DECIMOQUINTA CONVOCATORIA PARA EL FOMENTO DE LA INVESTIGACIÓN Y LA INNOVACIÓN 2020

Título del proyecto	
Inclusión de la lengua Wayuunaiki en el reconocimiento de comandos de voz del robot social Pepper empleando la metodología de transformación de modelos	
Campo de acción	**Transdisciplinariedad - Aporte al PIM**
Desarrollo tecnológico con apuesta social	Este proyecto aporta a la Línea 3 del PIM, Proyección Social e Investigación Pertinentes, ya que a través del desarrollo de la investigación científica del mismo se apunta a la obtención de resultados que brindan oportunidades para la reducción de brechas tecnológicas de una comunidad específica, como lo es la comunidad indígena colombiana wayúu.

Articulación con funciones sustantivas y el sector social y productivo

Docencia

A través de la realización del proyecto se da la apropiación de conocimiento que puede ser aplicado directamente en la actualización de syllabus de espacios académicos o en la presentación de cursos electivos que hacen parte del componente flexible del pensum; casos específicos de lo anterior son los espacios académicos del programa de pregrado en Ingeniería Electrónica “Señales y Sistemas”, “Identificación y Modelado de Sistemas”, y electivas como “Introducción a la Inteligencia Artificial” y “Deep Learning”; para el programa de posgrado Maestría en Ingeniería Electrónica se aporta a los espacios de “Optimización”, “Herramientas de Ingeniería” y “Electiva - Inteligencia Artificial”. Entre las temáticas específicas relacionadas al programa en las que se espera que aporte el proyecto se encuentran: Procesamiento de señales, Adquisición de datos, Programación y desarrollo de software, Identificación de sistemas, Reconocimiento de voz y Robótica social.

Investigación

El proyecto, a través de su desarrollo y los resultados esperados, generará a su vez nuevos proyectos que se enmarcan en las líneas de investigación del Grupo de Estudio y Desarrollo en Robótica - GED, en particular las líneas “Robótica” e “Inteligencia Computacional”.

Adicionalmente, los resultados del proyecto apuntan a la solución de retos de investigación planteados en la competencia Robocup 2019 y que se deben resolver para la Robocup 2020. Se proyecta participar en el simposio de la Robocup 2020 con una ponencia y artículo donde se presenten los resultados obtenidos.

Responsabilidad Social

La reducción de la brecha tecnológica a través del acercamiento de las herramientas desarrolladas para la robótica social a las comunidades indígenas es la apuesta social del proyecto. Para ello, el proyecto se apalanca en los resultados obtenidos por el proyecto Kailumá de la Universidad Santo Tomás, buscando impactar positivamente la comunidad wayúu. Como fruto del proyecto se espera proporcionar un conjunto de herramientas tecnológicas de fácil

implementación y mantenimiento con las cuales se pueda, en una primera etapa, familiarizar y enseñar sobre las nuevas tecnologías de la información y las comunicaciones, y la robótica social.

Sector social y productivo

Al ser producto de las actividades de la Alianza Sinfonía con el fin de cumplir con sus objetivos, el proyecto busca proporcionar soluciones y el desarrollo de nuevas tecnologías que sean directamente aplicables en lo social y el sector productivo.

Grupo de investigación	Línea de investigación en la que se inscribe el proyecto
Grupo de Estudio y Desarrollo en robótica GED - Inteligencia Computacional y robótica	

Nombre del Investigador principal	Enlace CvLAC	Enlace ORCID	Enlace Google Académico
Armando Mateus Rojas	https://scienti.colciencias.gov.co/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0000680630	https://orcid.org/0000-0002-2399-4859	https://scholar.google.es/citations?hl=es&pli=1&user=1az5o_IAAAAJ
División	**Facultad**	**Programa**	**Grupo de investigación**
Ingenierías	Ingeniería Electrónica	Ingeniería Electrónica	Grupo de Estudio y Desarrollo en Robótica - GED
Nombre del Co-investigador	**Enlace CvLAC**	**Enlace ORCID**	**Enlace Google Académico**
Sindy Paola Amaya	https://scienti.colciencias.gov.co/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0000796425	https://orcid.org/0000-0002-1714-1593	https://scholar.google.es/citations?hl=es&authuser=2&user=Gg2sofAAAAAJ
División	**Facultad**	**Programa**	**Grupo de investigación**
Ingenierías	Ingeniería Electrónica	Ingeniería Electrónica	Grupo de Estudio y Desarrollo en Robótica - GED

Resumen de la propuesta	Palabras clave
La presente propuesta busca dotar al robot social Pepper, con el que cuenta la Universidad Santo Tomás, de la capacidad de reconocimiento para comandos de voz en lengua Wayuunaiki. De esta forma, se permitirá la utilización de	

funciones de robótica social más avanzadas como la asistencia y el servicio. Así mismo, a través de los resultados y productos esperados se pretende disminuir la brecha tecnológica en la comunidad wayuu proveyendo nuevas capacidades al proyecto Kailumá de la Universidad Santo Tomás. Las herramientas de reconocimiento de voz disponibles están basadas en tecnologías que requieren de una gran cantidad de datos previos junto con un proceso de entrenamiento de dichas herramientas. Esto hace que el soporte de idiomas por parte de estas herramientas se dé para idiomas como el Español e Inglés pero no para lenguas con pocos hablantes, como es el caso de las lenguas indígenas colombianas. Por lo anterior, se propone una solución a esta problemática mediante la utilización de cadenas de transformación de modelos. Para esto, se modelan los comandos de voz en idioma Wayuunaiki que serán la entrada de la cadena de transformación; así, se requiere configurar, entrenar y modelar una herramienta de reconocimiento de voz (speech recognition) que servirá como salida de la cadena de transformación. La generación de dichos modelos surge del análisis del conjunto de comandos de voz para robótica social predefinidos para un entorno de servicio doméstico. La cadena de transformación será implementada como un nodo de ROS para ser habilitada en el robot Pepper.	Reconocimiento de voz, Speech recognition, Lenguas indígenas, Cadenas de transformación de modelos, Robótica social

Problema de investigación

Los desarrollos tecnológicos en robótica social, en especial en lo referente a reconocimiento de comandos de voz (*speech recognition*), no tienen en cuenta lenguas aborígenes como el Wayuunaiki. Por este motivo, no es posible realizar planes y programas (como la Etnoeducación) que acerquen las tecnologías relacionadas con la robótica social a comunidades indígenas para estar en línea con el espíritu expresado por la Constitución Política Colombiana (Art. 7) [1]. El Wayuunaiki, la lengua wayuu o guajira, es una lengua aborigen y oficial de la península de la Guajira, compartida por Colombia y Venezuela; es la lengua del pueblo indígena guajiro y hablada por aproximadamente 100.000 habitantes, siendo en su mayoría monolingües [2]. Este último hecho hace que la integración de la comunidad indígena con las TIC (tecnologías de la información y la comunicación) posea desafíos particulares que han contribuido a ensanchar la brecha social con el mundo moderno, ya que como se indica en [3] los sistemas actuales no pueden facilitar oportunidades a todas las comunidades. Así, al no poseer herramientas en su lengua nativa, las TIC no contribuyen a la preservación cultural de la comunidad wayuu siendo, en su mayoría, un factor que complica la adaptación a las necesidades del mundo actual.

El robot social Pepper [4], fabricado por la empresa Softbank - Aldebaran, cuenta con un sistema de sensores que le permite la interacción con seres humanos y su entorno. Los comandos de voz se constituyen en la forma natural de comunicación por medio de la cual una persona le indica al robot Pepper qué actividad debe realizar. Sin embargo, el robot Pepper solo posee la capacidad de reconocimiento de voz para los idiomas más representativos como Inglés, Español, Francés, etc. Lo anterior conlleva dos problemáticas tecnológicas que imposibilitan al robot Pepper la interacción con la comunidad wayúu: no se tiene soporte para lenguas indígenas, como el Wayuunaiki, esto debido al bajo número relativo de hablantes y su alta focalización geográfica; adicionalmente, si bien el Español es uno de los idiomas soportados, solo lo está en acento ibérico, lo que dificulta la utilización por hablantes latinoamericanos y hablantes no nativos.

Si bien las bases teóricas y los métodos empleados para el reconocimiento de comandos de voz cuentan con una alta divulgación, estos métodos se encuentran basados en aprendizaje automático (machine learning) y requieren de una enorme cantidad de datos estructurados y de una base de usuarios regulares que permita retroalimentar el sistema. Deep Speech (https://github.com/mozilla/DeepSpeech) es una iniciativa de la fundación Mozilla para proporcionar una herramienta universal de reconocimiento de voz basada en [5] y soportada por una comunidad de desarrolladores voluntarios; esta herramienta requiere de una gran cantidad de horas de grabación de audios con órdenes específicas para el entrenamiento; esto hace que la utilización del sistema para un determinado idioma sea altamente complejo como lo ejemplifica el caso de la no existencia aún de una adaptación al Español de esta herramienta. Diversos trabajos apuntan a la mejora del reconocimiento como [6]. Sin embargo, la tasa de éxito de reconocimiento en estos casos se ve afectada por factores culturales y personales como acentos, entonaciones, construcciones gramaticales y expresiones típicas.

Así, se plantea la siguiente pregunta de investigación:
¿Cómo incluir Wayuunaiki en el reconocimiento de comandos de voz del robot social Pepper para su uso por la comunidad indígena wayuu empleando la metodología de transformación de modelos?

Justificación

La robótica social tiene como principal objetivo la utilización de robots para el servicio y asistencia a seres humanos en entornos y situaciones cotidianas, siendo parte de diferentes ambientes que se comparten con personas como el hogar, hospitales, instituciones educativas, sitios públicos, locales comerciales, etc. [7]. Entre las tareas asignadas a los robots sociales se encuentran el cuidado a personas mayores, incapacitadas y/o enfermos, asistencia en tareas domésticas, atención en sitios públicos, acompañantes en la enseñanza y aprendizaje, entre otras. Las tareas anteriores implican que un robot social deba poder servir a las personas de forma tal que no se requiera algún entrenamiento especializado en robótica.

El habla es la representación acústica del lenguaje, siendo portadora de información y siendo así un campo de estudio indispensable en la interacción Humano – Robot (HRI por sus siglas en inglés) [8]. Siendo la comunicación oral una de las formas principales de comunicación entre seres humanos y la que menos exigencias impone a estos [9], se hace fundamental que los robots sociales cuenten con un sistema de reconocimiento de voz eficaz y eficiente. Soluciones

como “SIRI” y “Alexa” de las compañías Apple y Google entre otras, han familiarizado el uso de comandos de voz ampliando las funcionalidades de los sistemas de información y comunicación modernos a la vez que mejoran la experiencia de los usuarios.

Pese a lo anterior, las herramientas de reconocimiento de voz desarrolladas se han centrado en los idiomas más hablados a nivel mundial como el Inglés, el Español, el Francés y el Mandarín, entre otros. Esto ocasiona que estas nuevas tecnologías no se encuentren disponibles para idiomas y lenguas cuyo número de hablantes es comparativamente pequeño y altamente localizado, como es el caso de las lenguas indígenas colombianas. Lo anterior debido a la necesidad de disponer de una gran cantidad de audios con sus transcripciones que permitan al sistema el aprendizaje necesario [9].

Si bien, los avances en dispositivos, técnicas y métodos de reconocimiento de voz cuentan con una gran cantidad de documentación en publicaciones científicas y existen iniciativas de carácter abierto, como Deep Speech, estos avances se basan en su mayoría en técnicas de Machine Learning (aprendizaje automático) que requieren de un entrenamiento con una gran cantidad de datos disponibles para idiomas como el Inglés y el Español pero inexistentes para lenguas indígenas lo que hace inviable la utilización de las soluciones tradicionales. Un problema adicional que surge de la aproximación tradicional lo constituye la reducción en la tasa de éxito en la identificación de comandos de voz debido a las variaciones en la entonación, las expresiones y otros aspectos de carácter personal y/o cultural.

Se propone, por tanto, solucionar la problemática anterior mediante la implementación de una cadena de transformación de modelos que permita, partiendo de la definición de modelos de comando de voz en Wayuunaiki, Español e Inglés, la identificación de los comandos de voz en Wayuunaiki entendibles por el robot social Pepper.

Kailumá [10], proyecto de la Universidad Santo Tomás que busca entre otras acercar las nuevas tecnologías a la comunidad wayuu, se constituye en un referente para el proyecto a la vez que brinda los espacios para la aplicación del proyecto. En adición, los resultados y los desarrollos logrados con la ejecución de este proyecto responden a retos y desafíos presentados en la competencia Robocup. Se proyecta emplear los resultados del proyecto directamente en la competencia y realizar ponencias en el simposio del evento.

Objetivo general

Implementar el reconocimiento de comandos de voz en la lengua Wayuunaiki para el robot Pepper, empleando la metodología de transformación de modelos con miras a su utilización por parte de la comunidad wayúu.

Objetivos específicos

- Definir modelos y reglas de transformación entre comandos de voz para el robot social Pepper en idiomas Inglés, Español y Wayuunaiki.

- Implementar la cadena de transformación de modelos y las herramientas de reconocimiento de voz necesarias para un conjunto de comandos de voz limitado a un entorno doméstico.
- Validar y verificar el correcto funcionamiento de la implementación desarrollada mediante el desarrollo de casos de uso diseñados para un entorno doméstico y la utilización de métricas para la medición del desempeño.

Estado del arte y marco conceptual

MARCO CONCEPTUAL

Speech recognition

Dada su naturaleza, la robótica social se caracteriza por requerir de una comunicación humano - robot que se da a través de las interfaces enfocadas al lenguaje natural [11]; audios, gestos y expresiones faciales son combinados para generar comunicación multimodal (dada por señales de diferentes naturalezas) que tanto las personas como el robot social pueden comprender [12]. Adicionalmente, la comunicación multimodal y el lenguaje natural permiten que las personas puedan hacer uso de las funciones de los robots sociales sin que requiera un conocimiento especializado o algún tipo de entrenamiento previo [13].

Entre el conjunto de funciones que integran el lenguaje natural se encuentran las funciones de reconocimiento de voz que permiten, entre otras, el empleo de comandos de voz, la identificación de personas [14], sistemas de seguridad, reconocimiento de emociones [15] y sistemas conversacionales [16]. De entre las anteriores, los robots sociales requieren del reconocimiento de comandos de voz como función básica, permitiendo responder a solicitudes realizadas por parte de personas con características específicas, como es el caso de personas de la tercera edad [17].

El habla es una de las principales formas de interacción que posee el ser humano para su relacionarse con sus semejantes; por medio de esta, se genera una representación acústica [9] de los mensajes que transportan la información. Debido a esto, la tecnología de Speech Recognition (reconocimiento de habla) se ha transformado en un campo de alto interés en el estudio de la interacción Humano – Robot (HRI por sus siglas en inglés) [8]. Al ser la voz humana un fenómeno altamente dependiente del hablante, se tiene una alta incertidumbre dada por la diferencia entre estilos de hablar, diferencia entre tractos vocales y el modo de expresión [14].

En general, la tecnología de Speech recognition puede dividirse es cuatro componentes: análisis, extracción de características, modelado y pruebas. Estos cuatro componentes buscan extraer la información presente en el habla y que puede ser divida en dos tipos: información propia del mensaje (como palabras y frases) y la información que se deriva de las características propias de cada hablante (como la entonación y emociones) [9].

Ingeniería basada en modelos

La ingeniería basada en modelos (MDE por sus siglas en inglés) comprende las teorías, metodologías y marcos de trabajo desarrollados para el desarrollo de proyectos desde la perspectiva de la teoría de sistemas (omg.org). Para ello, MDE apunta a la definición de metamodelos con características necesarias y suficientes para la descripción total de modelos de objetos y de las relaciones entre estos. Esta categorización entre modelos y metamodelos posee la enorme ventaja de permitir la definición de lenguajes (de descripción de modelos o de programación) específicos de dominio (DSL por sus siglas en inglés), lo que permite crear lenguajes adaptados a las particularidades de las necesidades [18].

En adición a lo anterior, a través de la aplicación de la metodología guiada por modelo (MDD por sus siglas en inglés), es posible realizar transformaciones de las descripciones realizadas en un lenguaje particular hacia un lenguaje destino; así es posible obtener el código VHDL que describe un componente electrónico a partir de la definición en UML de su funcionamiento [19].

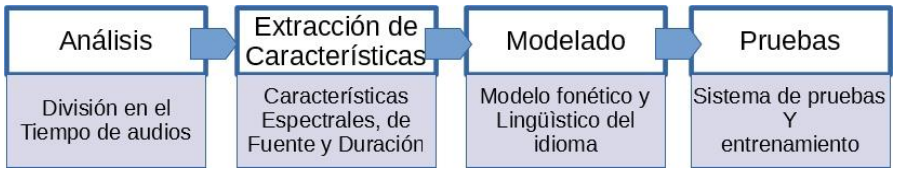

Figura 1. Proceso de Speech Recognition

ESTADO DEL ARTE

Diversas técnicas han sido desarrolladas para solventar inconvenientes que se tienen en el reconocimiento de voz, muchas de estas basadas en el procesamiento de señales y algoritmos de aprendizaje automático como en [20]; sin embargo, esta aproximación necesita una elevada potencia de cómputo que en la práctica implica el uso en paralelo de múltiples GPUs para lograr el aprendizaje con la gran cantidad de datos de que se dispone [5].

En general, los trabajos realizados abordan los componentes de la tecnología Speech Recognition: análisis, extracción de características, modelado y pruebas. El proceso de análisis se divide speech processing, speech analysis y speech recognition; speech analysis tiene como objetivo realizar el análisis de la segmentación requerida para el análisis del audio; la segmentación se realiza con tramas de duración entre 10ms y 30ms mientras que la subsegmentación se realiza con duraciones entre 3ms y 5ms; esta última aproximación se ejemplifica en [21], donde se presenta que las propiedades que tiene un hablante se dan con una base de cortos periodos de tiempo. Mediante la extracción de características se obtienen características que son agrupadas en tres grupos: características espectrales (ancho de banda y componentes principales entre otras), características de la fuente de excitación (como el tono) y características de larga duración (duración, energía, etc.) [22].

El modelamiento realiza modelos específicos de usuarios a partir de la definición de un "cluster" de características. Algunos de los tipos de modelos mayormente usados son los modelos acústicos – fonéticos, los de reconocimiento de patrón, modelos de deformación temporal y modelos de inteligencia artificial [9].

En [8], se plantea la caracterización de elementos auditivos (del sistema auditivo humano) a través de análisis de componente principal (PCA por sus siglas en inglés) para la implementación de un algoritmo robusto de reconocimiento de voz; de igual forma, en [22] se emplean datos médicos con el objetivo de obtener características de reconocimiento robusto de voz. Las particularidades propias de cada grupo de usuarios del reconocimiento de voz presentan también desafíos específicos como lo ejemplifica el estudio realizado en [23], que demuestra que la relación señal a ruido afecta el reconocimiento de palabras por parte de personas mayores influyendo en la forma en que dichas personas interactúan. Así mismo, el empleo de técnicas multimodales permite aumentar la tasa de efectividad en el reconocimiento de voz

Pocos proyectos encaminados principalmente a establecer funciones de traducción entre lenguas indígenas colombianas y el Español han sido desarrollados. En [24] se presenta el desarrollo de una API (Interfaz para programación de aplicaciones) con funciones de traducción Wayuunaiki empleando servicios en Internet; de esta forma se acercan los conceptos básicos de programación como palabras y cláusulas escritos en Inglés a la comunidad wayúu. También se han desarrollado diccionarios "online" como Glosbe (https://es.glosbe.com/es/cmi) que incluyen lenguas indígenas como el Embera. También se han adelantado proyectos como Kailumá de la Universidad Santo Tomás que buscan entre otros, acercar la tecnología a la comunidad Wayúu.

Metodología

La metodología de desarrollo consta de las siguientes etapas para la ejecución del proyecto:

1. **Etapa de levantamiento de información**

Se lleva a cabo la recopilación y análisis de información a partir de artículos científicos indexados en bases de datos (IEEE, Springer, Science Direct, etc.), manuales de usuario y repositorios de desarrollo. Con esta información se elabora un documento extendido de estado del arte que contiene métodos y tecnologías empleados para el reconocimiento de voz y las bases teóricas necesarias sobre la lengua Wayuunaiki. Así mismo se recopila información sobre los comandos de voz empleados en la robótica social estableciendo un conjunto de comandos a ser trabajados y el conjunto de casos de uso para validación.

2. **Desarrollo de modelos para comandos de voz**

Con la información y la teoría recopilada sobre la lengua Wayuunaiki y el conjunto de comandos de voz definido se desarrollan modelos para estos comandos en los idiomas Wayuunaiki, Español e Inglés; estos modelos deben seguir las reglas establecidas para MDE (Ingeniería basada en modelos). Se establecen casos de uso que permitan verificar la validez de los modelos.

3. **Configuración y entrenamiento de "speech recognition"**

Se realiza la evaluación y selección de la herramienta para "speech recognition" a ser empleada. La herramienta seleccionada es configurada de acuerdo a las características del Wayuunaiki y entrenada para el reconocimiento de palabras. Se establecen métricas que permitan desarrollar estadísticas para la mejora de la comprensión del reconocimiento.

4. **Desarrollo de cadena de transformación de modelos**

Una vez establecidos los modelos y la herramienta de reconocimiento de voz, se implementa la cadena de transformación de modelos que permita transformar órdenes dadas en lengua Wayuunaiki y conformes a los modelos desarrollados en órdenes entendibles por el robot Pepper.

5. **Implementación sobre el robot Pepper**

Luego de la validación de la cadena de transformación se desarrolla el nodo ROS que será implementado en un hardware que ejecute la cadena. El nodo ROS permitirá, por medio de tópicos y servicios, la integración del robot Pepper con la cadena de transformación.

6. **Evaluación de desempeño**

Con base en el conjunto de casos de uso, se desarrollan escenarios de prueba para la validación y la verificación de la implementación. Para ello, se utilizará la métrica WER (tasa de error de palabra) definida en [9] como:

$$WER = S + D + I/N$$

S: Número de sustituciones a la salida
D: Número de eliminaciones a la salida
I: Número de inserciones a la salida
N: Número real de palabras de entrada

Como audios de prueba se generarán traducciones de un conjunto seleccionado a partir del "2000 HUB5", que es un grupo de 40 transcripciones de audios telefónicos que sirven como estándar para pruebas de herramientas de speech recognition.

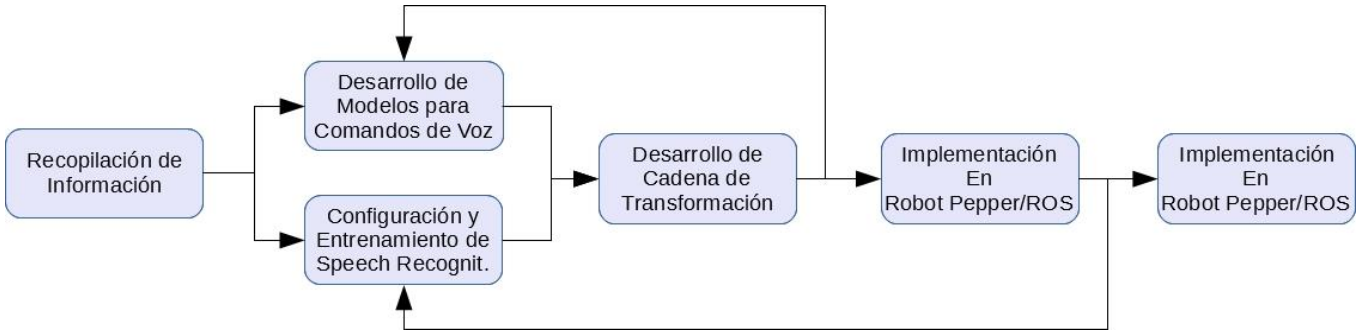

Figura 2. Etapas de desarrollo para la metodología propuesta

Resultados esperados

Los resultados esperados están relacionados con la generación de nuevo conocimiento y la apropiación del conocimiento que benefician a los investigadores y a la Universidad.

- Nuevo conocimiento: 1 artículo Q3.
- Desarrollo tecnológico: 1 registro de software.
- Producto de apropiación social del conocimiento: 1 participación en evento científico.
- Producto de formación de recurso humano: 2 direcciones de tesis pregrado.

	Presupuesto				
	Horas nómina				
Concepto	**Nombre**	**Escalafón**	**Horas mes**	**Sede / Seccional o Externo**	**Total ($)**
Horas Nomina (Investigador Principal)	Armando Mateus Rojas	2	50	Sede Principal	12.065.937,50
Horas Nomina (Co-Investigadores)	Sindy Paola Amaya	2	25	Sede Principal	6.032.968,75

Cronograma

ACTIVIDADES	RESPONSABLE	FECHA		FEBRERO				MARZO				ABRIL				MAYO				JUNIO				JULIO				AGOSTO				SEPTIEMBRE				OCTUBRE				NOVIEMBRE			
		Inicio	Fin.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	28	29	30	31	32	33	34	35	36	37	38	39	40	41
Revisión bibliográfica y determinación de estado de arte	Armando Mateus Sindy Amaya	01 - 02	28 - 02																																								
Diseño de modelos para comandos de voz	Armando Mateus Sindy Amaya	22 - 02	15 - 03																																								
Desarrollo de casos de uso/prueba	Armando Mateus	01 - 03	30 - 03																																								
Evaluación de herramientas de speech recognition	Armando Mateus Sindy Amaya	15 - 03	31 - 04																																								

Configuración y entrenamiento de herramienta de Speech recognition	Armando Mateus Sindy Amaya	01 – 04	30 - 06	
Implementación de cadena de transfromación de modelos	Armando Mateus Sindy Amaya	01 - 06	30 - 07	
Validación de cadena de transformación	Armando Mateus	25 - 07	30 - 09	
Implementación de nodo ROS	Armando Mateus Sindy Amaya	22 – 08	31 – 10	
Pruebas y verificación	Armando Mateus Sindy Amaya	22 - 08	30 – 11	

FINANCIACIÓN	RECURSO	DESCRIPCIÓN	Valor partida	Valor contrapartida (Externa)	Total ($)
RUBROS	**Servicios Técnicos**	NA			$ 0
	Salidas de campo	NA			$ 0
	Equipos	NA			$ 0
	Materiales, insumos y software	Costos asociados a los materiales, insumos y software que se pretende usar para el desarrollo y la implementación del proyecto propuesto: SDRs, Raspberry PI 3, FPGAs, routers, micrófonos, baterías.			$ 3.000.000
BOLSAS	**Papelería**	NA			$ 0
	Fotocopias	NA			$ 0
	Material bibliográfico	NA			$ 0
	Auxilio de transporte	Viaje a campo para estudio y verificación de resultados.			$ 1.000.000
	Movilidad	Costos relacionados con la participación en evento académico con el objetivo de generar el espacio para la apropiación social del conocimiento.			$ 14.000.000
	Publicaciones (Artículos, proceso editorial y traducción)	NA			$ 0
				TOTAL DEL PROYECTO:	18.000.000

Referencia bibliográficas

[1] Constitución Política de Colombia, Artículo 7. (1991).
[2] Mansen, Captain. Lenguas Indígenas de Colombia. Instituto Caro y Cuero. (2000).
[3] S. Roy, A. K. Maiti, I. Ghosh, I. Chatterjee, and K. Ghosh, A new assistive technology in Android platform to aid vocabulary knowledge acquirement in indian sign language for better reading comprehension in l2 and mathematical ability, 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), March 2019, pp. 408-413.
[4] How to Create a Great Experience with Pepper. SoftBank Robotics. (2017).
[5] Awni Hannun, Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, and Andrew Y. Ng, Deepspeech: Scaling up end-to-end speech recognition, (2014).
[6] R. Bolhassan, J. Craneeld, and D. Dorner, Indigenous knowledge sharing in sarawak: A system-level view and its implications for the cultural heritage sector, 2014 47th Hawaii International Conference on System Sciences, Jan 2014, pp. 3378-3388.
[7] Sebastian Schneider, Michael Goerlich, and Franz Kummert, A framework for designing socially assistive robot interactions, Cognitive Systems Research 43 (2017), 301 - 312.
[8] Y. Shi, J. Bai, P. Xue, and D. Shi, Fusion feature extraction based on auditory and energy for noise-robust speech recognition, IEEE Access 7 (2019), 81911-81922.
[9] T. Barman and N. Deb, State of the art review of speech recognition using genetic algorithm, 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Sep. 2017, pp. 2944{2946.
[10] El Buscador. Universidad Santo Tomás. Oct 2018.
[11] Sebastian Weigelt and Walter Tichy, Poster: Pronat: An agent-based system design for programming in spoken natural language, 05 2015, pp. 819{820.
[12] R. Mead, Semio: Developing a cloud-based platform for multimodal conversational ai in social robotics, 2017 IEEE International Conference on Consumer Electronics (ICCE), Jan 2017,m pp. 291-292.
[13] D. Zhang, L. Wu, S. Li, Q. Zhu, and G. Zhou, Multi-modal language analysis with hierarchical interaction-level and selection-level attentions, 2019 IEEE International Conference on Multimedia and Expo (ICME), July 2019, pp. 724{729.
[14] Tumisho Mokgonyane, Tshephisho Sefara, Thipe Modipa, Mercy Mogale, Madimetja Manamela, and Phuti Manamela, Automatic speaker recognition system based on machine learning algorithms, 01 2019.
[15] Jian-Hua Tao, Jian Huang, Ya Li, Zheng Lian, and Ming-Yue Niu, Semi-supervised ladder networks for speech emotion recognition, International Journal of Automation and Computing
[16] Y. Suh, Y. Kim, H. Lim, J. Goo, Y. Jung, Y. Choi, H. Kim, D. Choi, and Y. Lee, Development of distant multi-channel speech and noise databases for speech recognition by in-door conversational robots, 2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA), Nov 2017, pp. 1-4.
[17] Jianxin Peng, Lei Zhao, and Yanmei Jiang, Investigation of word recognition for the elderly in speech and noise spatial separation, Applied Acoustics 153 (2019), 48 - 52.
[18] David R. Appleton, What do we mean by a statistical model?, Statistics in Medicine 14 (1995), no. 2, 185-197.

[19] J. L. Mayorga, C. Dominguez-Bonilla, A. Gutierrez, F. Jimenez, and H. Chamorro, Development of real-time control emulator in fpga using hiles methodology, IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society, Nov 2015, pp. 004076-004081.
[20] P. Serai, P. Wang, and E. Fosler-Lussier, Improving speech recognition error prediction for modern and off
-the-shelf speech recognizers, ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, pp. 7255-7259.
[21] L. Li, D. Wang, Y. Chen, Y. Shi, Z. Tang, and T. F. Zheng, Deep factorization for speech
signal, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing
(ICASSP), April 2018, pp. 5094-5098.
[22] T. Athanaselis, S. Bakamidis, G. Giannopoulos, I. Dologlou, and E. Fotinea, Robust speech recognition in the presence of noise using medical data, 2008 IEEE International Workshop on Imaging Systems and Techniques, Sep. 2008, pp. 349-352.
[23] Jianxin Peng, Lei Zhao, and Yanmei Jiang, Investigation of word recognition for the elderly in speech and noise spatial separation, Applied Acoustics 153 (2019), 48 - 52.
[24] D. Iguaran Fernandez, A. Molina, O. Quintero and O. Bedoya, Design and implementation of an web api for the automatic translation Colombia's language pairs: Spanish-wayuunaiki case, 05 2013, pp. 1-9.
[25] Hlldobler, B. Rumpe, and I. Weisemller, Systematically deriving domain-specic transformation languages, 2015 ACM/IEEE 18th International Conference on Model Driven Engineering Languages and Systems (MODELS), Sep. 2015, pp. 136-145.
[26] E. Kasano, S. Muramatsu, A. Matsufuji, E. Sato-Shimokawara, and T. Yamaguchi, Estimation of speakers con dence in conversation using speech information and head motion, 2019 16th International Conference on Ubiquitous Robots (UR), June 2019, pp. 294-298.
[27] J. Loewen and Kinshuk, The need for technological innovations for indigenous knowledge transfer in culturally inclusive education, 2012 IEEE 12th International Conference on Advanced Learning Technologies, July 2012, pp. 577-578.
[28] M. Masinde and P. N. Thothela, Itiki plus: A mobile based application for integrating indigenous knowledge and scientic agro-climate decision support for Africas small-scale farmers, 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT),
March 2019, pp. 303-309.
[29] T. B. Mokgonyane, T. J. Sefara, T. I. Modipa, M. M. Mogale, M. J. Manamela, and P. J. Manamela, Automatic speaker recognition system based on machine learning algorithms, 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), Jan 2019, pp. 141-146.
[30] M. Pleva, J. Juhar, S. Ondas, C. R. Hudson, C. L. Bethel, and D. W. Carruth, Novice user experiences with a voice-enabled human-robot interaction tool, 2019 29th International Conference Radioelektronika (RADIOELEKTRONIKA), April 2019, pp. 1-5.

ANEXO 1

Codification			Choregraphe		NAOqi API			Basic Channel		Remote
Locale	Language	Dialog code	Tools	Content and Samples	Notifications	ASR Engine	TTS Engine	Activity launcher	Conversation	Remote ASR Engine
ja_JP	Japanese	jpj	✓	✓	✓	✓	AiTalk	✓	✓	✓
en_US	English	enu	✓	✓	✓	✓	Nuance	✓	✓	✓
fr_FR	French	frf	✓	✓	✓	✓	Nuance	✓	✓	⊘
it_IT	Italian	iti	✓	✓	✓	✓	Nuance	✓	⊘	⊘
de_DE	German	ded	✓	✓	✓	✓	Nuance	✓	⊘	⊘
es_ES	Spanish	spe	✓	✓	✓	✓	Nuance	✓	⊘	⊘
zh_CN	Chinese	mnc	✓	✓	·	✓	Nuance	✓	⊘	⊘
nl_NL	Dutch	dun	✓	✓	·	✓	Nuance	·	⊘	⊘
ar_SA	Arabic	arw	✓	✓	·	✓	Nuance	·	⊘	⊘
ko_KR	Korean	kok	✓	✓	·	⊘	⊘	·	⊘	⊘
pl_PL	Polish	plp	✓	✓	·	⊘	⊘	·	⊘	⊘
pt_BR	Brazilian	ptb	✓	✓	·	⊘	⊘	·	⊘	⊘
pt_PT	Portuguese	ptp	✓	✓	·	⊘	⊘	·	⊘	⊘
cs_CZ	Czech	czc	✓	✓	·	⊘	⊘	·	⊘	⊘
da_DK	Danish	dad	✓	✓	·	⊘	⊘	·	⊘	⊘
fi_FI	Finnish	fif	✓	✓	·	⊘	⊘	·	⊘	⊘
sv_SE	Swedish	sws	✓	✓	·	⊘	⊘	·	⊘	⊘
ru_RU	Russian	rur	✓	✓	·	⊘	⊘	·	⊘	⊘
tr_TR	Turkish	trt	✓	✓	·	⊘	⊘	⊘	⊘	⊘
nn_NO	Norwegian	nor	✓	⊘	·	⊘	⊘	⊘	⊘	⊘
el_GR	Greek	grg	✓	⊘	·	⊘	⊘	⊘	⊘	⊘

✓ OK
· Partial / Not tested
⊘ Not Supported Yet

Figura 3. Idiomas soportados por las diferentes herramientas proporcionadas por el fabricante del Robot Pepper (Tomado de http://doc.aldebaran.com/2-4/family/pepper_technical/languages_pep.html).