Procesamiento del Lenguaje Natural, Revista nº 63, septiembre de 2019, pp. 185-188 recibido 01-04-2019 revisado 16-04-2019 aceptado 17-05-2019

Asistente Conversacional Móvil para la Rehabilitación de la Tartamudez

Mobile Conversational Interface for Stuttering Treatment

David Griol[1], **Javier Font**[1], **Zoraida Callejas**[2]
[1]Univ. Carlos III de Madrid. Avda. de la Universidad, 30. Leganés (España)
[2]Univ. de Granada. Calle Periodista Daniel Saucedo Aranda, s/n. Granada (España)
dgriol@inf.uc3m.es, 100291364@alumnos.uc3m.es, zoraida@ugr.es

Resumen: Los dispositivos móviles se han convertido en una herramienta de uso diario que ha provocado cambios en las formas de interactuar de las personas gracias a la accesibilidad que ofrecen a grandes contenidos de información y a servicios de comunicación de forma continua y prácticamente ubicua. La combinación de estos dispositivos y los interfaces conversacionales ha facilitado el desarrollo de aplicaciones en el campo de la salud móvil (mHealth en inglés) cada vez más avanzadas. En este trabajo presentamos una app que utiliza el potencial de los interfaces conversacionales para ayudar a mejorar la fluidez en el habla en personas que padecen tartamudez.
Palabras clave: Tartamudez, salud móvil, interfaces conversacionales, reconocimiento de emociones

Abstract: Mobile devices have become a tool for daily use that has caused changes in people's ways of interacting, thanks to the accessibility they offer to large contents of information and communication services continuously, regardless of the place and the moment. Thanks to the development of mobile applications, advanced apps have been developed in the area of mobile health (mHealth), a field of eHealth in which the practice of medicine and public health is supported by mobile devices. In this paper we present an app that uses the potential of conversational interfaces for the development of mobile applications that help improve speech fluency in people who suffer from stuttering.
Keywords: Stuttering, mobile health, conversational interfaces, emotion recognition

1 Introducción

Uno de los trastornos de la comunicación más conocidos es la tartamudez o disfemia, que está asociada a las interrupciones involuntarias del habla. Se trata de un trastorno de la comunicación y no del lenguaje que, según la Fundación Española de la Tartamudez, afecta a un 2 % de adultos y a un 5 % de niños. Estos datos se traducen aproximadamente en 800.000 personas en España (Fundación Española de la Tartamudez, 2019).

Los síntomas de la tartamudez se manifiestan, sobre todo, en situaciones en las que un paciente tiene que interactuar en público. Sin embargo, si el paciente se encuentra en un entorno agradable y aislado, puede que sea capaz de hablar de forma fluida. Por esta razón, existen técnicas empleadas en Logopedia, que consisten en engañar al cerebro, de tal forma que el paciente piense que está estableciendo una conversación (Jackson, Rodgers, y Rodgers, 2019). Estas técnicas emplean dispositivos electrónicos que tienen la capacidad de aplicar un tiempo de retardo a la señal de voz y de modificar la frecuencia fundamental de esta señal, cuyo uso provoca una reducción de la velocidad en el habla y la prolongación de vocales, minimizando los efectos que produce la tartamudez (Jaramillo y Gil-Lozada, 2014; Salgado-Ruíz, 2005).

Los avances en las capacidades y sensores de los dispositivos móviles han posibilitado generar aplicaciones (Apps) destinadas a la salud y a los servicios sociosanitarios que han supuesto una revolución en los últimos años gracias a que pueden ofrecer funcionalidades como las mencionadas anteriormente, basadas en la retroalimentación auditiva con modificaciones del tono, y con cierto retardo de la señal de voz, ayudando de forma posi-

ISSN 1135-5948. DOI 10.26342/2019-63-26
© 2019 Sociedad Española para el Procesamiento del Lenguaje Natural

tiva a minimizar los efectos de la tartamudez a través de un dispositivo móvil.

En este artículo presentamos una aplicación para dispositivos móviles con sistema operativo Android basada en un interfaz conversacional hablado (McTear, Callejas, y Griol, 2016; Griol et al., 2014) que incorpora herramientas para mejorar la fluidez verbal de personas que padecen tartamudez. Se ha desarrollado un sistema con un funcionamiento sencillo e intuitivo que fomenta la comunicación verbal a través del habla. El interfaz conversacional es el elemento central de la aplicación, pues a partir de él, el usuario tendrá que interactuar de forma oral para poder ejecutar las diferentes funcionalidades facilitadas por la aplicación. Además, la aplicación dispone de un módulo basado en plantear una serie de preguntas que permitan evaluar el estado emocional del usuario. Esta información se facilita al agente conversacional para realizar recomendaciones al usuario entre el conjunto de actividades integradas en la aplicación, además de considerar las actividades completadas por el usuario en interacciones anteriores con la aplicación y los resultados obtenidos.

A partir de la búsqueda en Google Play Store del texto "Tartamudez", podemos encontrar un número muy pequeño de aplicaciones dedicadas a su tratamiento. Partiendo de la información sobre las funcionalidades que ofrecen, se puede comprobar que la mayoría de estas aplicaciones ofrecen funcionalidades basadas en la retroalimentación auditiva retardada con la posibilidad de modificar el tono de la señal de voz. Otras aplicaciones simplemente contienen artículos o permiten la realización de un test para la detección temprana de la tartamudez. La aplicación desarrollada, en cambio, proporciona las siguientes funcionalidades:

- Un módulo de aprendizaje con consejos y contenidos multimedia didácticos para mejorar la fluidez verbal.
- Un módulo de juegos para mejorar tanto la fluidez verbal, como el control de la respiración de una forma más práctica y entretenida.
- Un módulo de entrenamiento para la lectura de textos en voz alta, a partir de la selección de un sistema de reconocimiento, grabación y síntesis de de voz, o de un sistema de retroalimentación auditiva de la señal de voz con un tiempo de retardo.
- Un módulo de relajación que facilita mejorar el estado de ánimo.

Con la intención de conseguir una aplicación que consuma el menor número de recursos disponibles del dispositivo móvil, se ha integrado una base de datos en la nube, que almacena el mayor número de contenidos posibles que conforman la interfaz gráfica de la aplicación (por ejemplo, los textos que se muestran en pantalla). Además, esta base de datos almacena las consultas solicitadas al usuario, los diferentes estados de control y resultados procedentes de las actividades realizadas por el usuario, necesarias para que el interfaz conversacional pueda adaptar la interacción dependiendo de la evolución, preferencias y requisitos de cada usuario.

2 App desarrollada

La aplicación desarrollada está conformada por un total de siete módulos principales. En el módulo inicial es necesario permitir el acceso al micrófono para grabar los audios de los usuarios. En caso de que el usuario utilice la aplicación por primera vez, se le solicita unos datos para su registro en la base de datos de la aplicación en la plataforma para el desarrollo de aplicaciones web y aplicaciones móviles Firebase[1].

El segundo módulo de la aplicación está conformado por un agente conversacional diseñado en la plataforma Dialogflow[2], que interactúa con el usuario en función de los valores almacenados en la base de datos, la actividad previa del usuario y última fecha de uso de la aplicación, su estado emocional y el modelo de diálogo diseñado para que el agente interactúe con el usuario según el tipo de actividad propuesta por la aplicación.

El módulo de estado emocional se encuentra conectado al servicio web de analítica de textos proporcionado por Meaningcloud[3] para que el agente conversacional pueda estimar el estado de ánimo del usuario. El estado emocional se obtiene a partir del análisis de las respuestas proporcionadas por el usuario para una serie de preguntas extraídas del

[1] https://firebase.google.com/
[2] https://dialogflow.com/
[3] https://www.meaningcloud.com/es

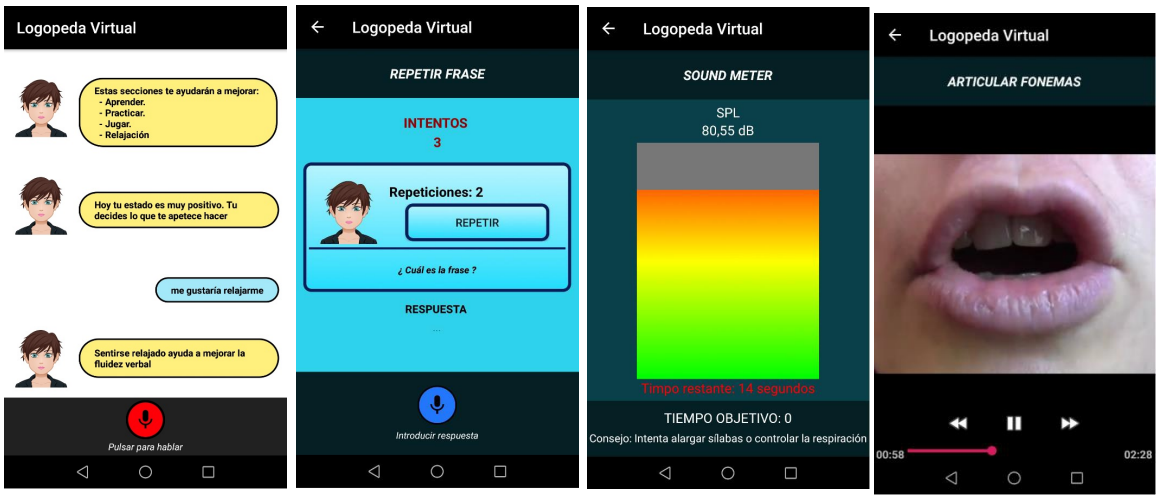

Figura 1: Capturas de pantalla de la app (I): asistente conversacional y módulo de práctica

Figura 2: Capturas de pantalla de la app (II): módulo de relajación, módulo de juegos y módulo de aprendizaje

nodo correspondiente en la base de datos. El usuario debe ir contestando a cada una de las preguntas por medio de la voz. Para garantizar que el sistema de transcripción de la voz ha funcionado de forma correcta, la app posibilita visualizar la transcripción y confirmar la respuesta antes de pasar a la siguiente pregunta. El resultado final de evaluación del estado de ánimo del usuario se extrae de una función interna de la aplicación, que se encarga de realizar un sistema de puntuaciones en función de la valoración del texto generada por Meaningcloud. A partir de los resultados obtenidos del test emocional, el asistente virtual recomendienda al usuario acciones adecuadas a su estado de ánimo.

Para los casos en los que el estado emocional sea negativo o muy negativo, el asistente conversacional inicia de forma automática el módulo de relajación, con el objetivo de mejorar el estado de ánimo del usuario. Este módulo está conectado al nodo de control de actividad y facilita ejercicios de relajación de corta duración basados en audios y vídeos. Una vez terminada la actividad, se redirige al usuario a la pantalla de interacción con el asistente conversacional.

El módulo de juegos está conformado por juegos que ayudan a mejorar la respiración y la fluidez verbal. Estos juegos extraen los enunciados y las soluciones del nodo de la base de datos que administra los contenidos de los mismos. Además, este bloque está conectado al nodo de control de actividad del

usuario que también, va a almacenar los resultados obtenidos tras su correspondiente finalización. Se han integrado tres tipos de actividades: ordenar palabras para formar una frase con sentido, sonómetro para solicitar al usuario mantener la presión sonora durante un determinado tiempo, repetición de la frase recitada por el asistente virtual. Para este tipo de actividades se han definido ejercicios de tres grados de dificultad (baja, media y alta). En todos ellos, el usuario interactúa con la aplicación a través de la voz.

El módulo de práctica permite almacenar textos introducidos por el usuario en la base de datos a través del nodo encargado de la gestión de textos. Estos textos se utilizan posteriormente en un sistema de grabación de voz o en un sistema de retroalimentación retardada, de tal modo que el usuario pueda realizar prácticas de lectura en voz alta para mejorar la fluidez verbal. En el modo de grabación, el usuario puede grabar su voz para posteriormente reproducirla. Cada vez que se utilice este modo se genera un archivo multimedia con formato 3GP que se almacena en el el dispositivo móvil, pero que puede ser reemplazado cada vez que se realice una nueva grabación del mismo texto, con el objetivo de no consumir grandes cantidades de memoria que limiten el dispositivo. En el modo de retroalimentación auditiva, el usuario podrá ir escuchando su propia voz con un tiempo de retardo mientras realiza la grabación de su mensaje. Para que este modo sea de utilidad, el usuario deberá de utilizar unos auriculares durante la práctica de lectura en voz alta.

Finalmente, el módulo de aprendizaje está conformado por actividades con contenidos multimedia que permitan al usuario mejorar su fluidez verbal y controlar la respiración. Este módulo también se encuentra conectado a un nodo de la base de datos que contiene una lista de consejos y con el nodo de la base de datos encargado de controlar la actividad del usuario. Se han incluido tres tipos de actividades: articulación de fonemas (contiene archivos de vídeo que muestran la articulación de los diferentes sonidos vocálicos y consonánticos que conforman el abecedario a través de ejemplos prácticos), respiración (contiene archivos de vídeo cuyos contenidos enseñan al usuario los beneficios de practicar una respiración correcta y controlada) y consejos prácticos. Las Figuras 1 y 2 muestran diferentes capturas de pantalla de los módulos de la app descritos en esta sección.

3 Conclusiones y trabajo futuro

En este artículo hemos descrito una aplicación móvil cuya funcionalidad es ayudar a mejorar los problemas de comunicación relacionados con la fluidez verbal debidos a la tartamudez o disfemia. Con respecto a las diferentes aplicaciones relacionadas con la tartamudez que se pueden encontrar en Google Play Store, la aplicación desarrollada incorpora funcionalidades novedosas y distintivas de las existentes, destacando la implementación de un agente conversacional con la capacidad de interactuar, motivar y basarse en el estado de ánimo del usuario para recomendar actividades de relajación, práctica de juegos y actividades de aprendizaje y mejora. Como trabajo futuro, además de extender el número de actividades de la aplicación, queremos validar la aplicación con usuarios que padecen tartamudez, pues hasta ahora solamente hemos realizado una evaluación preliminar con cuatro pacientes.

Bibliografía

Fundación Española de la Tartamudez. 2019. https://www.fundacionttm.org/la-tartamudez/mas-informacion/. (Consultado: Marzo 2019).

Griol, D., Z. Callejas, R. López-Cózar, y G. Riccardi. 2014. A domain-independent statistical methodology for dialog management in spoken dialog systems. *Computer Speech and Language*, 28(3):743–768.

Jackson, E., N. Rodgers, y D. Rodgers. 2019. An exploratory factor analysis of action responses to stuttering anticipation. *Journal of Fluency Disorders*, 60:1–10.

Jaramillo, J. H. y Y. Gil-Lozada. 2014. Efectos de la retroalimentación auditiva retardada en los patrones de tartamudez. *Revista Ciencias de la Salud*, 12(2):243–251.

McTear, M., Z. Callejas, y D. Griol. 2016. *The conversational interface: Talking to smart devices*. Springer.

Salgado-Ruíz, A. 2005. *Manual práctico tartamudez*. Síntesis.