



Universidad Internacional del Ecuador

Extensión Loja

ESCUELA DE INGENIERIA INFORMÁTICA Y MULTIMEDIA

Trabajo de Titulación previo a la obtención del título de

INGENIERO EN INFORMÁTICA Y MULTIMEDIA

“Evaluación de técnicas del procesamiento del lenguaje natural aplicadas en la pronunciación de niños de educación general básica de la escuela José Antonio Campos del cantón Zapotillo, provincia de Loja.”

AUTOR:

JEFFERSON VICENTE CORONEL REQUENA

DIRECTOR:

MGS. ING. LUIS ALBERTO CUENCA

MAYO 2017

LOJA - ECUADOR

Yo, **JEFFERSON VICENTE CORONEL REQUENA**, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional, y, que ha sido respaldada con su respectiva bibliografía.

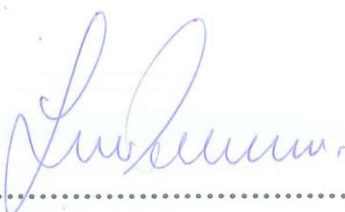
Cedo mis derechos de propiedad intelectual a la Universidad Internacional del Ecuador, para que el presente trabajo sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, reglamentos y leyes.



Jefferson Vicente Coronel Requena

C.C. 1104680028

Yo, **LUIS ALBERTO CUENCA MACAS**, certifico que conozco al autor del presente trabajo, siendo responsable exclusivo tanto en su originalidad, autenticidad, como en su contenido



Mgs. Ing. Luis Alberto Cuenca Macas

DIRECTOR DE TESIS

Agradezco a Dios, por regalarme la fortaleza, decisión e inteligencia para continuar con mis estudios superiores y permitirme subir un escalón en mi vida profesional.

A la Universidad Internacional del Ecuador Extensión Loja, forjadora de excelentes profesionales por brindarme la oportunidad de seguir creciendo profesionalmente.

Al Director y Estudiantes de la Escuela de Educación General Básica José Antonio Campos por su valiosa colaboración en la elaboración de esta investigación.

A los docentes de la U.I.D.E. por impartir sus sabios conocimientos y experiencia durante la carrera; y de manera muy especial al Mgs. Ing. Luis Alberto Cuenca Macas por su valioso aporte en la dirección y revisión de esta tesis.

A mi hermana Vanessa por su apoyo incondicional brindado durante el desarrollo de esta investigación.

Este trabajo se lo dedico a Dios por ser pilar fundamental en mi vida y permitirme alcanzar este objetivo.

A mis Abnegados Padres: Vicente y Janeth, quienes durante toda mi vida me han demostrado su amor a través de su apoyo incondicional, que este logro sea el reconocimiento de sus esfuerzos.

A mi pequeña Danna quien con sus inocentes palabras supo forjar en mí el trabajo y dedicación para alcanzar mis sueños.

A todas las personas que directamente e indirectamente han apoyado al desarrollo de esta Tesis.

Jefferson Vicente Coronel Requena

Resumen

Al desarrollar un proyecto de Inteligencia Artificial donde interviene el Procesamiento del Lenguaje Natural, surge la interrogante, ¿Qué técnica debo utilizar?, por este motivo, en esta investigación se seleccionó tres técnicas de procesamiento del lenguaje natural y se realizó la respectiva evaluación de las mismas, para lo cual fue necesario obtener muestras de audio de un grupo de 10 niños de la Escuela de Educación General Básica José Antonio Campos del cantón Zapotillo, provincia de Loja, durante 10 semanas, para con éstas muestras proceder a realizar la respectiva evaluación. Las técnicas seleccionadas fueron reconocimiento basado en el Conocimiento, reconocimiento basado en Modelos Ocultos de Markov y reconocimiento basado en Redes Neuronales Artificiales. Para realizar la evaluación se desarrolló un prototipo de aplicación en el Lenguaje de programación Java que integre estas técnicas de reconocimiento, y, tras analizar las muestras de audio con cada una de las técnicas, los resultados como palabras válidas reconocidas, tiempo de respuesta, y memoria RAM utilizada por cada técnica, fueron almacenados en una base de datos PostgreSQL, para en lo posterior a través del mismo prototipo de aplicación presentar los resultados obtenidos durante la evaluación. Como resultado de esta investigación se pudo determinar que la técnica de reconocimiento basado en Redes Neuronales Artificiales es la mejor de las tres técnicas evaluadas, por ser la que mayor número de palabras reconoció, en menor tiempo y haciendo uso de menos memoria RAM.

Palabras clave: Inteligencia Artificial, Procesamiento del Lenguaje Natural, Reconocimiento Basado en el Conocimiento, Reconocimiento Basado en Modelos Ocultos de Markov, Reconocimiento Basado en Redes Neuronales Artificiales, Java.

Abstract

When developing an Artificial Intelligence project where Natural Language Processing intervenes, the question arises, which technique should I use? for this reason, in this research we selected three techniques of natural language processing and, the respective evaluation of them, for which it was necessary to obtain audio samples from a group of 10 children of the Basic General Education School José Antonio Campos of canton Zapotillo , province of Loja for 10 weeks, with these samples proceed to make the respective evaluation. The selected techniques were recognition based on Knowledge, recognition based on Hidden Markov Models and recognition based on Artificial Neural Networks. In order to perform the evaluation, a prototype application was developed in the Java programming language that integrates these recognition techniques and after analyzing the audio samples with each of the techniques, the results as recognized valid words, response time, and RAM used by each technique, were stored in a PostgreSQL database to later on through the same application prototype present the results obtained during the evaluation. As a result of this research, it was possible to determine that the technique based on Artificial Neural Networks is the best of the three techniques evaluated, since it was the one with the highest number of words recognized, in less time and using less memory.

Keywords: Artificial Intelligence, Natural Language Processing, Knowledge Based Recognition, Recognition Based on Hidden Markov Models, Recognition Based on Artificial Neural Networks, Java.

**EVALUACIÓN DE TÉCNICAS DEL PROCESAMIENTO DEL LENGUAJE
NATURAL APLICADAS EN LA PRONUNCIACIÓN DE NIÑOS DE EDUCACIÓN
GENERAL BÁSICA DE LA ESCUELA JOSÉ ANTONIO CAMPOS DEL CANTÓN
ZAPOTILLO, PROVINCIA DE LOJA.**

Resumen	v
Abstract	vi
Índice de Tablas	ix
Índice de Figuras.....	x
Índice de Gráficos	xi
Índice de Anexos.....	xii
Introducción	1
Capítulo 1	2
1 Análisis	2
1.1 Planteamiento del Problema.....	2
1.2 Justificación.....	2
1.3 Objetivos del Proyecto	3
1.3.1 Objetivo General:.....	3
1.3.2 Objetivos específicos:	3
1.4 Hipótesis.....	3
1.4.1 Variables	4
1.5 Alcance.....	4
1.6 Metodología	6
1.6.1 Metodología ágil.	6
1.6.2 Programación Extrema(XP).....	8
1.7 Herramientas utilizadas en el proyecto.	11
1.8 Costo del proyecto.....	14
Capítulo 2	18
2 Marco Referencial	18
2.1 Definiciones	18
2.1.1 Lenguaje.....	18
2.1.2 Lenguaje Natural.....	18
2.1.3 Lengua y Habla	19

2.1.4	Lenguaje Formal	19
2.1.5	Lenguaje de programación.....	20
2.1.6	Procesamiento del Lenguaje Natural (PLN)	20
2.2	Reconocimiento Automático del Habla (ASR)	25
2.2.1	Factores que determinan la complejidad del reconocimiento del habla.	25
2.2.2	Áreas de trabajo que intervienen en el diseño y especificación de sistemas de reconocimiento del habla.	27
2.2.3	Técnicas de diseño de reconocedores del habla.....	31
2.2.4	Etapas de un Sistema de Reconocimiento Automático del Habla (ASR).....	39
2.2.5	Comparativa de las técnicas de procesamiento del lenguaje natural.	41
2.2.6	Proyectos de Procesamiento del Lenguaje Natural en el Ecuador.....	43
Capítulo 3		47
3	Materiales y Métodos	47
3.1	Localización de la zona de estudio.....	47
3.1.1	Ubicación Geográfica	47
3.2	Materiales	47
3.2.1	Materiales y Equipos de Campo	47
3.2.2	Materiales de Oficina	47
3.3	Selección de técnicas de procesamiento del lenguaje natural	47
3.4	Selección del Lenguaje de Programación	48
3.5	Cumplimiento de los objetivos.....	50
3.6	Tamaño de la población y muestra.....	51
3.7	Técnicas e Instrumentos de Investigación.....	52
Capítulo 4		55
4	Recolección y Análisis de Datos	55
4.1	Recolección de datos.....	55
4.1.1	Trabajo de Campo.....	55
4.1.2	Codificación de los datos.....	57
4.2	Análisis de datos.....	62
Capítulo 5		82
5	Resultados	82
Capítulo 6		83
6	Problemas encontrados en el proyecto	83
Conclusiones		85
Recomendaciones		87
Bibliografía		89
Anexos		93

Índice de Tablas

Tabla 1 Costo recursos humanos	15
Tabla 2 Costo recursos de hardware	15
Tabla 3 Costo recursos materiales.	16
Tabla 4 Costo de movilización.....	16
Tabla 5 Costo total de recursos.	16
Tabla 6 Comparativa de las técnicas de procesamiento del lenguaje natural	41
Tabla 7 Ranking IEEE Spectrum de lenguajes de programación	48
Tabla 8 Comparativa de los lenguajes de programación	49
Tabla 9 Tamaño de la población y muestra	52
Tabla 10 Registro de la toma de muestras de audio.....	56
Tabla 11 Codificación de los datos.	58
Tabla 12 Análisis de palabras reconocidas	62
Tabla 13 Mínimo, promedio y máximo de palabras reconocidas por cada técnica.	63
Tabla 14 Porcentaje promedio de reconocimiento por cada técnica.....	64
Tabla 15 Mínimo, promedio y máximo de tiempo utilizado por cada técnica.	65
Tabla 16 Mínimo, promedio y máximo de memoria RAM utilizada por cada técnica.	66
Tabla 17 Promedio de palabras pronunciadas por cada niño.....	68
Tabla 18 Nivel de pronunciamiento del niño 1.....	69
Tabla 19 Nivel de pronunciamiento del niño 2.....	70
Tabla 20 Nivel de pronunciamiento del niño 3.....	71
Tabla 21 Nivel de pronunciamiento del niño 4.....	72
Tabla 22 Nivel de pronunciamiento del niño 5.....	74
Tabla 23 Nivel de pronunciamiento del niño 6.....	75
Tabla 24 Nivel de pronunciamiento del niño 7.....	76
Tabla 25 Nivel de pronunciamiento del niño 8.....	77
Tabla 26 Nivel de pronunciamiento del niño 9.....	79
Tabla 27 Nivel de pronunciamiento del niño 10.....	80
Tabla 28 Glosario de Términos.....	93
Tabla 29 Nivel de Pronunciamiento.....	100
Tabla 30 Niños de la muestra y sus códigos respectivos.	101
Tabla 31 Requerimiento funcional DANNA	102
Tabla 32 Historias de usuario.....	104
Tabla 33 Tarjeta CRC de la clase Evaltech.....	106
Tabla 34 Tarjeta CRC de la clase principal	106
Tabla 35 Tarjeta CRC de la clase ReconocedorHMM	107
Tabla 36 Tarjeta CRC de la clase ReconocedorRNA	107
Tabla 37 Tarjeta CRC de la clase Resumen.....	108

Índice de Figuras

Figura 1 Actividades estructurales de XP	9
Figura 2 Lenguaje Natural.	19
Figura 3 Ejemplo de Gramática G.	24
Figura 4 Reconocimiento del habla.	25
Figura 5 Alineamiento temporal entre la Referencia y la palabra a reconocer.....	33
Figura 6 Modelo de Markov con n estados.....	35
Figura 7 Red Neuronal de 3 Capas.	38
Figura 8 Esquema genérico de un sistema ASR.	41
Figura 9 Sistema DANNA.	53
Figura 10 Diccionario Fonético de más de 20K palabras.	94
Figura 11 Modelo de Lenguaje Estadístico.	95
Figura 12 Diccionario Fonético propio.....	96
Figura 13 Gramática propia.	97

Índice de Gráficos

Gráfico 1 Tamaño de la población y muestra.	52
Gráfico 2 Media y mediana de palabras reconocidas por cada técnica.	63
Gráfico 3 Mínimo, promedio y máximo de palabras reconocidas por cada técnica.	64
Gráfico 4 Porcentaje de reconocimiento promedio de cada técnica.	65
Gráfico 5 Mínimo, promedio y máximo de tiempo utilizado por cada técnica.	66
Gráfico 6 Mínimo, promedio y máximo de memoria RAM utilizada por cada técnica.	67
Gráfico 7 Promedio de palabras pronunciadas por cada niño.....	68
Gráfico 8 Nivel de pronunciamiento del niño 1.....	70
Gráfico 9 Nivel de pronunciamiento del niño 2.....	71
Gráfico 10 Nivel de pronunciamiento del niño 3.....	72
Gráfico 11 Nivel de pronunciamiento del niño 4.....	73
Gráfico 12 Nivel de pronunciamiento del niño 5.....	75
Gráfico 13 Nivel de pronunciamiento del niño 6.....	76
Gráfico 14 Nivel de pronunciamiento del niño 7.....	77
Gráfico 15 Nivel de pronunciamiento del niño 8.....	78
Gráfico 16 Nivel de pronunciamiento del niño 9.....	80
Gráfico 17 Nivel de pronunciamiento del niño 10.....	81

Índice de Anexos.

Anexo A. Glosario de Términos.....	93
Anexo B. Diccionario y Modelo de Lenguaje de CMU utilizados en el Reconocimiento Basado en Modelos Ocultos de Markov.	94
Anexo C. Diccionario y gramática propios utilizados en el Reconocimiento Basado en el Conocimiento.....	96
Anexo D. Entrenar un modelo acústico para CMUSphinx.....	98
Anexo E. Lectura “El pajarito que no quería volar” de Sergio R. Peral.	99
Anexo F. Nivel de pronunciamiento.	100
Anexo G. Niños de la muestra.....	101
Anexo H. Metodología de desarrollo.	102
Anexo I. Solicitud entregada a la Escuela de Educación General Básica José Antonio Campos del cantón Zapotillo.	109

Introducción

La inteligencia artificial es una ciencia que día a día trata de formar parte importante de nuestras vidas, la encontramos en smartphones, computadores, servidores web, equipos para domótica entre otros equipos eléctricos.

Hoy en día es común utilizar el procesamiento del lenguaje natural, ya sea en el reconocimiento de voz, análisis de texto, recolección de información o síntesis de voz, al utilizar la búsqueda por voz de nuestros dispositivos inteligentes ya sea un smartphone con Android y las búsquedas de google, o con IOS y su asistente virtual Siri.

En el ámbito del desarrollo de software al momento de desarrollar una solución que utilice procesamiento del lenguaje natural, es común preguntarse ¿Qué técnica debo utilizar? razón por la cual a través del presente trabajo de investigación se pretende responder esta pregunta mediante la selección de tres técnicas de procesamiento del lenguaje natural, y su respectiva evaluación en aspectos como: palabras reconocidas, tiempo de ejecución y memoria RAM utilizada, logrando determinar luego de este análisis ¿cuál de las tres técnicas seleccionadas es la mejor?, contestando la pregunta antes mencionada y de esta manera ayudar a tomar la mejor decisión al momento de desarrollar un proyecto que haga uso del Procesamiento del Lenguaje Natural.

Capítulo 1

1 Análisis

1.1 Planteamiento del Problema

Existe variedad de técnicas para el procesamiento del lenguaje, pero al momento de realizar un proyecto resulta difícil inclinarse hacia una u otra, por desconocimiento de las mismas o porque no existe un análisis a fondo de estas; mediante este proyecto se planea realizar la evaluación de estas técnicas, determinando así la técnica que ofrece mejores resultados, además de los beneficios de la utilización de una u otra técnica. Para lograrlo se tomará una muestra de 10 niños de la Escuela de Educación General Básica José Antonio Campos del Cantón Zapotillo, provincia de Loja; a los cuales se les hará la toma de muestras periódicamente durante el transcurso del proyecto, para el posterior análisis y evaluación.

1.2 Justificación

El siguiente proyecto es factible económicamente porque se dispone de los recursos materiales y económicos para la realización de esta investigación, además se pretende utilizar en su mayoría licencias GPL (General Public License).

Se justifica científicamente pues se puede evaluar la eficiencia de las diferentes técnicas del procesamiento del lenguaje natural. Además, se obtendrá una recopilación de las mejores técnicas de procesamiento del lenguaje natural, cantidad de recursos que utilizan, conociendo más a fondo cada una de estas para a futuro tomar la mejor decisión al momento de utilizar una u otra técnica, ahorrando recursos monetarios, de tiempo y de personal en proyectos que utilizan estas técnicas.

Se justifica técnicamente pues se cuenta con el hardware que soporte el prototipo de aplicación, también el software o aplicativo va a ser desarrollado durante el transcurso del proyecto, y porque se puede incluir las técnicas en un aplicativo.

Se justifica operativamente porque la persona que desarrollará el proyecto, en este caso mi persona, está apta para el desarrollo del mismo.

Se justifica legalmente porque se pretende utilizar software y herramientas de desarrollo con licencias GPL (General Public License) o de código abierto.

1.3 Objetivos del Proyecto

1.3.1 Objetivo General:

Evaluación de técnicas del procesamiento del lenguaje natural aplicadas en la pronunciación de niños de educación general básica de la escuela José Antonio Campos del cantón Zapotillo, provincia de Loja.

1.3.2 Objetivos específicos:

- Medir el nivel de pronunciamiento de la población seleccionada periódicamente.
- Determinar las técnicas de procesamiento del lenguaje natural que van a ser evaluadas durante el presente proyecto.
- Comparar como las diferentes técnicas influyen en el reconocimiento de la pronunciación de los niños.
- Desarrollar una aplicación para la captura de datos y presentación de resultados, utilizando las diferentes técnicas del procesamiento del lenguaje natural seleccionadas.

1.4 Hipótesis

Se espera obtener mejores resultados de reconocimiento mediante la utilización de la técnica de procesamiento del lenguaje natural basada en Modelos Ocultos de Markov frente a las otras técnicas.

1.4.1 Variables

1.4.1.1 Variables Independientes

- a) Técnica Seleccionada 1
- b) Técnica Seleccionada 2
- c) Técnica Seleccionada 3
- d) Niños

1.4.1.2 Variables Dependientes

- a) Nivel de Pronunciación de palabras
- b) Palabras Reconocidas Válidas
- c) Tiempo de Respuesta
- d) Cantidad de Memoria RAM utilizada

1.5 Alcance

Mediante la utilización de técnicas del procesamiento del lenguaje natural implementadas en una aplicación, se pretende realizar la evaluación de éstas técnicas, tomando muestras de audio a un grupo de 10 niños de la Escuela de Educación General Básica José Antonio Campos del Cantón Zapotillo.

La toma de muestras de los niños se realizará semanalmente a través de un dispositivo de grabación portátil, estas muestras serán procesadas y los resultados serán almacenados en una base de datos para su posterior análisis y evaluación correspondiente.

Para seleccionar las técnicas de procesamiento del lenguaje natural a ser evaluadas se realizará una extensa búsqueda bibliográfica para tener información completa de cada una de las técnicas, permitiendo así realizar la correcta selección de las mismas.

La aplicación será multiplataforma, para sistemas operativos de escritorio y al final debe presentar los resultados obtenidos e información como: tiempo de respuesta, exactitud de reconocimiento y memoria RAM utilizada durante el proceso.

Para el cumplimiento del objetivo planteado medir periódicamente el nivel de pronunciamiento de la población seleccionada, se utilizará la mejor técnica de acuerdo a los resultados obtenidos, el número de palabras reconocidas por esta técnica, para cada niño, durante cada semana, será el valor para medir el nivel de pronunciamiento, para una mejor interpretación, en la pantalla de presentación de resultados se realizará una gráfica por cada niño con esta información.

Para comparar como las diferentes técnicas influyen en el reconocimiento de la pronunciación de los niños, se procederá a tomar muestras de audio de una lectura antes seleccionada realizada por cada niño, luego estos audios se convertirán a un formato común aceptado por el prototipo de aplicación que integra las técnicas para estar en igualdad de condiciones y el resultado de palabras reconocidas y palabras validas junto con el tiempo de ejecución y la memoria utilizada será almacenado en la base de datos.

Para desarrollar la aplicación para la captura de datos y presentación de resultados, utilizando las diferentes técnicas del procesamiento del lenguaje natural seleccionadas, se realizará una búsqueda extensa de librerías que utilicen estas técnicas, permitan la integración de las mismas, que en lo posible sean de código abierto y se adapten al desarrollo multiplataforma.

1.6 Metodología

En este apartado se pretende explicar la metodología a utilizar para la consecución de los objetivos de este proyecto.

Para este proyecto se utilizará la experimentación científica, basándose en la toma de muestras, y, experimentando con estas con las diferentes técnicas de procesamiento del lenguaje natural para lograr así realizar la correcta evaluación de las mismas, para ello se realizará primero el estudio y la selección de las técnicas a ser evaluadas, posteriormente se realizará el prototipo de aplicación que permita la integración de las diferentes técnicas previamente seleccionadas, luego se realizará la toma de muestras de audio de los niños y se experimentará con estas a través de la aplicación anteriormente desarrollada, toda la información de la evaluación será almacenada en una base de datos, para a continuación realizar el análisis de datos y obtener así los resultados como: tiempo de respuesta, exactitud al reconocer palabras, consumo de recursos.

Para la generación de resultados se realizará una ventana dentro del prototipo de aplicación con la información antes mencionada.

1.6.1 Metodología ágil.

Del 11 al 13 de febrero de 2001, en el Lodge en Snowbird en las montañas Wasatch de Utah, diecisiete personas (grupo conocido como la “Alianza Ágil”) se reunieron para hablar, esquiar, relajarse y tratar de encontrar terreno común y, por supuesto, comer. Lo que surgió fue el “Manifiesto por el Desarrollo ágil de Software”. El cual es una alternativa a las metodologías tradicionales de desarrollo de software, el manifiesto establece lo siguiente (Beck et al., 2001):

“Estamos descubriendo formas mejores de desarrollar software tanto por nuestra propia experiencia como ayudando a terceros. A través de este trabajo hemos aprendido a valorar:

Individuos e interacciones sobre procesos y herramientas.

Software funcionando sobre documentación extensiva.

Colaboración con el cliente sobre negociación contractual.

Respuesta ante el cambio sobre seguir un plan.

Esto es, aunque valoramos los elementos de la derecha, valoramos más los de la izquierda.”

“Un manifiesto normalmente se asocia con un movimiento político emergente: ataca a la vieja guardia y sugiere un cambio revolucionario (se espera que para mejorar). En cierta forma, de eso es de lo que trata el desarrollo ágil.” (Pressman, 2010).

En base a estos conceptos se puede definir como metodología ágil a todas las metodologías que cumplan lo expuesto en el manifiesto por el desarrollo ágil de software en el año 2001.

1.6.1.1 Principios ágiles

Los siguientes principios ágiles (Alliance, 2001) se basan en el manifiesto Ágil:

1. La prioridad más alta es satisfacer al cliente a través de la entrega pronta y continua de software valioso.
2. Son bienvenidos los requerimientos cambiantes, aun en una etapa avanzada del desarrollo. Los procesos ágiles dominan el cambio para provecho de la ventaja competitiva del cliente.
3. Entregar con frecuencia software que funcione, de dos semanas a un par de meses, de preferencia lo más pronto que se pueda.
4. Las personas de negocios y los desarrolladores deben trabajar juntos, a diario y durante todo el proyecto.

5. Hay que desarrollar los proyectos con individuos motivados. Debe darse a éstos el ambiente y el apoyo que necesiten, y confiar en que harán el trabajo.
6. El método más eficiente y eficaz para transmitir información a los integrantes de un equipo de desarrollo, y entre éstos, es la conversación cara a cara.
7. La medida principal de avance es el software que funciona.
8. Los procesos ágiles promueven el desarrollo sostenible. Los patrocinadores, desarrolladores y usuarios deben poder mantener un ritmo constante en forma indefinida.
9. La atención continua a la excelencia técnica y el buen diseño mejora la agilidad.
10. Es esencial la simplicidad: el arte de maximizar la cantidad de trabajo no realizado.
11. Las mejores arquitecturas, requerimientos y diseños surgen de los equipos con organización propia.
12. El equipo reflexiona a intervalos regulares sobre cómo ser más eficaz, para después afinar y ajustar su comportamiento en consecuencia.

No toda metodología aplica estos doce principios, algunas inclusive pueden obviar algunos de ellos, pero si poseen un espíritu ágil. (Pressman, 2010)

1.6.2 Programación Extrema(XP)

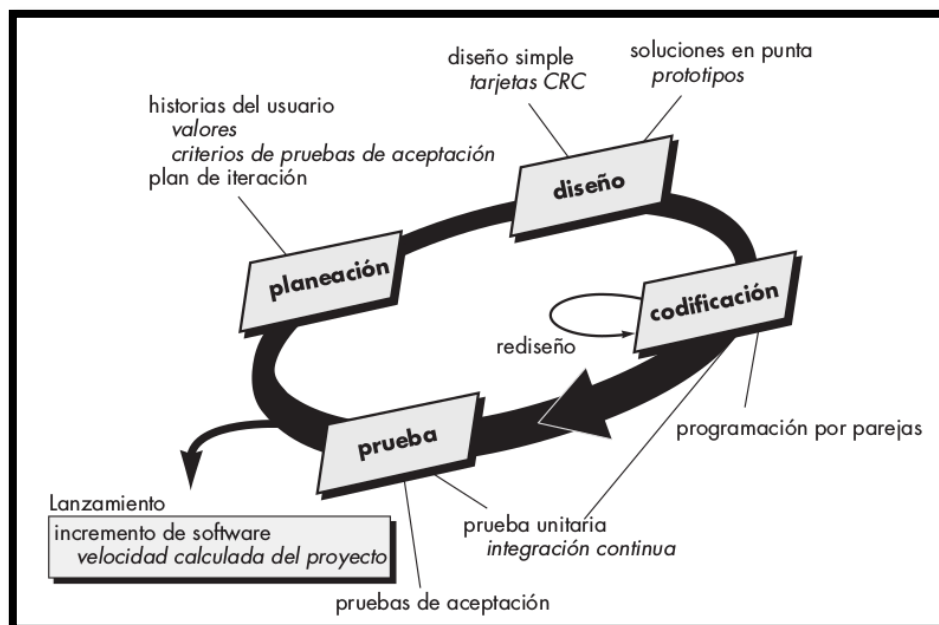
Según Beck (1999) XP es una metodología ligera para equipos de pequeño a mediano tamaño que desarrollan software frente a requisitos vagos o que cambian rápidamente.

XP es un estilo de desarrollo de software centrado en la aplicación excelente de técnicas de programación, comunicación clara y trabajo en equipo que permite realizar cosas que previamente no se podía ni siquiera imaginar. (Beck, 2004)

1.6.2.1 El Proceso XP

La programación extrema usa un enfoque orientado a objetos como paradigma preferido de desarrollo, y engloba un conjunto de reglas y prácticas que ocurren en el contexto de cuatro actividades estructurales: planeación, diseño, codificación y pruebas. (Pressman, 2010). La Figura 1. Ilustra el proceso XP y resalta algunas de las ideas y tareas clave que se asocian con cada actividad estructural.

Figura 1 Actividades estructurales de XP



Fuente: (Pressman, 2010)

Elaborado por: Roger S. Pressman, Ph.D.

A continuación, se describen las actividades estructurales de XP:

- **Planeación:** En la etapa de planeación se levantan requerimientos, se crean historias de usuario a las cuales el cliente le asigna la prioridad durante el proceso, estas son evaluadas por el equipo XP y posteriormente le asignan un

costo, medido en semanas. Los clientes y desarrolladores deciden como agrupar las historias para la siguiente entrega (Incremento de software). Después de la primera entrega del proyecto, el equipo calcula la velocidad de éste. Las historias pueden ser modificadas durante el transcurso del proyecto, por lo cual, los planes del mismo deberán ser modificados.

- **Diseño:** En Xp el diseño sigue el principio MS (Mantenlo Sencillo), el diseño guía la implementación de una historia tal cual está escrita: sin más ni menos. En esta fase se estimula el uso de tarjetas CRC(Clase-Responsabilidad-Colaborador) las cuales identifican y organizan las clases relevantes para el incremento actual del software. Si se encuentra un problema de diseño difícil, XP recomienda la creación inmediata de un prototipo operativo de esa porción de diseño, denominado solución en punta. En esta fase también se utiliza el rediseño, el cual es un método para la optimización del diseño. El diseño en XP ocurre tan antes como después de que comienza la codificación.
- **Codificación:** En ésta etapa se desarrolla una serie de pruebas unitarias para cada historia que se va a incluir en la entrega en curso. Esta prueba sirve para que el desarrollador pueda centrarse en lo que realmente debe implementar para pasar la prueba. Una vez terminado el código se le aplica esta prueba, esto sirve como retroalimentación para el equipo de desarrollo. Es fundamental la programación en parejas. A medida que las parejas terminan su trabajo, el código se integra con el trabajo de los demás. Esto lo realiza el equipo de integración o la pareja de desarrolladores.
- **Pruebas:** Se implementan las pruebas unitarias anteriormente creadas, esto estimula una estrategia de regresión siempre que se modifique el código. Las

pruebas de aceptación o pruebas del cliente, son especificadas por el cliente y se centran en características y funcionalidades pueden ser revisadas por este.

Considerando la dificultad del proyecto y lo cambiante que este puede ser, se cree conveniente utilizar la metodología XP por ser la que mejor se adapta a este tipo de proyecto y por la posibilidad de efectuar cambios durante el transcurso del mismo sin que esto signifique un costo mayor o que se extienda demasiado el tiempo de entrega del mismo, permitiendo además entregar un producto de calidad acorde a las necesidades del proyecto. Los artefactos de trabajo de la metodología XP se encuentran en el Anexo H.

1.7 Herramientas utilizadas en el proyecto.

Para cumplir con los objetivos del proyecto y realizar una aplicación multiplataforma se consideró la utilización de herramientas de software libre o en su defecto herramientas gratuitas que permitan el desarrollo multiplataforma y ofrezcan una relación costo-beneficio razonable en el desarrollo del proyecto. El desarrollador posee un nivel intermedio en el manejo de estas herramientas, pero en Internet existe gran documentación sobre las mismas y como utilizarlas.

Para el desarrollo del proyecto se utilizaron las siguientes herramientas de software:

- **NetBeans IDE 8.1:** es un entorno de desarrollo libre y gratuito sin restricciones de uso. Se basa en las licencias Common Development and Distribution License (CDDL) v1.0 and the GNU General Public License (GPL) v2. Está escrita en java y será la principal herramienta de desarrollo durante todo el proyecto.
- **Apache Maven Plug-in 4:** Maven es una herramienta de gestión y comprensión del desarrollo de proyectos. Utilizada para crear y administrar cualquier proyecto basado en java. La publicación de sitios y la distribución se controlan

desde el archivo declarativo pom.xml. Se utilizará para la importación de las librerías o dependencias del proyecto.

- **JDK 1.8:** La escritura de aplicaciones y applets de Java necesita herramientas de desarrollo como JDK. JDK incluye Java Runtime Environment, el compilador Java y las API de Java. Familiarizarse resulta fácil para los programadores nuevos y con experiencia. Es la base en el desarrollo de aplicaciones java.
- **JAVA:** Java es un lenguaje de programación y una plataforma informática comercializada por primera vez en 1995 por Sun Microsystems. Java está diseñado para permitir el desarrollo de aplicaciones portátiles de elevado rendimiento para el más amplio rango de plataformas informáticas posible. Se utiliza este lenguaje de desarrollo porque permite el desarrollo multiplataforma, la integración de las técnicas en una sola aplicación y por la amplia documentación que posee.
- **PostgreSQL 9.5:** PostgreSQL es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD (*Berkeley Software Distribution*) y con su código fuente disponible libremente. Utiliza un modelo cliente-servidor, ofrece estabilidad, robustez y facilidad de implementación ideal para el desarrollo del proyecto.
- **pgAdmin 1.22:** Es una herramienta de administración de código abierto. Incluye una interfaz gráfica de administración, una herramienta de consulta SQL, un editor de código de procedimiento y mucho más. Se utilizará para la creación y manejo de la base de datos simplificando mucho la tarea en lugar de hacerlo mediante consola.
- **Audacity 2.1.2:** Audacity es una aplicación libre, de código abierto, multiplataforma para grabación y edición de sonido. Se utilizará esta aplicación

para convertir las muestras de audio a un formato y codificación soportada por la aplicación desarrollada en el proyecto.

- **Gedit 3:** Es un editor de texto pequeño y ligero para el escritorio de gnome. Se utilizará principalmente en la elaboración de la gramática y del diccionario fonético.
- **Sphinx4:** Sphinx4 es una biblioteca de reconocimiento de voz en Java. Proporciona una API rápida y fácil para convertir las grabaciones de voz en texto con la ayuda modelos acústicos CMUSphinx. Se la utilizará para implementar algunas de las técnicas seleccionadas.
- **CMUSphinx-es 5.2:** es el modelo acústico en español, que será implementado en algunas de las técnicas seleccionadas. Este modelo posee un diccionario de 20000 palabras y está altamente entrenado.
- **Google Cloud SDK 154:** El Cloud SDK es un conjunto de herramientas para Cloud Platform. Las cuales permitirán acceder a los servicios en la nube de google en este caso Google Speech API y Google Cloud Storage. Utilizados por una técnica de reconocimiento.
- **Google Speech API:** Librería de Google utilizada para el reconocimiento de las palabras mediante la utilización de redes neuronales y machine learning. Además, ofrece la posibilidad de utilizar archivos de Google Cloud Storage.
- **Google Cloud Storage:** Google Cloud Storage es un almacenamiento de objetos unificado para desarrolladores y empresas, desde la administración de datos en vivo hasta el análisis de datos y el archivado de datos. Será utilizado para almacenar las muestras de audio a ser analizadas que superen el minuto de duración.

- **Google API Client Library for JAVA:** La Biblioteca de cliente de API de Google para Java proporciona funcionalidad común a todas las API de Google, por ejemplo, transporte HTTP, tratamiento de errores, autenticación, carga y descarga de medios. Se la utilizará para la autenticación con los servidores de Google.
- **jFreeChart 1.0.19:** es una librería 100% libre para la elaboración de gráficos en JAVA de manera fácil y con un estilo profesional. Es de código abierto y se distribuye bajo términos de licencia LGPL (*Lesser General Public License*) el cual permite su uso en aplicaciones propietarias.
- **JSpeech(JSGF):** es un formato de gramática independiente de la plataforma, utilizado por los reconocedores de voz para determinar lo que el reconocedor debe escuchar. JSGF adopta el estilo y las convenciones del lenguaje de programación JAVA además del uso de anotaciones gramaticales tradicionales.

1.8 Costo del proyecto.

Este proyecto tiene una duración de 12 meses laborando 20 horas por semana y como el software utilizado es software libre no se considera el costo del software dentro del presupuesto, solo se toma en cuenta el hardware, recursos humanos, recursos materiales y de movilización. La remuneración a considerar por el tiempo utilizado en el proyecto será el valor por hora en base de una remuneración básica unificada mensual según el acuerdo ministerial vigente para el año 2016 (Trabajo, 2016).

Tabla 1 Costo recursos humanos

Concepto	Nro. de horas	Valor Unitario	Valor Total
Investigador y Desarrollador del proyecto.	960	2,29	2198,40
Total			2198,40

Elaborado por: El Autor.

Tabla 2 Costo recursos de hardware

Concepto	Cantidad	Valor Unitario	Valor Total
Computador Portátil	1	1200	1200
Grabadora de Voz	1	70	70
Impresora Epson L355	1	400	400
Memoria USB 16GB HP	1	20	20
Total			1690

Elaborado por: El Autor.

Tabla 3 Costo recursos materiales.

Concepto	Cantidad	Valor Unitario	Valor Total
Servicios Básicos	12	10	120
Internet	12	25	300
Recarga de Tinta	4	12	48
Copias	200	0,01	2
Resma de papel A4	1	3	3
Total			473

Elaborado por: El Autor.

Tabla 4 Costo de movilización.

Concepto	Cantidad	Valor Unitario	Valor Total
Pasaje Loja – Zapotillo	10	10	100
Pasaje Zapotillo – Loja	10	10	100
Total			200

Elaborado por: El Autor.

Tabla 5 Costo total de recursos.

Concepto	Subtotal
----------	----------

Recursos Humanos	2198,40
Recursos de Hardware	1690,00
Recursos Materiales	473,00
Movilización	200,00
Total	4561,40

Capítulo 2

2 Marco Referencial

2.1 Definiciones

2.1.1 Lenguaje

Un lenguaje se puede definir de diferentes formas: desde el punto de vista funcional lingüístico se define como una función que expresa pensamientos y comunicaciones entre la gente. Esta función puede realizarse mediante signos escritos (escritura) o mediante señales y vocales (voz). Desde un punto de vista formal se define como un conjunto de frases, que generalmente es infinito y se forma con combinaciones de elementos tomados de un conjunto (usualmente finito) llamado alfabeto, respetando un conjunto de reglas de formación (sintácticas o gramaticales) y de sentido (semánticas). Además de las características fundamentales del lenguaje debe considerarse que sea funcional, es decir, el lenguaje debe permitirnos expresar nuestras ideas.(Sethi, 1992).

2.1.2 Lenguaje Natural

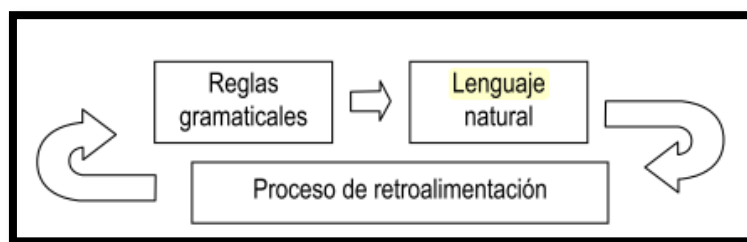
Un lenguaje natural es aquel que ha evolucionado con el tiempo para fines de comunicación humana, como el español o alemán (Brookshear, 1993). Estos lenguajes continúan su evolución sin considerar la gramática, cualquier regla se desarrolla después de sucedido el hecho.

En un primer resumen, los lenguajes naturales se caracterizan por las siguientes propiedades (Cortez, Vega, & Pariona, 2009):

1. Un lenguaje natural se define a partir de una gramática G , sin embargo, este se enriquece progresivamente modificando así también la gramática que la define. Esto dificulta la formalización de la definición de G .

2. Un LN tiene un gran poder expresivo debido a la riqueza del componente semántico (poli semántica). Esto dificulta aún más la formalización completa de su gramática.

Figura 2 Lenguaje Natural.



Fuente: (Cortez et al., 2009)

Elaborado por: (Cortez et al., 2009).

2.1.3 Lengua y Habla

La lengua no es función del sujeto hablante, sino el producto que el individuo registra pasivamente. Nunca supone premeditación y la reflexión no interviene en ella más que para la actividad de clasificar.

El habla es el acto individual de voluntad y de inteligencia, ya que supone composición premeditada haciendo uso de la lengua (Cortez et al., 2009).

2.1.4 Lenguaje Formal

El lenguaje formal es aquel que el hombre ha desarrollado para expresar las situaciones que se dan en específico en cada área del conocimiento científico. Los lenguajes formales pueden ser utilizados para modelar una teoría de la mecánica, física, matemática, ingeniería eléctrica, o de otra naturaleza, con la ventaja de que en estos toda ambigüedad es eliminada. Revisten especial importancia los lenguajes de programación de computadoras, y estas se definen considerando un conjunto de componentes léxicos, reglas gramaticales y una delimitación semántica.

En resumen, las características de los lenguajes formales son las siguientes:

1. Se desarrollan a partir de una gramática G preestablecida.
2. Componente semántico mínimo.
3. Posibilidad de incrementar el componente semántico de acuerdo con la teoría a formalizar.
4. La sintaxis produce oraciones no ambiguas.
5. Completa formalización y por esto, el potencial de la construcción computacional(Cortez et al., 2009).

2.1.5 Lenguaje de programación

Un lenguaje de programación es un lenguaje formal definido como un conjunto de elementos (componentes léxicos) organizados a través de constructores (reglas gramaticales) que permiten escribir un programa y que éste sea entendido por el computador y pueda ser trasladado a computadores similares para su funcionamiento en otros sistemas. Un programa es una secuencia de instrucciones ordenadas correctamente que permiten realizar una tarea o trabajo específico. Un lenguaje de programación se basa en dos elementos muy importantes:

- **Sintaxis:** que se encarga del orden correcto de los componentes léxicos
- **Semántica:** se encarga de que cada “oración” del lenguaje de programación utilizado tenga un significado correcto.

2.1.6 Procesamiento del Lenguaje Natural (PLN)

Una de las tareas fundamentales de la Inteligencia Artificial (IA) es la manipulación de lenguajes naturales usando herramientas de computación, en esta, los lenguajes de programación juegan un papel importante, ya que forman el enlace necesario entre los lenguajes naturales y su manipulación por una máquina. El PLN consiste en la utilización de un lenguaje natural para comunicarse con la computadora, debiendo ésta entender las oraciones que le sean proporcionadas, el uso de estos lenguajes naturales,

facilita el desarrollo de programas que realicen tareas relacionadas con el lenguaje o bien, desarrollar modelos que ayuden a comprender los mecanismos humanos relacionados con el lenguaje.(Cortez et al., 2009).

2.1.6.1 Aplicaciones del PLN

Las aplicaciones del PLN son muy variadas, ya que su alcance es muy grande, algunas de las aplicaciones son(Cortez et al., 2009):

- Traducción automática
- Recuperación de la información
- Extracción de Información y Resúmenes
- Resolución cooperativa de problemas
- Tutores inteligentes
- Reconocimiento de Voz

2.1.6.2 Arquitectura de un sistema de PLN

La arquitectura de un sistema de PLN se sustenta en una definición del LN por niveles: estos son (Cortez et al., 2009): fonológico, morfológico, sintáctico, semántico, y pragmático.

- A. Nivel Fonológico: trata de cómo las palabras se relacionan con los sonidos que representan.
- B. Nivel Morfológico: trata de cómo las palabras se construyen a partir de unas unidades de significado más pequeñas llamadas morfemas.

- C. Nivel Sintáctico: trata de cómo las palabras pueden unirse para formar oraciones, fijando el papel estructural que cada palabra juega en la oración y que sintagmas son parte de otros sintagmas.
- D. Nivel Semántico: trata del significado de las palabras y de cómo los significados se unen para dar significado a una oración, también se refiere al significado independiente del contexto, es decir de la oración aislada.
- E. Nivel Pragmático: trata de cómo las oraciones se usan en distintas situaciones y de cómo el uso afecta al significado de las oraciones. Se reconoce un subnivel recursivo: discursivo, que trata de cómo el significado de una oración se ve afectado por las oraciones inmediatamente anteriores.

La arquitectura del sistema de procesamiento del lenguaje natural muestra como la computadora interpreta y analiza las oraciones que le sean proporcionadas.

La siguiente es una breve explicación del sistema:

- a. El usuario expresa al computador lo que desea hacer.
- b. El computador realiza el análisis morfológico y sintáctico.
- c. Luego analiza las oraciones semánticamente, es decir saber cuál es el significado de la oración.
- d. Ahora procede al análisis pragmático, es decir analiza las oraciones en contexto.
- e. Y finalmente devuelve al usuario el resultado de la expresión final.

2.1.6.3 Sintaxis y Gramática

La sintaxis se define como la disposición de palabras en una oración para mostrar su relación. Describe la secuencia de símbolos que constituyen programas validos (Cortez,

2002; Hopcroft, Motwani, & Ullman, 1993). La sintaxis muestra al hablante la forma como debe escribir buenas oraciones. La sintaxis es más útil al usuario del lenguaje que al sistema de PLN.

Una gramática G es un modelo lingüístico - matemático que describe el orden sintáctico que deben cumplir las frases bien formadas de un lenguaje (Brookshear, 1993; Ravi Sethi, Ullman, & Aho, 1990). Una gramática se define formalmente de la siguiente forma:

$G = (VT, VN, P, S)$ donde:

VT: conjunto finito de símbolos terminales del lenguaje

VN: conjunto finito de símbolos no terminales

P: conjunto finito de reglas de producción

S: Símbolo distinguido o axioma inicial a partir del cual se reconocerán las secuencias de L aplicando sucesivamente las reglas de producción.

Se considera la siguiente gramática $G (VN, VT, S, P)$

donde $VN = \{S, NOMBRE, VERBO, ADJ, NOMB-SING, VERBO-SING, ADVERBIO\}$

$VT = \{El, La, Los, Las, Pequeño, traviesa, niño, niña, estudia, corre, juega, salta\}$

$P = \{S \rightarrow NOMBRE VERBO$

$NOMBRE \rightarrow ADJ NOMBRE$

$NOMBRE \rightarrow ADJ NOMB-SING$

$VERBO \rightarrow VERB-SING ADVERBIO$

$ADJ \rightarrow El /La /Los /Las /Ellos /Ellas$

ADJ → Pequeño / traviesa / inquieto

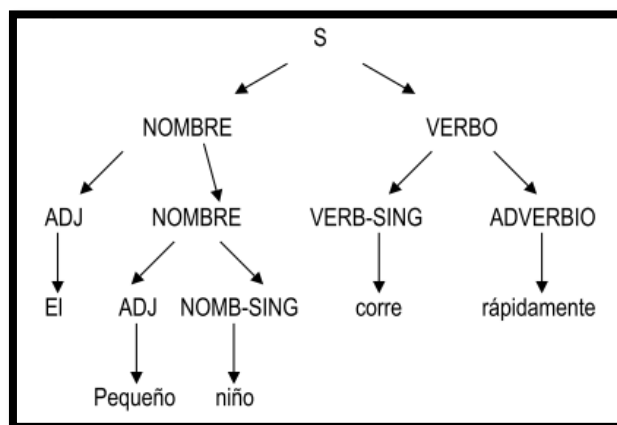
NOMB-SING → niño / niña

VERB-SING → estudia / corre / juega / salta

ADVERBIO → rápidamente / despacio / mucho

}

Figura 3 Ejemplo de Gramática G.



Fuente: (Cortez et al., 2009)

Elaborado por: (Cortez et al., 2009)

luego $w = \text{'El Pequeño niño corre rápidamente'} \in L(G)$

Durante el análisis sintáctico, se realizan derivaciones (de reglas gramaticales) a partir de un símbolo distinguido, para verificar si una frase pertenece al lenguaje definido por la gramática. A este proceso para determinar si es gramaticalmente correcta o no se le conoce como análisis sintáctico (parsing). Los árboles de análisis sintáctico muestran la sintaxis concreta de un lenguaje (Cortez, 2002; Sethi, 1992). Sin embargo, para aplicar esta gramática de forma mecánica y automatizada a una oración, es necesario contar con un lexicón que ofrezca información al analizador sintáctico (parser) sobre las categorías gramaticales que están asociadas a las palabras que

aparecen en la oración que se desea analizar. El análisis semántico es a la vez la fase medular de las instrucciones. Aquí se procesan las estructuras sintácticas reconocidas por el analizador sintáctico. Un analizador semántico puede estar constituido por un conjunto de analizadores semánticos más pequeños. Cada uno de los cuales, maneja un tipo particular de construcción. Estos interactúan entre ellos mismos a través de información que se guarda en estructuras de datos.

2.2 Reconocimiento Automático del Habla (ASR)

Reconocimiento de voz es la tarea de identificar una secuencia de palabras pronunciadas por un hablante, dada la señal acústica. Se ha convertido en una de las principales aplicaciones de la IA. Millones de personas interactúan con los sistemas de reconocimiento de voz todos los días para navegar por los sistemas de correo de voz, la búsqueda Web de teléfonos móviles y otras aplicaciones. El habla es una opción atractiva cuando se necesita el funcionamiento manos libres, como cuando se utiliza maquinaria (Russell & Norvig, 2009).

Figura 4 Reconocimiento del habla.



Fuente: (Russell & Norvig, 2009)

Elaborado por: El Autor.

2.2.1 Factores que determinan la complejidad del reconocimiento del habla.

Cinco son los factores que determinan la complejidad del reconocimiento del habla (Fadiño Rodríguez, 2005):

2.2.1.1 El Locutor

Es quizás el aspecto que introduce mayor variabilidad en la forma de onda entrante, y por tanto requiere que el sistema de reconocimiento sea altamente robusto. Una persona no pronuncia siempre de la misma forma, debido a distintas situaciones físicas y psicológicas (es la llamada variabilidad intralocutor).

Existe además gran variedad entre distintos locutores (hombres, mujeres, niños), diferencias según la edad o la región de origen (variabilidad interlocutor). Es mucho más sencillo que un sistema funcione para un determinado locutor y que este lo haya entrenado previamente (se dice que el sistema es dependiente del locutor), a que un sistema funcione para cualquier locutor (sistema independiente del locutor).

2.2.1.2 La forma de hablar

Es el segundo factor que determina la complejidad de un reconocedor de habla. El hombre pronuncia las palabras de una forma continua, y debido a la inercia de los órganos articulatorios, que no pueden moverse instantáneamente, se producen efectos coarticulatorios. Ello, unido a las variaciones introducidas por la prosodia, hace que una palabra al principio de una frase sea diferente que cuando se dice en medio, o que sea diferente dependiendo de qué es lo que le precede o le sigue.

2.2.1.3 El vocabulario

Se conoce por tal el número de palabras diferentes que debe reconocer el sistema. Mientras mayor es el número de palabras más difícil es el reconocedor, por dos motivos. El primero porque al aumentar el número de palabras es más fácil que aparezcan palabras parecidas entre sí, y el segundo porque el tiempo de tratamiento aumenta al aumentar el número de palabras con las que comparar.

2.2.1.4 La gramática

Es el conjunto de reglas que limita el número de combinaciones permitidas de las palabras del vocabulario. En general la existencia de una gramática en un reconocedor ayuda a mejorar la tasa de reconocimiento, al eliminar ambigüedades y puede ayudar a disminuir la necesidad de cálculo, al limitar el número de palabras en una determinada fase del reconocimiento.

2.2.1.5 El entorno físico

Es una parte tan importante como las anteriores para definir el reconocedor. No es lo mismo un sistema que funciona en un ambiente poco ruidoso, como puede ser el despacho de un médico, o el que tiene que funcionar en un coche o en una fábrica. O, por ejemplo, el que debe de funcionar a través de la línea telefónica, con la consiguiente reducción de banda o el que recibe la voz a través de un micrófono, que tiene mayor ancho de banda que la línea telefónica.

2.2.2 Áreas de trabajo que intervienen en el diseño y especificación de sistemas de reconocimiento del habla.

Las siguientes áreas de trabajo intervienen en el diseño y especificación de sistemas de reconocimiento del habla (Fadiño Rodríguez, 2005):

2.2.2.1 Proceso de la señal de la voz.

La primera operación que debe realizar un reconocedor es procesar la señal de voz de entrada al sistema, con objeto de extraer la información acústica relevante para la tarea que va realizar.

2.2.2.2 Técnicas de reconocimiento de patrones.

El reconocimiento de patrones es la técnica más específica de todo sistema de reconocimiento. De ahí que muchos reconocedores se identifiquen a partir de la técnica de reconocimiento de patrones que incorporan. A partir de la representación paramétrica

de la voz, este módulo realiza un proceso de clasificación utilizando una serie de patrones. Estos patrones se obtienen en una fase de entrenamiento del sistema y son representativos de un conjunto de unidades lingüísticas (palabras, sílabas, sonidos, fonemas). La peculiaridad más característica de este proceso, que marca su dificultad, es la variabilidad temporal que puede presentar una misma unidad lingüística al ser producida por diferentes modos y/o velocidades de habla. Así pues, las primeras técnicas de reconocimiento de patrones utilizadas fueron las basadas en un Alineamiento Temporal a través de algoritmos de Programación Dinámica, técnicas DTW (Rabiner, 1989). Posteriormente se recurrió a la mayor flexibilidad que el modelado de procesos estocásticos permite para representar secuencias de duración variable. Concretamente la alternativa a las técnicas DTW fueron los Modelos Ocultos de Markov (Gorin & Mammone, 1994), (HMM), que pueden verse como una generalización de algoritmos DTW y han demostrado mejores prestaciones en multitud de sistemas de reconocimiento. También hay que mencionar que, recientemente, la potencia y excelentes capacidades de clasificación mostradas por las denominadas Redes Neuronales Artificiales (RNA) las sitúa como posible alternativa frente a los HMM (Soong & Huang, 1990).

2.2.2.3 Modelado dependiente del estilo del habla.

Se distinguen tres modos fundamentales de hablar frente a un sistema de reconocimiento:

Palabras aisladas

Supone que el usuario pronuncia una sola palabra o comando que el sistema deberá reconocer.

Habla conectada

El usuario pronuncia de forma fluida un mensaje utilizando un vocabulario muy restringido; el ejemplo más típico sería la pronunciación de un número telefónico.

Habla continua

Corresponde al modo más avanzado de funcionamiento de un reconocedor, y supone la pronunciación de frases de forma natural para un vocabulario amplio de palabras.

Además de los tres modos fundamentales anteriores, los reconocedores de voz tienen que afrontar, para un modelado robusto del habla, los tres aspectos siguientes (Fadiño Rodríguez, 2005):

1) Reconocimiento en contexto o "word spotting"

Técnica especialmente utilizada en reconocimiento de palabras aisladas, encaminada a detectar la presencia de palabras del vocabulario a reconocer en el contexto de otras palabras o pronunciaciones. La mayoría de las veces el contexto es resultado de la dificultad que encuentra el usuario para ceñirse a la pronunciación de una única palabra aislada. En otras ocasiones, el reconocimiento en contexto es la solución apropiada para robustecer el reconocimiento en ambientes acústicamente hostiles; por ejemplo, cuando la palabra que pronuncia el usuario viene acompañada de ruidos telefónicos, urbanos, etc. En cualquier caso, se trata de una técnica importante para robustecer los sistemas en aplicaciones reales.

2) Rechazo

Otro efecto de la presencia de sonidos indeseados (ruidos, sonidos o palabras fuera del vocabulario), es provocar el reconocimiento de palabras que realmente no han sido pronunciadas. Los procedimientos conocidos como técnicas de rechazo tienen como objetivo permitir incluir entre los resultados de reconocimiento la identificación de esos sonidos indeseados.

3) Múltiples candidatos

El proceso de reconocimiento de patrones que realiza un reconocedor se basa en identificar el patrón que ofrezca la puntuación más alta para decidir cuál es la mejor palabra o secuencia de palabras reconocida. Este proceso se basa en información exclusivamente acústica, sin tener en consideración otras posibles fuentes de conocimiento que podrían utilizarse para completar las puntuaciones de las diferentes palabras o secuencias candidatas. En la mayoría de los casos, la aplicación en que se encuentra el reconocedor es la que posee la información necesaria que permitiría seleccionar entre varias hipótesis de reconocimiento.

2.2.2.4 Dependencia del locutor

El grado de dependencia del locutor define si el sistema incorpora patrones de unidades lingüísticas adaptados a un locutor determinado, y, por tanto, sólo funcionará correctamente para él, o si los patrones pretenden ser válidos para cualquier hablante. En el primer caso se habla de reconocimiento dependiente del locutor, mientras que en el segundo de reconocimiento independiente del locutor.

2.2.2.5 Dependencia del vocabulario

Las prestaciones de un reconocedor dependen fuertemente del tamaño y grado de dificultad del vocabulario. Es decir, del número de palabras que el sistema es capaz de reconocer, y de la mayor o menor dificultad de su reconocimiento en base a las relaciones de similitud fonética entre palabras. En la actualidad se diseñan sistemas tanto para vocabularios pequeños (menos de 50 palabras) y medios (entre 50 y 500 palabras), como para grandes vocabularios (más de 500 palabras), llegándose hasta 50.000 palabras para aplicaciones de dictado o acceso a bases de datos mediante lenguaje natural.

2.2.2.6 Gramáticas de reconocimiento

Según aumenta el número de palabras del vocabulario, el número de posibles combinaciones crece exponencialmente. Por tanto, se hace imprescindible la incorporación de restricciones, en cuanto al número de combinaciones válidas, según la tarea en que se inserte el sistema. Restricciones que suelen incorporarse en forma de gramáticas basadas en reglas sintácticas y/o semánticas destinadas a reducir el número de palabras susceptibles de ser reconocidas en cada momento.

2.2.3 Técnicas de diseño de reconocedores del habla

A continuación, se van a estudiar cuatro técnicas distintas que se utilizan o se han utilizado para el diseño de reconocedores de habla.

Las técnicas son:

Técnicas topológicas: Dynamic Time Warping (DTW), basado en el cálculo y comparación de distancias.

Técnicas probabilísticas: Modelos ocultos de Markov (HMM), que son modelos generativos de las palabras del vocabulario.

Redes neuronales artificiales. A través de la interconexión de unidades de procesamiento en paralelo.

Sistemas basados en el conocimiento: reconocedores por reglas o sistemas expertos.

En los cuatro casos se puede hablar de una fase de "entrenamiento" (cálculo de los patrones de referencia, cálculo de los parámetros de los modelos de Markov, entrenamiento de las redes neuronales o creación de estructuras de datos para los sistemas expertos) y de otra fase de "reconocimiento" propiamente dicho. Y también en los cuatro casos el primer proceso necesario es la "parametrización" o transformación de

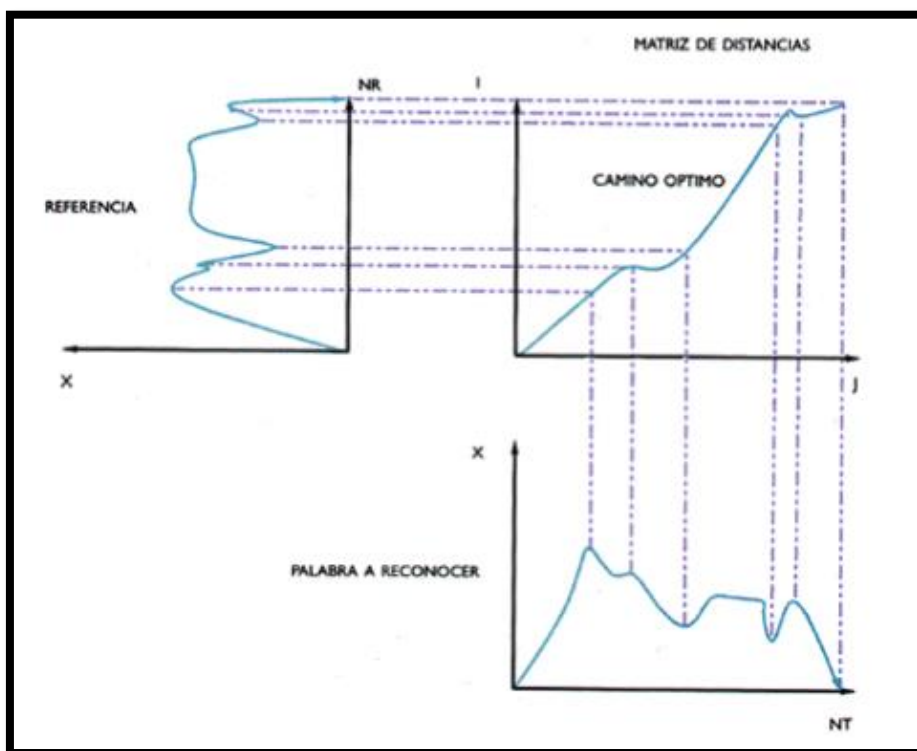
la forma de onda de la señal entrante en un conjunto de parámetros o características adecuadas a cada reconocedor (Fadiño Rodríguez, 2005).

2.2.3.1 Dynamic Time Warping

Los reconocedores de habla basados en técnicas de Dynamic Time Warping (DTW) han sido los primeros que han alcanzado un nivel de fiabilidad suficientemente alto como para dar lugar al desarrollo de productos comerciales.

Los sistemas de reconocimiento basados en DTW funcionan de la siguiente manera: Primero se parametriza la señal de voz a reconocer; para ello se divide en pequeñas ventanas de análisis (unos 20 mseg), y sobre cada una de esas ventanas se realiza un proceso de análisis que extrae un conjunto de parámetros (que pueden ser acústicos o coeficientes espectrales). Ese conjunto o vector de parámetros se puede ver como un punto en un espacio n -dimensional. El conjunto de todas las ventanas de análisis se convertirá así en una secuencia de puntos en ese espacio, y esa secuencia de puntos es lo que se llama "patrón" o "plantilla".

Figura 5 Alineamiento temporal entre la Referencia y la palabra a reconocer.



Fuente: (Fandiño Rodríguez, 2005)
Elaborado por: Alexander Fandiño

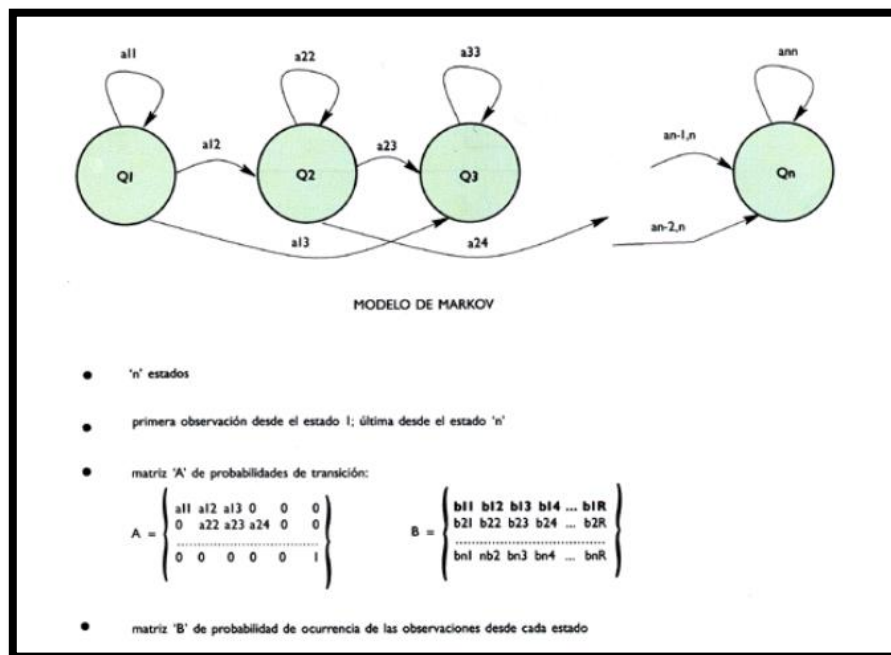
2.2.3.2 Modelos ocultos de Markov

Otro enfoque alternativo al de medir distancias entre patrones enfoque topográfico es el de adoptar un modelo estadístico (paramétrico) para cada una de las palabras del vocabulario de reconocimiento, como son los modelos ocultos de Markov (HMM, del inglés 'Hidden Markov Models') (Huang, Alleva, Hwang, & Rosenfeld, 1993).

Un HMM se puede ver como una máquina de estados finitos en que el siguiente estado depende únicamente del estado actual, y asociado a cada transición entre estados se produce un vector de observaciones o parámetros. Se puede así decir que un modelo de Markov lleva asociados dos procesos: uno oculto (no observable directamente) correspondiente a las transiciones entre estados, y otro observable (y directamente relacionado con el primero), cuyas realizaciones son los vectores de parámetros que se producen desde cada estado y que forman la plantilla a reconocer.

Para aplicar la teoría de los HMM en reconocimiento de voz, se representa cada palabra del vocabulario del reconocedor con un modelo generativo (que se calculara en la fase de entrenamiento) y posteriormente, se calcula la probabilidad de que la palabra a reconocer haya sido producida por cada uno de los modelos de la base de datos del reconocedor. Para ello, se asume que durante la pronunciación de una palabra, el aparato fonador puede adoptar sólo un número (finito de configuraciones articulatorias (o estados), y que desde cada uno de esos estados se producen uno o varios vectores de observación (puntos de la plantilla), cuyas características espectrales dependerán (probabilísticamente) del estado en el que se hayan generado Así vista la generación de la palabra, las características espectrales de cada fragmento de señal dependen del estado activo en cada instante, y la evolución del espectro de la señal durante la pronunciación de una palabra depende de la ley de transición entre estados. La representación más usual de un HMM es la utilizada para máquinas de estados finitos, es decir, conjuntos de nodos (que representan a los estados) y arcos (transiciones permitidas entre los estados). Un tipo de HMMs especialmente apropiado para reconocimiento de voz son los modelos "de izquierda a derecha"; modelos en los que una vez que se ha abandonado un estado, ya no se puede volver a él. La figura 6 representa un modelo con 'n' estados en el que desde cada estado sólo se permiten tres tipos de transición: al propio estado, al estado vecino y a dos estados más allá (este tipo de saltos queda recogido en una matriz de transiciones tridiagonal).

Figura 6 Modelo de Markov con n estados.



Fuente: (Fadiño Rodríguez, 2005)
Elaborado por: Alexander Fadiño

En cuanto a la generación de puntos de la plantilla, en estos modelos se asume que el primer vector de observaciones se produce desde el primer estado, y el último se emite desde el último estado. Recuérdese que la secuencia de estados es la parte oculta del modelo: se conocen los vectores de parámetros, pero no desde que estado se han producido.

2.2.3.2.1 Definición formal de un HMM

Un modelo M viene determinado por los siguientes parámetros (Fadiño Rodríguez, 2005):

- N- Número de estados del modelo.
- Matriz de transiciones, de dimensión (N x N). Define la estructura del modelo: cada uno de sus elementos, A_{ij} , define la probabilidad de pasar del estado i al estado j. Normalmente A será bidiagonal o

tridiagonal, significando que desde cada estado se pueden producir dos o tres tipos distintos de transición.

- c) Conjunto de parámetros B que definen, para cada estado funciones de densidad de probabilidad (fdp) que modelan estadísticamente las observaciones producidas desde cada estado. Habrá pues tantas fdps como estados.
- d) P - Vector de dimensión N . Cada uno de sus elementos, P_i indica la probabilidad de encontrarse inicialmente en el estado i . Para modelos de izquierda a derecha, $P_1 = 1$, y $P_j = 0$ para los demás estados.

Como en el caso de DTW, la señal de voz viene representada por una plantilla o secuencia de vectores de características $O = [O_1, O_2, \dots, O_T]$, donde cada O_j es un conjunto de parámetros (coeficientes LPC, Cepstrum, log área ratios...) que caracteriza la señal de voz en una ventana de tiempo centrada en $t = i$, y T es el número total de puntos de la plantilla. Los modelos HMM basados en este tipo de observaciones se llaman HMM continuos (Kirschning, Aguas, & Ahuactzin, 2000), y B será un conjunto de fdps continuas.

2.2.3.3 Redes Neuronales Artificiales

Los modelos computacionales basados en redes neuronales surgieron hace ya relativamente bastante tiempo, pero se abandonó su estudio por no disponer de algoritmos eficientes de entrenamiento. Ahora ya no existe esa dificultad, y se ha demostrado ampliamente su enorme potencia computacional.

Los sistemas de reconocimiento basados en redes neuronales pretenden, interconectando un conjunto de unidades de proceso (o neuronas) en paralelo (de forma similar que en la mente humana), obtener prestaciones de reconocimiento similares a las

humanas, tanto en tiempo de respuesta como en tasa de error. Esa forma de interconexión de las unidades de proceso es especialmente útil en aplicaciones que requieren una gran potencia de cálculo para evaluar varias hipótesis en paralelo, como sucede en los problemas de reconocimiento de voz.

Las unidades de proceso pueden ser de varios tipos; las más simples (y utilizadas) disponen de varias entradas, y la salida es el resultado de aplicar alguna transformación no lineal a la combinación lineal de todas las entradas. Otro tipo de neuronas un poco más elaborado se caracteriza por disponer de memoria; en ellas la salida en cada momento depende de entradas anteriores en el tiempo.

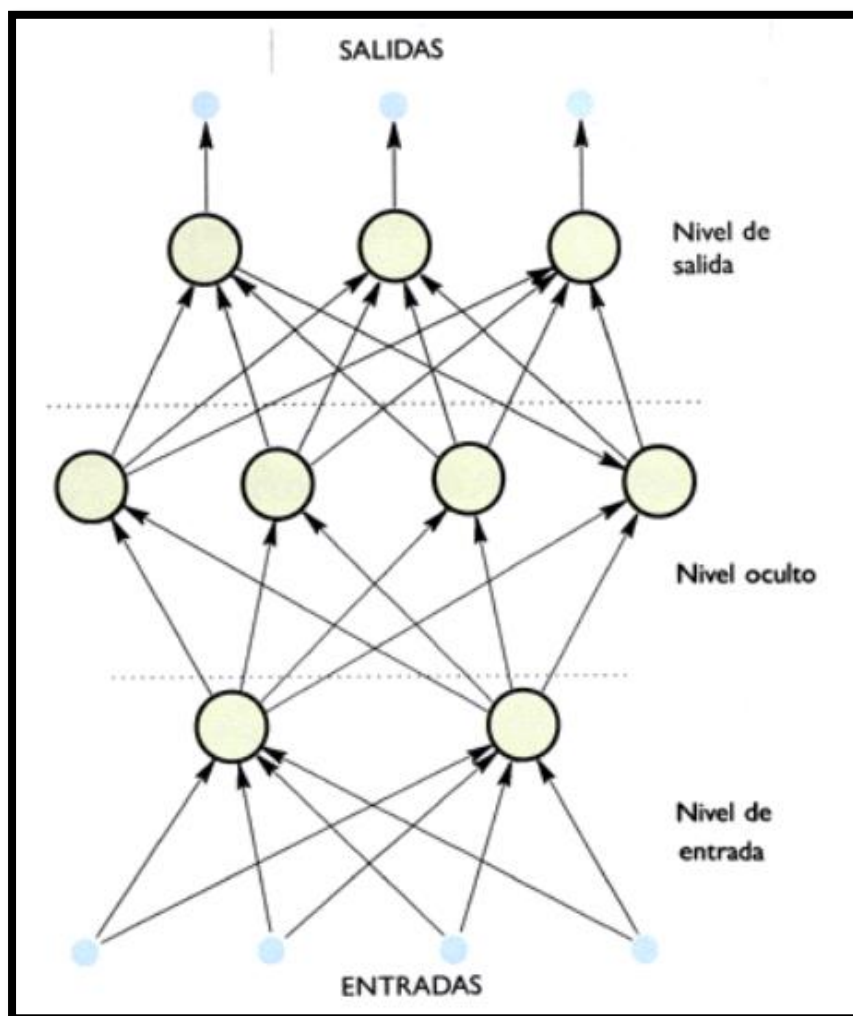
La forma en que las neuronas se conectan entre si define la topología de la red, y se puede decir que el tipo de problemas que una red neuronal particular soluciona de forma eficiente, depende de la topología de la red, del tipo de neuronas que la forman, y la forma concreta en que se entrena la red.

Igual que se dijo para las técnicas anteriores, una red neuronal debe ser entrenada para resolver un tipo determinado de problemas. El algoritmo particular de entrenamiento dependerá de la estructura interna de las neuronas (Ney, Haeb-Umbach, Tran, & Oerder, 1992), pero, en cualquier caso, el entrenamiento se llevará a cabo a partir de una base de datos etiquetada, como sucedía con los modelos de Markov, y será un proceso iterativo en el que se modifican los parámetros de la red para que ante un conjunto determinado de estímulos (plantillas), produzca una respuesta determinada: la palabra del vocabulario representada por esas plantillas.

La red neural que mejores resultados está dando hasta este momento en reconocimiento automático del habla es la denominada "perceptrón multicapa". La figura 7 muestra su topología: las neuronas se disponen por "capas"; hay una capa de entrada, que opera directamente sobre los vectores de observación o puntos de las

plantillas, una capa de salida que apunta la palabra reconocida, y una o más capas intermedias. Cada capa está compuesta por varias unidades de proceso, que se conectan con la siguiente capa por una serie de enlaces a los que se da un cierto peso específico W_{ij} .

Figura 7 Red Neuronal de 3 Capas.



Fuente: (Fandiño Rodríguez, 2005)
Elaborado por: Alexander Fandiño

El conjunto de vectores de características entra en la capa de neuronas de entrada, y posteriormente es propagado a las capas siguientes. En cada célula de proceso se calcula la suma ponderada (por los pesos W_{ij}) de las señales de entrada, y posteriormente se

procesa en la neurona con su sistema no lineal. Si el resultado de esta operación supera un cierto umbral, la neurona reacciona, transmitiendo señal a las neuronas siguientes de la capa superior.

En la fase de entrenamiento, dada una entrada conocida (ej. un conjunto de vectores que representa el dígito 1), la salida de la red es comparada con la salida esperada (y conocida de antemano), calculándose un error. Ese error se propaga hacia abajo, ajustándose de esta manera los pesos de las conexiones entre neuronas. Efectuándose este proceso varias veces se consigue que la red "aprenda" que respuesta debe dar para cada entrada en la fase de reconocimiento.

2.2.3.4 Reconocimiento basado en el conocimiento

Los métodos de reconocimiento descritos hasta ahora funcionan bastante bien cuando se trata de reconocer palabras aisladas. Cuando el sistema debe reconocer frases o habla continua, es necesario acudir a otras fuentes de conocimiento además de las puramente matemáticas y acústicas. Estas son por lo general reglas de tipo lingüístico, como se va a ver a continuación. Con este tipo de sistemas se llegará a tener no solo un reconocedor de habla sino un sistema de "comprensión" de habla.

La razón por la que a estos sistemas avanzados de reconocimiento se les llama Sistemas basados en el Conocimiento, se debe al uso de otras fuentes, otras disciplinas, otros conocimientos para llegar al entendimiento de la frase. En definitiva, lo que se trata es que una máquina llegue a tener y utilizar los conocimientos que tiene una persona humana, para entender un mensaje.

2.2.4 Etapas de un Sistema de Reconocimiento Automático del Habla (ASR)

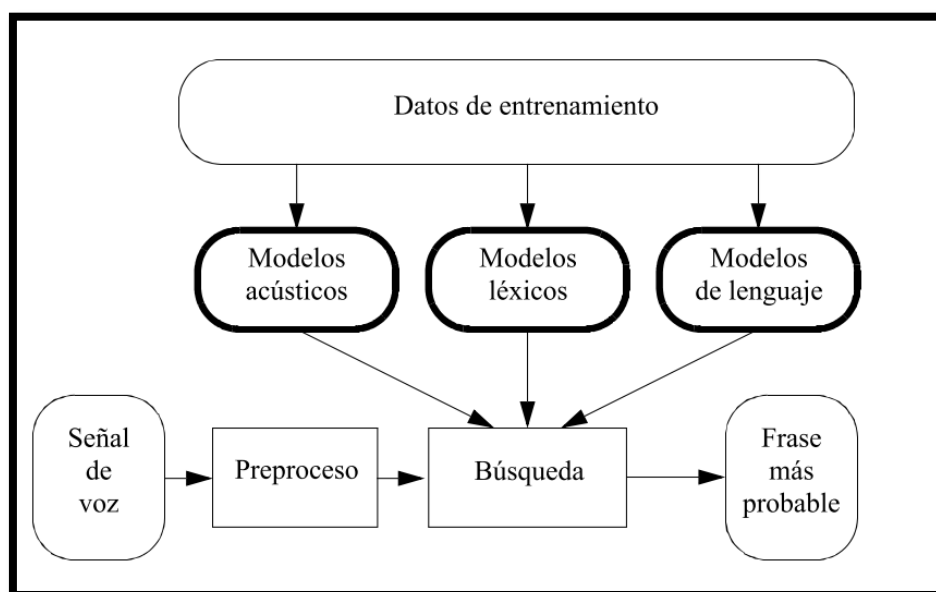
La estructura general de los sistemas de ASR tiene esencialmente tres módulos o etapas (Rufiner & Milone, 2004):

1. Procesamiento o análisis del habla (en inglés se conoce como front- end):

en esta etapa se realiza algún tipo de análisis de la señal de voz en términos de la evolución temporal de parámetros espectrales (previa conversión analógica/digital de la señal). Esto tiene por función hacer más evidentes las características necesarias para la etapa siguiente y a veces también limpiar y reducir la dimensión de los patrones para facilitar su clasificación.

2. Clasificación de unidades fonéticas o modelo acústico: esta etapa clasifica o identifica los segmentos de voz ya procesados con símbolos fonéticos (fonemas, dífonos o sílabas). A veces se puede asociar una probabilidad con este símbolo fonético, lo que permite ampliar la información presentada al siguiente módulo.

3. Análisis en función de reglas del lenguaje o modelo del lenguaje: en esta última etapa se pueden aprovechar las reglas utilizadas en la codificación del mensaje contenido en la señal para mejorar el desempeño del sistema y producir una transcripción adecuada. Aquí se utilizan otras fuentes de conocimiento como la ortográfica, la sintáctica, la prosódica, la semántica o la pragmática.

Figura 8 Esquema genérico de un sistema ASR.

Fuente: (Macías Guarasa, 2001)

Elaborado por: Javier Macías

2.2.5 Comparativa de las técnicas de procesamiento del lenguaje natural.

Luego de la extensa revisión bibliográfica se realiza la comparativa de las diferentes técnicas de procesamiento del lenguaje natural, esta información se encuentra resumida en la tabla 6.

Tabla 6 Comparativa de las técnicas de procesamiento del lenguaje natural

	DTW	RBC(Conocimiento)	HMM(Modelos)	RNA(Redes)
Dependiente	Palabras	Palabras Aisladas	Palabras Aisladas	Palabras
del Estilo	Aisladas	Habla Conectada	Habla Conectada	Aisladas
del Habla		Habla continua	Habla continua	Habla
				Conectada
				Habla continua
Dependiente	Dependiente	Dependiente/	Dependiente/	Dependiente/
del Locutor		Independiente	Independiente	Independiente

Dependiente del Entorno	Dependiente	Dependiente/Independiente	Dependiente/Independiente	Dependiente/Independiente
Dependiente del Vocabulario	Dependiente / Vocabularios Pequeños	Dependiente / Vocabularios Grandes	Dependiente / Vocabularios Grandes	Dependiente / Vocabularios Grandes
Gramática	No	Si	Si	Si

Elaborado por: El Autor.

La técnica de DTW posee un estilo de habla ideal para el reconocimiento de palabras aisladas, mientras que las técnicas de reconocimiento basado en el conocimiento, modelos ocultos de Markov y Redes Neuronales Artificiales, además de este estilo poseen los estilos de habla conectada y habla continua.

En la técnica DTW el reconocimiento es dependiente del locutor, a diferencia de las demás técnicas en las cuales el reconocimiento puede ser dependiente o independiente.

En la técnica DTW el reconocimiento es dependiente del entorno mientras que en las demás técnicas puede ser dependiente o independiente.

Todas las técnicas aquí comparadas son dependientes del vocabulario, con la diferencia que la técnica DTW se usa con vocabularios pequeños mientras que con las demás los vocabularios pueden ser pequeños o grandes.

En la técnica DTW al ser de vocabulario pequeño no es necesario el uso de una gramática mientras que en las demás por ser de vocabularios grandes, se puede hacer uso de una gramática para mejorar el reconocimiento.

2.2.6 Proyectos de Procesamiento del Lenguaje Natural en el Ecuador.

Loja, Loja

“Procesamiento del Lenguaje Natural Aplicado a la Interacción con un Call Center utilizando una Gramática Predefinida y Basado en Algoritmos de Búsqueda Inteligentes.”

Según la autora (Coronel, 2006) en este proyecto se desarrolló una aplicación para la interacción con un call center, utilizando algoritmos de búsqueda inteligentes en una red neuronal con el uso de palabras restringidas en una gramática predefinida.

Riobamba, Chimborazo

“Estudio en Implementación de Domótica Activado por Comandos de Voz y Comunicación en Zigbee”

Según los autores (Falconí & Jiménez, 2009) en este proyecto se implementó un sistema que permite controlar la mayoría de aspectos relacionados con la automatización del hogar, en especial de una habitación, a través de redes de comunicación inalámbrica y cuyo control se realice en forma de comandos de voz.

Quito, Pichincha

“Diseño y Construcción de un Sistema para el Control de Encendido y Apagado de Dispositivos Eléctricos por Medio de Comandos de Voz”

Según el autor (Romero Fonseca, 2010) este proyecto es el resultado de la investigación y el afán de suministrar a las personas con discapacidad motriz una herramienta que facilite y mejore su estilo de vida, de tal forma que aquellas personas a las cuales les es difícil o en extremo imposible ejecutar acciones típicas dentro de la vida cotidiana como el encender y apagar luminarias, ventilación, abrir puertas, cortinas y manejar las

opciones básicas de control de una televisión lo puedan hacer a través de comandos de voz.

Quito, Pichincha

“Diseño e Implementación de un sistema móvil anti-robo comandado por voz a un sistema de seguridad electrónica para vehículo.”

Según los autores (Báez & Cabrera, 2010) en este proyecto se implementó un sistema biométrico de reconocimiento de voz a una alarma de un vehículo para así mejorar la seguridad del mismo puesto que posee un patrón único de voz a reconocer, por lo que aporta mayor seguridad que una alarma convencional.

Guayaquil, Guayas

“Sistema que simule conocimiento humano con Voz SEUZ”

Según el autor (Alarcon, 2011) el proyecto pretende ayudar a la forma de estudio y tratar de implementar un modelo distinto de aprendizaje, el cual es escuchar información en lugar de estar leyendo, porque este método no funciona para las personas que no tienen mucho tiempo, ya sea por su trabajo o tareas diarias, en cambio, por medio de este modelo de aprendizaje, el cual es escuchar, se puede utilizar así sea cuando se está realizando cualquier labor, como por ejemplo manejar un vehículo, este modelo puede ser aplicado con cualquier tipo de información. También implementa dos nuevas tecnologías no explotadas hasta el momento que son el reconocimiento de voz y la síntesis de voz.

Quito, Pichincha

“DISEÑO E IMPLEMENTACIÓN DE UN DISPOSITIVO QUE PERMITA EL ENCENDIDO DE LOS AUTOMÓVILES DE SISTEMA CONVENCIONAL,

UTILIZANDO UNA PLACA ELECTRÓNICA QUE CONTROLA EL SISTEMA DE ENCENDIDO COMANDADO POR VOZ PROGRAMADA.”

En este proyecto según el autor (Pusay, 2014) se diseñó e implementó un dispositivo capaz de controlar el encendido y arranque del vehículo mediante comandos de voz, con la finalidad de optimizar el tiempo y comodidad a los usuarios al manipular las llaves del switch, se emplearon técnicas de investigación, métodos lógicos, y sistemáticos para la construcción de la placa electrónica del dispositivo, mediante la determinación de componentes y la relación que existe entre usuario y módulo de reconocimiento de voz VRbot, LCD para la visualización de los mensajes, también los diferentes tipos de encendidos y los diferentes elementos que conforman el circuito.

Quito, Pichincha

“ANÁLISIS, DISEÑO, CONSTRUCCIÓN Y EVALUACIÓN DE UN SISTEMA EXPERTO TRADUCTOR DE LOS IDIOMAS: CASTELLANO A KICHWA ECUATORIANO BASADO EN REGLAS DE INFERENCIA Y RAZONAMIENTO APROXIMADO.”

Según los autores (Gómez & Freire, 2014) el producto realizado es un software de traducción de los idiomas Castellano – Kichwa ecuatoriano, dicho software cuenta con una interfaz simple la cual facilita la comunicación entre éste y el usuario, el mismo que simplemente ingresa texto en el idioma Castellano, pulsa el botón traducir, y obtiene la respuesta de traducción al idioma Kichwa ecuatoriano más acertada y entendible posible.

Latacunga, Cotopaxi

“Aplicación Móvil Intérprete entre Personas con Discapacidad Visual y Auditiva a partir de Técnicas de Procesamiento de Lenguaje Natural.”

Según los autores (Carrillo & Criollo, 2016) el proyecto consiste en el desarrollo de una aplicación móvil intérprete en el sistema operativo Android, para facilitar la comunicación entre una persona con discapacidad auditiva y una persona con discapacidad visual de la Unidad Educativa Especializada de Cotopaxi de la ciudad de Latacunga, mediante técnicas de procesamiento de lenguaje natural que son interacciones entre las máquinas y el lenguaje humano sea escrito o hablado.

Una vez analizados estos proyectos de investigación se concluye que son proyectos donde se desarrollan aplicaciones que implementan las diferentes técnicas de procesamiento del lenguaje natural como Comparación de Plantillas, Redes Neuronales y Reconocimiento Basado en el Conocimiento, pero la mayoría hacen uso de comparación de plantillas porque reconocen palabras aisladas para comandos simples, y algunos como el traductor de Español a Kichwa que hace uso de una Red Bayesiana y un Sistema Experto para la traducción.

Capítulo 3

3 Materiales y Métodos

3.1 Localización de la zona de estudio

3.1.1 Ubicación Geográfica

La presente investigación se realizó en el cantón Zapotillo, provincia de Loja; en la Escuela de Educación General Básica José Antonio Campos

3.2 Materiales

3.2.1 Materiales y Equipos de Campo

Entre los materiales que se utilizaron para documentar y realizar el trabajo de tesis constan: lectura para los niños, grabadora portátil y libreta de campo.

3.2.2 Materiales de Oficina

Computadora, grabadora portátil, impresora, Memoria USB, calculadora, papel A4 de 75 g y bolígrafos.

3.3 Selección de técnicas de procesamiento del lenguaje natural

Luego de la extensa revisión bibliográfica realizada previamente, y en base a la comparativa de las técnicas de procesamiento del lenguaje natural resumidas en la Tabla 6 de este documento, y en base a que para la evaluación de las técnicas, el prototipo a desarrollar debe cumplir las siguientes características: debe reconocer habla continua, ser independiente del locutor y con un vocabulario mediano de palabras; las técnicas que cumplen estos requisitos son: Reconocimiento Basado en el Conocimiento, Reconocimiento Basado en Modelos Ocultos de Markov y Reconocimiento Basado en Redes Neuronales Artificiales, por este motivo, éstas son las técnicas a ser evaluadas.

3.4 Selección del Lenguaje de Programación

Una vez seleccionadas las técnicas a ser evaluadas, hay que determinar el lenguaje de programación con el cual desarrollar el prototipo de aplicación para la evaluación de las técnicas. Para esto se tomó en cuenta el Ranking de IEEE Spectrum de Lenguajes de Programación para el año 2016 reflejado en la tabla 7 (Diakopoulos & Cass, 2016).

Tabla 7 Ranking IEEE Spectrum de lenguajes de programación.

Lenguaje	IEEE Spectrum
C	1
Java	2
Python	3
C++	4
R	5

Fuente: IEEE Spectrum
Elaborado por: El Autor

De este ranking se seleccionó los 3 primeros lenguajes, se realizó la tabla comparativa de los mismos, y, de esta manera poder seleccionar el que mejor se adapte a nuestro proyecto. La información está reflejada en la tabla 8.

De esta tabla comparativa se observa que los lenguajes Java y Python, son más portables que C; Java y Python son orientados a objetos mientras que C es estructurado; Java incluye librerías para interface gráfica de usuario mientras que C y Python deben hacer uso de librerías externas; Java y Python poseen gran variedad de librerías de reconocimiento de voz a diferencia de C en cuyo lenguaje están limitadas. Se aprecia una ventaja mínima de Java frente a los otros dos lenguajes para este tipo de proyecto y

para la selección del lenguaje, a más de lo ya expuesto se considera que el desarrollador en este caso mi persona posee conocimiento intermedio de programación en el lenguaje Java, que existen las librerías para el desarrollo de un prototipo multiplataforma, portable que integre las técnicas antes seleccionadas, en base a esto se selecciona el lenguaje Java como lenguaje de desarrollo para el prototipo de aplicación de este proyecto.

Tabla 8 Comparativa de los lenguajes de programación

	C	Java	Python
Sistema Operativo	Multiplataforma	Multiplataforma	Multiplataforma
Portabilidad	Si compilado por plataforma	Si compilado una sola vez	Si Interpretado
Paradigma	Imperativo Estructurado	/ Imperativo / Orientado a Objetos	Multiparadigma
Tipado	Débil / Estático	Fuerte / Estático	Fuerte / Dinámico
Interface gráfica	Si / Externas	Si / Incluida	Si / Externas
Librerías de reconocimiento	HTK	CMU Sphinx4 Google Speech Recognition Google Cloud Speech API Java Speech API	CMU Sphinx pocketsphinx Google Speech Recognition Google Cloud Speech API Wit.ai

Elaborado por: El Autor

Para reducir los costos se seleccionó una base de datos PostgreSQL y pgadmin como gestor de la misma.

3.5 Cumplimiento de los objetivos

En el proyecto de investigación para dar cumplimiento al primer objetivo específico “Medir el nivel de pronunciamiento de la población seleccionada periódicamente.”, se procedió a tomar las muestras de audio a cada uno de los estudiantes, con el fin de entrevistar y observar el desenvolvimiento de cada uno de ellos, de la misma manera se construyó una matriz de registro para ir guardando manualmente los detalles de las interacciones de cada estudiante y poder llevar un registro ordenado de los mismos.

Para el cumplimiento del segundo objetivo, “Determinar las técnicas de procesamiento del lenguaje natural que van a ser evaluadas durante el presente proyecto”, se realizó revisión extensa de literatura sobre las técnicas de reconocimiento para en base a ello elegir las tres más convenientes a aplicar en la investigación. Para determinar las técnicas se empleó el método analítico y sintético. Obteniendo como resultado lo siguiente:

- **Técnica 1:** Reconocimiento Basado en el Conocimiento.
- **Técnica 2:** Reconocimiento Basado en Modelos Ocultos de Markov.
- **Técnica 3:** Reconocimiento Basado en Redes Neuronales Artificiales.

Para el cumplimiento del tercer objetivo “Comparar como las diferentes técnicas influyen en el reconocimiento de la pronunciación de los niños”, se desarrolló un prototipo de aplicación que utiliza cada una de las técnicas seleccionadas, se introdujo los archivos o grabaciones de audio al programa para ser procesados y la información

obtenida fue almacenada en una base de datos, para a través del mismo prototipo de aplicación obtener los resultados. Mediante un constante monitoreo, se verificó el ingreso de todos los audios en su respectivo orden.

Para el cumplimiento del cuarto objetivo “Desarrollar una aplicación para la captura de datos y presentación de resultados, utilizando las diferentes técnicas del procesamiento del lenguaje natural seleccionadas.”, se investigó los Frameworks que utilizan cada una de estas técnicas, se seleccionó el motor de base de datos más idóneo y en base a esto se procedió a desarrollar la aplicación. Además, se elaboró un diccionario fonético y una gramática para ser utilizada en la comparativa.

La información consolidada en los cuatro objetivos permite estructurar el objetivo general de la investigación “Evaluación de técnicas del procesamiento del lenguaje natural aplicadas en la pronunciación de niños de educación general básica de la escuela José Antonio Campos del cantón Zapotillo, provincia de Loja.”

3.6 Tamaño de la población y muestra

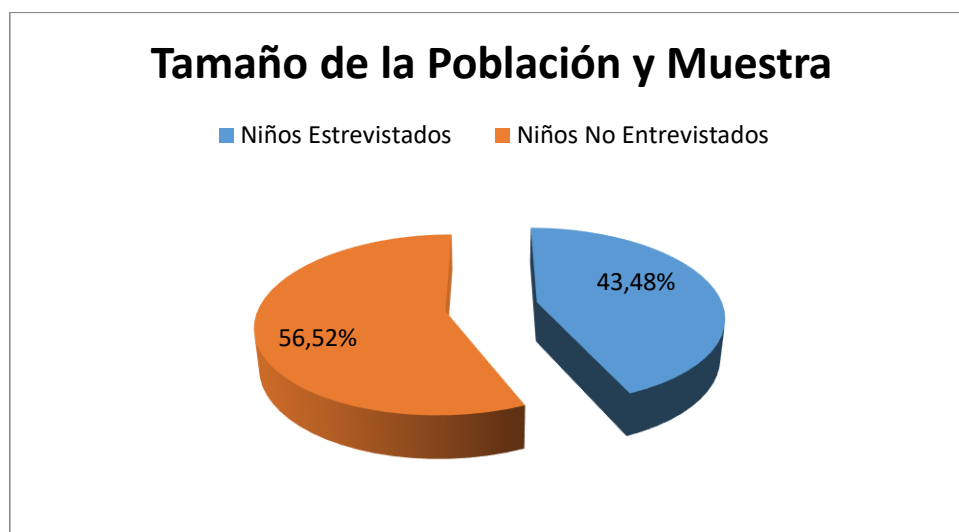
Para determinar la población objetivo, así como el tamaño óptimo de la muestra se requirió conocer con precisión el universo a investigarse, de donde se consideró mediante un muestreo aleatorio simple como muestra a 10 niños de los 23 niños que integran el séptimo año de educación general básica. En total se recolectaron 100 muestras de audio durante 10 semanas. En los cuadros siguientes se muestra la recopilación de datos y su análisis para alcanzar el objetivo general y los objetivos específicos de la investigación.

Tabla 9 Tamaño de la población y muestra

Tamaño de la Población y Muestra		
	Frecuencia	Porcentaje
Niños Entrevistados	10	43,48%
Niños No Entrevistados	13	56,52%
Total Niños	23	100%

Fuente: Investigación de campo

Elaboración: El Autor

Gráfico 1 Tamaño de la población y muestra.

Fuente: Investigación de campo

Elaboración: El Autor

En la Tabla No. 9 se recogen el número de niños con los que se va a desarrollar la propuesta, que tras un proceso de depuración entraron a formar parte de la población objetivo a entrevistar. Una vez calculado el número de niños a encuestar, se observa en el Gráfico 1 que la muestra representa el 43,48% de la población.

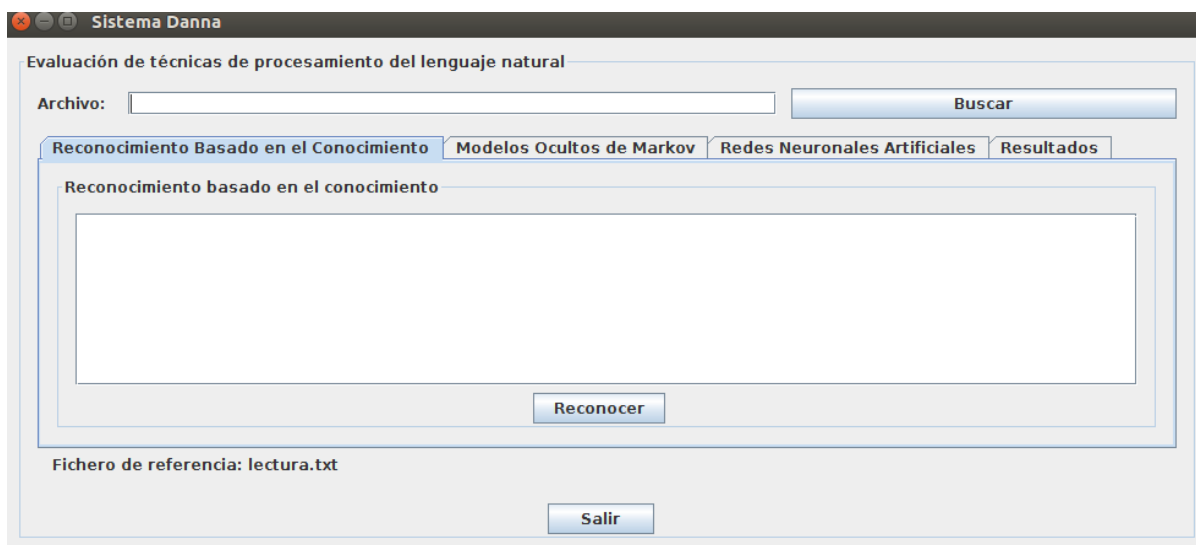
3.7 Técnicas e Instrumentos de Investigación

- **Los Experimentos:** según Real (2016) constituye el método de investigación con mayores garantías de control y es el único que permitirá establecer relaciones de causa-efecto entre variables. Para la evaluación de las técnicas de

procesamiento de lenguaje natural seleccionadas, se experimentará con el prototipo de aplicación y las muestras de audio.

- **Evaluación mediante escala:** Para medir el nivel de pronunciamiento de la población seleccionada periódicamente se realizó una valoración mediante escala, la cual fue elaborada tomando en cuenta el número de palabras reconocidas utilizando la mejor de las tres técnicas de reconocimiento.
- **Revisión Bibliográfica:** a través de la revisión de artículos científicos, libros, publicaciones, entre otros, se determinó las técnicas de procesamiento del lenguaje natural que van a ser evaluadas durante el proyecto.
- **Análisis Comparativo:** se realizará la comparación de los resultados obtenidos para cada una de las variables dependientes, logrando determinar cómo las diferentes técnicas influyen en el reconocimiento de la pronunciación de los niños.
- **Danna:** es el prototipo de aplicación desarrollada que permitirá la captura de datos y presentación de resultados, utilizando las diferentes técnicas del procesamiento del lenguaje natural seleccionadas.

Figura 9 Sistema DANNA.



Elaborado por: El Autor.

Danna utiliza diferentes librerías para integrar las diferentes técnicas de procesamiento del lenguaje natural a ser evaluadas, la librería Sphinx4 junto con el modelo acústico CMUSphinx-es, el cual posee un diccionario fonético con un corpus de más de 20000 palabras (Anexo B) y un modelo de lenguaje probabilístico altamente entrenado (Anexo B), permitirá el reconocimiento basado en Modelos Ocultos de Markov. Ésta librería Sphinx4 junto con el modelo acústico CMUSphinx-es, con un diccionario fonético propio (Anexo C) más una gramática generada especialmente para esta investigación (Anexo C), permitirá el reconocimiento basado en el conocimiento. El Kit de desarrollo de software Google Cloud SDK junto a las librerías Google Speech API, Google Cloud Storage y Google Client Library for JAVA permitirán el reconocimiento basado en redes neuronales artificiales. Éstas librerías permitirán realizar el reconocimiento de palabras a través de las diferentes técnicas seleccionadas, esta información resultante del reconocimiento será almacenada en una Base de Datos PostgreSQL, para comunicar el prototipo de aplicación con la base de datos se utilizó la librería JDBC for PostgreSQL. Para obtener los resultados de reconocimiento y poder realizar la evaluación fue necesaria la librería JFreeChart, la cual permite visualizar la información mediante Gráficos lo que resulta conveniente al momento de comparar las diferentes técnicas. El Hardware utilizado fue: Computador Portátil con Procesador Intel i7-3630QM, 8GB de memoria RAM, 1TB de disco duro, SO: Ubuntu 16.04 Xenial Xerus, y una conexión a Internet de 10Mbps con compartición 2:1.

Capítulo 4

4 Recolección y Análisis de Datos.

En este capítulo se da a conocer el proceso de recolección y análisis de datos de la investigación.

4.1 Recolección de datos.

La recolección de datos se constituye como una etapa muy importante dentro de cualquier investigación, los datos permitirán obtener información sobre el problema que nos interesa y ayudarán a contestar las interrogantes que se había planteado responder con la investigación.

4.1.1 Trabajo de Campo.

En esta investigación se realizó la toma de muestras de audio durante 10 semanas a los 10 niños seleccionados. Para este proceso se le entregó la lectura seleccionada a cada niño y de uno en uno mientras leían se les fue grabando un audio para su posterior análisis mediante el prototipo de aplicación. En total se grabaron 100 muestras de audio, es decir diez por cada niño, una por semana. La toma de muestras se realizó en un ambiente propicio evitando en la mayoría del tiempo cualquier clase de ruido que pudiera interferir en la obtención de las muestras. Por cada semana se llevó un registro con la toma de muestras. Este registro se incluyó el Nro. de semana y los niños. Respetando el artículo 52 del Código de la niñez y adolescencia se asignó un código a cada niño para que estos permanezcan en el anonimato. Para llevar un registro del trabajo de campo se realizó la Tabla 10.

Tabla 10 Registro de la toma de muestras de audio.

	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6	Semana 7	Semana 8	Semana 9	Semana 10
Niño 1	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado
Niño 2	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado
Niño 3	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado
Niño 4	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado
Niño 5	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado
Niño 6	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado
Niño 7	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado
Niño 8	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado
Niño 9	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado
Niño 10	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado	Realizado

Elaborado por: El Autor.

4.1.2 Codificación de los datos.

La codificación de los datos permitirá en lo posterior realizar el respectivo análisis. El formato en el que se dispondrán los datos es en forma de una matriz bidimensional donde los niños ocuparan las filas y las variables las columnas. Como resultado de la codificación se obtuvo la Tabla 11.

Tabla 11 Codificación de los datos.

Niños	Semanas	Técnica 1			Técnica 2			Técnica 3		
		Palabras	Tiempo	Memoria	Palabras	Tiempo	Memoria	Palabras	Tiempo	Memoria
Niño 1	Semana 1	97	101	467	50	88	209	132	24	49
	Semana 2	93	81	406	52	93	790	131	19	28
	Semana 3	90	63	433	43	87	515	130	17	30
	Semana 4	90	68	654	57	84	465	130	17	29
	Semana 5	92	73	782	43	78	994	130	15	42
	Semana 6	84	61	631	38	105	509	125	15	55
	Semana 7	92	65	397	44	75	527	127	16	18
	Semana 8	91	73	335	41	95	286	127	19	143
	Semana 9	83	68	759	34	76	1174	124	14	62
	Semana 10	84	68	754	27	81	527	129	20	29
Niño 2	Semana 1	93	60	632	62	80	555	130	20	29
	Semana 2	93	46	334	62	74	710	120	15	26
	Semana 3	87	48	491	68	85	1109	127	15	26
	Semana 4	103	58	504	58	96	1194	126	17	18
	Semana 5	99	53	742	70	87	710	127	23	82
	Semana 6	97	48	509	54	86	688	132	16	55
	Semana 7	92	53	527	53	80	911	127	15	62
	Semana 8	89	48	598	67	86	694	127	16	18
	Semana 9	77	33	771	65	78	474	124	13	67
	Semana 10	73	46	647	58	75	535	126	17	138
Niño 3	Semana 1	108	73	577	76	93	280	126	24	32
	Semana 2	107	49	469	77	86	963	122	19	30
	Semana 3	101	35	534	64	72	1389	126	17	29
	Semana 4	102	52	514	59	86	617	119	20	32
	Semana 5	96	39	420	67	75	519	113	16	54

	Semana 6	88	40	698	63	93	585	116	19	19
	Semana 7	105	50	640	71	79	1340	118	19	20
	Semana 8	93	42	545	55	66	920	122	16	55
	Semana 9	100	52	667	60	80	1257	124	23	83
	Semana 10	101	61	557	52	76	739	128	23	66
Niño 4	Semana 1	72	106	703	24	93	796	125	20	30
	Semana 2	64	75	522	18	67	705	123	18	30
	Semana 3	68	89	442	22	74	755	128	19	29
	Semana 4	73	83	786	22	94	951	129	15	25
	Semana 5	73	85	341	20	92	553	125	17	139
	Semana 6	77	77	578	35	88	495	117	15	56
	Semana 7	85	96	555	25	89	1367	125	19	56
	Semana 8	83	81	717	38	96	1153	124	15	55
	Semana 9	85	85	747	32	93	325	123	16	55
	Semana 10	82	100	819	24	90	647	129	21	81
Niño 5	Semana 1	97	75	706	52	78	807	126	20	29
	Semana 2	93	70	404	48	74	485	122	20	30
	Semana 3	103	58	694	50	76	1052	126	20	30
	Semana 4	103	77	707	55	75	1017	118	20	29
	Semana 5	99	74	390	49	77	693	126	23	81
	Semana 6	97	69	487	48	81	1377	122	25	85
	Semana 7	102	61	694	58	71	724	125	24	105
	Semana 8	92	65	771	56	71	696	121	22	20
	Semana 9	88	62	858	42	61	1559	109	22	42
	Semana 10	96	53	457	56	74	706	120	20	84
Niño 6	Semana 1	87	72	462	42	94	487	137	20	29
	Semana 2	90	76	738	25	84	562	126	19	30
	Semana 3	86	63	848	33	79	293	127	15	30
	Semana 4	89	67	886	31	81	294	129	16	26

	Semana 5	83	62	586	39	85	485	128	20	81
	Semana 6	81	55	434	40	79	766	125	13	53
	Semana 7	82	66	415	27	68	1386	124	16	18
	Semana 8	90	61	442	32	74	1171	129	15	18
	Semana 9	82	55	536	44	85	726	128	14	50
	Semana 10	83	51	673	39	72	208	121	14	17
Niño 7	Semana 1	93	77	497	38	95	610	123	21	30
	Semana 2	79	71	321	36	88	1024	119	20	30
	Semana 3	84	84	382	34	92	528	123	19	29
	Semana 4	77	75	726	29	82	1236	115	17	29
	Semana 5	82	64	437	42	92	885	116	23	76
	Semana 6	78	60	839	33	101	593	120	22	79
	Semana 7	81	60	702	31	75	1491	107	16	18
	Semana 8	73	63	880	31	82	1268	106	15	55
	Semana 9	76	63	888	35	89	1542	113	22	76
	Semana 10	90	78	932	30	90	660	119	21	19
Niño 8	Semana 1	89	114	444	41	108	351	131	24	28
	Semana 2	96	90	546	36	97	1350	119	22	20
	Semana 3	91	84	593	39	103	807	125	21	29
	Semana 4	98	81	448	38	94	678	123	22	20
	Semana 5	87	89	452	42	93	1557	125	26	87
	Semana 6	94	83	475	41	98	525	120	24	83
	Semana 7	90	87	956	45	100	639	126	26	99
	Semana 8	92	81	724	38	92	1165	129	22	56
	Semana 9	89	87	826	33	86	1024	127	20	50
	Semana 10	80	83	830	37	86	1056	129	24	79
Niño 9	Semana 1	94	82	610	50	78	848	119	23	33
	Semana 2	109	78	492	58	88	483	114	24	35
	Semana 3	95	74	406	42	79	388	122	21	29

	Semana 4	102	76	643	60	82	688	127	22	29
	Semana 5	98	71	737	48	81	898	115	26	91
	Semana 6	105	70	919	40	77	879	120	26	86
	Semana 7	105	66	802	56	81	1141	117	27	113
	Semana 8	107	66	447	55	75	584	110	23	83
	Semana 9	101	62	859	48	72	564	105	21	85
	Semana 10	98	58	581	49	73	861	122	23	81
Niño 10	Semana 1	61	81	773	20	68	488	121	22	32
	Semana 2	77	106	347	23	104	391	120	22	30
	Semana 3	69	100	536	28	90	996	128	21	28
	Semana 4	80	110	803	23	100	1396	122	21	30
	Semana 5	73	102	430	22	90	746	126	26	80
	Semana 6	79	103	810	32	98	1320	122	28	85
	Semana 7	89	100	771	27	100	441	127	21	78
	Semana 8	85	108	901	33	106	887	123	21	20
	Semana 9	87	110	477	33	102	576	117	24	90
	Semana 10	85	93	913	31	97	1373	122	24	76

Elaborado por: El Autor.

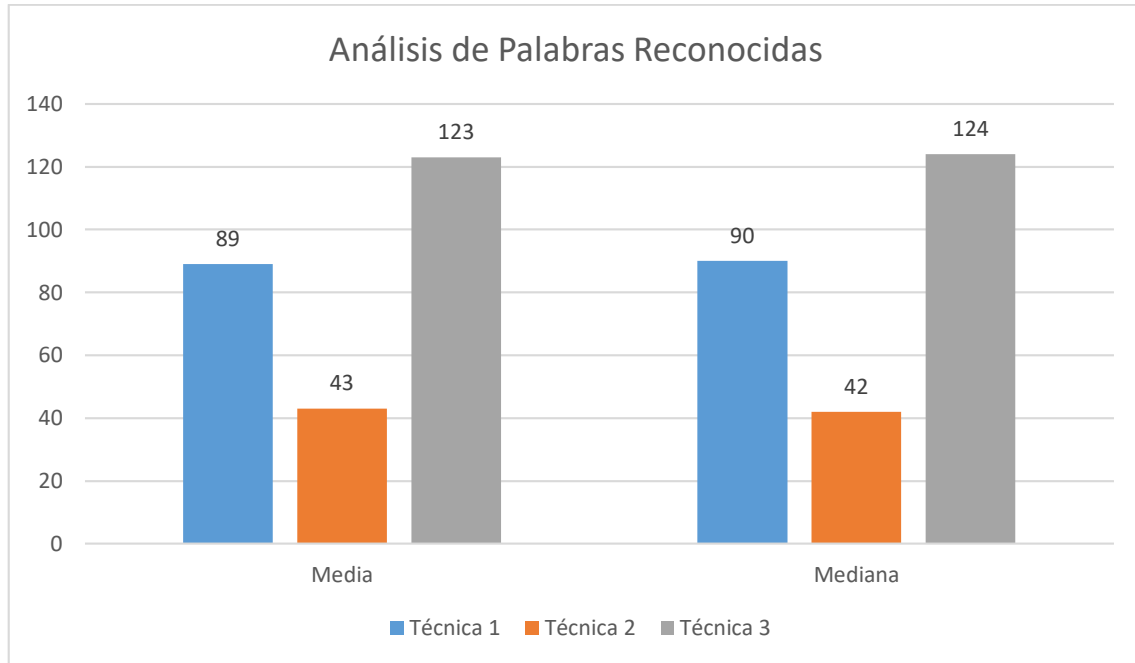
4.2 Análisis de datos.

Una vez recabados los datos y codificados, se procede al análisis de los mismos. Para el número de palabras reconocidas correctamente por cada técnica se procedió a calcular la media y mediana de los datos recabados para así tener información más cercana a la realidad de la investigación. El resultado de estas operaciones estadísticas se observa en la Tabla 12. Para tener una mejor interpretación se procedió a graficar la información analizada reflejada en el gráfico 2, donde se utiliza la media o promedio como medida de referencia y la mediana para validar que los valores son correctos, se observa claramente que la técnica 3 es decir reconocimiento basado en Redes Neuronales Artificiales es la que mayor número de palabras reconoció, por esta razón en cuanto al número de palabras reconocidas la técnica 3 es la mejor. Cabe resaltar que mientras más alto es el valor es mejor.

Tabla 12 Análisis de palabras reconocidas

	Técnica 1	Técnica 2	Técnica 3
Media	89	43	123
Mediana	90	42	124

Elaborado por: El Autor.

Gráfico 2 Media y mediana de palabras reconocidas por cada técnica.

Elaborado por: El Autor.

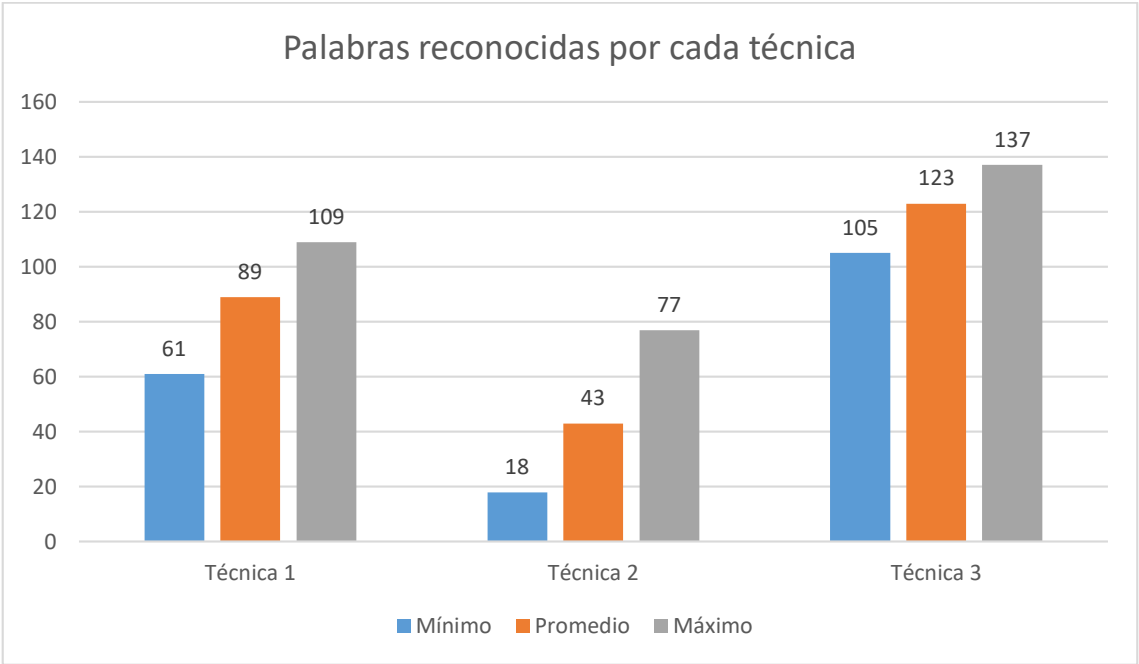
Además, se procedió a analizar los valores mínimos, promedio, y máximo de cada una de las técnicas seleccionadas, reflejados en la Tabla 13 y Gráfico 3, donde nuevamente se observa a la técnica 3 sobresalir frente a las demás tanto en mínimo, promedio y máximo de palabras reconocidas.

Tabla 13 Mínimo, promedio y máximo de palabras reconocidas por cada técnica.

	Técnica 1	Técnica 2	Técnica 3
Mínimo	61	18	105
Promedio	89	43	123
Máximo	109	77	137

Elaborado por: El Autor.

Gráfico 3 Mínimo, promedio y máximo de palabras reconocidas por cada técnica.



Elaborado por: El Autor.

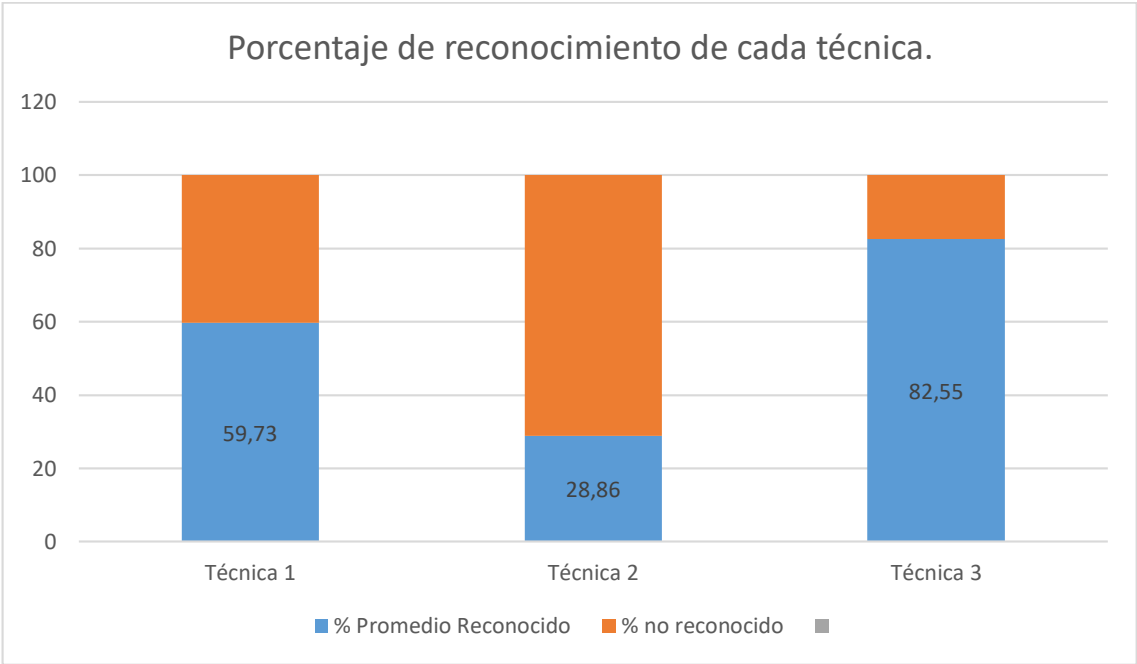
También, en base a estos resultados se calculó el porcentaje promedio de reconocimiento de cada técnica. Reflejados en la Tabla 14 y Gráfico 4, donde la técnica 3 es la que tiene mejor porcentaje promedio de reconocimiento con un 82,55%.

Tabla 14 Porcentaje promedio de reconocimiento por cada técnica.

	Técnica 1	Técnica 2	Técnica 3
% Promedio	59,73	28,86	82,55

Elaborado por: El Autor.

Gráfico 4 Porcentaje de reconocimiento promedio de cada técnica.



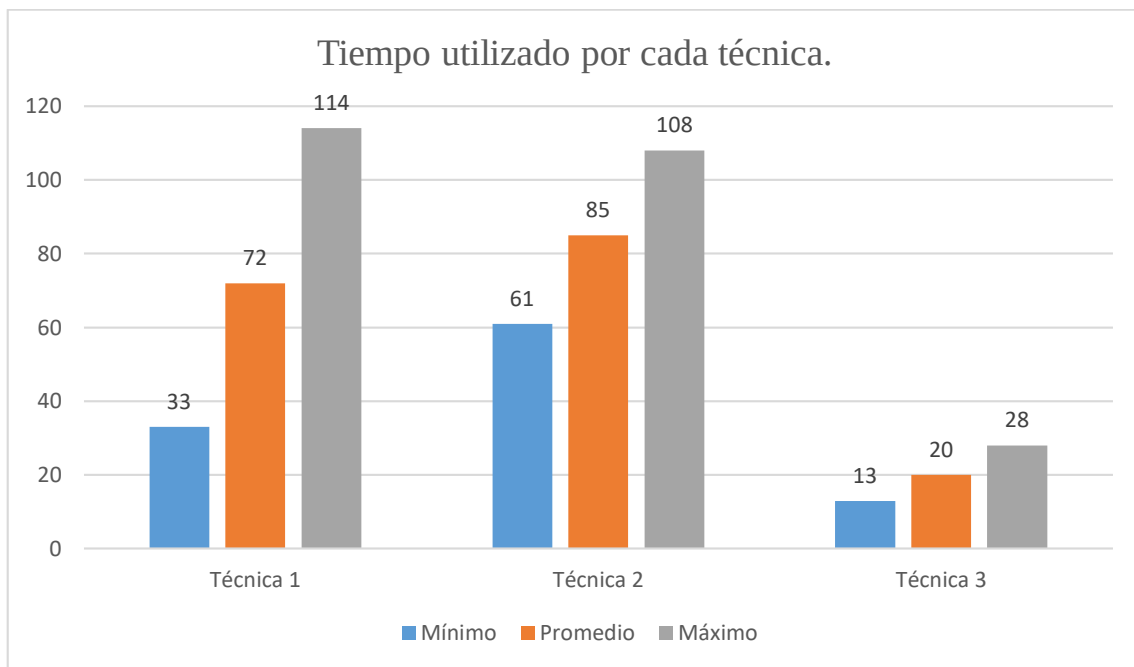
Elaborado por: El Autor.

Luego, se procedió a analizar el tiempo de respuesta, medido en Segundos para cada técnica. Los resultados obtenidos fueron representados en la Tabla 15 y para una mejor interpretación de esta información se procedió a realizar el Gráfico 5. Cabe resaltar que, para esta medida, mientras menor es el valor obtenido es mejor. Por ende, se observa que la técnica 3 es la que menor tiempo obtuvo, así como: mínimo, promedio y máximo.

Tabla 15 Mínimo, promedio y máximo de tiempo utilizado por cada técnica.

	Técnica 1	Técnica 2	Técnica 3
Mínimo	33	61	13
Promedio	72	85	20
Máximo	114	108	28

Elaborado por: El Autor.

Gráfico 5 Mínimo, promedio y máximo de tiempo utilizado por cada técnica.

Elaborado por: El Autor.

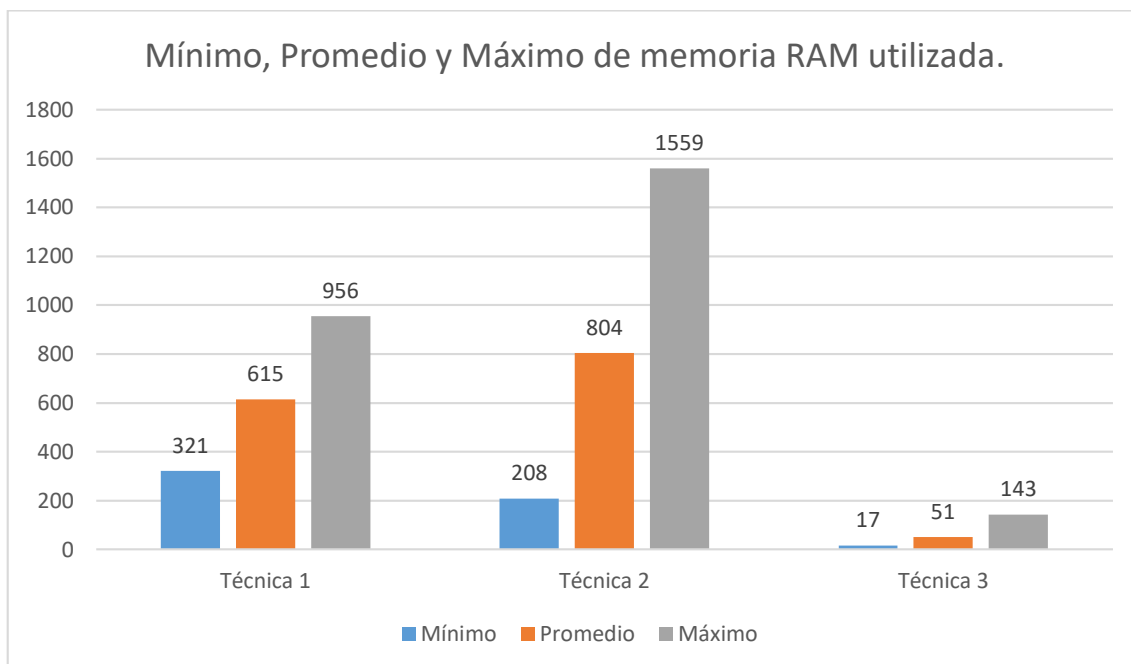
Se analizó también la cantidad de memoria RAM en MB utilizada en el reconocimiento por cada técnica, hay que tener en cuenta que mientras menor sea la memoria utilizada se considera mejor. La información analizada está en la Tabla 16 y su respectivo gráfico para una mejor interpretación es el Gráfico 6, donde se aprecia que la técnica 3 es la que mejor resultado obtuvo, tanto en el valor mínimo, promedio y máximo de memoria utilizada.

Tabla 16 Mínimo, promedio y máximo de memoria RAM utilizada por cada técnica.

	Técnica 1	Técnica 2	Técnica 3
Mínimo	321	208	17
Promedio	615	804	51
Máximo	956	1559	143

Elaborado por: El Autor.

Gráfico 6 Mínimo, promedio y máximo de memoria RAM utilizada por cada técnica.



Elaborado por: El Autor.

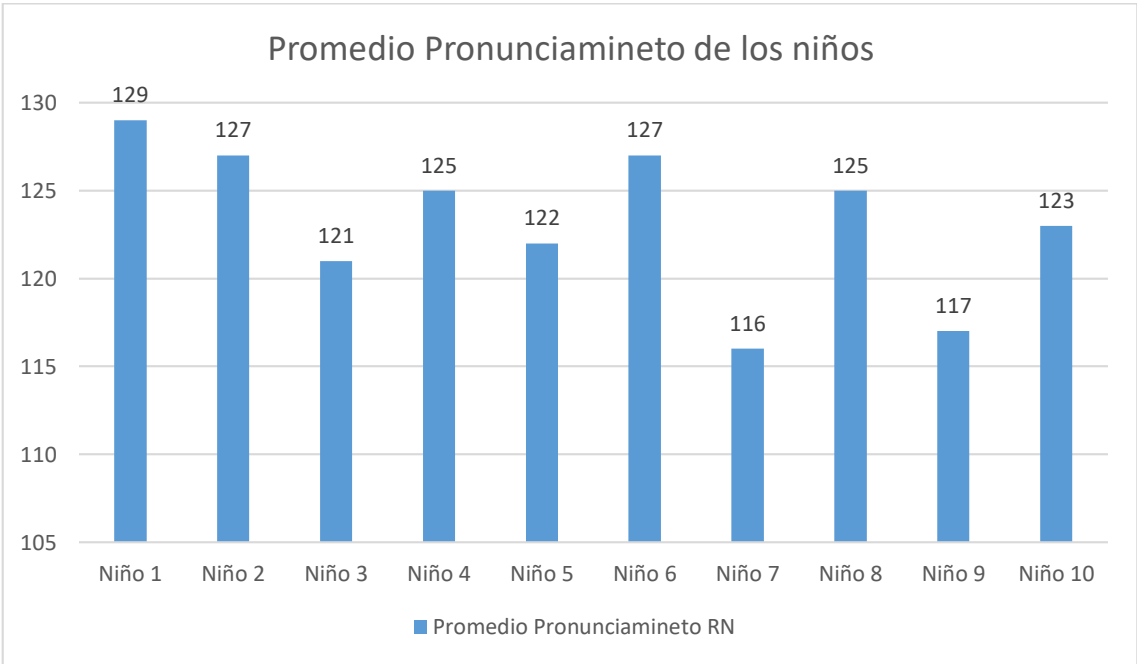
Llegado a este punto del análisis, y, con la información antes recabada, codificada y analizada, se puede determinar a la técnica 3 como la mejor de las tres técnicas seleccionadas, razón por la cual en base a esta se procederá a realizar el análisis respectivo de los niños. Para ello se procedió a analizar el promedio de palabras pronunciadas por los niños en esta técnica. Los resultados de éste análisis están reflejados en la Tabla 17 y Gráfico 7 respectivamente, donde se aprecia que el niño 1 es quien tiene el promedio más alto, lo que se interpreta como el niño que posee mejor nivel de lectura; además se observa que 5 niños tienen un promedio mayor o igual a 125 palabras, que se considera como excelente; 3 niños tienen un promedio mayor o igual 120 palabras considerado como muy bueno, y, que 2 niños poseen un promedio mayor o igual a 115 palabras que se considera como bueno.

Tabla 17 Promedio de palabras pronunciadas por cada niño.

Niños	Promedio Palabras Pronunciadas.
Niño 1	129
Niño 2	127
Niño 3	121
Niño 4	125
Niño 5	122
Niño 6	127
Niño 7	116
Niño 8	125
Niño 9	117
Niño 10	123

Elaborado por: El Autor.

Gráfico 7 Promedio de palabras pronunciadas por cada niño.



Elaborado por: El Autor.

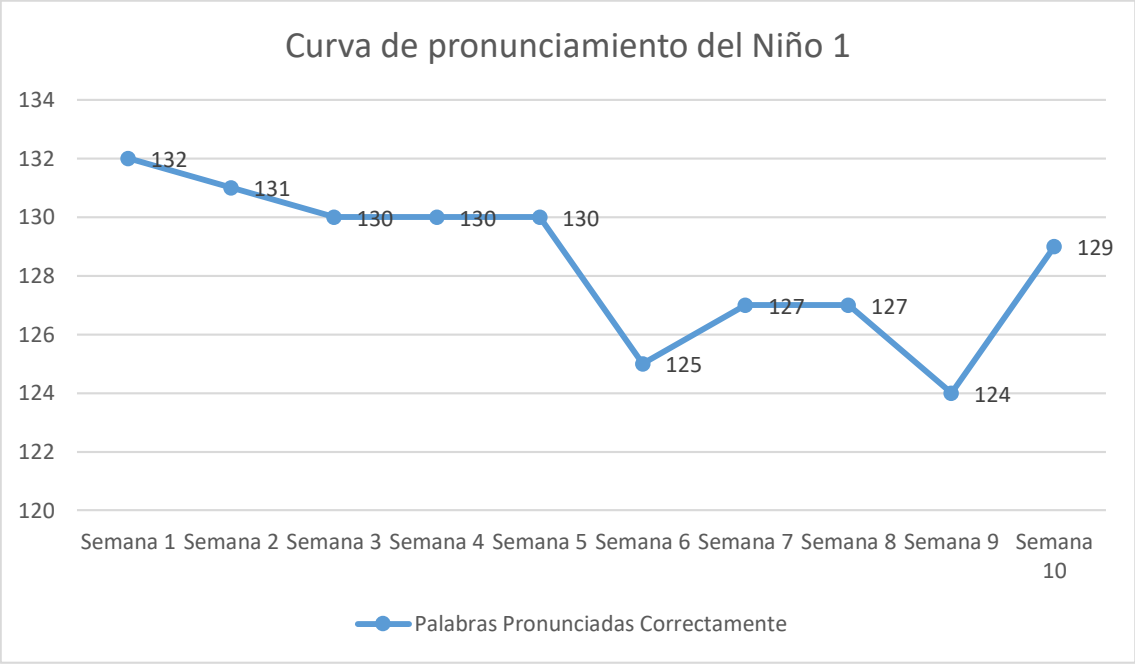
Luego se procedió a analizar el nivel de pronunciamiento de cada niño durante las 10 semanas; para ello se realizó una tabla y un gráfico correspondiente a cada niño. Donde se observa que el niño 1 tienen un nivel de pronunciamiento excelente, porque su pronunciamiento es mayor o igual a 125 palabras en 9 semanas y mayor a 120 palabras en 1 semana. Prueba de esto es la tabla 18 y gráfico 8.

Tabla 18 Nivel de pronunciamiento del niño 1.

Semana	Palabras Pronunciadas Correctamente
Semana 1	132
Semana 2	131
Semana 3	130
Semana 4	130
Semana 5	130
Semana 6	125
Semana 7	127
Semana 8	127
Semana 9	124
Semana 10	129

Elaborado por: El Autor.

Gráfico 8 Nivel de pronunciamiento del niño 1.



Elaborado por: El Autor.

El niño 2 tiene un nivel de pronunciamiento excelente, porque su pronunciamiento es mayor a 125 palabras en 8 semanas y mayor o igual a 120 palabras en 2 semanas. Prueba de esto es la tabla 19 y gráfico 9.

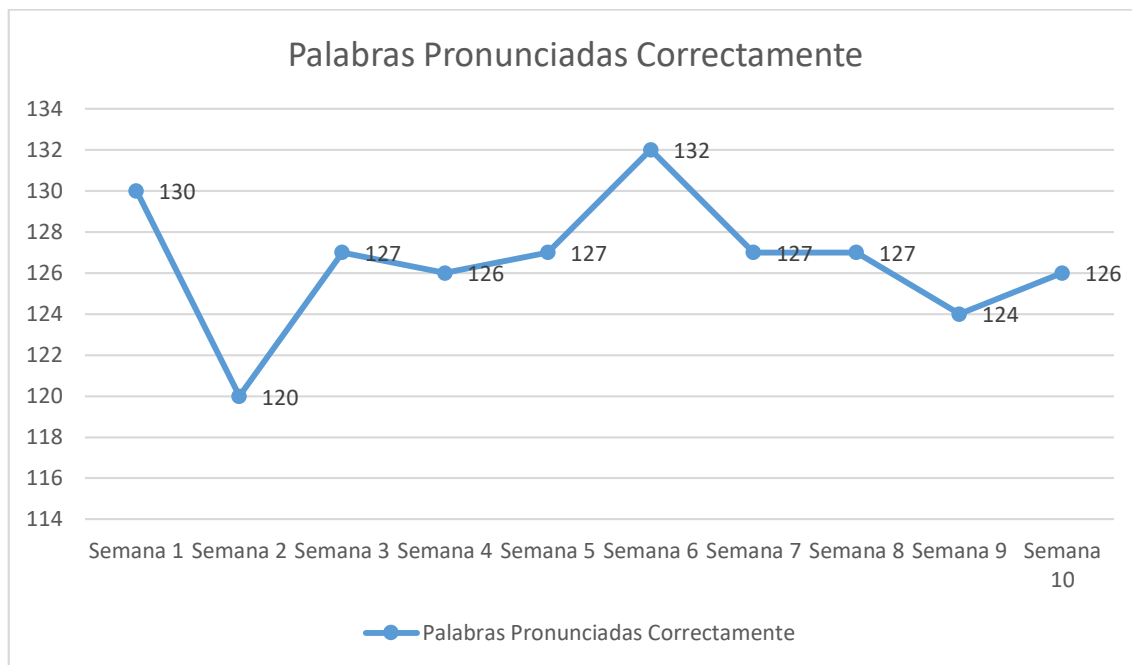
Tabla 19 Nivel de pronunciamiento del niño 2.

Semana	Palabras Pronunciadas Correctamente
Semana 1	130
Semana 2	120
Semana 3	127
Semana 4	126
Semana 5	127
Semana 6	132
Semana 7	127
Semana 8	127
Semana 9	124

Semana 10

126

Elaborado por: El Autor.

Gráfico 9 Nivel de pronunciamiento del niño 2.

Elaborado por: El Autor.

El niño 3 tiene un nivel de pronunciamiento muy bueno, porque su pronunciamiento es mayor a 120 palabras en 6 semanas, mayor a 115 palabras en 3 semanas y mayor a 110 palabras en 1 semana. Prueba de esto es la tabla 20 y gráfico 10.

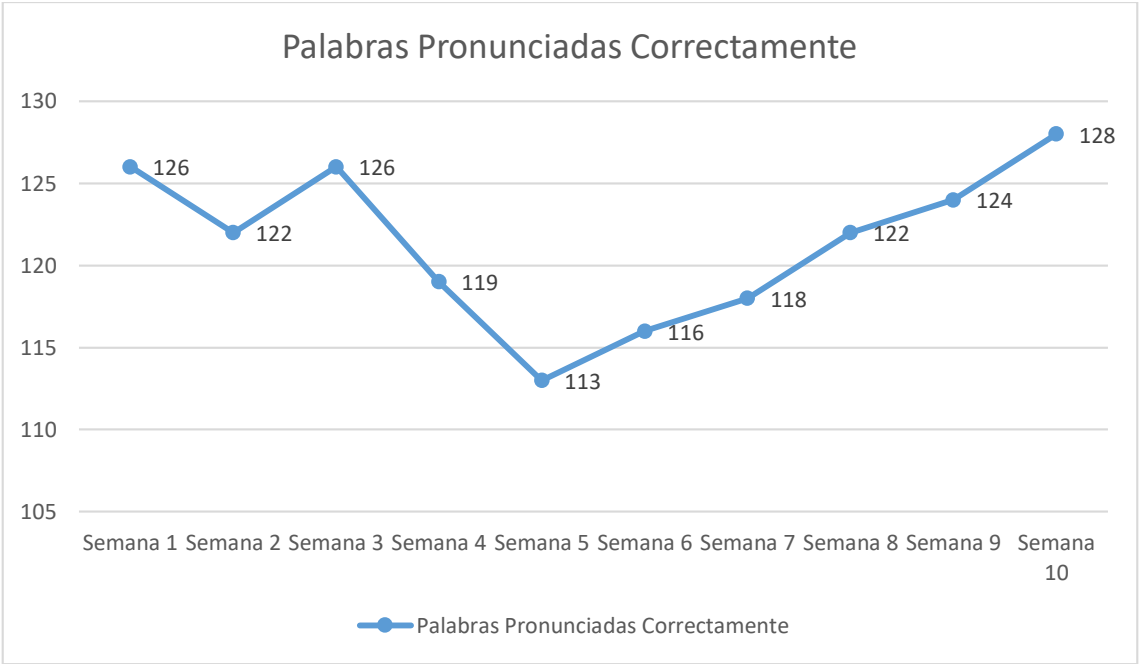
Tabla 20 Nivel de pronunciamiento del niño 3.

Semana	Palabras Pronunciadas Correctamente
Semana 1	126
Semana 2	122
Semana 3	126
Semana 4	119
Semana 5	113
Semana 6	116

Semana 7	118
Semana 8	122
Semana 9	124
Semana 10	128

Elaborado por: El Autor.

Gráfico 10 Nivel de pronunciamiento del niño 3.



Elaborado por: El Autor.

El niño 4 tiene un nivel de pronunciamiento excelente, porque su pronunciamiento es mayor o igual a 125 palabras en 6 semanas, mayor de 120 palabras en 3 semanas y mayor de 115 palabras en 1 semana. Prueba de esto es la tabla 21 y gráfico 11.

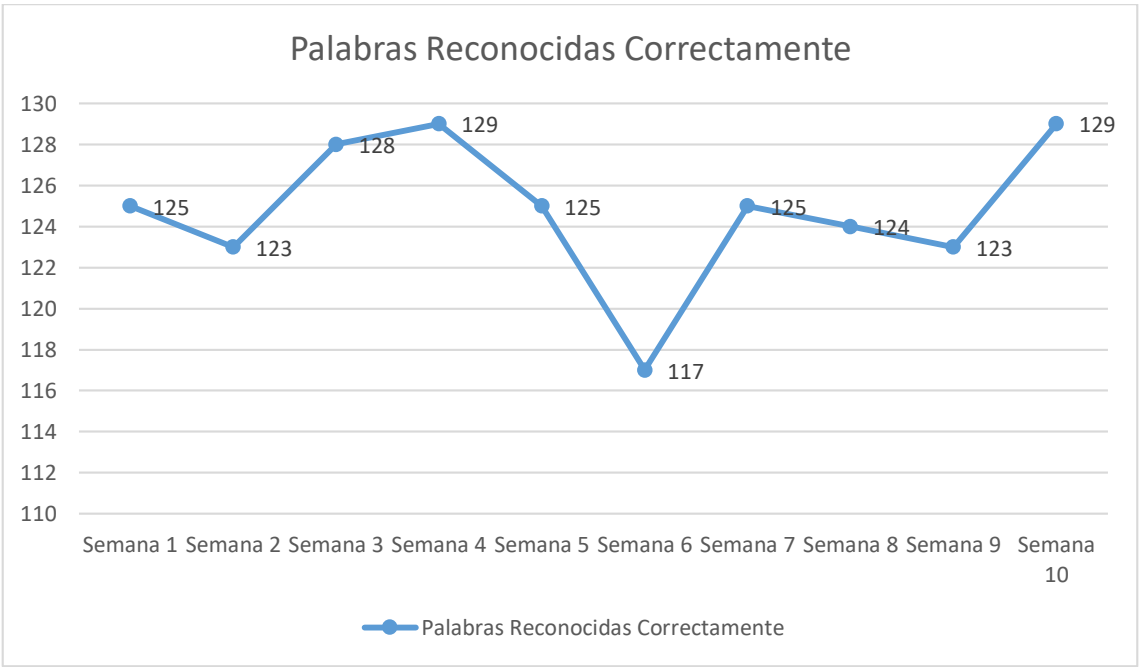
Tabla 21 Nivel de pronunciamiento del niño 4.

Semana	Palabras Pronunciadas Correctamente
Semana 1	125

Semana 2	123
Semana 3	128
Semana 4	129
Semana 5	125
Semana 6	117
Semana 7	125
Semana 8	124
Semana 9	123
Semana 10	129

Elaborado por: El Autor.

Gráfico 11 Nivel de pronunciamiento del niño 4.



Elaborado por: El Autor.

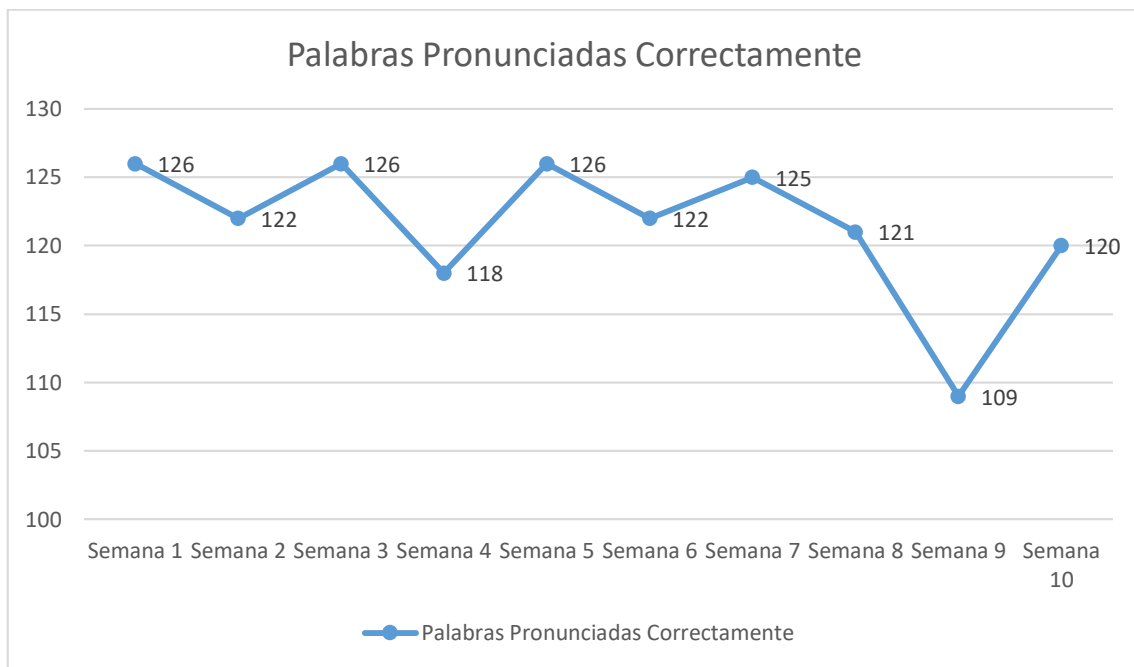
El niño 5 tiene un nivel de pronunciamiento muy bueno, porque su pronunciamiento es mayor o igual a 125 palabras en 4 semanas, mayor o igual a 120 palabras en 4 semanas,

mayor a 115 palabras en 1 semana y mayor a 105 palabras en 1 semana. Prueba de esto es la tabla 22 y gráfico 12.

Tabla 22 Nivel de pronunciamiento del niño 5.

Semana	Palabras Pronunciadas Correctamente
Semana 1	126
Semana 2	122
Semana 3	126
Semana 4	118
Semana 5	126
Semana 6	122
Semana 7	125
Semana 8	121
Semana 9	109
Semana 10	120

Elaborado por: El Autor.

Gráfico 12 Nivel de pronunciamiento del niño 5.

Elaborado por: El Autor.

El niño 6 tienen un nivel de pronunciamiento excelente, porque su pronunciamiento es mayor o igual a 125 palabras en 8 semanas y mayor a 120 palabras en 2 semanas. Prueba de esto es la tabla 23 y gráfico 13.

Tabla 23 Nivel de pronunciamiento del niño 6.

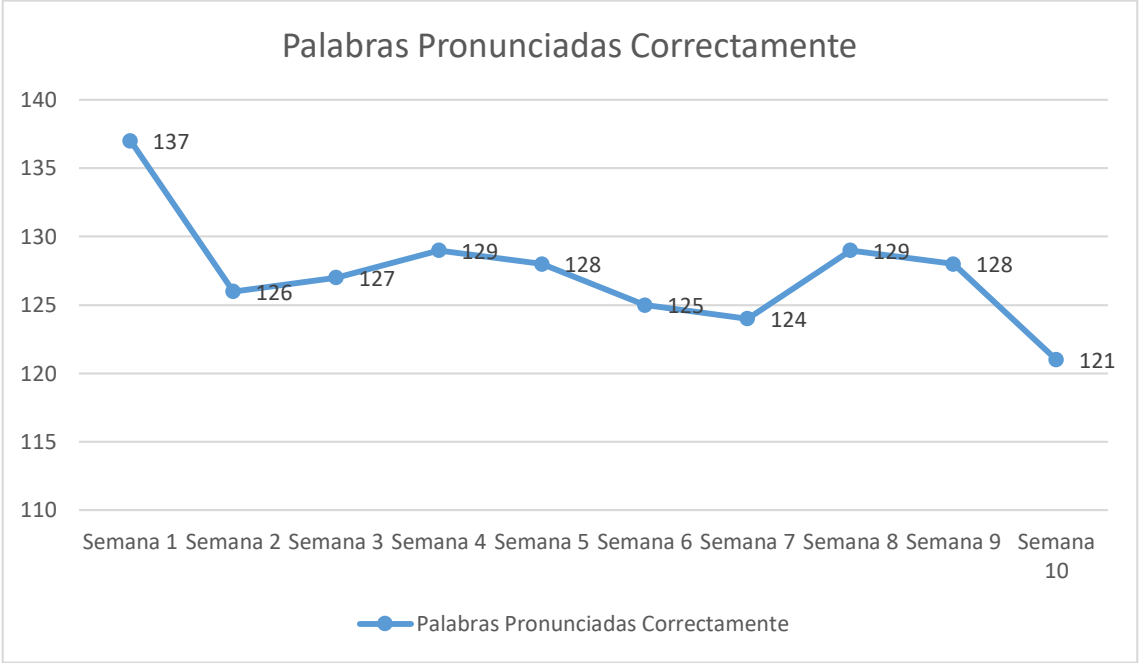
Semana	Palabras Pronunciadas Correctamente
Semana 1	137
Semana 2	126
Semana 3	127
Semana 4	129
Semana 5	128
Semana 6	125
Semana 7	124
Semana 8	129
Semana 9	128

Semana 10

121

Elaborado por: El Autor.

Gráfico 13 Nivel de pronunciamiento del niño 6.



Elaborado por: El Autor.

El niño 7 tiene un nivel de pronunciamiento bueno, porque su pronunciamiento es mayor o igual a 120 palabras en 3 semanas, mayor o igual a 115 palabras en 4 semanas, mayor a 110 palabras en 1 semana y mayor a 105 palabras en 2 semanas. Prueba de esto es la tabla 24 y gráfico 14.

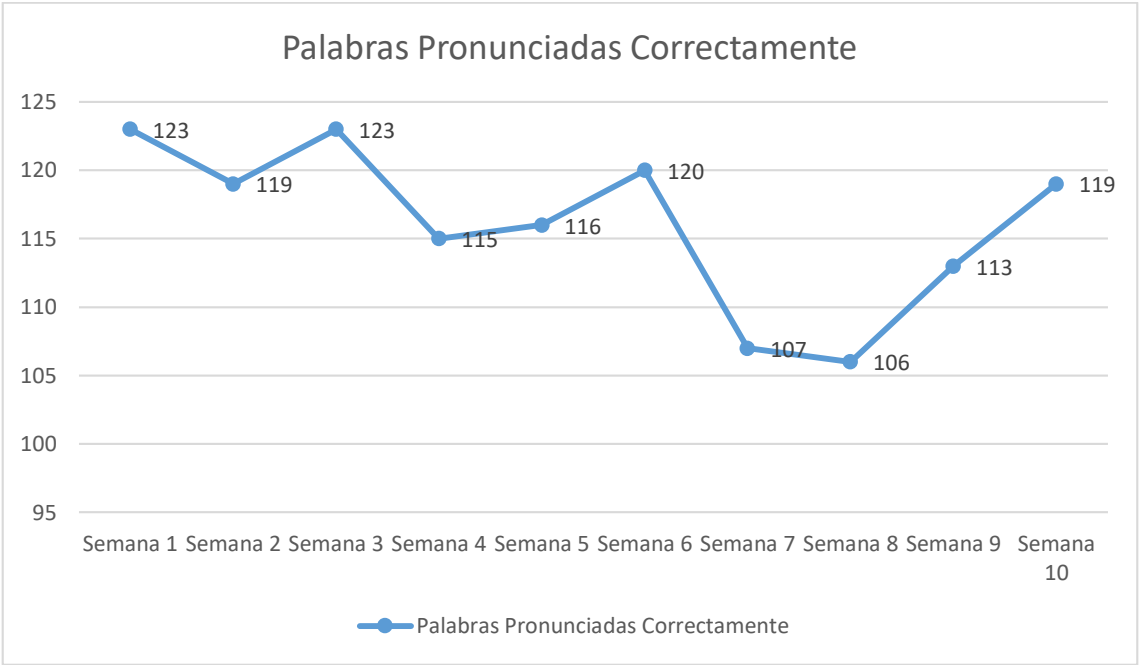
Tabla 24 Nivel de pronunciamiento del niño 7.

Semana	Palabras Pronunciadas Correctamente
Semana 1	123
Semana 2	119
Semana 3	123
Semana 4	115
Semana 5	116

Semana 6	120
Semana 7	107
Semana 8	106
Semana 9	113
Semana 10	119

Elaborado por: El Autor.

Gráfico 14 Nivel de pronunciamiento del niño 7.



Elaborado por: El Autor.

El niño 8 tiene un nivel de pronunciamiento excelente, porque su pronunciamiento es mayor o igual a 125 palabras en 7 semanas, mayor o igual a 120 palabras en 2 semanas y mayor a 115 palabras en 1 semana. Prueba de esto es la tabla 25 y gráfico 15.

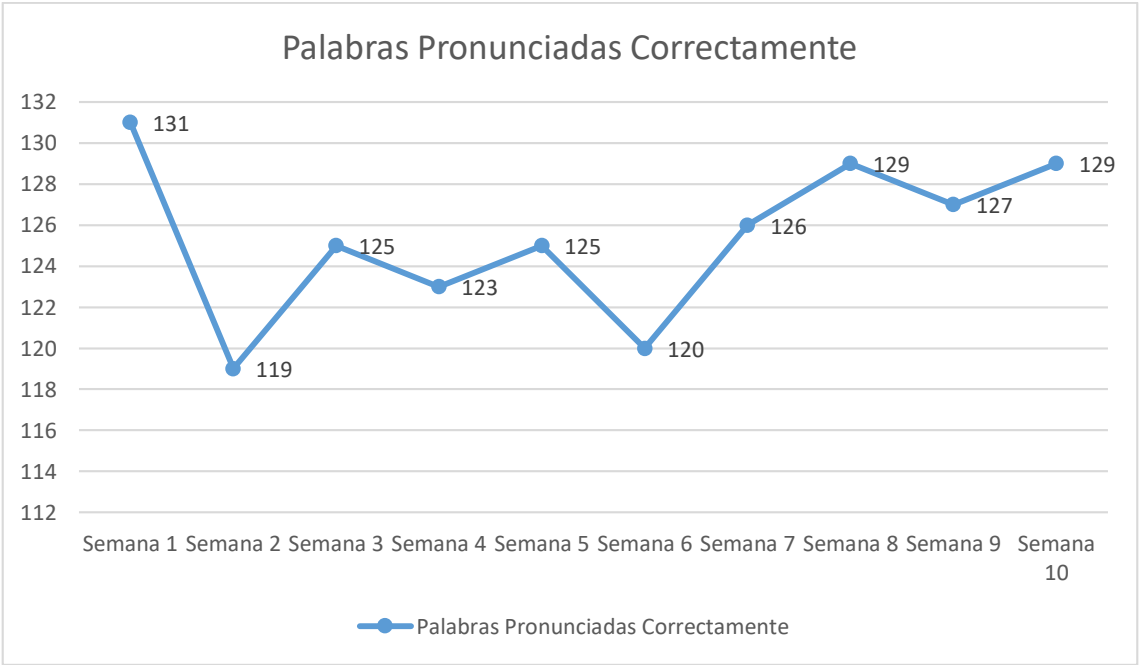
Tabla 25 Nivel de pronunciamiento del niño 8.

Semana	Palabras Pronunciadas Correctamente
--------	-------------------------------------

Semana 1	131
Semana 2	119
Semana 3	125
Semana 4	123
Semana 5	125
Semana 6	120
Semana 7	126
Semana 8	129
Semana 9	127
Semana 10	129

Elaborado por: El Autor.

Gráfico 15 Nivel de pronunciamiento del niño 8.



Elaborado por: El Autor.

El niño 9 tiene un nivel de pronunciamiento bueno, porque su pronunciamiento es mayor a 125 palabras en 1 semana, mayor o igual a 120 palabras en 3 semanas, mayor o

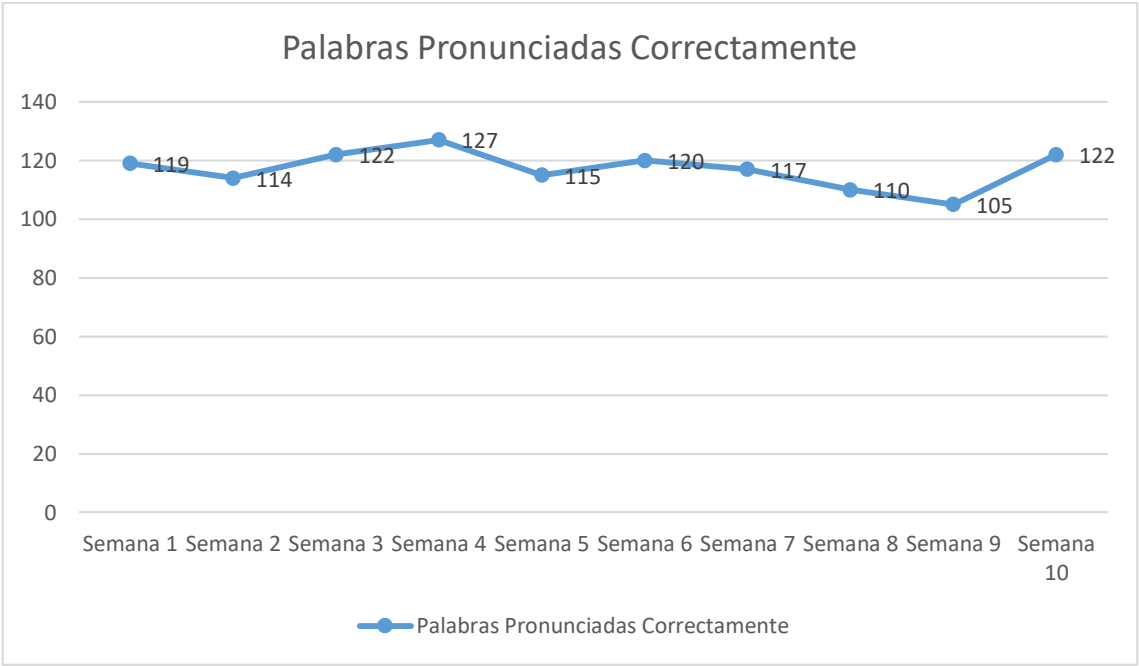
igual a 115 palabras en 3 semanas, mayor o igual a 110 palabras en 2 semanas y mayor o igual a 105 palabras en 1 semana. Prueba de esto es la tabla 26 y gráfico 16.

Tabla 26 Nivel de pronunciamiento del niño 9.

Semana	Palabras Pronunciadas Correctamente
Semana 1	119
Semana 2	114
Semana 3	122
Semana 4	127
Semana 5	115
Semana 6	120
Semana 7	117
Semana 8	110
Semana 9	105
Semana 10	122

Elaborado por: El Autor.

Gráfico 16 Nivel de pronunciamiento del niño 9.



Elaborado por: El Autor.

El niño 10 tiene un nivel de pronunciamiento muy bueno, porque su pronunciamiento es mayor a 125 palabras en 3 semanas, mayor o igual a 120 palabras en 6 semanas y mayor a 115 palabras en 1 semana. Prueba de esto es la tabla 27 y gráfico 17.

Tabla 27 Nivel de pronunciamiento del niño 10.

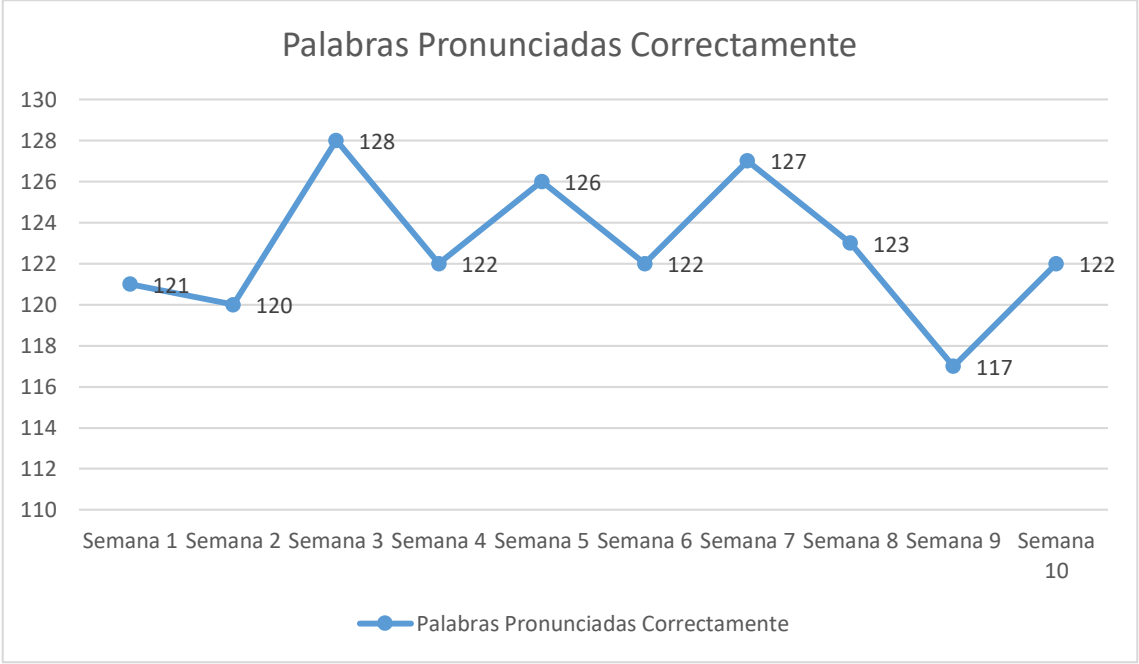
Semana	Palabras Pronunciadas Correctamente
Semana 1	121
Semana 2	120
Semana 3	128
Semana 4	122
Semana 5	126
Semana 6	122
Semana 7	127
Semana 8	123
Semana 9	117

Semana 10

122

Elaborado por: El Autor.

Gráfico 17 Nivel de pronunciamiento del niño 10.



Elaborado por: El Autor.

Capítulo 5

5 Resultados

Luego de haber analizado los datos y en base a los resultados obtenidos, como: número de palabras reconocidas, en donde la técnica 3 (Reconocimiento Basado en Redes Neuronales) es la que mayor número de palabras reconocidas posee; tiempo de respuesta, donde la técnica 3 sale ganadora al tener menor tiempo de respuesta por ende resulta ser mucho más rápida al reconocer las palabras; y memoria RAM utilizada, donde, el consumo de memoria utilizada por la técnica 3, resulta ser significativamente menor que el consumo de memoria de las demás técnicas, entonces como la técnica 3 reconoce más palabras, en menos tiempo, y, utiliza menos memoria RAM se considera esta técnica como la mejor de las tres.

Se pudo apreciar también que las técnicas seleccionadas tienen gran influencia en el reconocimiento de la pronunciación de los niños, puesto que de utilizar una a otra técnica el número de palabras reconocidas por estas técnicas cambia de manera considerable.

En cuanto a los resultados obtenidos del análisis del nivel promedio de pronunciamiento de los niños se puede decir que 8 de los diez niños poseen un nivel promedio de pronunciamiento superior a las 120 palabras, considerándose como muy bueno, mientras que solamente dos niños poseen un nivel promedio de pronunciamiento menor a 120 y mayor a 115 palabras considerándose como bueno. Hay que destacar que el niño 1 posee un nivel promedio de pronunciamiento de 129 palabras siendo éste el que alcanzó el nivel promedio más alto de palabras pronunciadas correctamente.

Como resultado también hay que destacar el prototipo de aplicación desarrollado para la evaluación de las técnicas.

Capítulo 6

6 Problemas encontrados en el proyecto.

Durante el desarrollo del proyecto de investigación se presentaron algunos problemas, que de ser cierto fueron solucionados, pero tuvieron repercusiones en el transcurso del mismo. Estos problemas se detallan a continuación:

- Se pretendió utilizar un modelo acústico propio para este proyecto, pero los recursos necesarios para la generación del mismo, el tiempo que lleva entrenarlo, y la cantidad de horas de audio necesarias para esto, llevó a la utilización de un modelo de VoxForge, pero este a su vez no fue integrado correctamente por este motivo se recurrió al modelo de la Universidad de Carnegie Mellon.
- Al ser las librerías de Google Cloud de prueba resultó tediosa la configuración de autenticación del prototipo de aplicación con los servidores de google, llevando más tiempo del esperado, al final se cambió el tipo de autenticación para poder realizar la correcta comunicación con los servidores de Google Cloud.
- Las librerías de Google al estar en constante actualización, en el mes de febrero hicieron cambios en el mensaje que devolvía el servidor, cuando un archivo de audio a analizar superaba el minuto de duración, lo cual generó retraso en el desarrollo del prototipo de aplicación y en el desarrollo del proyecto de investigación puesto que se utilizó bastante tiempo y recursos encontrar el problema y poder darle solución.
- En un principio el reconocimiento basado en Redes Neuronales Artificiales se pensó realizar sin cargar el archivo en un servidor de Google Cloud Storage ni hacer uso de este servicio, pero al tener la limitante de que cuando un archivo

supera el minuto de duración, para que este pueda ser procesado, debe estar subido primero en los servidores de Google Cloud Storage, fue obligatorio el uso de esta librería junto con el tiempo adicional que llevó el documentarse sobre la misma y lograr la correcta integración con los demás módulos desarrollados.

Conclusiones

Una vez realizado el trabajo de investigación se puede concluir lo siguiente:

- La selección de las técnicas a ser evaluadas fue la adecuada para el tipo de investigación realizado, lo que facilitó la correcta evaluación de las mismas.
- Se logró desarrollar un prototipo de aplicación que integre las técnicas de procesamiento del lenguaje natural para su evaluación, este prototipo facilitó esta tarea, el análisis de datos y la presentación de resultados. Además de servir como base para futuras investigaciones correspondientes al procesamiento del lenguaje natural.
- Se logró medir el nivel de pronunciamiento de la población seleccionada periódicamente, además se determinó el niño cuyo nivel promedio de pronunciamiento fue el mejor de los 10 niños seleccionados. Como se observa en la tabla 17 y gráfico 7.
- Se determinó a la técnica de reconocimiento basada en Redes Neuronales Artificiales como la mejor de las tres técnicas evaluadas, porque reconoce mayor número de palabras como se observa en las tablas 12, 13 y 14 y en los gráficos 2, 3 y 4, utiliza menos tiempo en el reconocimiento como se evidencia en la tabla 15 y gráfico 5, también porque utiliza menos memoria como se observa en la tabla 16 y gráfico 6, lo que sirve para futuros proyectos basados en el procesamiento del lenguaje natural siendo la principal opción a utilizar.
- El uso de tecnologías OpenSource, como Ubuntu y el lenguaje de programación Java permitieron la integración de las tres técnicas en un único prototipo de aplicación, además se logró que la aplicación sea multiplataforma y que el coste de recursos monetarios para el desarrollo de este prototipo sea bajo.

- El uso del modelo acústico CMUSphinx-es facilitó el desarrollo del proyecto, puesto que, al poseer un diccionario grande junto con un modelo de lenguaje altamente entrenado, se evitó de entrenar un modelo propio, lo cual redujo considerablemente el número audios, de recursos humanos, materiales y de tiempo que lleva entrenar un modelo propio.
- La utilización de una gramática propia hizo posible adquirir un alto nivel de conocimiento sobre gramáticas de reconocimiento, los formatos que estas admiten, su composición y los beneficios que el uso de estas otorga, además, se determinó que es mejor crear una gramática desde cero con todo el proceso que esto conlleva, que crear un modelo de lenguaje y entrenarlo desde cero.

Recomendaciones

El desarrollo de este proyecto marca un avance significativo en cuanto al uso de técnicas de procesamiento del lenguaje natural, y también muestra los aspectos a mejorar para futuros proyectos de este tipo, por lo cual se deben tener en cuenta las siguientes recomendaciones:

- Analizar adecuadamente el tipo de proyecto a realizar y en base a este seleccionar la técnica que mejor se adapte, es posible que la técnica que se determinó en este proyecto como mejor de las tres, no sea la adecuada para utilizar en otro proyecto, debido a las características del mismo.
- Para proyectos en los que una conexión a internet no sea posible, las técnicas de reconocimiento Basado en el Conocimiento y la que utiliza Modelos ocultos de Markov son alternativas importantes a considerar. En contraste con las que utilizan Redes Neuronales Artificiales y sobre todo las librerías de Google Cloud, las cuales hacen uso de Internet.
- De disponer los recursos necesarios se recomienda entrenar un modelo de lenguaje para aumentar la precisión al momento de reconocer las palabras.
- Se considera para futuros proyectos de investigación de este tipo, basarse en muestras de audio de sujetos que pronuncien mejor las palabras y brinden resultados que puedan ser comparados con los obtenidos durante este proceso de investigación.
- En cuanto al desarrollo del prototipo de aplicación se recomienda realizar respaldos previos a la integración de un nuevo módulo puesto que de no funcionar o de no pasar las pruebas de regresión se pueda volver al estado anterior con facilidad y poder depurar los problemas encontrados.

- Se recomienda en lo posible mantener la versión de las librerías con las que se empezó a desarrollar puesto que la actualización de las mismas puede contener cambios en los procedimientos y métodos, lo cual terminaría generando un problema de compatibilidad y este a su vez acarrearía pérdida de recursos, tiempo y dinero. Dificultando el proceso de desarrollo.

Bibliografía

- Alarcon, D. (2011). *SISTEMA QUE SIMULE CONOCIMIENTO HUMANO CON VOZ "SEUZ."* Tesis. Universidad de Guayaquil.
- Báez, L. S., & Cabrera, F. P. (2010). *Diseño e Implementación de un sistema móvil anti-robo comandado por voz a un sistema de seguridad electrónica para vehículo .*
- Brookshear, J. G. (1993). *Teoría de la Computación - Lenguajes Formales, automatas y complejidad.*
- Carrillo, E., & Criollo, E. (2016). *Aplicación Móvil Intérprete entre Personas con Discapacidad Visual y Auditiva a partir de Técnicas de Procesamiento de Lenguaje Natural.* Universidad Técnica de Cotopaxi.
- CMUSphinx Project. (2009). Training Acoustic Model. Retrieved from <https://cmusphinx.github.io/wiki/tutorialam/>
- Coronel, R. del C. (2006). *Procesamiento del Lenguaje Natural Aplicado a la Interacción con un Call Center utilizando una Gramática Predefinida y Basado en Algoritmos de Búsqueda Inteligentes.* Universidad Técnica Particular de Loja.
- Cortez, A. (2002). *Lenguajes y compiladores.* UNMSM.
- Cortez, A., Vega, H., & Pariona, J. (2009). Procesamiento de lenguaje natural. *Revista de Ingenieria de Sistemas E Informática*, 6(2), 45–54.
- Diakopoulos, N., & Cass, S. (2016). Interactive: The Top Programming Languages 2016. Retrieved from <http://spectrum.ieee.org/static/interactive-the-top-programming-languages-2016>
- Fadiño Rodríguez, D. A. (2005). *Estado del arte en el reconocimiento Automático de voz.* Universidad Nacional de Colombia.

- Falconí, L. F., & Jiménez, C. (2009). *Estudio en Implementación de Domótica Activado por Comandos de Voz y Comunicación en Zigbee*. Infoplcn.Net. Escuela Superior Politécnica de Chimborazo. Retrieved from http://www.infoplcn.net/files/descargas/schneider/infoplcn_net_18t00436.pdf
- Gómez, P., & Freire, P. (2014). *ANÁLISIS, DISEÑO, CONSTRUCCIÓN Y EVALUACIÓN DE UN SISTEMA EXPERTO TRADUCTOR DE LOS IDIOMAS: CASTELLANO A KICHWA ECUATORIANO BASADO EN REGLAS DE INFERENCIA Y RAZONAMIENTO APROXIMADO*. Univeridad Politécnica Salesiana sede Quito.
- Gorin, A., & Mammone, R. J. (1994). Introduction to the special issue on neural networks for speech processing. *IEEE Transactions on Speech and Audio Processing*, 2(1), 113–114. <https://doi.org/10.1109/89.260355>
- Hopcroft, J., Motwani, R., & Ullman, J. (1993). *Introducción a la teoría de autómatas, lenguajes y computación*.
- Huang, X., Alleva, F., Hwang, M.-Y., & Rosenfeld, R. (1993). An overview of the SPHINX-II speech recognition system. *Proceedings of the Workshop on Human Language Technology - HLT '93*, 81. <https://doi.org/10.3115/1075671.1075690>
- Kirschning, I., Aguas, N., & Ahuactzin, A. (2000). Aplicación de tecnología de voz en la enseñanza del español. *HAVOL 2000. 1er Taller Internacional de Tratamiento Del Habla, Procesamiento de Voz Y El Lenguaje*. Retrieved from <http://ict.udlap.mx/people/ingrid/ingrid/HAVOL2000a.pdf>
- Macías Guarasa, J. (2001). *Arquitecturas y métodos en sistemas de reconocimiento automático de habla de gran vocabulario*.

- Ney, H., Haeb-Umbach, R., Tran, B.-H., & Oerder, M. (1992). Improvements in beam search for 10000-word continuous speech recognition. In *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 9–12 vol.1). IEEE. <https://doi.org/10.1109/ICASSP.1992.225985>
- Peral, S. R. (2016). El pajarito que no quería volar.
- Pressman, R. (2010). *Ingeniería Del Software Un enfoque práctico* (Séptima).
- Pusay, E. (2014). *DISEÑO E IMPLEMENTACIÓN DE UN DISPOSITIVO QUE PERMITA EL ENCENDIDO DE LOS AUTOMÓVILES DE SISTEMA CONVENCIONAL, UTILIZANDO UNA PLACA ELECTRÓNICA QUE CONTROLA EL SISTEMA DE ENCENDIDO COMANDADO POR VOZ PROGRAMADA*. Universidad Tecnológica Equinoccial. Retrieved from http://repositorio.ute.edu.ec/bitstream/123456789/11938/1/58621_1.pdf
- Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*. <https://doi.org/10.1109/5.18626>
- Ravi Sethi, Ullman, J. D., & Aho, A. V. (1990). *Compiladores: principios, técnicas y herramientas*. 1990. Retrieved from https://books.google.com.ec/books?id=yG6qJBAnE9UC&pg=PA262&lpg=PA262&dq=recuperacion+de+errores+analisis+sintactico&source=bl&ots=rs-QNX4-jL&sig=dJ6B08joyFjvt1SgZg-uyY3DO3M&hl=es&sa=X&redir_esc=y#v=onepage&q=recuperacion de errores analisis sintactico&f=f
- Real, E. (2016). *Manual de Investigación para Ciencias Sociales y de la Salud en Grado y Posgrado*.

- Romero Fonseca, F. F. (2010). *Diseño y Construcción de un Sistema para el Control de Encendido y Apagado de Dispositivos Eléctricos por Medio de Comandos de Voz. Micro*. Escuela Politécnica Nacional.
- Rufiner, H. L., & Milone, D. H. (2004). Sistema de reconocimiento automático del habla. *Ciencia, Docencia Y Tecnología*, 15(28), 151–177.
- Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*, 3rd edition. Prentice Hall. <https://doi.org/10.1017/S0269888900007724>
- Sethi, R. (1992). *Lenguajes de programación: conceptos y constructores*. (E. García, Ed.). Addison-Wesley Iberoamericana.
- Soong, F. K., & Huang, E.-F. (1990). A tree-trellis based fast search for finding the N Best sentence hypotheses in continuous speech recognition. *Proceedings of the Workshop on Speech and Natural Language - HLT '90*, 12–19.
<https://doi.org/10.3115/116580.116591>
- Trabajo, M. de. Acuerdo Ministerial Nro: MDT-2015-0291 (2016).

Anexos

Anexo A. Glosario de Términos.

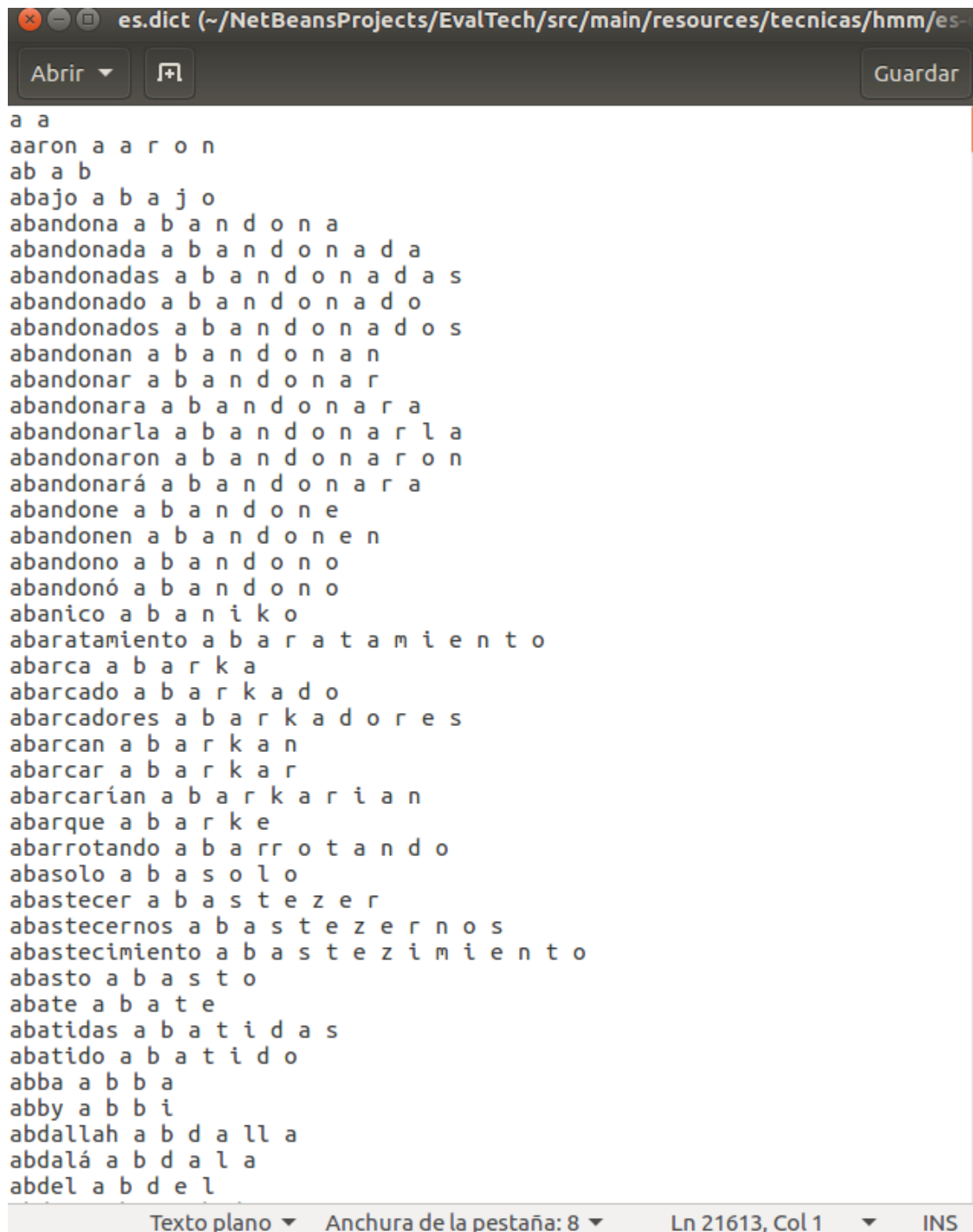
Tabla 28 Glosario de Términos

Término	Descripción
API	Interface de Programación de Aplicaciones
ASR	Reconocimiento Automático del Habla
CMU	Carnegie Mellon University
CRC	Clase Responsabilidad Colaborador
DTW	Dynamic Time Warping
FDP	Funciones de Densidad de Probabilidad
Fonema	Unidad Fonológica Mínima
Framework	Entorno de Trabajo
GPL	General Public Licence
Hardware	Parte física de un computador
HMM	Modelos Ocultos de Markov
HTTP	Protocolo de Transferencia de Hipertexto
IA	Inteligencia Artificial
IDE	Entorno de Desarrollo Integrado
IEEE	Instituto de Ingenieros Eléctricos y Electrónicos
IOS	Sistema operativo móvil de Apple
JDK	Kit de Desarrollo de Java
JS GF	Jspeech Grammar Format
LGLP	Licencia Publica General Reducida
LN	Lenguaje Natural
LPC	Codificación Predictiva Lineal
MS	Mantenlo Sencillo
MSEG	Milisegundos
PLN	Procesamiento del Lenguaje Natural
RAM	Random Access Memory
RNA	Redes Neuronales Artificiales
SDK	Kit de Desarrollo de Software
SO	Sistema Operativo
Software	Parte intangible del computador
Ubuntu	Distribución del SO Linux.
XP	Programación Extrema

Elaborado por: El Autor.

Anexo B. Diccionario y Modelo de Lenguaje de CMU utilizados en el Reconocimiento Basado en Modelos Ocultos de Markov.

Figura 10 Diccionario Fonético de más de 20K palabras.



Elaborado por: El Autor.

Figura 11 Modelo de Lenguaje Estadístico.

```

es-20k.lm (~/.NetBeansProjects/Ada/src/ada/idioma/es-es) - gedit
Abrir ▾
*Documento sin título 1

\data\
ngram 1=23500
ngram 2=1463799
ngram 3=1109989

\1-grams:
-2.073001      </s>
-99          <s> -1.427248
-2.259718      a -1.224971
-4.164376      aaron -0.4051057
-3.869848      ab -0.6230895
-3.397116      abajo -0.6270864
-4.393895      abandona -0.6053905
-4.354497      abandonada -0.6307656
-4.91154       abandonadas -0.3971842
-4.081063      abandonado -0.6111054
-4.709945      abandonados -0.5459952
-4.508602      abandonan -0.603159
-4.020785      abandonar -0.7337242
-4.959351      abandonara -0.3819644
-4.976534      abandonarla -0.3738776
-4.760175      abandonaron -0.3871547
-5.013084      abandonará -0.3263336
-4.782016      abandone -0.4081999
-4.91154       abandonen -0.4210792
-4.552521      abandono -0.41703
-4.334062      abandonó -0.5317633
-5.231411      abanico -0.3720719
-6.218355      abaratamiento
-5.013084      abarca -0.2143985
-6.218355      abarcado
-6.218355      abarcadores
-5.172546      abarcan -0.4215826
-5.172546      abarcar -0.2585831
-6.063494      abarcarían -0.161455
-5.965824      abarque -0.161455
-5.691341      abarrotoando -0.2045087
-6.063494      abasolo -0.161455
-5.013084      abastecer -0.3190961
-5.608813      abastecerse -0.2665800

```

Elaborado por: El Autor.

Anexo C. Diccionario y gramática propios utilizados en el Reconocimiento Basado en el Conocimiento.

Figura 12 Diccionario Fonético propio.

8227.dict (~/.NetBeansProjects/EvalTech/src/main/resources/tecni		
Abrir ▾ [🔍]		
a	a	
ahora	a o r a	
ala	a l a	
alas	a l a s	
aletear	a l e t e a r	
aprendió	a p r e n d i o	
aquel	a k e l	
así	a s i	
atacó	a t a k o	
ataque	a t a k e	
azul	a z u l	
bandada	b a n d a d a	
batió	b a t i o	
blanca	b l a n k a	
cada	k a d a	
color	k o l o r	
como	k o m o	
con	k o n	
contento	k o n t e n t o	
cuando	k u a n d o	
círculos	z i r k u l o s	
de	d e	
del	d e l	
demás	d e m a s	
derecho	d e r e c h o	
dos	d o s	
día	d i a	
el	e l	
empujaba	e m p u j a b a	
en	e n	
enorme	e n o r m e	
era	e r a	
especial	e s p e z i a l	
está	e s t a	
evitando	e b i t a n d o	
feliz	f e l i z	
fuerzas	f u e r z a s	
giro	j i r o	
grande	g r a n d e	
hizo	i z o	
iban	i b a n	
izquierdo	i z k i e r d o	

Elaborado por: El Autor.

Figura 13 Gramática propia.

```
#JSGF V1.0;

/**
 * JSGF Grammar
 */

grammar grammar;

//public <oracion> = <simple>;

public <simple> = <sujeito> <predicado> | <predicado> <sujeito>;

public <sujeito> = ([<adverbio>] [<articulo> | <articulo> <adjetivo>] <nucleo> [<articulo> | <adjetivo> | <complementoN>]);

public <predicado> = ([<negacion> | <conjuncion> ] <verbo> <complemento>) | ([<conjuncion>] <complemento> <verbo>);

public <negacion> = no ;

public <complemento> = (<complementoD> [<complementoI> | <complementoN>]);

public <complementoD> = <complementoN> | <articulo> <nucleo> | <nucleo>;

public <complementoN> = <preposicion> <articulo> <nucleo> | (<contraccion> <nucleo> [<adjetivo>]) | (<contraccion> <nucleo> ) |
(<adverbio> <adjetivo>);

public <contraccion> = del | la;

public <complementoI> = les | se;

public <nucleo> = ( ala | alas | ataque | bandada | círculos | color | día | evitando | fuerzas | giro | juntos | lado |
lanzándose | línea | menos | pajarito | pajaritos | picado | tener | vez | volaba | volar | águila );

public <verbo> = ( aletear | aprendió | atacó | batió | empujaba | era | está | hizo | iban | podía | quería | reían | tenía |
tener );

public <articulo> = <determinado> | <indeterminado>;

public <determinado> = (el | la | los | las );

public <indeterminado> = ( un | una ); // uno, una, unos, unas
```

Texto plano Anchura de la pestaña: 8 Ln 18, Col 1 INS

Elaborado por: El Autor.

Anexo D. Entrenar un modelo acústico para CMUSphinx.

El proyecto CMUSphinx cuenta con modelos acústicos altamente entrenados y precisos al momento de reconocer, pero se presentarán ocasiones en las cuales se tenga que entrenar un modelo (CMUSphinx Project, 2009), para esto se debe preparar el modelo de lenguaje y tener los recursos para hacerlo.

Cuando necesita entrenar.

Cuando desea crear un modelo acústico nuevo, o necesita un modelo especializado para la aplicación de un vocabulario pequeño y posee los datos para entrenar:

- 1 hora de grabación para comando y control para un solo hablante.
- 5 horas de grabaciones de 200 hablantes para comando y control para muchos hablantes.
- 10 horas de grabaciones para el dictado de un solo hablante.
- 50 horas de grabaciones de 200 hablantes para dictado de muchos hablantes.
- Y tiene tiempo para entender la estructura fonética del lenguaje.
- Y tienen tiempo para entrenar el modelo y optimizar los parámetros. (Un mes como mínimo).

Cuando no necesita entrenar.

- Necesita mejorar la precisión, en este caso haga la adaptación.
- No se dispone de los suficientes datos, en este caso haga la adaptación.
- No tiene suficiente tiempo.
- No tienen suficientes conocimientos.

Las cantidades de datos antes mencionadas son para entrenar un buen modelo. Si no posee los suficientes datos no podrá entrenar un buen modelo.

Anexo E. Lectura “El pajarito que no quería volar” de Sergio R. Peral.

Aquel pajarito tenía dos preciosas alas de color azul. Pero el pajarito no quería volar porque el ala del lado derecho era más grande que el ala del lado izquierdo. Y cada vez que volaba, solamente podía volar en círculos porque el ala grande empujaba más que la pequeña y los demás pajaritos se reían de él. Poco a poco aprendió que para volar en línea recta tenía que aletear un poco menos con el ala grande. Un día, cuando iban todos juntos en bandada, una enorme águila blanca les atacó lanzándose en picado sobre el pajarito que tenía el ala más grande. El pajarito batió las alas con todas sus fuerzas y, como el ala grande empujaba mucho, hizo un rápido giro evitando así el ataque del águila. Ahora el pajarito está muy contento y feliz de tener un ala tan especial.(Peral, 2016)

Anexo F. Nivel de pronunciamiento.**Tabla 29 Nivel de Pronunciamiento.**

Nro. Palabras	Nivel de pronunciamiento
Mayor o igual a 125 palabras	Excelente
Mayor o igual a 120 palabras	Muy Bueno
Mayor o igual a 115 palabras	Bueno
Mayor o igual a 110 palabras	Regular
Mayor o igual a 105 palabras	Insuficiente

Elaborado por: El Autor.

Anexo G. Niños de la muestra.**Tabla 30 Niños de la muestra y sus códigos respectivos.**

Niño	Código
Niño 1	CERL
Niño 2	DRJR
Niño 3	EKMM
Niño 4	ENOG
Niño 5	ESGV
Niño 6	JCRR
Niño 7	JSVC
Niño 8	KGRV
Niño 9	LAMR
Niño 10	MCRR

Elaborado por: El Autor.

Anexo H. Metodología de desarrollo.

1. Requerimientos

a. Requerimientos Funcionales

En la tabla 31 se describen los requerimientos funcionales del prototipo de aplicación.

Tabla 31 Requerimiento funcional DANNA

Código	Requerimiento
RF001	Mostrar la ventana principal con un label, un cuadro de texto, un botón para buscar el archivo de audio a reconocer, un botón para salir y cuatro pestañas para el reconocimiento y la presentación de resultados: • Reconocimiento Basado en el Conocimiento: Pestaña para realizar el reconocimiento basado en el conocimiento, con un área de texto para la presentación de información del proceso de reconocimiento y un botón de ejecución. • Reconocimiento Basado en Modelos Ocultos de Markov: Pestaña para realizar el reconocimiento basado en modelos ocultos de Markov, con un área de texto para la presentación de información del proceso de reconocimiento y un botón de ejecución. • Reconocimiento Basado en Redes Neuronales Artificiales: Pestaña para realizar el reconocimiento basado en redes neuronales artificiales, con un área de texto para la presentación de información del proceso de reconocimiento y un botón de ejecución.

- **Resultados:** Pestaña para la presentación de resultados, con una lista desplegable y un botón para la presentación de los resultados del proceso de evaluación y un panel para los gráficos.

RF002	Al hacer clic en el botón buscar de la ventana principal se desplegará una ventana para seleccionar el archivo de audio a reconocer, solo debe permitir seleccionar archivos en formato wav, y debe devolver la ruta completa del archivo.
RF003	Al iniciar el prototipo de aplicación se debe establecer la conexión con la base de datos.
RF004	Al hacer clic en el botón reconocer de la pestaña reconocimiento basado en el conocimiento, se ejecuta el reconocimiento del archivo seleccionado, la información se muestra en el área de texto correspondiente y el resultado es almacenado en la tabla correspondiente de la base de datos.
RF005	Al hacer clic en el botón reconocer de la pestaña reconocimiento basado en modelos ocultos de Markov, se ejecuta el reconocimiento del archivo seleccionado, la información se muestra en el área de texto correspondiente y el resultado es almacenado en la tabla correspondiente de la base de datos.
RF006	Al hacer clic en el botón reconocer de la pestaña reconocimiento basado en redes neuronales artificiales, se ejecuta el reconocimiento del archivo seleccionado, la información se muestra en el área de texto correspondiente y el resultado es almacenado en la tabla correspondiente

de la base de datos.

RF007 Al hacer clic en el botón mostrar resultados de la pestaña resultados, se muestra los resultados de acuerdo a la selección de la lista desplegable, la información es consultada a la base de datos y se muestra en el panel correspondiente.

RF008 Al hacer clic en el botón salir el prototipo de aplicación debe finalizar la conexión a la base de datos y cerrarse.

Elaborado por: El Autor.

2. Historias de usuario

Las historias de usuario del prototipo de aplicación se encuentran en la tabla 32.

Tabla 32 Historias de usuario

Identificador de la historia	Rol	Característica	Finalidad	Número de escenario
HU001	Usuario de la aplicación	Necesito seleccionar un archivo en formato wav a ser reconocido.	Con la finalidad de reconocerlo utilizando las diferentes técnicas de procesamiento del lenguaje natural.	1

HU002	Usuario	de	Necesito	Con	la	2
	la		reconocer	un	finalidad	de
	aplicación.		archivo de audio	evaluar	las	
			utilizando	la	técnicas	de
			técnica	de	procesamiento	
			Reconocimiento	del	lenguaje	
			basado en el	natural		
			conocimiento.	seleccionadas.		
HU003	Usuario	de	Necesito	Con	la	3
	la		reconocer	un	finalidad	de
	aplicación.		archivo de audio	evaluar	las	
			utilizando	la	técnicas	de
			técnica	de	procesamiento	
			Reconocimiento	del	lenguaje	
			basado en	natural		
			modelos ocultos	seleccionadas.		
			de Markov.			
HU004	Usuario	de	Necesito	Con	la	4
	la		reconocer	un	finalidad	de
	aplicación.		archivo de audio	evaluar	las	
			utilizando	la	técnicas	de
			técnica	de	procesamiento	
			Reconocimiento	del	lenguaje	
			basado en redes	natural		

neuronales seleccionadas.
artificiales.

HU005 Usuario de Necesito guardar Con la 5
la los resultados en finalidad de ver
aplicación. una base de datos los resultados
para visualizarlos de evaluación.
en lo posterior.

HU006 Usuario de Necesito graficar Con la 6
la los resultados de finalidad de
aplicación. la evaluación analizar los
guardados en la datos
base de datos. almacenados
en la base.

Elaborado por: El Autor.

3. Tarjetas Clase Responsabilidad Colaborador

Aquí se describen las tarjetas clase responsabilidad colaborador o CRC.

Tabla 33 Tarjeta CRC de la clase Evaltech

EvalTech	
Lanzar ventana principal	• Principal

Elaborado por: El Autor.

Tabla 34 Tarjeta CRC de la clase principal

Principal	
Lanzar ventana de selección de archivo	• ReconocedorHMM

Reconocer archivo de audio

- ReconocedorRNA

Guardar información en la base de datos

- Resumen

Mostrar resultados de evaluación de la base de datos.

Elaborado por: El Autor.

Tabla 35 Tarjeta CRC de la clase ReconocedorHMM

ReconcedorHMM			
Reconocer	audio	mediante	• Sphinx4-core
reconocimiento	basado	en el	
conocimiento.			
Reconocer	audio	mediante	
reconocimiento	basado	en modelos	
ocultos de Markov.			

Elaborado por: El Autor.

Tabla 36 Tarjeta CRC de la clase ReconocedorRNA

ReconcedorRNA			
Reconocer	audio	mediante	• Google Cloud Speech
reconocimiento	basado	en redes	• Google Cloud Storage
neuronales artificiales.			• Google Cloud Auth
Autenticarse con los servidores de			
google cloud.			
Subir archivos a google cloud.			

Elaborado por: El Autor.

Tabla 37 Tarjeta CRC de la clase Resumen

Resumen	
Graficar los resultados de evaluación de	• JFreeChart
las técnicas.	

Elaborado por: El Autor.

4. Pruebas

Conforme se desarrolló el proyecto se fue probando, primero se realizó pruebas unitarias por cada módulo o característica programada y luego se realizó las respectivas pruebas de regresión para comprobar que los demás módulos seguían funcionando. Una vez probado se prosiguió con el desarrollo hasta completar el prototipo de aplicación.



UIDE

Anexo I. Solicitud entregada a la Escuela de Educación General Básica José

Antonio Campos del cantón Zapotillo.

Zapotillo, 17 de Febrero del 2017

Lic.

Washington Rodríguez Ramírez

DIRECTOR DE LA UNIDAD EDUCATIVA JOSÉ ANTONIO CAMPOS

Ciudad.-

De mi especial consideración:

Por medio del presente, me dirijo a su autoridad para saludar, desear éxitos en sus funciones y a la vez exponer y solicitar lo siguiente:

Como egresado de la carrera de Informática y Multimedia de la Universidad Internacional del Ecuador Sede Loja, me permito dirigirme a usted en mi calidad de Tesista, para poner en su conocimiento la construcción de mi proyecto de tesis "Evaluación de Técnicas del Procesamiento del Lenguaje Natural aplicadas en la pronunciación de niños de educación general básica de la escuela José Antonio Campos del Cantón Zapotillo, Provincia de Loja", mismo que he creído conveniente desarrollarlo es la Escuelita bajo su dirección como un aporte al establecimiento educativo en el cual aprendí mis primeras letras y a mi Cantón en general.

Con este antecedente, me permito dirigir a usted, con la finalidad de solicitar su apoyo y apertura para el desarrollo del mismo, que incluirá entre otras actividades la toma de muestras de lectura de 10 niños y que al establecimiento educativo a su cargo le permitirá medir el nivel de progreso de destrezas lingüísticas de los estudiantes al final del proceso.

En espera de una respuesta favorable, por el adelante y progreso de nuestros pueblos, desde ya le anticipo mis sinceros agradecimientos.

Atentamente,

Egdo. Jefferson Coronel Requena
TESISTA UIDE-LOJA

Recibido: 17-02-2017

DIRECTOR FEGBJAC2.
AUTORIZADO.