

PROYECTO "VoiceMe"

(Intérprete de voz y sonido)





ÍNDICE:

1. Proyecto VoiceMe	3
2. Antecedentes del proyecto	3
3. Equipo	4
4. Producto y objetivos	5
4.1.Producto: visión general	5
4.2.Puntos fuertes y ventajas	5
4.3.El cliente o usuario final.	7
5. Dafo del proyecto	7
6. Recursos	8
6.1.Recursos humanos Semilla	8
6.2.Recursos humanos Profesionalizados	8
6.3.Recursos técnicos	8
6.4.Recursos tecnológicos: Uso de la IA	8
8. Condiciones del proyecto	9
8.1.Requisitos del proyecto	9
8.2.Partes interesadas (stakeholders)	9
9. Planes operativos	10
9.1.Estructura de desglose de trabajo	10
9.2.Plazo y cronograma del proyecto	11
10. Inversiones, financiación y análisis financiero	12
10.1.Inversión Inicial	12
10.2.Gastos e ingresos	13
11. Riesgos	14
11.1.Riesgos posibles	14
12. Aspectos legales	14
13. Competencia y colaboradores	15
14. Conclusiones finales	15



1. PROYECTO VOICEME

Muchas personas con dificultad con el habla o incapacidad para ello (por ejemplo: por parálisis cerebral, ELA, esclerosis múltiple o trastorno del espectro autista) en muchas ocasiones son capaces de emitir ciertos sonidos que se pueden reconocer como específicos y concretos, entendiendo este lenguaje individual como un lenguaje propio.

Estas personas suelen tener dificultades para hacerse entender por quienes no pertenecen a su entorno cotidiano. Si bien, las personas habituadas a escucharles son capaces de diferenciar y reconocer patrones de sonidos con intencionalidad que siempre se repiten y se identifican con una acción o expresión de la persona que los emite.

El objetivo final del proyecto es aportar una solución tecnológica a personas con problemas en la comunicación verbal (afasia o disartria). El proyecto actual consiste en detectar estas barreras comunicativas y ofrecer una solución de bajo coste y alta tecnología mediante el desarrollo de un traductor personalizado basado en Inteligencia Artificial y reconocimiento de patrones de voz. La aplicación a desarrollar será personalizable y de uso exclusivo para cada usuario.

El equipo de desarrollo está compuesto por seis miembros de los que cuatro tienen discapacidad. De esos cuatro, tres de ellos presentan dificultad en el habla o tienen ausencia de ella. Sus capacidades de aprendizaje no están afectadas. Es por ello que son conocedores de primera mano del reto al que nos enfrentamos, son promotores de la idea y podrán testar la aplicación por ellos mismos.

2. ANTECEDENTES DEL PROYECTO

A pesar de vivir en un mundo hiperconectado y con grandes avances en el reconocimiento de voz (asistentes virtuales, dictado automático), estos sistemas comerciales están entrenados con voces y pronunciaciones normativas. Todo este crecimiento tecnológico no se ha visto reflejado en herramientas útiles para personas con discapacidad del habla o problemas de articulación. Como por ejemplo personas traqueostomizadas, con daño o



parálisis cerebral, ELA o con autismo, para quienes los reconocedores de voz estándar son ineficaces.

3. EQUIPO

El Equipo de desarrolladores está formado por seis jóvenes con experiencia en proyectos similares ya que en 2023 participaron en el proyecto SUPERMAKERS: Pequeña Fábrica Social que resultó ser uno de los proyectos ganadores en la primera edición de XTalento Challenge.

VoiceMe es el resultado de una necesidad detectada por el equipo, concretamente por varios de sus componentes que presentan la dificultad que se podrá resolver con la app y que esperamos supondrá un gran impacto para otras personas con dificultad o ausencia del habla.

Pedro Antonio: Es un chico con movilidad reducida producida por una encefalopatía que también afecta a la capacidad de articular palabras. Utiliza un tablero de comunicación para deletrear lo que quiere decir pero es capaz de emitir ciertos sonidos concretos, por ejemplo “ah-ah” para referirse al agua. La aplicación grabará ese sonido específico que a través de IA se guardará como onda de sonido para diferenciarlo de otros sonidos similares, se etiquetará y se relacionaría con la palabra escrita “agua”. De tal forma que cuando a través de la aplicación diga “ah-ah”, se reproducirá con voz natural la palabra “agua” para que cualquier persona pueda entenderlo..

Rafa y Paula (Síndrome de Moebius y Parálisis Cerebral) tienen una forma similar de hablar aunque no pronuncian correctamente. Pero hay muchas palabras que pueden sonar igual por la dificultad que tienen al vocalizar, por ejemplo “casa” y “caja”. A través de la IA se discriminará el espectro de onda de sonido para poder etiquetar cada palabra y entrenar al programa.

Sonia trabaja con niños con autismo que en muchos casos no tienen lenguaje verbal pero sí que articulan sonidos o gritos concretos para referirse a acciones, por ejemplo, cuando quieren ir al parque. VoiceMe ayudará a identificar esta intencionalidad gracias a que la sus familiares ha grabado este sonido y han escrito su significado.



Pablo y Hugo son los dos componentes del equipo con más conocimiento técnico y sus aportaciones serán muy valiosas para el desarrollo de este proyecto.

El Equipo RAFAPUEDE está tutorizado por Antonio Marroquí, es Graduado en ingeniería Mecánica y con Máster en Industria 4.0

4. PRODUCTO Y OBJETIVOS

4.1 Producto: Visión general

Vamos a desarrollar una aplicación para móvil, con entorno visual sencillo, que permita grabar voz y sonidos concretos de los usuarios, asociarlos a su significado y entrenar un modelo de IA que reconozca esos sonidos. El sistema aprenderá estos patrones acústicos, los guardará y los relaciona con una palabra o frase, que podrá ser en castellano o cualquier otro idioma, traduciendo en tiempo real a texto y a voz natural comprensible por cualquier persona, facilitando así la comunicación.

La aplicación está pensada para que el propio usuario o su familia puedan entrenarla sin conocimientos técnicos.

El funcionamiento consta de las siguientes fases:

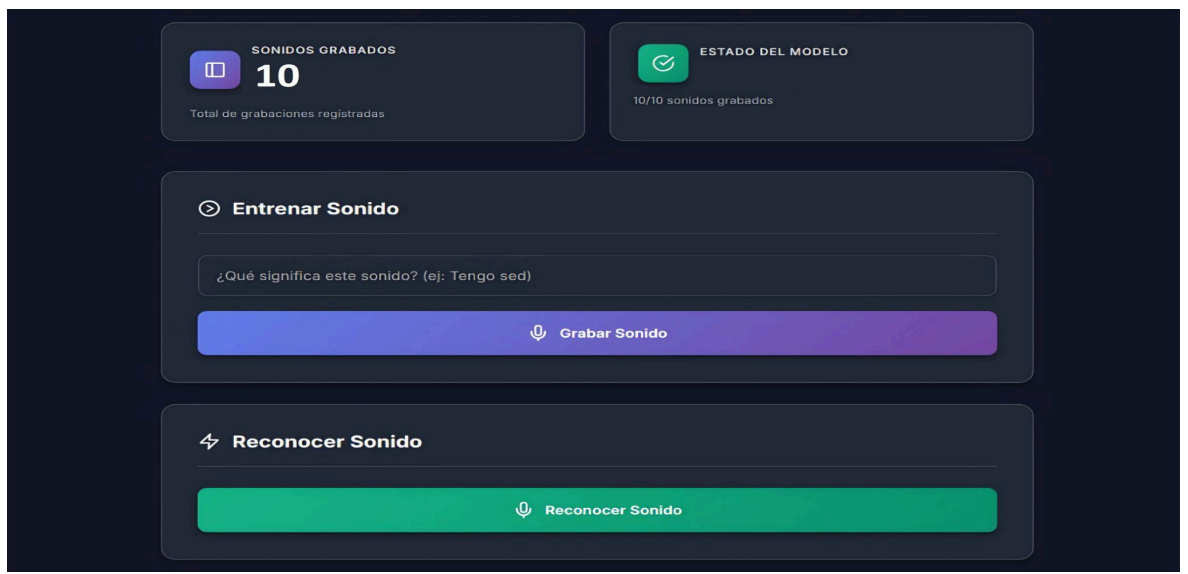
- Grabación y Entrenamiento: El usuario graba su forma particular de pronunciar determinadas palabras, frases **e incluso sonidos** en la app. Estas grabaciones se asocian a un texto escrito con relación a la palabra que quiere decir.
- Reconocimiento de Patrones: La aplicación entrena un modelo de Inteligencia Artificial para reconocer las peculiaridades fonéticas únicas de ese usuario.
- Traducción y Reproducción: Cuando el usuario habla, la app busca en su biblioteca de patrones, identifica la palabra pregrabada que quiere decir y la reproduce de forma clara mediante un sistema de Texto a Voz (TTS - Text-to-Speech).

4.2 Puntos fuertes y ventajas

- La aplicación y su uso serán GRATUITOS para los usuarios finales atendiendo a los fines de la Fundación RafaPuede como promotor.



- No requiere hardware específico costoso, solo un dispositivo con conexión a internet y un micrófono de cualquier dispositivo con el que grabar el sonido o palabra.
- Es altamente personalizable. El modelo IA aprende de ese usuario específico y guarda la información en su biblioteca personal.
- Fomenta la autonomía del usuario en entornos sociales, educativos y laborales.
- Será adaptativa para diferentes dispositivos móviles y web y contará con funciones de accesibilidad universal como lectura de etiquetas de las funciones, alto contraste y colores amigables con personas daltónicas.
- Ya hemos desarrollado un **versión Alpha** que, comprobando su potencial, nos permite ser optimistas con este proyecto.



```
return conn
except PysocksOperationalError as e:
    print(f"⚠️ Intento {attempt + 1}/{retries} fallido: {e}")
    if attempt < retries - 1:
        time.sleep(delay)
    else:
        raise

class AudioCNNRNIN(nn.Module):
    def __init__(self, num_classes=20):
        super().__init__()
        self.conv1 = nn.Conv2d(1, 32, kernel_size=3, padding=1)
        self.conv2 = nn.Conv2d(32, 64, kernel_size=3, padding=1)
        self.pool = nn.MaxPool2d(2, 2)
        self.dropout = nn.Dropout(0.5)
        self.rnn = nn.LSTM(64 * 64, 128, batch_first=True, bidirectional=True, dropout=0.3)
        self.fc = nn.Linear(256, num_classes)
```

```
def forward(self, x):
    x = torch.relu(self.conv1(x))
    x = self.pool(torch.relu(self.conv2(x)))

    batch_size = x.size(0)
    x = x.permute(0, 3, 1, 2)
    x = x.contiguous().view(batch_size, x.size(1), -1)

    out, (hn, _) = self.rnn(x)
    hidden = torch.cat((hn[-2], hn[-1]), dim=1)
    hidden = self.dropout(hidden)

    return self.fc(hidden)

def hash_password(password: str) -> str:
    return hashlib.sha256(password.encode()).hexdigest()
```

versión Alpha del producto



4.3 El cliente o Usuario Final.

Nuestro público objetivo es cualquier persona que presente dificultades en la articulación del habla o pronunciación no normativa y que requiera de un traductor adaptado a sus propios patrones vocales para comunicarse de manera efectiva con su entorno.

El propio usuario, su familiar o cuidador podrá grabar un sonido con intencionalidad comunicativa y relacionarlo con la palabra que, por su significado, contexto o experiencias previas, quiere referir.

5. DAFO DEL PROYECTO

Se ha utilizado el método DAFO para identificar y proponer alternativas a posibles problemas asociados al proyecto.

En el análisis de contexto sus elementos principales:

DEBILIDADES <ul style="list-style-type: none">• Requiere tiempo inicial de grabación por parte del usuario para entrenar el modelo.• Necesita conexión a internet para el procesamiento de IA.• Problemas en la página web que requieran soporte técnico.	AMENAZAS <ul style="list-style-type: none">• Altos costes de servidores en la nube si hay un elevado número de usuarios concurrentes.• Falta de patrocinios o financiación a largo plazo.
--	---

FORTALEZAS <ul style="list-style-type: none">• Producto 100% personalizado a cada usuario.• Solución basada en software, muy escalable.• Aplicación gratuita para el usuario final.	OPORTUNIDADES <ul style="list-style-type: none">• Mejora directa y radical de la calidad de vida e integración social.• Fomentar el uso de IA para fines sociales.• Potencial de crear una aplicación móvil nativa en el futuro.
--	---



6. RECURSOS

Los recursos a emplear son recursos humano, técnicos y tecnológicos:

6.1 Recursos humanos “semilla”

El equipo estará compuesto por jóvenes con y sin discapacidad de entre 17 y 21 años que recibirán formación y asesoramiento durante todo el proceso de desarrollo. Generando oportunidades de conocimiento y empleabilidad, además de fortalecer su autoestima al ser partícipes de un proyecto que puede tener un gran impacto social.

6.2 Recursos humanos “profesionalizados”

El equipo estará supervisado por desarrolladores, diseñadores de UI/UX y el coordinador voluntario de la Fundación RafaPuede. Se buscará la colaboración con universidades e institutos tecnológicos para el apoyo en la arquitectura de Machine Learning.

6.3 Recursos técnicos

Los recursos técnicos necesarios serán los detallados a continuación:

- Ordenadores para el equipo de desarrollo.
- Conexión a internet
- Contratación de Servidores en la Nube (AWS, Google Cloud o Azure) para alojar y procesar los modelos de IA.
- Licencias de software, APIs de TTS (Text-to-Speech) y dominios web. -
- Micrófonos de calidad para las primeras pruebas de concepto.

6.4 Recursos tecnológicos: Uso de IA

La inteligencia artificial es el componente que permite **interpretar y traducir patrones de voz no estándar en lenguaje comprensible**, mediante modelos de aprendizaje automático entrenados con datos personalizados del usuario.

La IA en VoiceMe actúa como un sistema de reconocimiento de voz adaptado, basado en técnicas de *machine learning*. A diferencia de los sistemas tradicionales, el modelo se entrena con muestras específicas de cada usuario, permitiendo identificar patrones fonéticos únicos asociados a dificultades en el habla.

El sistema procesa el audio capturado, extrae características acústicas (como coeficientes MFCC) y utiliza un modelo de clasificación para asociar cada patrón a palabras o expresiones



previamente entrenadas. Este enfoque permite convertir vocalizaciones no normativas en comunicación comprensible en tiempo real.

8. CONDICIONES DEL PROYECTO

8.1 Requisitos del proyecto

- El proyecto será sin coste para el usuario final.
- La financiación será procedente de **patrocinios** de empresas, subvenciones o donaciones y contará con el soporte de la Fundación RafaPuede como promotor y propietario del proyecto.

8.2 Partes interesadas (Stakeholders)

Evaluación de interesados (Stakeholders) en el proyecto:

Nombre	Rol	Expectativas	Influencia
Empresas Patrocinadoras	Financiación	Buenas	Alta
Universidades/Institutos	Colaboración en Desarrollo IA	Buenas	Alta
Proveedores Cloud (AWS/Google)	Infraestructura de Servidores	Estabilidad/Pagos	Media
Colectivos con problemas de habla	Usuarios Finales	Altas	Alta



9 PLANES OPERATIVOS

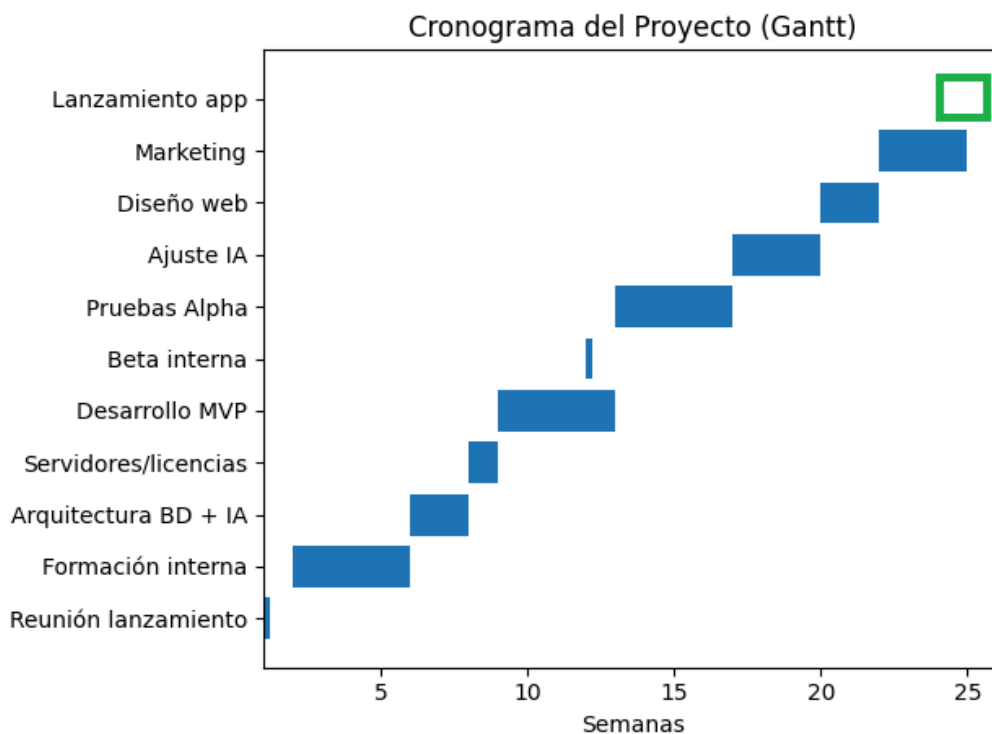
9.1 Estructura de Desglose de Trabajo

En la siguiente Estructura de desglose del Trabajo, EDT, se han identificado las principales actividades e hitos en función del alcance del proyecto “VoiceMe”.

Estructura de Desglose de Trabajo, EDT.	
Código	Actividad
1	Reunión lanzamiento del proyecto
2	Resolver necesidades de formación o asesoramiento especializado de nuestro equipo.
3	Diseño de la arquitectura de la Base de Datos y la IA
4	Adquisición de servidores, dominios y licencias
5	Desarrollo del MVP (Aplicación Web y motor de reconocimiento)
6	Pruebas Alpha: Grabación de vocabulario con los primeros 4 usuarios reales: Síndrome de Moebius, Parálisis Cerebral, Comunicación No Verbal y TEA.
7	Campaña de Marketing para captación de Patrocinadores
8	Lanzamiento de la aplicación para descarga gratuita para Android y Apple Store

9.2 Plazo y Cronograma del Proyecto

Código	Actividad e Hitos de Control	Duración Estimada	Periodo de Ejecución
1	Reunión de lanzamiento del proyecto y asignación de tareas	1 día	Semana 1
2	Formación interna	4 semanas	Semanas 2 a 5
3	Diseño de la arquitectura de la Base de Datos y la IA	2 semanas	Semanas 6 y 7
4	Adquisición de servidores, dominios y licencias necesarias	1 semana	Semana 8
5	Hito: Desarrollo del MVP (Aplicación Web y motor de reconocimiento)	4 semanas	Semanas 9 a 12
6	Hito: Lanzamiento interno de versión Beta	1 día	Semana 12
7	Pruebas Beta: Grabación y entrenamiento con usuarios reales	4 semanas	Semanas 13 a 16
8	Análisis de resultados de la IA, corrección de errores y ajuste del modelo	3 semanas	Semanas 17 a 19
9	Diseño de página web	2 semana	semana 20 y 21
10	Campaña de Marketing para la captación de Patrocinadores	3 semanas	Semanas 22 a 24
11	Lanzamiento de la aplicación	1 semana	Semana 25



10. INVERSIONES, FINANCIACIÓN Y ANÁLISIS FINANCIERO

10.1 inversión inicial

Las necesidades financieras para el desarrollo de este proyecto se basa en la necesidad de adquirir el material informático, contar con formación y asesoramiento especializado, crear un página Web y subir la aplicación a AppStore y Play Store.

El presupuesto total de partida del proyecto es de 5.000€, correspondientes al premio del concurso XTalento Challenge. Los gastos principales recaerán en infraestructura digital, formación y asesoramiento, computación en la nube para entrenar la Inteligencia Artificial y la creación de la plataforma web.



10.2 Gastos e ingresos

Gastos

Actividad	Costes Estimados (€)
Desarrollo web, licencias y APIs	- 1.400
Formación y asesoramiento	-1.200
Servidores Cloud y almacenamiento (1 año/100 usuarios)	- 500
Adquisición de equipos informáticos/Micrófonos	- 1.500
Coste Total de Actividades	- 4.600
Reserva de contingencia	- 400
PRESUPUESTO TOTAL	5.000 €

Ingresos

Ingresos de la Actividad	Ingresos Estimados (€)
Premio XTalento Challenge	+ 5.000
Ingreso por Patrocinio (escalable según alcance y número de usuarios)	+700
Donaciones de usuarios	+200
PRESUPUESTO TOTAL	5.900 €

La estimación de beneficio que repercutirá en los gastos de 2027 de acceso y mantenimiento online es de **+900 €**

11. RIESGOS

No hemos detectado un riesgo de magnitud que suponga la cancelación de este proyecto.

11.1 Riesgos posibles

Dentro de los riesgos de este proyecto se han identificado una serie de actividades que podrían interferir en el desarrollo del proyecto. Se han creado planes de contingencia para cada uno de ellos, que se pondrán en marcha en caso de producirse.

Para poder hacer frente a algunos de estos riesgos, se reservará un fondo en la tesorería del proyecto, con el que poder actuar en caso de ser necesario.

Por tanto, los riesgos más probables pueden ser:

Riesgo	Plan de contingencia
Aumento en los costes originalmente estimados debido a factores externos al proyecto.	Podrá ser cubierto con la reserva de contingencia contemplada en el presupuesto.
Elevado número de usuarios que implicaría la contratación extra de más espacio de almacenamiento en servidores Cloud.	Contratación de más espacio iCloud que podrá ser cubierto con la reserva de contingencia contemplada en el presupuesto, con las donaciones y patrocinadores.
Dificultades en la obtención de patrocinio.	La Fundación RafaPuede, como promotor, asume el coste de mantenimiento del proyecto.

12. ASPECTOS LEGALES

Se pondrá especial énfasis en el cumplimiento de la Ley Orgánica de Protección de Datos (LOPD) y el RGPD, garantizando que las grabaciones de voz de los usuarios estén anonimizadas ya que se consideran datos biométricos de carácter sensible.

13. COMPETENCIA Y COLABORADORES



Si bien existen múltiples proyectos de envergadura dedicados a learn machine y reconocimiento de voz, no hemos encontrado ninguna app o software para uso de un usuario único. A diferencia de otras herramientas, esta solución es personalizable, entrenable por la familia o usuarios y gratuita respondiendo a la acción social de la Fundación RafaPuede.

Nuestros colaboradores serán entidades sociales, asociaciones de personas con discapacidad, centros educativos, familias, voluntariado y logopedas. Se buscarán alianzas con organizaciones locales para pruebas piloto y difusión, así como con profesionales que aporten conocimiento y acompañamiento.

14. CONCLUSIONES FINALES

Después de valorar los riesgos del proyecto, así como los beneficios que supondrá para la autonomía e independencia de sus usuarios finales, la conclusión es que este proyecto es altamente innovador y necesario.

Traducir la voz no normativa a un lenguaje comprensible para todos mediante IA tiene el potencial de cambiar por completo la integración social, educativa y laboral de muchas personas.

Somos un equipo consciente del potencial de la aplicación que vamos a desarrollar y con la herramientas suficientes para llevarlo a cabo si fuésemos uno de los proyectos ganadores.