Lumen Gentium Vol. 4 No. 2
ISSN 2539 - 0678
e-ISSN 2590-8714X

Sistema de reconocimiento de voz para controlar la aplicación WhatsApp orientado a personas con limitaciones motrices*

Voice recognition system to control the WhatsApp application for people with motor disabilities

PP. 101-115

José Mosquera De la Cruz[1]
Carlos Ferrin Bolaños[2]
Daniel Santander Ariza[3]
Kevin Rosero Ramos[4]
Miguel Libreros Segura[5]
Humberto Loaiza Correa[6]

Rec: 23/09/2020
Acep: 22/11/2020

Resumen

El trabajo describe una interfaz que permite la interacción entre una persona con limitaciones motrices y la aplicación WhatsApp Desktop mediante comandos de voz. Los alcances y limitaciones que se cuantificaron, establecieron resultados significativos en las pruebas realizadas; del 91.7%, en desempeño del sistema en reconocimiento de voz; 93% en indagación para dictados y la prueba de rendimiento de la interfaz sobre la aplicación WhatsApp Desktop, obtuvo un porcentaje de acierto del 91,16 %. Finalmente, se realizó una encuesta para cuantificar la percepción de los usuarios luego de interactuar con la interfaz

* Artículo de investigación, proyecto "Asistente virtual para personas tetrapléjicas controlado por voz y gestos faciales (fase I)", Fundación Universitaria Católica Lumen Gentium.
1 Magister Ingeniería Electrónica Universidad del Valle, Ingeniero Electrónico Universidad del Valle. Correo electrónico: jhmosquerad@unicatolica.edu.co; Orcid: https://orcid.org/0000-0002-9324-1692
2 Magister Ingeniería Electrónica de la Universidad del Valle, Ingeniero Físico de la Universidad del Cauca. Correo electrónico: cdferrinb@unicatolica.edu.co; Orcid: https://orcid.org/0000-0002-2739-9205
3 Estudiante de Ingeniería de Sistemas en la Fundación Universitaria Católica Lumen Gentium. Miembro del grupo de investigación Khimera, Facultad de Ingeniería de la Fundación Universitaria Católica Lumen Gentium, Unicatólica, Cali, Colombia. Correo electrónico: daniel.santandero1@unicatolica.edu.co
4 Estudiante de Ingeniería de Sistemas en la Fundación Universitaria Católica Lumen Gentium. Miembro del grupo de investigación Khimera, Facultad de Ingeniería de la Fundación Universitaria Católica Lumen Gentium, Unicatólica, Cali, Colombia. Correo electrónico: kevin.rosero01@unicatolica.edu.co
5 Estudiante de Ingeniería de Sistemas en la Fundación Universitaria Católica Lumen Gentium. Miembro del grupo de investigación Khimera, Facultad de Ingeniería de la Fundación Universitaria Católica Lumen Gentium, Unicatólica, Cali, Colombia. Correo electrónico: miguel.librero501@unicatolica.edu.co
6 PhD. en Robótica de la Université d Evry Val d Essonne, M.Sc. en Automática de la Universidad del Valle, Ingeniero Eléctrico de la Universidad del Valle. Director del grupo de investigación Percepción y Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad del Valle. Cali, Colombia. Correo electrónico: humberto.loaiza@correounivalle.edu.co; Orcid: https://orcid.org/0000-0001-7206-7333

desarrollada, en general un 40,6 % y un 23,2 % de los sujetos de prueba estuvieron de acuerdo y totalmente de acuerdo en que la interfaz es una herramienta útil y mejora la calidad de vida de personas con limitaciones motrices.

Palabras clave: comandos de voz, interfaz humano-computador, rehabilitación, sintetizador de voz.

Abstract

The article describes an interface that allows interaction between a person with motor limitations and the WhatsApp Desktop application through voice commands. The scope and limitations that were quantified established significant results in the tests carried out; 91.7%, in performance of the voice recognition system; 93% in inquiry for dictations and the performance test of the interface on the WhatsApp Desktop application, obtained a success rate of 91.16%. Finally, a survey was conducted to quantify the perception of users after interacting with the developed interface, in general 40.6% and 23.2% of the test subjects agreed and totally agreed that the interface it is a useful tool and improves the quality of life of people with motor limitations.

Keywords: human-computer interface, rehabilitation, voice command, voice synthesizer.

Introducción

En la actualidad, las computadoras aportan soluciones para una gran parte de las necesidades del ser humano; las computadoras están casi en todos los ámbitos de nuestras vidas, empresas, bancos, escuelas, universidades y hogares, con lo cual suponen una interacción y un flujo de datos mediante un sistema de cómputo. La gran acogida de la computadora ha sido posible gracias a la disminución de costos y a la gran acogida del internet, el cual permite una comunicación sin barreras y facilita la expansión del conocimiento al llegar hasta las zonas más remotas del mundo (Chayapathy *et al*., 2018; Bose *et al*., 2017).

Lo anterior demuestra que el desempeño de los equipos de cómputo ha mejorado exponencialmente a la par con los avances en ingeniería electrónica. Sin embargo, se puede denotar la poca actualización de los periféricos con los cuales el humano interactúa con la máquina, aún se conserva el *mouse* y teclado como principales medios de comunicación con el ordenador, lo cual desencadena una experiencia de uso poco natural y se convierte en una limitante para personas con alguna discapacidad que les impida tener contacto físico con estos periféricos (Mosquera-De la Cruz *et al*., 2017). Por esta razón es necesario concebir nuevos mecanismos de interacción, como los comandos de voz que permitan una comunicación natural entre las personas con limitaciones motrices y la tecnología (Mosquera-De la Cruz *et al*., 2020).

Antecedentes

El desarrollo de un asistente virtual basado en voz orientado a la manipulación del computador por personas con limitaciones motrices (específicamente pacientes con lesiones medulares a nivel cervical C3-C6) presenta varios desafíos (Kepuska y Bohouta, 2018), algunos se mencionan a continuación.

El primer desafío está relacionado con la adaptación de los asistentes virtuales a las nuevas formas de pronunciación de

los fonemas por parte de los pacientes con tetraplejia, en algunos casos los pacientes deben contar con una traqueotomía que ayuda a la respiración cuando el camino habitual presenta alguna complicación (Mertl *et al*., 2017) (Pampoulou, 2018). Sin embargo, este procedimiento quirúrgico modifica el flujo de aire a través de las cuerdas bucales y cambia las características frecuenciales y de amplitud durante el habla, lo cual se reduce a una mayor robustez en el sistema de reconocimiento de voz (Lu y Renals, 2017).

Un segundo desafío se encuentra en la calibración o configuración de los asistentes virtuales, especialmente en el caso de los pacientes tetrapléjicos se procura tener una manipulación exacta del cursor a partir de la voz y es un gran reto debido a la gran cantidad de posibilidades y tiempo en la ubicación a detalle requeridos al generar posiciones efectivas. Esto bajo el entendimiento de que una calibración supervisada de los movimientos del cursor requiere un lazo cerrado de realimentación para alcanzar la exactitud, por esto es una necesidad latente concebir asistentes virtuales que calibren con la menor cantidad de datos posible y el mejor tiempo la ubicación del cursor para optimizar esta funcionalidad de forma no supervisada (Basharirad y Moradhaseli, 2017; Nur *et al*., 2018).

Un tercer desafío está relacionado con la realimentación que existe entre el paciente y el sistema de inteligencia detrás de los asistentes virtuales. Esta tarea es muy importante pues permite corregir o dar cuenta a la máquina de falsos positivos o verdaderos negativos buscando un aprendizaje activo de la máquina que permita perfeccionarse mientras interactúa con el usuario (Pardowitz *et al*., 2007).

En esta investigación se parte de los desafíos mencionados y de la necesidad de implementar nuevos mecanismos de interacción entre las personas y los computadores, se propone desarrollar una interfaz de computadora dirigida por comandos de voz que permita a las personas con limitaciones motrices realizar tareas básicas de interacción con la aplicación WhatsApp Desktop. Para lograr este objetivo se implementó el sistema *software* y se propuso un protocolo de pruebas (Hartson y Pyla, 2012) que después de ser ejecutado permitió establecer los alcances y limitaciones de la interfaz propuesta.

Referencias teóricas

Interacción humano-máquina

La interacción que se lleva a cabo entre humano y máquina supone una doctrina que se enfoca en definir sistemas computacionales que apoyan en las tareas diarias del ser humano. El mundo de la tecnología avanza muy rápido, por lo tanto, trae consigo cambios importantes con los cuales se han podido desarrollar varias formas de interacción, tales como interfaces táctiles, hápticas, gestuales y de reconocimiento de voz. En otras palabras, la interacción humano-máquina brinda al usuario una forma de interactuar más natural y novedosa, a diferencia del formato clásico como el *mouse* y teclado.

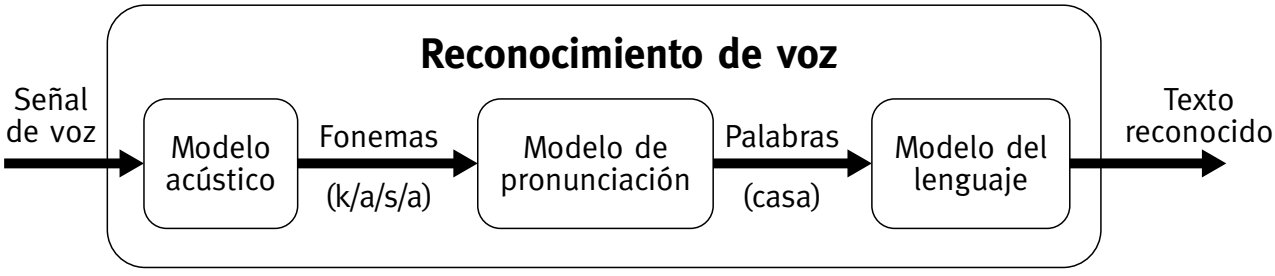
La interacción entre el humano y las máquinas depende de un intercambio de información en ambas direcciones entre el operario y el sistema, usualmente se considera

que el operario controla todas las acciones de la máquina por medio de la información que introduce y las acciones que realiza sobre esta, pero también es necesario considerar que el sistema retroalimenta con cierta información al usuario, por medio de distintas señales, para indicar el estado del proceso o las condiciones del sistema (Chizari *et al*., 2013; Hartson y Pyla, 2012; Oliveira *et al*., 2017).

Reconocimiento de voz

Corresponde al proceso de capturar señales de voz generadas por el usuario a través de un micrófono y transformarlas a texto. Este sistema se compone de tres modelos como se presenta en la figura 1. El modelo acústico inicia con un comportamiento probabilístico al recibir la señal de voz y como resultado muestra la distribución de los fonemas reconocidos. Seguido de ello, el modelo de pronunciación toma estos fonemas y realiza una búsqueda dentro de un diccionario con un idioma establecido para formar las palabras. Finalmente, para que el texto reconocido tenga un sentido lógico, el modelo del lenguaje organiza las palabras y muestra la secuencia que más se ajusta a la señal de voz (Anand *et al*., 2013; Sharma y Kumari, 2014).

Figura 1.
Reconocimiento de voz

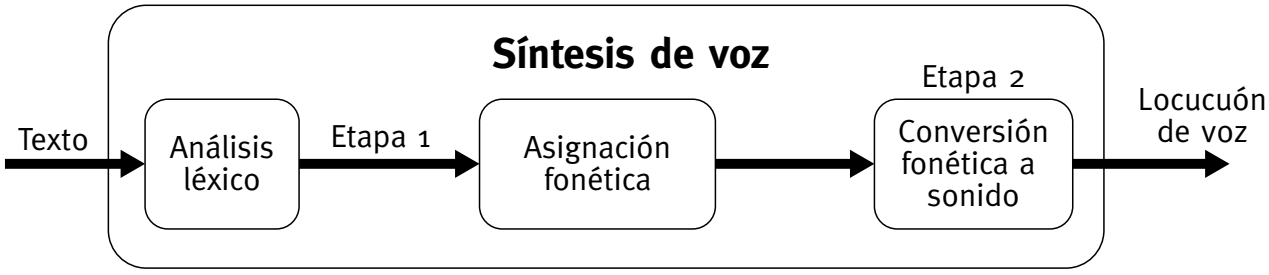


Fuente: Mosquera-DeLaCruz *et al*. (2017)

Síntesis de voz

Su funcionalidad es la de convertir un texto escrito en una locución artificial de voz mediante una concatenación de sonidos pregrabados. La efectividad de lograr una similitud en las características de la voz depende de su almacenamiento de palabras enteras u oraciones que dan el contexto a la frase. Su estructura normalmente está compuesta por dos etapas internas como se presenta en la figura 2 en la cual se puede observar que la primera etapa se divide en dos bloques: el texto ingresado se convierte en subunidades fonéticas gracias a un análisis léxico, seguido del bloque de asignación fonética en el que se reconocen las sílabas, hiatos, diptongos, etc. Finalmente, la segunda etapa unifica las divisiones fonológicas para formar las palabras y convertirlas en sonidos controlando el tono y la duración de los fonemas para dar una síntesis de mayor calidad (Bose *et al*., 2017).

Figura 2.
Síntesis de voz



Fuente: Mosquera-DeLaCruz *et al*., (2017)

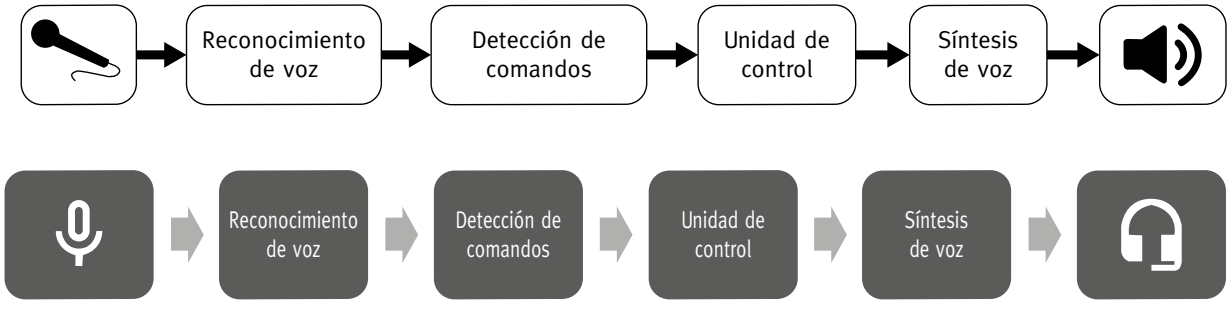
Acceso a la información para personas con limitaciones motrices

El desarrollo de este proyecto de investigación es fundamental para mejorar la calidad de vida de las personas tetrapléjicas que han perdido una de las partes más importantes de la vida: "la autonomía", la libertad de poder realizar acciones valiéndose por sí mismas y de acceder a la información del mundo mediante internet (Boman *et al*., 2015; Lancioni *et al*., 2018; Lopes *et al*., 2014; Wang *et al*., 2017).

Metodología

La figura 3 presenta el diagrama de bloques de la interfaz propuesta. El primer bloque representa el reconocimiento de voz que se encarga de convertir la señal de voz en una cadena de texto. El segundo bloque detecta si la cadena de texto contiene o no los comandos establecidos para el manejo de la aplicación, de lo contrario, la interfaz asimila que es un texto libre que se activa durante un dictado. En el bloque de unidad de control se ejecuta el comando de forma automática en el sistema operativo. Finalmente, el bloque de síntesis de voz genera una realimentación al usuario que confirma el comando identificado y ejecutado.

Figura 3.
Diagrama de bloques de la interfaz propuesta



Fuente: Santander-Ariza *et al*. (2020)

Reconocimiento de voz

Este es el bloque que se encarga de ser intermediario entre el usuario y el aplicativo. Para su principal funcionamiento se utiliza la librería *Speech Recognition*, la cual constantemente escucha al usuario y reconoce el contenido del mensaje, retornando una cadena de texto que se utiliza para reconocer los comandos presentes en el mensaje. En la figura 4 se presenta la configuración de la librería utilizada.

Figura 4.
Librería Speech Recognition

```
import speech_recognition as sr

global r
r = sr.Recognizer()
global source
with sr.Microphone() as source:
    r.adjust_for_ambient_noise(source, duration = 1)
    audio = r.listen(source,phrase_time_limit = None)
    try:
        response = r.recognize_google(audio, language="es-CO")
        print("Entendi: '" + response + "'")
    except sr.UnknownValueError:
        print("No te entendi")
    except sr.RequestError as e:
        print("GSR; {0}".format(e))
        print("No logre reconocer ningun audio")
```

Fuente: Santander-Ariza *et al*., (2020)

La figura 4 muestra que para hacer uso del reconocimiento de voz es necesaria la librería *PyAudio*, la cual se encarga de crear los canales para *PortAudio* y le permite a *Python* grabar o reproducir audio. Esta herramienta empieza su reconocimiento con una toma de muestra del ruido ambiente mediante la función *adjust_for_ambient_noise* que cumple una función importante, ya que escucha el espectro de energía en el ambiente y permite tener una referencia de ruido que se debe excluir más adelante. Al ser una acción parametrizable, se estableció un segundo (1s) de escucha, pues se identificó que, entre más tiempo, el proceso de reconocimiento de voz se torna lento y no ofrece una buena experiencia al usuario. La función *listen* obtiene el audio del micrófono y la función *recognize_google* es la encargada de reconocer el mensaje pronunciado por el usuario. Esta necesita de una conexión a internet para poder reconocer la voz, ya que funciona mediante el API de Google. Una vez ejecutado, este retorna la cadena de caracteres que corresponde a las palabras pronunciadas por el usuario "El texto reconocido".

Detección de comandos

Para la detección de comandos el primer paso es recibir la cadena de texto y compararla con los comandos establecidos. En esta investigación se desarrolló un diccionario de 17 comandos, los cuales se presentan en la primera columna de la tabla 1.

Tabla 1.
Lista de comandos y acciones

Reconocimiento de voz	Síntesis de voz	Comando ejecutado
Abrir whatsapp	"abriendo WhatsApp"	os.system ('WhatsApp.exe')
Arriba	"subiendo al siguiente chat"	'ctrl' + 'shift' + 'tab'
Abajo	"bajando siguiente chat"	'ctrl' + 'shift'
Bajar	"bajando"	auto.scroll(-10)
Subir	"subiendo"	auto.scroll(10)
Minimizar	"minimizando WhatsApp"	'win' + 'm'
Restaurar o maximizar	"abriendo WhatsApp"	'win' + 'shift' + 'm'
Aumentar letra o acercar	"aumentando letra"	'ctrl' + '+'
Reducir letra o alejar	"reduciendo letra"	'ctrl' + '-'
Nuevo	"creando nuevo chat"	'ctrl' + 'n'
Perfil	"mostrando perfil"	'ctrl' + 'p'
Escribir + texto	"claro, escribiendo" + texto	auto.typewrite (texto)
Atrás o escape	"atrás"	'esc'
Eliminar conversación	"eliminando conversación"	ctrl' + 'backspace'
Aceptar o Enter	"aceptar"	'enter'
Chat de + nombre de la persona	"muy bien"	'ctrl' + 'f' auto.typewrite (nombre)
Borrar todo	"muy bien"	'ctrl' + 'a' 'backspace'

Fuente: Santander-Ariza *et al.* (2020)

A nivel algorítmico se usó la función *in()*, esta función es nativa de *Python* y no pertenece a ninguna librería, se utilizó con el objetivo de comparar una palabra o carácter en una cadena, esta retorna un valor booleano, donde *true* es si la palabra o carácter está contenida en la cadena o *false* si no lo está, un ejemplo de cómo funciona la función se presenta en la figura 5.

Figura 5.
Implementación de la función in()

```
texto1 = "manzana"
texto2 = "manzana"
resultado = texto1 in texto2;
print(resultado)
 >>> True
```

Fuente: Santander-Ariza *et al.* (2020)

Control del sistema operativo

Este módulo de permite comandar el computador sin tener contacto físico, emula el presionar combinaciones de teclas (las cuales se generadas como falsas interrupciones al sistema operativo). Las combinaciones de teclas asignadas para cada uno de los comandos implementados se presentan en la columna 3 de la tabla 1.

Para implementar el control en el sistema operativo a nivel algorítmico se utilizaron dos librerías. La primera es la librería OS que pertenece a la biblioteca estándar de *Python*. Una de sus funciones llamada *system* permitió ejecutar la aplicación de WhatsApp Desktop. Luego de tener la aplicación de escritorio abierta se utilizó la librería *pyautogui*, de la cual se utilizaron las funciones *press* y *hotkey* que emulan la acción de presionar las teclas de forma virtual.

Para ejemplificar mejor el uso de estas librerías, en la figura 6 se presenta un ejemplo algorítmico el cual realiza tres acciones: la primera es abrir la aplicación WhatsApp, la segunda es presionar la tecla de 'esc' (*escape*) para retroceder en el aplicativo y por último, una combinación de teclas para eliminar.

Figura 6.
Automatización del Sistema Operativo

```
def abrir1():
    os.system('WhatsApp.exe')

def escape():
    auto.press('esc')

def borrar():
    auto.hotkey('ctrl', 'backspace')
```

Fuente: Santander-Ariza *et al*. (2020)

En la figura anterior la función *os.system()* permite ejecutar comandos dados en forma de cadena de texto en una *subshell* y definir la ruta de la aplicación como una variable del sistema. La función *pyautogui.press()* permite ejecutar de manera remota sin presionar físicamente una tecla específica, se pude presionar también un conjunto de teclas, esto se realiza mediante la función *hotkey* con la cual se puede emular todo el teclado del computador.

Síntesis de voz

Este algoritmo es el encargado de dar una realimentación auditiva del aplicativo y va dirigida al usuario que le da a entender el comando que fue reconocido y ejecutado. La lista de frases sintetizadas por cada comando reconocido se presenta en la columna 2 de la tabla 1. Su funcionamiento necesita de una cadena de texto la cual pronuncia el sintetizador. En la figura 7 se presenta un ejemplo de configuración de la librería utilizada.

Figura 7.
Implementación de la Librería Pyttsx3

```
import pyttsx3
tts = pyttsx3.init()

def hablar(nombre):
  tts.say("Hola")
  tts.say("como estas "+nombre+"?")
  tts.runAndWait()

hablar('Daniel')
```

Fuente: Santander-Ariza *et al.* (2020)

Como se observa en la figura anterior, para hacer uso de la librería es necesario realizar su inicialización con la función *pyttsx3.init( ),* esta permite crear un entorno de vocalización, luego la función *say()* permite vocalizar y estructurar el texto que se va a sintetizar. También se observa que esta herramienta permite pronunciar más de una línea usando la función *say( )*, e internamente crea una cola de textos a la espera de su reproducción por la función *runAndWait()*. Esta última realiza la confirmación que libera el texto sintetizado, si existe una cola de textos se reproduce una pequeña pausa entre cada uno de los textos pronunciados.

Pruebas y resultados

Inicialmente se realizó un análisis de los sujetos de prueba mediante encuestas para definir el grado de familiaridad con la navegación en internet, la aplicación WhatsApp y el uso de reconocimiento de voz.

La primera prueba evaluó el desempeño del reconocimiento de comandos, se estudió un total de 26 comandos repetidos tres veces por cada usuario, se analizó el número de comandos acertados y a su vez, se buscó el comando con menos aciertos. La segunda permitió evaluar el reconocimiento de dictados. Se definió previamente un mensaje que cada usuario repitió tres veces y permitió identificar las limitaciones de dicha herramienta, a su vez dio a conocer posibles errores futuros y las palabras con dificultad de reconocimiento.

Una tercera evaluó el comportamiento de los usuarios ante una situación controlada, se busca una integración entre todos los bloques de la aplicación y simular un uso cotidiano. Se pidió a los usuarios que siguieran una rutina en la que ellos pudieran abrir la aplicación, enviar mensajes y navegar entre las distintas conversaciones.

Finalmente, se buscó realizar un análisis cualitativo en el que se implementó una encuesta de diez preguntas dirigida a los sujetos de prueba para evaluar su experiencia de usuario y percepción sobre el funcionamiento de la interfaz desarrollada.

Sujetos de prueba

Las pruebas se realizaron a un grupo de siete personas, cada una aprobó y firmó el formato de consentimiento informado utilizado. Inicialmente se construyó una encuesta para medir el grado de familiaridad de los usuarios con las temáticas tratadas y se evaluaron mediante una escala tipo Likert (Matas, 2018; Sánchez y Terrats, 2011) en la cual las posibles opciones de respuesta eran: nunca, casi nunca, algunas veces, casi siempre y siempre.

Por otro lado, las preguntas realizadas incluían edad, uso diario de internet, WhatsApp y sistemas de reconocimiento de voz. En los resultados obtenidos se observó que el rango de edad de los usuarios oscilaba entre 20 y 54 años, además se calculó el porcentaje de familiaridad de los usuarios con internet, WhatsApp y reconocimiento de voz, la información se presenta en la tabla 2.

Tabla 2.
Resultado porcentual sobre el grado de familiaridad de los usuarios de prueba

Usuario	Casi nunca (%)	Algunas veces (%)	Casi siempre (%)	Siempre (%)
Internet	0	14,29	28,57	57,14
WhatsApp	0	14,29	14,29	71,43
Reconocimiento de voz	57,14	42,86	0	0

Fuente: Tomado de Santander-Ariza *et al*. (2020)

En los resultados obtenidos se observó que el grupo de personas cuya edad es menor a 25 años tienen un uso de internet más amplio, con un total del 57,14 % respecto al 42,86 % del segundo grupo que tienen una edad superior a 38 años y que la usan algunas veces en su día a día.

Las pruebas muestran que la mayoría de los usuarios usan siempre y casi siempre la aplicación de WhatsApp, lo que da un total del 85,59 %, pues se comunican a diario por motivos personales o laborales; el resto de las personas lo usan algunas veces, pero no dejan de usar la aplicación.

En los resultados se identifica que los usuarios no usan con frecuencia los comandos de voz. Un 42,86 % de los usuarios votaron algunas veces para referenciar que la aplicación la usan en ocasiones específicas y esto se evidencia en personas mayores de edad que constituyen el 28,57 % de la opinión respecto al 14,28 % que son los jóvenes.

Prueba de reconocimiento de comandos

En esta prueba se solicitó a los voluntarios pronunciar los 26 comandos utilizados en la interfaz propuesta y repetir tres veces cada uno de ellos, se evaluó un total de 546 comandos.

En los resultados se identificó que los comandos “activar modo prueba” y “desactivar modo prueba” presentaron el menor desempeño con un 57,14 % y 52,38 % respectivamente. Una hipótesis sobre este resultado está ligada a que presentan una estructura fonética similar que posibilita que el algoritmo confundiera entre sí las palabras “desactivar” y “activar”, además se observó una falencia en el reconocimiento al presentar palabras terminadas en “ar”. En la prueba se identificó que a medida que los usuarios repetían los comandos se empezaban a obtener mejores porcentajes de acierto; el primer usuario tuvo un 82,05 % de aciertos y el último un 96,15 %, esto muestra que las personas tomaban como referencia el resultado de los primeros voluntarios y mejoraban el tono y tiempo de su pronunciación.

En esta primera prueba se observó que 14 comandos de los 26 totales lograron un 100 % de aciertos y subieron mucho el promedio

general de la prueba. Obtuvieron así un 91,76 % que se ve reflejado en 504 comandos acertados frente a los 42 que fallaron.

Prueba de reconocimiento de dictados

En esta prueba se solicitó a los voluntarios pronunciar un mensaje ya definido previamente y repetirlo tres veces, el dictado evaluado fue:

"Mantén tu teléfono conectado, WhatsApp se conecta a tu teléfono para sincronizar los mensajes. Para reducir el consumo de tus datos, conecta tu teléfono a una red Wi-Fi".

Es necesario mencionar que esta prueba se realizó bajo las mismas condiciones ambientales de 53,7 dB medidos con la aplicación Sonómetro (Sound Meter) (Google Play, 2020). El dictado fue evaluado comparando palabra por palabra entre las reconocidas por el sistema y el dictado original que contaba con 28 palabras.

En los resultados obtenidos por el reconocimiento del dictado de prueba se observó un acierto del 93 % con una desviación estándar del 2 %. Se evidenció que el voluntario con menor desempeño tuvo un resultado de 91 % de aciertos respecto al usuario con mejor desempeño que fue de 95 %. Por otro lado, se identificó que "Wi-Fi" se reconoció muy pocas veces debido a que está compuesta por siglas, se evidencian así las dificultades que presenta la herramienta para reconocer abreviaturas o palabras en inglés si el usuario no tiene una buena pronunciación.

Prueba de la interfaz para controlar la aplicación WhatsApp Desktop

En esta prueba se solicitó a los voluntarios abrir la aplicación de mensajería mediante el comando "abrir WhatsApp", el aplicativo responde de forma verbal que el proceso de abrir la aplicación se está ejecutando. Como siguiente paso el usuario dictó el comando "enviar mensaje a Daniel", allí el aplicativo responde preguntando por el mensaje que desea enviarle al contacto. Seguido de esto se pidió al usuario pronunciar la frase "hola, ¿cómo estás?" Con ella, el aplicativo filtra por el nombre "Daniel" en la lista de contactos, pega el mensaje dicho por el usuario y finalmente envía el mensaje automáticamente.

Para un segundo caso se pidió al usuario eliminar la conversación mediante el comando "eliminar conversación", el aplicativo responde de forma verbal que la conversación fue eliminada. Luego se pidió navegar entre los distintos chats que tenían iniciados, mediante el comando "ir al chat de Daniel" o simplemente "chat de Daniel", por consiguiente, el usuario era dirigido hacia el chat del contacto. Ya dentro del chat presente se pidió usar la opción de enviar mensajes cortos con el comando "escribir" y seguido, dictar el mensaje "Hola, ¿cómo estás?" El aplicativo añade el mensaje dicho por el usuario y con el comando "aceptar" se confirma y envía el mensaje para así finalizar la prueba. A continuación, en la tabla 3 se muestra el protocolo de pruebas.

Tabla 3.
Protocolo de pruebas con la aplicación WhatsApp Desktop

Tipo de interacción	Contenido de la interacción
Comando	Abrir WhatsApp
Comando	Enviar mensaje a Daniel
Dictado	¿Hola, cómo estás?
Comando	Eliminar conversación
Comando	Ir al chat de Daniel / Chat de Daniel
Dictado	Escribir, Hola, ¿cómo estás?
Comando	Aceptar / Enter

Fuente: Santander-Ariza *et al*. (2020)

Los resultados muestran que las acciones que obtuvieron un mejor desempeño fueron: "Abrir WhatsApp", "Eliminar conversación",

"Escribir" y "Aceptar". Por el contrario, los comandos que obtuvieron menor resultado fueron: "Enviar mensaje a" e "Ir al chat de". Esto demuestra que si el usuario tiene dos contactos con el mismo nombre el aplicativo selecciona al primero de la lista y provoca el envío del mensaje a una persona incorrecta.

Finalmente, en la prueba se obtuvo un resultado general del 91,16 % de aciertos frente al 8,84 % de error promedio, el mayor número de comandos erróneos por usuario fueron dos. También se observó que la primera persona en realizar la prueba tuvo un menor resultado de acierto con un 80,95 % debido a que presentó un ritmo acelerado al pronunciar los comandos frente a la última persona que tuvo un acierto del 95,24 % porque realizó las pruebas con calma.

Evaluación de la experiencia de usuario

Una vez terminadas las pruebas de interacción, se le solicitó a cada usuario de prueba responder las preguntas de una encuesta para medir la experiencia del usuario al interactuar con la interfaz propuesta. Se utilizó como métrica de evaluación la misma escala de Likert (Matas, 2018; Sánchez y Terrats, 2011). La encuesta que se realizó fue la siguiente:

1. La interfaz reconoce los comandos de voz.
2. La interfaz reconoce los dictados.
3. La interfaz ejecuta bien los comandos.
4. La interfaz es de fácil uso para el usuario.
5. La interfaz se puede utilizar en cualquier ambiente social sin interferencia en el reconocimiento de voz.
6. La interfaz cumple el objetivo de realizar tareas básicas de navegación en la aplicación WhatsApp Desktop.
7. Con miras a las necesidades de las personas con tetraplejia, la interfaz es una herramienta que se puede recomendar.

En los resultados obtenidos se identificó que en promedio un 85,71 % de los usuarios encuestados estaban de acuerdo en que la interfaz reconoce los comandos y dictados de forma correcta, además el 91,62 % estuvo de acuerdo en que la interfaz ejecuta bien los comandos y el 78,49 % de los usuarios estuvo de acuerdo en que la interfaz es de fácil uso.

Se observó que el 62,85 % de las personas afirmaron que no es posible usar la aplicación en cualquier ambiente, recomiendan un espacio silencioso que les permita tener un uso estable de la aplicación (que sea sensible al ruido ambiente).

También se observó que el 90,17 % de los usuarios estuvo totalmente de acuerdo en que la interfaz permite realizar tareas básicas de navegación en la aplicación WhatsApp Desktop. Por otro lado, un 85,71 % estuvieron de acuerdo y totalmente de acuerdo en que el aplicativo puede ayudar a las personas con tetraplejía y reducir la limitación de comunicación con la tecnología.

La encuesta reunió observaciones constructivas expresados por los sujetos de prueba, un ejemplo de estas puede ser un soporte para *emojis* que les permita tener conversaciones más naturales y les permita expresar ciertos estados de ánimo en sus mensajes. Otra observación recomendó que los mensajes no se envíen de forma automática si no que permita confirmar para enviar el mensaje, ya que, si la interfaz filtró otro contacto por error o con

el mismo nombre, no hay opción de reversar dicha acción y afecta la experiencia de usuario. Finalmente, se recibieron muchos comentarios alentadores y positivos que afirmaban que para una persona que no tengan la posibilidad de usar un periférico tradicional la interfaz mejora su calidad de vida y es de mucha utilidad.

Conclusiones y recomendaciones

Se desarrolló una interfaz que mediante comandos de voz permite una emulación del teclado sin tener un contacto físico con el computador, dicha interfaz puede ser usada por personas con discapacidades motoras o que deseen controlar de forma verbal la aplicación WhatsApp Desktop. El sistema de reconocimiento de voz permitió realizar las tareas de navegación más comunes en la aplicación: enviar un mensaje, ir a una conversación en específico o crear una nueva conversación. Se realizaron pruebas con un grupo compuesto por siete usuarios (con diferentes edades y hábitos de uso de internet. En las pruebas se utilizaron 672 comandos de voz y 42 dictados, se obtuvo un mayor porcentaje de acierto en los dictados con un 93 % y un menor porcentaje en los comandos con un 91,76 % de aciertos.

En los comandos de voz se observó que las palabras con estructura fonética similar presentan fallos pues se confundían dos comandos entre sí con facilidad, un ejemplo es "activar" y "desactivar", al ser muy similares, si el usuario habla antes de que el micrófono se active puede llegar a confundir los comandos. También se observó como los que terminaban con "ar" tuvieron más fallos con un 7,70 % de las pruebas realizadas.

En los dictados se observó que no se reconocieron los signos de puntuación, pero sí se reconoció el texto pronunciado, a su vez, se pudo observar que existe una relación directamente proporcional en el reconocimiento de palabras en inglés y la pronunciación del usuario, ello evidencia una ayuda con poca asistencia a un lenguaje diferente al parametrizado.

El sistema de interacción por comandos de voz lo evaluaron por personas que no tenían limitaciones motrices y con base en su experiencia con el aplicativo, se concluyó que la herramienta puede mejorar la calidad de vida de personas con limitaciones motrices, pues le permite controlar la aplicación de mensajería sin utilizar el teclado o *mouse*, adicionalmente, tener más privacidad al momento e interactuar con otras personas sin asistencia de terceros.

Referencias

Anand, G., Geethamsi, S., Chary, R. V. R., y Madhu Babu, C. (2013). *Email access by visually impaired* [ponencia]. Proceedings - 2013 International Conference on Communication Systems and Network Technologies, CSNT, 597–601. https://doi.org/10.1109/CSNT.2013.128

Basharirad, B., y Moradhaseli, M. (2017). *Speech Emotion Recognition Methods : A Literature Review* [ponencia]. The 2nd International Conference on Applied Science and Technology 2017 (ICAST'17). https://doi.org/10.1063/1.5005438

Boman, T., Kjellberg, A., Danermark, B., y Boman, E. (2015). Employment opportunities for persons with different types of disability. *Alter, 9*(2), 116–129. https://doi.org/10.1016/j.alter.2014.11.003

Bose, P., Malpthak, A., Bansal, U., y Harsola, A. (2017, abril). *Digital assistant for the blind* [ponencia]. 2nd International Conference for Convergence in Technology, I2CT 2017 (2015), 1250–1253. https://doi.org/10.1109/I2CT.2017.8226327

Chayapathy, V., Anitha, G. S., y Sharath, B. (2018). *IOT based home automation by using perso-*

nal assistant [ponencia]. Proceedings of the 2017 International Conference On Smart Technology for Smart Nation, SmartTechCon 2017, 385–389. https://doi.org/10.1109/SmartTechCon.2017.8358401

Chizari, H., Lalanne, D., y Schwaller, M. (2013). *Combining Voice and Gesture for Human Computer Interaction*. Departement fur Informatik, Universite de Fribourg. https://diuf.unifr.ch/main/diva/sites/diuf.unifr.ch.main.diva/files/Haleh_Chizari_Master_report.pdf

Google Play. (2020). Sonómetro (Sound Meter) - Apps en Google Play. https://play.google.com/store/apps/details?id=com.gamebasic.decibel&hl=es_CO

Hartson, R., y Pyla, P. S. (2012). The UX book, process and guidelines for ensuring a quality user experience. *ACM SIGSOFT Software Engineering Notes, 37*(5). https://doi.org/10.1145/2347696.2347722

Kepuska, V., y Bohouta, G. (2018). *Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)* [ponencia]. 2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018, 99–103. https://doi.org/10.1109/CCWC.2018.8301638

Lancioni, G. E., Singh, N. N., Reilly, M. F. O., Sigafoos, J., Perilli, V., Chiariello, V., Grillo, G., Turi, C., Lancioni, G. E., Singh, N. N., Reilly, M. F. O., Sigafoos, J., Alberti, G., Perilli, V., Chiariello, V., Grillo, G., y A, C. T. (2018). A tablet-based program to enable people with intellectual and other disabilities to access leisure activities and video calls. *Disability and Rehabilitation: Assistive Technology, 15*(1), 14-20. https://doi.org/10.1080/17483107.2018.1508515

Lopes, N. V., Pinto, F., Furtado, P., y Silva, J. (2014). *IoT architecture proposal for disabled people* [ponencia]. International Conference on Wireless and Mobile Computing, Networking and Communications, 152–158. https://doi.org/10.1109/WiMOB.2014.6962164

Lu, L., y Renals, S. (2017). Small-footprint highway deep neural networks for speech recognition. *IEEE/ACM Transactions on Audio Speech and Language Processing, 25*(7), 1502–1511. https://doi.org/10.1109/TASLP.2017.2698723

Matas, A. (2018). Diseño del formato de escalas tipo Likert: Un estado de la cuestión. Revista Electronica de Investigacion Educativa, 20(1), 38–47. https://doi.org/10.24320/redie.2018.20.1.1347

Mertl, J., á ková, E., y epová, B. (2017). Quality of life of patients after total laryngectomy: the struggle against stigmatization and social exclusion using speech synthesis. *Disability and Rehabilitation: Assistive Technology, 13*(4), 342-352. https://doi.org/10.1080/17483107.2017.1319428

Mosquera-De La Cruz, J. H., Loaiza-Correa, H., Nope-Rodríguez, S. E., y Restrepo-Girón, A. D (2017). Identifying facial gestures to emulate a mouse: navigation application on Facebook. *IEEE Latin America Transactions, 15*, 121–128.

Mosquera-De La Cruz, J. H., Loaiza-Correa, H., Nope-Rodríguez, S. E., y Restrepo-Girón, A. D. (2020). Disability and Rehabilitation: Assistive Technology Human-computer multimodal interface to internet navigation Human-computer multimodal interface to internet navigation. *Disability and Rehabilitation: Assistive Technology, 0*(0), 1–14. https://doi.org/10.1080/17483107.2020.1799440

Mosquera-De La Cruz, J. H., Loaiza-Correa, H., y Nope-Rodríguez, S. E. (2018). Sistema de Interacción Humano-Máquina Audiovisual [trabajo de Investigación de Maestría en Ingeniería con Énfasis en Ingeniería Electrónica, Escuela de Ingeniería Eléctrica y Electrónica, Universidad del Valle].

Nur, S., Mohamad, A., e Isa, K. (2018). *Assistive Robot for Speech Semantic Recognition System* [ponencia]. 2018 7th International Conference on Computer and Communication Engineering (ICCCE), 50–55.

Oliveira, L. P. De, Wehrmeister, M. A., y Oliveira, A. S. De. (2017*). Systematic Literature Review on Automotive Diagnostics* [ponencia]. Brazilian Symposium on Computing System Engineering, SBESC, 2017-Novem, 1–8. https://doi.org/10.1109/SBESC.2017.7

Pampoulou, E. (2018). Speech and language therapists' views about AAC system acceptance by people with acquired communication disorders. *Disability and Rehabilitation: Assistive Technology, 14*(5), 471-478. https://doi.org/10.1080/17483107.2018.1463401

Pardowitz, M., Knoop, S., Dillmann, R., y Zollner, R. D. (2007). Incremental learning of tasks from user demonstrations, past experiences, and vocal comments. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 37*(2), 322–332. https://doi.org/10.1109/TSMCB.2006.886951

Sánchez, J. G., & Terrats, J. A. (2011). Guía técnica para la construcción de escalas de actitud. https://www.odiseo.com.mx/2011/8-16/garcia-aguilera-castillo-guia-construccion-escalas-actitud.html

Santander-Ariza, D. A., Rosero-Ramos, K. S., Libreros-Segura, M., Mosquera-De La Cruz, J. H., y Ferrín-Bolaños, C. D. (2020). *Reconocimiento de voz para un sistema de interacción humano máquina orientado a personas con limitaciones motrices* [trabajo de grado, Ingeniería de Sistemas, Facultad de Ingeniería, Fundación Universitaria Lumen Gentium].

Sharma, M. K., y Kumari, O. (2014). *Speech Recognition: A Review* [ponencia]. National Conference on Cloud Computing & Big Data, International Journal of Advanced Networking and Applications (IJANA), 62–71. https://doi.org/10.9790/0661-1804020109

Wang, E., Zhou, L., Chen, S. K., Hill, K., Zhou, L., Chen, S. K., y Hill, K. (2017). Development and evaluation of a mobile AAC: a virtual therapist and speech assistant for people with communication disabilities. *Disability and Rehabilitation: Assistive Technology, 13*(8), 731-739. https://doi.org/10.1080/17483107.2017.1369592