivo desarrollo tecnológico y teniendo en cuenta que el método de comunicación más natural entre los seres humanos es el oral, no es de extrañar el intento de desarrollar nuevas formas de comunicación oral con las máquinas. Para que este tipo de interfaces se desarrollen correctamente es necesario disponer de sistemas de conversión de texto a habla de calidad y de herramientas que permitan evaluar dicha calidad.

En la evaluación de cualquier aspecto de las señales de voz, pueden distinguirse dos tipos de procesos diferentes: objetivos y subjetivos, cada uno de ellos con sus ventajas e inconvenientes [1].

Las técnicas objetivas están basadas en la comparación con una referencia [2] y su principal inconveniente es que no proporcionan información alguna sobre la satisfacción del usuario final. Pero tienen la ventaja de que son rápidas de realizar y sus resultados son repetibles.

Los procesos subjetivos consisten en reunir al mayor número de personas posibles para que den su opinión sobre diferentes aspectos del sistema. Los resultados obtenidos mediante estas técnicas proporcionan información sobre cuál puede ser el nivel de aceptación del sistema de conversión de texto a habla, pero tienen la desventaja de que sus resultados no son repetibles y de ser muy costosas de realizar, ya que es necesario llevar a todos los evaluadores delante del sistema y someterles al test.

En este artículo se describe una herramienta capaz de generar pruebas que permiten realizar evaluaciones subjetivas vía Internet, de forma que conseguir que un elevado número de personas colabore en la evaluación resulta menos costoso.

En la siguiente sección se describen los objetivos del sistema. Seguidamente, en la sección III, se detalla el funcionamiento del sistema de evaluación subjetiva vía Internet. Posteriormente, en la sección IV, se analiza una evaluación realizada y los resultados obtenidos en la misma. Por último, en la sección V, se recogen las conclusiones principales de este trabajo.

II. OBJETIVOS PRINCIPALES DEL SISTEMA

El objetivo principal de este trabajo ha sido proporcionar una herramienta que permita realizar evaluaciones subjetivas sobre diferentes aspectos de sistemas de conversión de texto a habla y de voz natural, de manera que los resultados obtenidos sean suficientemente fiables como para poder determinar el grado de aceptación que la voz evaluada va a tener.

Otro de los objetivos principales ha sido conseguir que el sistema sea fácilmente configurable, ya que la persona encargada de diseñar y realizar una prueba no tiene por qué ser la misma que diseñó el sistema de evaluación.

Las pruebas diseñadas con esta herramienta deben llegar al mayor número de evaluadores posible de forma fácil y cómoda, para ello se han realizado mediante formularios Web.

Al tratarse de una herramienta Web no existe contacto directo con el evaluador, y por ello el objetivo a la hora de realizar el diseño del interfaz ha sido que éste proporcione un entorno sencillo y amigable, de manera que facilite la tarea de evaluación y que no sea necesario que el usuario tenga ningún conocimiento técnico.

Un sistema de evaluación además de ser una herramienta para realizar una evaluación propiamente dicha, también debe proporcionar la posibilidad de visualizar los resultados obtenidos de forma gráfica. Por tanto la capacidad del

sistema para realizar cuadros de resultados ha sido otro de los objetivos considerados, permitiendo que el diseñador de la prueba pueda acceder fácilmente a esta información y la interpretación de los resultados le resulte más sencilla.

III. DESCRIPCIÓN DEL SISTEMA DE EVALUACIÓN

Las señales de voz que van a ser analizadas se almacenan en archivos de audio que son escuchados y calificados por los evaluadores. La herramienta diseñada permite evaluar diversos aspectos de la señal de voz, mediante la selección de un determinado tipo de test, que posteriormente los evaluadores deben completar.

A. Características técnicas

El sistema de evaluación subjetiva es una aplicación Web que ha sido programada con PHP. Dicho lenguaje de programación se escribe directamente en las páginas HTML y el servidor Web es quien lo interpreta. El procedimiento es el siguiente; el navegador envía una solicitud HTTP al servidor Web, éste reacciona, lee la página PHP, ejecuta los comandos y envía el documento compilado para que el navegador lo lea.

Para el uso de PHP es necesario que el proveedor ofrezca un servidor Web con soporte PHP, pero tiene la ventaja de que independientemente del navegador que se use en el cliente, el código se va a interpretar correctamente, ya que el servidor es el encargado de realizar dicha tarea.

La información relativa a los evaluadores y los resultados obtenidos en cada evaluación se almacenan en una base de datos. La gestión de dicha base de datos se realiza mediante MySQL, que es un sistema de gestión de bases de datos [3] considerado rápido y robusto y con el que es posible manejar grandes cantidades de datos.

Una gran ventaja de MySQL es que funciona en diversas plataformas y que es compatible con el lenguaje de programación PHP, de forma que la gestión de la base de datos puede realizarse a través de otros programas.

B. Aspectos evaluables

Las voces sintéticas pueden resultar frías e impersonales, haciendo que el oyente no se sienta cómodo al escucharlas. Por ello los principales aspectos a evaluar para determinar su calidad son la inteligibilidad y la naturalidad. Pero para conseguir una mayor satisfacción del oyente, también deben considerarse parámetros como la emotividad, es decir, debe conseguirse que la voz sintética sea capaz de transmitir estados anímicos tales como tristeza, enfado, alegría, etc.

Todos estos aspectos pueden ser evaluados con la herramienta diseñada.

C. Tipos de formularios

Para posibilitar la evaluación de diferentes aspectos, el sistema de evaluación consta de tres tipos de test, que pueden ser fácilmente combinados en función de las necesidades del diseñador de la prueba, quien es el encargado de decidir qué tipo de formularios van a emplearse y qué archivos de audio se van a necesitar. La posibilidad de combinar diferentes tipos de test hace que el

sistema sea totalmente flexible, ya que permite evaluar más de un aspecto en la misma prueba.

El objetivo del primer tipo de test es realizar una comparación entre dos señales de voz. Para ello el evaluador debe escuchar ambas y seleccionar la que considere más adecuada de acuerdo al criterio que se esté evaluando en el test. En la Fig.1 se muestra el formato de este tipo de formulario 1.

Fig. 1. Tipo de formulario 1.

En el segundo tipo de test el evaluador debe escuchar el archivo de audio y seleccionar una de las posibles opciones que se le ofrecen. Este formulario se utilizará para que el evaluador asigne una puntuación a cada señal según el criterio que se indique en el test, o para que asigne cada señal a una categoría de las ofrecidas. En la Fig. 2 se muestra su formato.

Fig. 2. Tipo de formulario 2.

El tercer tipo de formulario consiste en realizar al evaluador preguntas abiertas en las que puede escribir lo que quiera. Este tipo de formularios son empleados para medir aspectos tales como la inteligibilidad, en los que el evaluador tras escuchar cada señal deberá escribir lo que haya entendido. En la Fig. 3 puede observarse el formato de este tipo de test.

Fig. 3. Tipo de formulario 3.

D. Diseño de una prueba

La herramienta diseñada es totalmente flexible y configurable, ya que el diseñador de las pruebas con la simple modificación de unos ficheros de configuración puede adaptar la evaluación a sus necesidades. A continuación se describe brevemente cómo se realiza el diseño de una prueba:

En primer lugar el diseñador de la prueba debe decidir qué aspectos desea evaluar, y por tanto qué tipo de formularios tiene que utilizar. El sistema permite la combinación de diferentes tipos de test en una misma evaluación, para ello el diseñador tiene que escribir en un fichero cuántos formularios va a utilizar de cada tipo y el orden en el que éstos se vayan visualizando. Una prueba puede estar constituida por tantos formularios como el diseñador de la misma desee, no existe limitación alguna por parte del software.

En segundo lugar el diseñador tiene que seleccionar las señales de voz con las que va a realizar la evaluación, y escribir en el fichero correspondiente a cada formulario cuáles de todas las señales escogidas van a oírse en dicho formulario. Las señales que se escuchan en cada uno de los formularios son siempre las mismas, pero el orden en el que se presentan es aleatorio.

Para los tipos de test 1 y 3 no es necesario definir ningún otro fichero, en cambio para el tipo 2 existe un archivo en el que el diseñador de la prueba debe escribir cuáles son las opciones que desea darle al evaluador. Cada test de este tipo tiene su propio fichero de configuración de categorías, así que es posible usar en una misma evaluación más de un formulario del tipo 2 y que en algunos de ellos se pida realizar una asignación de puntuación y en los otros una asignación de categoría.

El aspecto de los formularios es también configurable y se puede modificar fácilmente a gusto del diseñador de la prueba.

Una vez configurados todos los ficheros la prueba está montada, por lo que el diseñador de la misma debe avisar a los evaluadores de que cuando lo deseen pueden acceder a la página Web y realizar la evaluación.

E. Proceso de evaluación

Cuando un usuario accede por primera vez a la página Web del sistema de evaluación debe registrarse haciendo clic en el botón correspondiente. Una vez realizado el registro se le permite acceder a los formularios y si lo desea puede empezar a completarlos.

No es necesario que el evaluador complete todos los formularios de una sola vez, ya que el sistema permite abandonar la evaluación y retomarla en otro momento. Los resultados obtenidos hasta ese momento son almacenados y cuando el evaluador vuelve a acceder a la página, introduciendo su nombre y su contraseña, el sistema lo lleva al formulario en el que abandonó la evaluación.

El sistema permite a los evaluadores escuchar cada señal de voz tantas veces como sea necesario, para ello simplemente tienen que hacer clic en el botón correspondiente.

Tanto los datos que el evaluador proporciona al registrarse como los resultados de la evaluación se almacenan en una base de datos, para su posterior utilización.

F. Fiabilidad de los resultados obtenidos

Para que los resultados obtenidos sean fiables es necesario que el número de personas que realiza la evaluación subjetiva sea suficientemente elevado, en caso contrario éstos no serán significativos.

El hecho que el sistema diseñado permita realizar la evaluación subjetiva vía Web permite obtener un mayor número de colaboradores, ya que las molestias causadas a los mismos son menores. Además, gracias a la universalización de Internet, es posible contar con la participación de personas de muy distintas aptitudes y capacidades, consiguiendo de este modo resultados de mayor fiabilidad.

IV. REALIZACIÓN DE UNA EVALUACIÓN SUBJETIVA

Esta herramienta de generación de pruebas subjetivas ha sido ya utilizada con éxito en la evaluación de una base de datos de habla emocional en la que una actriz había simulado seis emociones diferentes al leer los textos del corpus. El objetivo de la evaluación subjetiva fue probar si los oyentes eran capaces de distinguir dichas emociones en las señales de voz.

A. Proceso de evaluación

Para la realización de la evaluación todos los formularios empleados fueron del tipo 2, es decir, aquéllos en los que el evaluador debe seleccionar una de las opciones que se le ofrecen. En este caso debía escoger entre una de las seis posibles emociones a simular, y un estilo neutro. Las citadas emociones eran las siguientes: enfado, asco, miedo, alegría, tristeza y asombro. La evaluación constaba de un total de 13 formularios, y cada uno de ellos contenía 10 señales de voz.

Una vez finalizada la prueba se preguntó a los evaluadores cuáles les habían parecido las emociones más difíciles de reconocer y cuál era su opinión acerca del sistema de evaluación utilizado.

B. Resultados

Los resultados de la evaluación subjetiva mostraron que las emociones fueron correctamente identificadas. La tabla 1 muestra la matriz de confusión obtenida en el test.

Los evaluadores respondieron que el diseño del sistema de evaluación les había parecido adecuado y que les resultó fácil de utilizar.

Tabla 1. Matriz de confusión obtenida en el test. Las columnas contienen las respuestas correctas y las filas las seleccionadas por los evaluadores.

Emoción	Enfado	Asco	Miedo	Alegria	Neutro	Tristeza	Asombro
Enfado	85.7%	4.0%	0.3%	0.0%	7.3%	0.0%	3.3%
Asco	9.0%	51.3%	1.0%	0.0%	3.3%	0.3%	0.0%
Miedo	0.0%	7.0%	80.3%	0.0%	0.7%	5.7%	0.7%
Alegria	0.3%	0.3%	0.0%	82.0%	1.3%	0.0%	4.7%
Neutro	3.3%	23.0%	0.0%	14.0%	78.7%	17.7%	7.0%
Tristeza	0.0%	12.7%	18.0%	0.0%	0.0%	73.7%	0.0%
Asombro	1.7%	1.7%	0.3%	4.0%	8.7%	2.7%	84.3%

V. CONCLUSIONES

Se ha diseñado un sistema de evaluación subjetiva vía Internet con el que es posible evaluar diferentes aspectos de sistemas de conversión de texto a habla y de voz natural.

El hecho de que sea una aplicación Web hace posible contar con la colaboración de un mayor número de personas,

por tanto los resultados obtenidos son más fiables y estadísticamente más significativos.

La herramienta diseñada es totalmente flexible y configurable y se ha comprobado que el interfaz con el usuario resulta agradable y amigable. Además permite al diseñador de la prueba visualizar los resultados obtenidos en forma de tabla, para facilitar su interpretación.

AGRADECIMIENTOS

Los autores quieren agradecer la colaboración de todos los evaluadores que participaron en el test.

Este proyecto ha sido parcialmente financiado por la universidad del País Vasco (UPV-0147.345-E-14895/2002) y por el Ministerio de Ciencia y Tecnología (TIC C2000-1669-C0403 y TIC2003-08382-C05-03).

REFERENCIAS

- [1] D. Hirst, A. Rilliard, and V. Aubergé, "Comparison of subjective evaluation and an objective evaluation metric for prosody in text-to-speech synthesis", *3rd ESCA/COCOSDA Workshop on Speech Synthesis*, Blue Mountain, Australia, 1-4, 1998
- [2] R. A. J. Clark, and K. E. Dusterhoff, "Objective methods for evaluating synthetic intonation", *Eurospeech '99*, Budapest, pp.1623-1626, 1999
- [3] Petra Bilke, "PHP y MySQL páginas Web dinámicas", PC Cuadernos técnicos, p.25, 2001