

# Speech to Text

Autor 1: Sebastián Lopera Osorio

Autor 2: Luis David Restrepo C.

Autor 3: Edwin Alexander Enciso

*Risaralda, Pereira, Universidad Tecnológica de Pereira*

Correo-e: sebastian.lopera@utp.edu.co / david95rc@utp.edu.co /  
edwinenciso@utp.edu.co

**Resumen—** Descripción del funcionamiento de un reconocedor de audio que convierte lo dicho por el micrófono del teléfono a texto en pantalla, basado en el funcionamiento del reconocedor de voz de Google.

**Palabras clave—** Visual, voz, reconocimiento, lenguaje natural.

**Abstract—** Description of the operation of an audio recognizer that converts what is said by the microphone of the phone to text on the screen, based on the operation of the Google voice recognizer.

**Key Word —** Visual, voice, recognition, natural language.

## I. INTRODUCCIÓN

Cada vez podemos observar que esas películas de ficción donde los humanos se comunican con los diferentes elementos de la casa tales como las neveras, relojes o carros para saber la hora, el tráfico o la situación del clima a través de solo usar la voz sin necesidad de tocar botones ni usar otras herramientas es cada vez más común en pleno inicio del siglo XXI, hoy en día ya existen sistemas con los que nos comunicamos diariamente un gran ejemplo de esto fue el lanzamiento de Siri en el año 2011 un asistente personal que acepta y reconoce comandos por voz y no solo eso, también es capaz de responder en tiempo real con una precisión casi humana, ahora Microsoft compite con “Cortana” una asistente virtual capaz de reconocer más de 8 idiomas y que sin duda llegó al mercado para cambiar nuestra manera de comunicarnos.



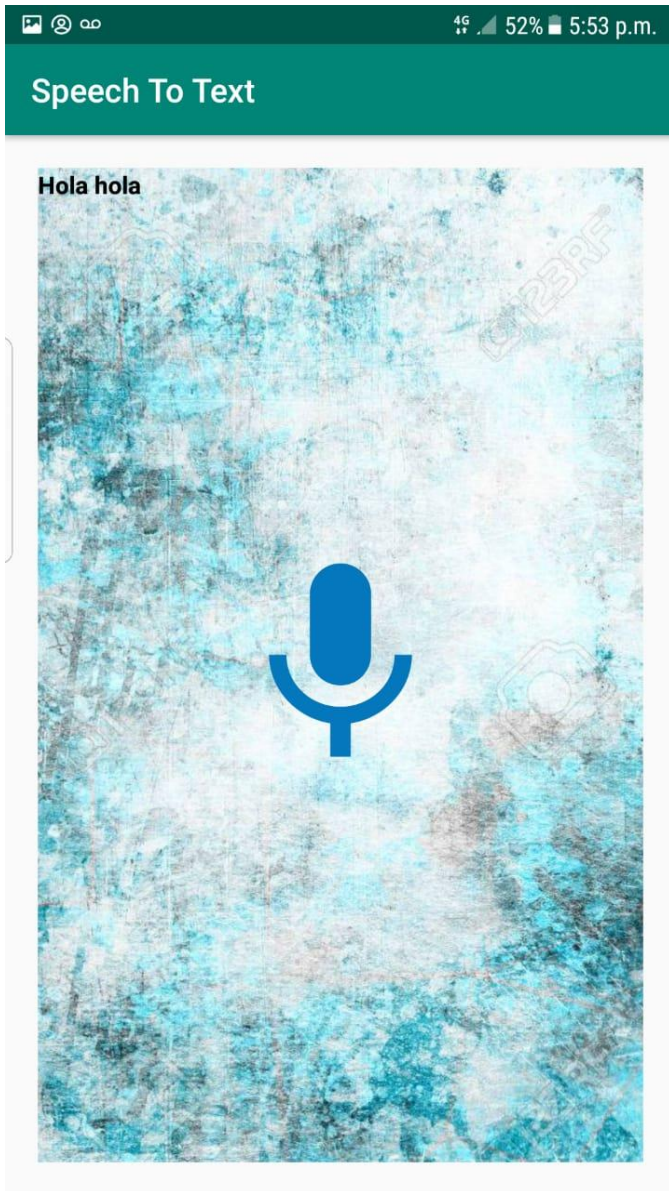
## II. ¿Cómo funciona?

El reconocimiento de voz en un dispositivo y después su paso a texto utiliza un modelo multicapa en donde la primera capa es el modelo acústico, esta capa determina el canal de comunicación para establecer el grado de distorsión, una segunda capa implementa un modelo lingüístico para determinar el idioma y los distintos acentos que pueda tener este mismo, esta capa se comunica con un transcriptor de texto para poder depurar el mensaje y transcribir lo dicho por voz, sistemas inteligentes tales como Siri o Cortana, usan un motor estadístico para realizar una búsqueda en una base de datos con las palabras relevantes que ingresen por el micrófono.

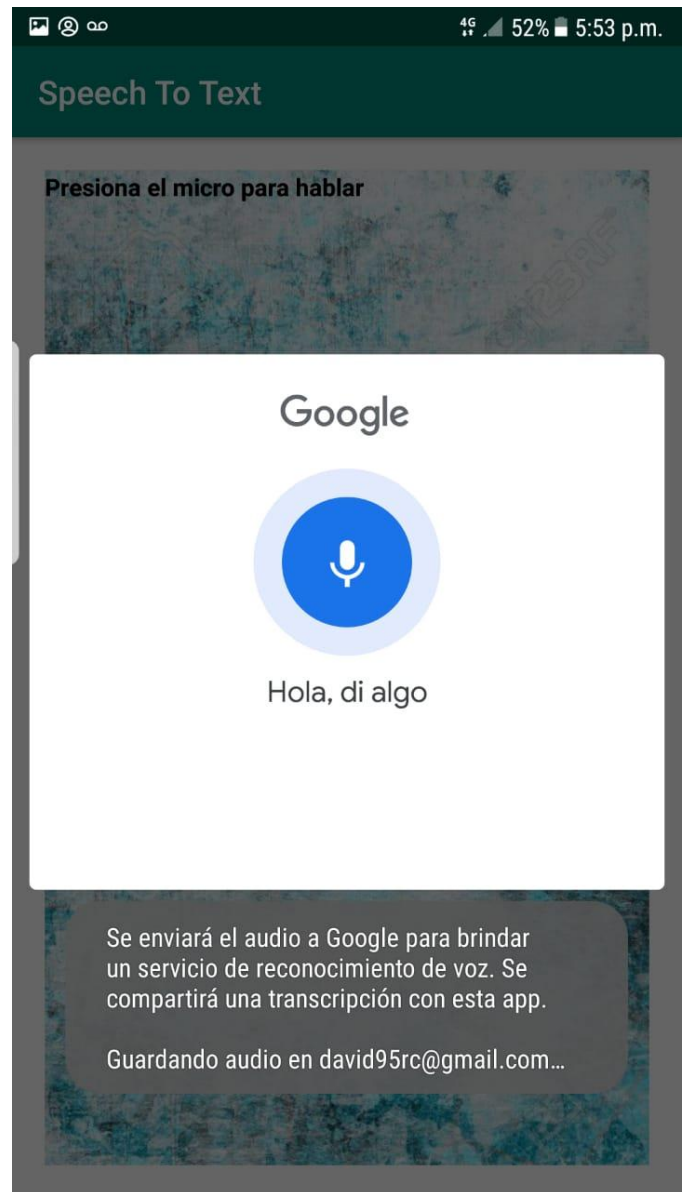


### III. DESCRIPCIÓN

El transcriptor de audio a texto que desarrollamos y expondremos en el presente documento es mucho más simple pero igual de funcional pues reconoce por medio del micrófono del dispositivo el audio y muestra en pantalla el texto.



El programa esta dividido en dos partes principales, en la primera tenemos la interfaz gráfica donde tenemos los elementos como el botón que activa el micrófono, un texto inicial que da la bienvenida a la aplicación y un label en donde se muestra en texto lo que ingresa el usuario por audio.



En la segunda esta la codificación e implementación de librerías que permiten el funcionamiento tanto con acceso a internet o sin acceso a este, para la recepción del audio y la conversión a texto, usamos librerías tales como androidx y android.speech, la aplicación se programa con el paradigma orientado a objetos en la herramienta android studio.

### REFERENCIAS

[https://www.eldiario.es/turing/reconocimiento-voz-biometria\\_0\\_201230680.html](https://www.eldiario.es/turing/reconocimiento-voz-biometria_0_201230680.html)

<https://developer.android.com/reference/android/speech/package-summary>