Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Estudios de Postgrado
Maestría de Tecnologías de la Información y Comunicación

PROTOTIPO PARA MEJORAR EL TONO Y LA DURACIÓN DE UNIDADES FONÉTICAS DE LA VOZ ARTIFICIAL, PARA EL IDIOMA ESPAÑOL EN GUATEMALA

Ing. Angel Eduardo Sic Morales
Asesorado por el Inga. Ms. María Elizabeth Aldana Díaz

Guatemala, septiembre de 2017

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

PROTOTIPO PARA MEJORAR EL TONO Y LA DURACIÓN DE UNIDADES FONÉTICAS DE LA VOZ ARTIFICIAL, PARA EL IDIOMA ESPAÑOL EN GUATEMALA

TRABAJO DE GRADUACIÓN

PRESENTADO A JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

ANGEL EDUARDO SIC MORALES
ASESORADO POR EL ING. MS. MARÍA ELIZABETH ALDANA DÍAZ

AL CONFERÍRSELE EL TÍTULO DE

MAESTRO EN TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN

GUATEMALA, SEPTIEMBRE DE 2017

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Pedro Antonio Aguilar Polanco
VOCAL I	Ing. Ángel Roberto Sic García
VOCAL II	Ing. Pablo Christian de León Rodríguez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Jurgen Andoni Ramírez Ramírez
VOCAL V	Br. Oscar Humberto Galicia Núñez
SECRETARIA	Inga. Lesbia Magalí Herrera López

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Pedro Antonio Aguilar Polanco
EXAMINADOR	Ing. Murphy Paiz Recinos
EXAMINADORA	Ing. Marlon Antonio Pérez Türk
EXAMINADORA	Ing. Luis Fernando Espino Barrios
SECRETARIA	Inga. Lesbia Magalí Herrera López

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

PROTOTIPO PARA MEJORAR EL TONO Y LA DURACIÓN DE UNIDADES FONÉTICAS DE LA VOZ ARTIFICIAL, PARA EL IDIOMA ESPAÑOL EN GUATEMALA

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha junio de 2016.

Angel Eduardo Sic Morales

FACULTAD DE
INGENIERÍA - USAC
ESCUELA DE
ESTUDIOS DE POSTGRADO

Escuela de Estudios de Postgrado
Facultad de Ingeniería
Teléfono 2418-9142 / 24188000 Ext. 86226

APT-2017-014

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Postgrado, al Trabajo de Graduación de la Maestría en Artes en Tecnologías de la Información y la Comunicación titulado: ***"PROTOTIPO PARA MEJORAR EL TONO Y LA DURACIÓN DE UNIDADES FONÉTICAS DE LA VOZ ARTIFICIAL, PARA EL IDIOMA ESPAÑOL DE GUATEMALA"*** presentado por el ingeniero en Ciencias y Sistemas **Angel Eduardo Sic Morales**, procede a la autorización para la impresión del mismo.

IMPRÍMASE.

"Id y Enseñad a Todos"

Ing. Pedro Antonio Aguilar Polanco
Decano
Facultad de Ingeniería
Universidad de San Carlos de Guatemala

UNIVERSIDAD DE SAN CARLOS
DECANO
FACULTAD DE INGENIERIA
DE GUATEMALA

Guatemala, agosto de 2017.

Cc: archivo/la

FACULTAD DE
INGENIERÍA - USAC
ESCUELA DE
ESTUDIOS DE POSTGRADO

Escuela de Estudios de Postgrado
Facultad de Ingeniería
Teléfono 2418-9142 / 24188000 Ext. 86226

APT-2017-014

El Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen y dar el visto bueno del revisor y la aprobación del área de Lingüística del Trabajo de Graduación titulado ***"PROTOTIPO PARA MEJORAR EL TONO Y LA DURACIÓN DE UNIDADES FONÉTICAS DE LA VOZ ARTIFICIAL, PARA EL IDIOMA ESPAÑOL DE GUATEMALA"*** presentado por el Ingeniero en Ciencias y Sistemas **Angel Eduardo Sic Morales,** correspondiente al programa de Maestría en Artes en Tecnología de la Información y la Comunicación; apruebo y autorizo el mismo.

Atentamente,

"Id y Enseñad a Todos"

UNIVERSIDAD DE SAN CARLOS
ESCUELA DE POSTGRADO
FACULTAD DE INGENIERÍA
DE GUATEMALA

MSc. Ing. Murphy Olympo Paiz Recinos
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería
Universidad de San Carlos de Guatemala

Guatemala, agosto de 2017.

Cc: archivo/la

FACULTAD DE
INGENIERÍA - USAC
ESCUELA DE
ESTUDIOS DE POSTGRADO

Escuela de Estudios de Postgrado
Facultad de Ingeniería
Teléfono 2418-9142 / 24188000 Ext. 86226

APT-2017-014

Como Coordinador de la Maestría en Artes en Tecnologías de la Información y la Comunicación del Trabajo de Graduación titulado ***"PROTOTIPO PARA MEJORAR EL TONO Y LA DURACIÓN DE UNIDADES FONÉTICAS DE LA VOZ ARTIFICIAL, PARA EL IDIOMA ESPAÑOL DE GUATEMALA"*** presentado por el Ingeniero en Ciencias y Sistemas **Angel Eduardo Sic Morales,** apruebo y recomiendo la autorización del mismo.

Atentamente,

"Id y Enseñad a Todos"

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA
COORDINACIÓN TRANSFERENCIA TECNOLÓGICA
ESCUELA DE POSTGRADO

MSc. Marlon Antonio Pérez Türk
Coordinador de Maestría
Escuela de Estudios de Postgrado
Facultad de Ingeniería
Universidad de San Carlos de Guatemala

Guatemala, agosto de 2017.

Cc: archivo/la

ACTO QUE DEDICO A:

Dios	Por brindarme la vida y las fuerzas, para seguir adelante y ser mi apoyo incondicional.
Mi madre	Midiam Damaris Castillo DEP, por todo el amor que me brindó, por su comprensión y por los valores que me enseñó, y que son el resultado de lo que ahora soy.
Mi padre	Angel Roberto Sic García, por todo el apoyo y amor que me brinda todos los días, por ser un gran soporte en mi vida personal como en la profesional, porque es una inspiración en mi vida, las metas que uno se proponga se pueden lograr.
Mis hermanos	Byron Sic, José Sic, Roberto Carlos Sic, por ser una parte muy importante en mi ser, por todo el apoyo y cariño que me brindan en todo momento.
Mi novia	Sharon Abigail Aragón, por todo su amor, paciencia, apoyo y comprensión a lo largo de mis estudios, por ser una parte muy importante en mi vida.
Mi familia	Gracias por el apoyo que me brindan, y por estar presentes en los momentos más difíciles y felices de mi vida.

Mis amigos

Por todas las ocasiones que me brindaron su apoyo durante los estudios de maestría, a mis amigos de la licenciatura y amigos de la infancia, por su apoyo y tantas risas hasta en los momentos difíciles.

AGRADECIMIENTOS A:

Dios	Porque con su apoyo y amor se pueden lograr las metas y sueños.
Mi asesora	Inga. María Aldana, por su tiempo, dedicación y gran apoyo que permitió culminar mi trabajo de graduación, le agradezco de todo corazón.
Mis catedráticos	Por compartir sus conocimientos y experiencia en todos los cursos de la maestría, en especial a la Inga. María Elizabeth Aldana Díaz y la Inga. Mildred Caballeros, por su ayuda brindada en la revisión y corrección del trabajo final de graduación.
Universidad de San Carlos de Guatemala	Por ser mi alma mater y permitirme cumplir otra meta educativa, asimismo agradezco a la Facultad de Ingeniería y Escuela de Postgrado, por la tarea que realizan para brindar mejores procesos y servicios educativos.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

FIGURAS

TABLAS

LISTA DE SÍMBOLOS

Símbolo	Significado
HMM	*Hidden Markov Models*
HTTP	*Hyper Text Transfer Protocol*
JSON	*Java Script Object Notation*
PSOLA	*Pitch Synchronous OverLap- Add*

GLOSARIO

Algoritmo sentimental	Tratamiento computacional de opiniones, sentimientos y fenómenos subjetivos del texto.
Análisis cualitativo	Este se centra en la calidad y diferencia de las características.
Applet	Se denomina así a un componente de una aplicación que se ejecuta en el contexto de otro programa, por ejemplo en un navegador web.
Arquitectura	Conjunto de patrones que proporcionan un marco referencial para guiar la construcción de un programa.
Autómata	Máquina que imita figuras y movimientos.
Capa	Modelo de desarrollo de software que cumple el objetivo de separar los componentes de software.
Conector de base de datos	Estándar de acceso a la base de datos.
Corpus de voz	Es la colección que contiene uno o más elementos de audio.
Corpus lingüístico	Es un conjunto amplio y estructurado de ejemplos reales de uso de la lengua.

Estructura de datos	Medio para manejar grandes cantidades de datos de manera eficiente.
Etiquetas	Son representaciones para determinar componentes en diferentes lenguajes o estructuras de datos.
Fonética	Parte de la lingüística que estudia los sonidos que se generan de un dialecto o lengua.
Java	Lenguaje multiplataforma y multipropósito orientado a objetos.
Librería	Conjunto de implementaciones funcionales, codificadas en un lenguaje de programación.
No estructurado	Son repositorios en los que se consolida y ordena información.
Nosql	Amplia clase de sistemas de gestión de base de datos no transaccionales.
Programación	Proceso de diseñar, codificar, depurar y mantener código fuente de programas de computadoras.
Prototipo	Representación limitada de un producto en situaciones reales o para explorar su uso.
Psola	Técnica de procesamiento de señal utilizada para la síntesis de voz.
Semántica	Parte de la lingüística que estudia los sonidos que se generan de un dialecto o lengua.

Servicio web	Se denomina así a la tecnología que utiliza un conjunto de protocolos y estándares para realizar intercambio de datos entre aplicaciones de software.
Sintetizador de voz	Es la producción artificial del habla, a partir de un texto escrito.
Tupla	Es un listado ordenado de elementos.
Variables lingüísticas	Esta se refiere a las variables de la lengua, las cuales posee rangos léxicos, fonéticos y morfosintácticos particulares.
Variable diacrónica o histórica	Son variedades que presentan la lengua a lo largo de la historia y su evolución.
Variable estilística o situacional	Es aquella que determina la diferencia en la lengua, causadas por el estilo o manera de expresarse.
Variable geográfica o diatópica	Son aquellas que relacionan al hablante con su origen territorial.
Variable social o diastráticas	Se relaciona con la distribución y estratificación de los hablantes, diferencias de expresión en los individuos pertenecientes a las distintas clases sociales.
Variación de tono	Es la altura de tono de voz que pronuncia un enunciado.
Wav	Formato de audio que admite diversas señales como mono y estéreo.

RESUMEN

El idioma español es uno de los idiomas más hablados alrededor del mundo, este se habla desde México hasta la Patagonia en América y España en Europa. Esto muestra que la comunicación juega un papel muy importante en las relaciones humanas y resalta como una lengua puede estar presente en más de 450 millones de habitantes. Aunque el idioma sea el mismo cada región cuenta con distintas características como las expresiones, pronunciación de las palabras, el vocabulario, modismos, entre otros. En Guatemala, el acento es distinto y se utilizan palabras como patojo, para referirse a un niño; chilero para identificar algo bonito, chispudo para referirse a alguien inteligente.

Con la intención de motivar el conocimiento y apoyar el aprendizaje del idioma español que se habla en Guatemala, se desarrolló en la presente investigación un prototipo sintetizador de voz de idioma español para Guatemala, que se enfoca en resaltar mediante la voz artificial las características de pronunciación y el acento utilizado en la capital de Guatemala, técnicamente se resaltan las características del tono y la duración de las unidades fonéticas de la voz artificial. Para esto se desarrollaron 3 técnicas dentro del prototipo: la primera es la generación de voz de síntesis por concatenación PSOLA; la segunda es la síntesis por concatenación PSOLA en combinación de un algoritmo semántico, y la tercera técnica el algoritmo de autómatas adaptativos por sílabas.

El prototipo sintetizador de voz se enfoca principalmente en la aplicación de un algoritmo semántico a un algoritmo de síntesis por concatenación PSOLA, con la intención de brindar emoción al audio que se quiere producir. El

algoritmo semántico implementa un algoritmo sentimental que se consume por un servicio web que proporciona *Meaning cloud* en su versión gratuita, este permite reconocer 3 sentimientos de un texto, dentro de estos se resalta la tristeza, la felicidad y el estado neutro o normal. Luego de reconocer los sentimientos del texto de entrada, estos se aplican al algoritmo de clasificación y evaluación de texto para identificar segmentos de texto y buscar los audios relacionados, se aplica una optimización a cada audio; por último, se concatenan todos los audios obtenidos por cada bloque de texto clasificado y se presenta un resultado de voz artificial en formato WAV.

Adicionalmente, se realizó un experimento que resalta la influencia de las características de tono y unidades fonéticas de la voz sobre los audios generados. El experimento consistió en generar voz artificial, mediante cada técnica del prototipo, luego se presentó un cuestionario a un grupo de personas para que resaltarán las características del audio artificial. De dicho experimento se concluyó que la aplicación de un algoritmo semántico sobre un algoritmo de síntesis por concatenación PSOLA resalta la calidad del tono y la influencia de naturalidad en las unidades fonéticas de la voz artificial, además de mostrar algún tipo de emoción en los audios generados por el prototipo.

PLANTEAMIENTO DEL PROBLEMA Y FORMULACIÓN DE PREGUNTAS ORIENTADORAS

Las herramientas de síntesis de voz son un aporte importante para la comunicación entre hombre y computadora. El sintetizador de voz tiene como propósito emular la voz humana, por medios artificiales a partir de una entrada de texto generando como resultado una cadena oral, esto sin intervención directa del hablante. La aplicación de síntesis de voz se utiliza en varios campos como la lingüística, validación de teorías, apoyo a personas con discapacidades vocal y auditiva, información telefónica y la música (De la Vega y Camacho, 2007). A lo largo del tiempo se han propuesto diferentes técnicas y modelos para la generación de voz artificial, a partir de la sintetización articulatoria, por formantes, por concatenación, por HMM (hidden markov models), por predicción lineal (Agüero, 2012) otras técnicas se apoyan en modelos del lenguaje como difonemas, semisílabas, PSOLA (Pitch Synchronous Overlap Add) y Vocoders (Roca, 1990).

Estudios anteriores comparten el desarrollo de las características del lenguaje, los cuales permite el análisis y representación de elementos de la expresión oral, tales como el acento, la entonación y los tonos, a pesar de estos avances aún se tiene limitante en la generación del tono y la duración de las unidades de voz para interpretar un sistema de texto a voz. Dichos estudios proponen mejorar la naturalidad y expresividad de la conversión de texto a voz, el entendimiento de la oración, la acentuación de las palabras, así como la realización de pausas entre palabras o signos de puntuación, variaciones de ritmo, entre otros aspectos que podrían mejorar el entendimiento y pronunciación de un contexto (Agüero, 2012). Con estos estudios se ha logrado

la implementación de múltiples sintetizadores de voz que aunque logran resultados aceptables aún tienen deficiencias que no permiten que la voz artificial generada muestre características como emotividad, entonación o que se reconozca como una voz más natural (de la Vega y Camacho, 2007).

Por lo tanto, el presente trabajo de graduación propone la siguiente pregunta principal:

¿Se puede desarrollar un algoritmo por concatenación PSOLA que mejore el tono y la duración de la voz en idioma español para Guatemala?

Adicionalmente, se plantearon las siguientes preguntas auxiliares:

1. ¿Qué tipo de variables lingüísticas se pueden utilizar para mejorar el tono y la duración de la voz generada por un sintetizador de voz?
2. ¿La implementación de algoritmos PSOLA en combinación con un algoritmo semántico puede apoyar a mejorar el entendimiento de la voz artificial que produce el sintetizador de voz?
3. ¿Puede la aplicación de autómatas adaptativos mejorar el tono y la duración de la fonética del sintetizador de voz?

OBJETIVOS

General

Implementar un algoritmo de síntesis por concatenación PSOLA en combinación con un algoritmo semántico, para mejorar el tono y la duración de las unidades fonéticas de síntesis de voz en idioma español para Guatemala.

Específicos

1. Determinar el tipo de variables lingüísticas que se pueden utilizar, para mejorar la calidad del tono y la duración de voz generada por un sintetizador de voz.

2. Evaluar el valor que un algoritmo semántico aporta en el algoritmo PSOLA al entendimiento de la voz artificial producida por un sintetizador de voz.

3. Comparar si los autómatas adaptativos pueden mejorar el tono y la duración de la fonética del sintetizador de voz.

RESUMEN DE MARCO METODOLÓGICO

El presente trabajo de graduación muestra una investigación experimental y cualitativa, que como objetivo principal generó un prototipo de software que permite el ingreso de texto plano en idioma español, genera una voz artificial como resultado.

Para la investigación cualitativa, se utilizó un estudio de caso, la cual apoyó en la búsqueda de las variables lingüísticas que aportan mayor valor a las características de la voz artificial, estas son el tono y la duración de las unidades fonéticas de la voz.

Para alcanzar los objetivos descritos, se desarrolló un prototipo que cuenta con las siguientes funcionalidades:

- Evaluación y clasificación de texto
- Generación de audio artificial
- Optimización del tono y duración de unidades fonéticas
- Comparación de resultados.

Por otra parte, para llevar a cabo la investigación experimental, se realizó una investigación cuantitativa donde se analizaron los resultados de los experimentos, parte de una técnica de recolección de datos, a través de encuestas.

El método utilizado para realizar la presente investigación consta de seis fases, en donde tres fases se enfocan en el desarrollo del prototipo, una de experimentación, otra de recolección y evaluación de resultados, y por último, una fase de redacción de resultados.

Dichas fases son:

- **Fase 1. Investigación documental**: se realizó la búsqueda y recolección de información sobre los tipos de sintetizador de voz y su funcionamiento en general. Además, se resaltaron las características que apoyan al incremento de la calidad de la voz artificial, como las variables lingüísticas.
- **Fase 2. Diseño de prototipo**: en esta fase se realizó un plan para estructurar un prototipo funcional que permite soportar los objetivos principales. Diseño de corpus de voz, evaluación y clasificación de texto de entrada, implementación de algoritmo semántico y de estados del autómata adaptativo.
- **Fase 3. Desarrollo de prototipo**: se realizó el desarrollo de código fuente del prototipo que contiene las librerías que convierten el texto de entrada a estructuras de datos. Uso de Java y MongoDB como principales herramientas tecnológicas. También se desarrollaron 3 técnicas para la generación de voz artificial, las cuales son: técnica de síntesis por concatenación PSOLA, técnica de síntesis por concatenación PSOLA en combinación de un algoritmo semántico y técnica de autómatas adaptativos.
- **Fase 4. Experimentación**: en esta fase se presentaron audios artificiales generados por el prototipo, luego se puso a prueba una comparación entre ellos y a base de encuestas, se determinó qué características de la voz apoyan a mejorar la naturalidad de la voz y cuáles no lo hacen.
- **Fase 5. Recolección y evaluación de resultados**: en esta fase se realizaron las encuestas mediante formularios de google. Luego de recibir las encuestas terminadas, se procedió a realizar el conteo de los datos, utiliza estadística descriptiva con análisis de frecuencias. Por otra parte, se evaluaron los resultados de la investigación cualitativa sobre un

estudio de caso que resalta la influencia de las variables lingüísticas sobre las características de tono y duración de unidades fonéticas.

- **Fase 6. Redacción de informe final**: como conclusión de los experimentos realizados, se redactó el presente informe final, el cual muestra las bases utilizadas que apoyaron al prototipo desarrollado, el aporte de aplicar un algoritmo semántico a la síntesis por concatenación PSOLA y las conclusiones de comparar las técnicas de generación de voz artificial del prototipo.

Para la recolección de información, se utilizaron artículos científicos, tesis de maestrías y doctorados, documentaciones de software y datos de encuestas realizadas. Dichas encuestas se realizaron con el tipo de muestreo no probabilístico, donde se desconoce la muestra, dado que es un procedimiento informal y más natural, este es casual por tener mayor acceso a los encuestados. Se realizaron 5 encuestas: la primera con el objetivo de recolectar información sobre un estudio de caso relacionado a las variables lingüísticas; la segunda, tercera y cuarta encuesta se utilizaron para obtener información sobre las tres técnicas de generación de voz individualmente, y por último, la quinta encuesta para obtener información sobre las tres técnicas en conjunto.

INTRODUCCIÓN

La comunicación es un proceso utilizado por el ser humano para transmitir información, el habla es el principal medio para llevar a cabo dicha interacción. Por lo cual con el pasar de los años y el incremento de la tecnología los científicos se han interesado por la comunicación hombre-máquina para facilitar la realización de las tareas y lograr una comunicación más directa. Pero a pesar de los adelantos que se han realizado con respecto a la generación de voz artificial, esta aún tiene deficiencias de calidad en lo que a tono y duración de unidades fonéticas se refiere.

Para mejorar la calidad de la voz que se genera por medios digitales se enfocaron en crear una solución que apoya el incremento de la calidad de la voz artificial en términos del tono y duración de las unidades fonéticas, dicho prototipo utiliza dos algoritmos que en combinación presentan una mejora en los aspectos anteriormente descritos, dentro de los algoritmos se puede mencionar el algoritmo de síntesis de voz por concatenación PSOLA y un algoritmo semántico que permite aportar mejoras al entendimiento del resultado final. Como parte del experimento de la presente investigación, se realizó una comparación del resultado del prototipo y el resultado de voz de una implementación realizada con autómatas adaptativos que comprueban cuál de los resultados cumple con mejorar el tono y la duración de unidades fonéticas de la calidad de la voz y cuál de los dos resultados presenta una voz más natural.

Como parte de la investigación se desarrollaron los siguientes capítulos:

En el capítulo uno, se presentan los antecedentes de estudios pasados que se han realizado con el pasar de los años, estos resaltan los problemas y las diferentes soluciones que se han propuesto, relacionadas a la generación de voz artificial.

En el capítulo dos, se describe la importancia y justificación para realizar la presente investigación y los medios utilizados para la experimentación.

En el capítulo tres, se explica el alcance que tiene la presente solución, alcances investigativos, técnicos y resultados que se generan mediante el prototipo propuesto.

En el capítulo cuatro, se incluyen todas las definiciones que apoyan a la investigación, brinda una visión general para comprender diferentes aspectos que se relacionan en el desarrollo del prototipo.

En el capítulo cinco, se explica la metodología que se llevó a cabo para la realización de la presente investigación, así como las variables e indicadores que se utilizan para el desarrollo del mismo. Se propuso el diseño del prototipo el cual consta de la presentación y explicación del esquema general de la arquitectura que se implementó, así como los componentes que interactúan para cumplir los objetivos propuestos. Se muestra el proceso de experimentación, estos son la evaluación y clasificación de texto, la generación de audio artificial y la optimización de los aspectos que mejoran la calidad de la voz. Por último, dentro de este capítulo se resaltan los métodos para la recolección y análisis de datos.

En el capítulo seis, se presentan los resultados obtenidos del análisis cuantitativo y cualitativo, con interpretaciones gráficas para mejorar el entendimiento de los mismos.

En el capítulo siete, se realiza una discusión de resultados para validar qué tan efectivo fue el experimento que se realizó y se presenta en qué medida fueron alcanzados los objetivos propuestos de la investigación.

1. ANTECEDENTES

La voz es una de las principales herramientas de la comunicación, por lo cual es de gran importancia el desarrollarla para lograr una interacción clara y completa con los demás individuos. Por esta razón, ha surgido el interés de ampliar la comunicación, cómo convertir el texto a voz artificial con diferentes técnicas tecnológicas, como lo propone el autor Alexandre Trilla (2009) en su artículo: "*Natural Language Processing techniques in Text-To-Speech synthesis and Automatic Speech Recognition"* en donde, describe técnicas de procesamiento del lenguaje natural para producir voz de un texto, conocido como síntesis de texto a voz.

Para evaluar la utilidad de las técnicas de procesamiento del lenguaje natural se analiza un marco genérico para el procesamiento de síntesis de texto en inglés, y se realiza una revisión de la transcripción fonética correcta del texto de entrada. En donde, se muestra más atención a técnicas de clasificación, resumen y traducción de texto. Por otro lado, se identifica que los estudios basados en reglas de transcripción fonética cuentan con problemas de lentitud y son muy tediosos.

En dicho estudio, se muestra una ciencia asociada al procesamiento del lenguaje humano, NLP (natural language processing, por sus siglas en inglés) que presenta una serie de técnicas y estrategias que se implementan para lograr el resultado requerido. Se explican técnicas como el procesamiento de texto genérico para la síntesis de voz en inglés TTS (Text to Speech, por sus siglas en inglés) las herramientas para obtener una transcripción fonética correcta del texto y las aplicaciones ASR (Automatic Speech Recognition, por

sus siglas en inglés) centradas en el uso correcto de la gramática. En conclusión se determina la importancia que tiene el procesamiento de lenguaje natural de síntesis de texto a voz, que refleja la naturalidad de expresiones de voz producidas por la aplicación de dichas técnicas que mejoran el rendimiento de los módulos de procesamiento y principalmente las de procesamiento del lenguaje natural que se definen, en donde se determina que la calidad de voz puede mejorarse al aplicar técnicas que se centren en la naturalidad del módulo de procesamiento.

Por otro lado, se proporciona un estudio de los autores Claudia V. Correa, Hoover F. Rueda y Henry Arguello (2010) que tiene por nombre: "*Síntesis de Voz por Concatenación de Difonemas para el Español de Colombia*", que determina técnicas de concatenación de difonemas que son aplicables para mejorar la calidad de voz en términos generales, las cuales pueden ser aplicadas a diferentes acentos del lenguaje español. Con el motivo de desarrollar voces sintéticas para diferentes países y regiones en donde el lenguaje es el mismo, por tal motivo se presenta una evaluación de sintetizador de voz que este enfocado al idioma español de Colombia, por medio de la unión de los segmentos acústicos que permite la transición entre dos fonemas subsiguientes conocido como concatenación de difonemas, dado que en un mismo idioma existen variaciones dependiendo del país, y por la falta de una herramienta de síntesis de voz enfocada en la tonalidad y pronunciación de las palabras en las regiones de Colombia. Utiliza un estudio experimental, donde se experimenta con las condiciones actuales de un hecho, en este caso, la calidad de la voz enfocada al idioma español de Colombia. Se obtiene como resultado el desarrollo de un corpus de voz para el español de Colombia, el cual proporciona características específicas que resaltan los sonidos propios y acentos de dicho país. Proporciona una diferenciación notoria con respecto a la calidad de voz.

Asimismo, como se ha hecho notoria la evolución de los sistemas de síntesis de texto a voz, se han obtenido contribuciones en el campo médico para mejorar la calidad de vida de los pacientes con problemas del habla como lo explican los autores Junichi Yamagishi, Christophe Veaux, Simon King y Steve Renals (2012) en su estudio: "*Speech synthesis technologies for individuals with vocal disabilities: Voice banking and reconstruction*", con el motivo de implementar la tecnología para mejorar la calidad de vida de los pacientes con enfermedades degenerativas, tales como enfermedades de neuronas motoras (MND por sus siglas en inglés) y Parkinson. En donde el problema que produce la degeneración del habla no solo en el sentido de la comunicación, sino que también en la expresión vocal del individuo y la identidad social del mismo. Aunque existe tecnología que ayuda a la comunicación de voz de salida (VOCA por sus siglas en inglés), ésta es muy costosa y para la manufactura no es rentable crear síntesis de voz personalizadas. Como conclusión, se determina que el estudio para mejorar la calidad de voz puede tener un beneficio social al apoyar a pacientes con enfermedades degenerativas neurológicas del habla. Para dicho estudio, se utiliza una metodología experimental, con la cual se apoyan de conceptos para desarrollar el proyecto de banco de voz en Edimburgo.

Otros factores que tienen relevancia con respecto al estudio que se realiza en el presente trabajo, es la implementación de una solución que contemple diversos lenguajes, genera un sistema de conversión de texto a voz (CTV) con voz femenina para tener un caso de éxito, si se desea incursionar en mejorar la calidad de la voz femenina, se tiene en cuenta un punto de partida. Como lo proponen los autores Agustín Alonso, Iñaki Sainz, Daniel Erro, Eva Navas e Inma Hernaez (2013) en el artículo científico "*Sistema de conversión texto a voz de código abierto para Lenguas Ibéricas*", con el motivo de desarrollar un CTV que sea de código abierto y que funcione para lenguas Ibéricas, tales como el

catalán, el gallego y el inglés. Se permite que éste pueda ser funcional para voces, tanto femeninas como masculinas. Por la necesidad de contar con un CTV multilingüe, que permita la generación de voz femenina y masculina. Toma en cuenta que la generación de voz femenina tiene mayor grado de dificultad y su frecuencia fundamental es mucho más alta, por el contrario la voz masculina ofrece mejor calidad sonora.

En dicho artículo, se identifica una metodología experimental donde se obtienen diversos sistemas a los cuales se le aplica un motor de síntesis estadístico para mejorar la salida de voz. En conclusión, se presenta un sistema CTV de código abierto que integra la voz femenina para los diferentes lenguajes y un API (Application Programming Interface) de desarrollo para apoyar a otras aplicaciones.

Por último, se deben de tener claras las métricas que se pueden aplicar para reconocer que un sistema de síntesis de texto a voz puede cumplir con altos estándares de calidad, para aplicarlo a un sintetizado de texto a voz para Guatemala. Para lo cual se toma en cuenta un estudio realizado por el autor Simon King (2014) que lleva por título: "*Measuring a decade of progress in Text-to-Speech*", en donde se toman en cuenta la gran cantidad de técnicas que se utilizan para realizar el proceso de conversión de texto a voz, el cual permite tener un abanico de posibilidades para obtener el mayor beneficio depende del ámbito de aplicación. Por tal motivo, se presenta la necesidad de realizar un experimento auditivo a gran escala y comparar el rendimiento de cada uno de los proyectos propuestos desde el 2005. De manera que se identifica una metodología cualitativa donde se enmarcan las características de cada sistema para evaluar su efectividad y métricas que incrementan su valor. Al hablar de métricas se refiere a la naturalidad, inteligibilidad y similitud de voz. En conclusión, se determinó que los sistemas destacan en diferentes ámbitos,

como las habilidades que garantizan la naturaleza del habla, integridad e inteligibilidad, destaca al método HMM (hidden markov models), el cual tiene aplicación en diferentes campos.

Finalmente, se consta que los estudios realizados aportan valor en diferentes aspectos sobre el estudio que se encuentran en el presente trabajo, permite brindar un apoyo con casos pasados.

2. JUSTIFICACIÓN

El presente trabajo tiene el propósito de contribuir en la línea de investigación de la Tecnología de la Información y la Comunicación para apoyo a la Educación, propone mejorar el tono y la duración de la voz artificial para apoyar a la comprensión del lenguaje español en Guatemala.

De acuerdo con estudios anteriores (Correa, Rueda y Arguello, 2010) sobre la síntesis de voz parte de un texto se presentan soluciones aceptables, pero presentan debilidades tecnológicas como el desempeño sobre el espectro del tono de voz en lo que a frecuencia y amplitud se refiere, por otra parte, se tiene deficiencia con el acceso al corpus de voz, el cual al mantener una base de datos muy grande se crea latencia en la conexión con los datos. Dichas deficiencias pueden mejorarse al implementar un algoritmo semántico que permita obtener características vocales que denoten el contenido emocional para mejorar el resultado en el tono de voz, con la implementación de una base de datos NoSQL, que permite facilitar el acceso a los datos no estructurados, por lo cual se considera de valor al mejorar un algoritmo que pueda incrementar la calidad de los resultados obtenidos.

Desde la perspectiva tecnológica es de importancia mejorar la forma en la que se generan los resultados de voz y la implementación de algoritmos que apoyen al entendimiento y regulación de tono de voz, en cuanto innovación se refiere, se presenta la implementación de una base de datos NoSQL, que permita el acceso a los corpus de voz para reducir el tiempo de respuesta.

La educación es una de las principales herramientas con las que cuenta el ser humano (Amar, 2011), para lograr un desarrollo integral y con herramientas como los sintetizadores de voz se permite facilitar el entendimiento del mismo.

Los principales beneficiados serán las personas con conocimiento en el idioma español y las personas con algún tipo de discapacidades del habla y del lenguaje (Fresneda y Mendoza, 2005), permite facilitar la comunicación y el entendimiento del lenguaje español, en lo que a tono y duración de voz se refiere (Jun, 2014). Como resultado de un algoritmo que permita simular la voz natural del ser humano con alta efectividad, se puede apoyar en la creación de aplicaciones educativas, para mejorar la lectura del idioma español, resalta características como la tonalidad, el acento, entre otros factores lingüísticos. Asimismo apoyar a personas con dislexia o con algún desorden de lectura para mejorar su comprensión.

Por último, se considera que la aplicación de dicha mejora en la calidad del tono y la duración de voz del lenguaje español puede apoyar a mejorar la comunicación entre máquina-hombre, permite brindar información de una forma más amigable y facilitar el entablar conversaciones (Fong, Thorpe & Baur, 2003).

.

3. ALCANCES

3.1. Investigativos

Con el desarrollo del prototipo de escritorio de síntesis de voz, se tomaron en cuenta factores de importancia para la investigación, se definen lo siguientes alcances:

- La investigación se realizó y se tomó en cuenta los tipos de variables lingüísticas geográficas, diacrónicas, las sociales o diastráticas y las diafásicas, en donde se evalúa el lugar de donde proviene la lengua, el tiempo, la cultura y los estratos sociales como la edad, sexo, profesión, entre otros, con un enfoque léxico, fonético y morfológico.

- La investigación se enfocó en proponer una mejora en el tono y duración de las unidades fonéticas de la voz, por medio del algoritmo de síntesis, por concatenación PSOLA en combinación con un algoritmo semántico.

- La investigación propone una comparación entre autómatas adaptativos y el algoritmo de síntesis de concatenación PSOLA en combinación con un algoritmo semántico que valide cuál es más entendible de acuerdo a las características de tono y duración de unidades fonéticas.

3.2. Técnicos

Para la lectura del texto de entrada, se utilizó un grupo de librerías desarrolladas en JAVA, tales como: LinkedList, variables de tipo nativo, métodos estáticos, entre otras; las cuales dividen la información. Posterior a

esto se obtendrán difonemas como textos reducidos y se realizará la concatenación de la información por medio del algoritmo de concatenación PSOLA (Pitch Synchronous Overlap-Add, por sus siglas en inglés) y se procederá a buscar dichas unidades en un corpus de voz en español, agiliza la búsqueda al implementar una base de datos NoSql (Not Only Structure Query Language, por sus siglas en inglés).

Para el apoyo al incremento del sentido del texto se aplica un algoritmo semántico sentimental, que brinda un análisis estructurado de opinión y que permite mayor precisión en la clasificación de polaridad de texto y magnifica el sentido emocional. Con dicho algoritmo semántico se identifican los textos y se establecen los rangos de selección de las unidades de voz que concatena para generar un audio resultante, con esto aporta un cambio significativo a las características de tono y duración de las unidades fonéticas, toma en cuenta las variables lingüísticas geográficas, diacrónicas, las sociales o diastráticas y las diafásicas.

Dentro del desarrollo del prototipo se presentan las siguientes funcionalidades:

- Desarrollo de aplicación de escritorio que permita generar un audio artificial como resultado.

- Ingreso de texto por medio de un cuadro de texto o por la carga directa de un documento con extensión txt.

- Generación de audio como resultado del procesamiento del texto obtenido de entrada, el cual se podrá guardar para su posterior uso.

- Aplicación de algoritmo semántico para mejorar el tono y la duración de las unidades fonéticas de la voz resultante.

- Diseño de la base de datos NoSQL para almacenar el corpus de voz.

3.3. Resultados

Se desarrolló un prototipo de escritorio de síntesis de voz que apoya a mejorar el tono y la duración de las unidades fonética del lenguaje en español para Guatemala, a través del conocimiento que genera el algoritmo semántico que permite identificar los cambios y configuraciones necesarias con respecto a las variables lingüísticas geográficas, diacrónicas, las sociales y las diafásicas para obtener un resultado que produce una voz artificial más natural sin intervención directa del usuario por medio de la entrada de texto. Dicho texto es ingresado de forma manual dentro de un cuadro de texto que presentará el prototipo en la interfaz o por medio de la carga de un archivo de texto en formato txt. Asimismo, sólo se tomará en cuenta el texto que sea escrito en idioma español y en esta primera versión no valida signos de puntuación.

En resumen, el prototipo genera como resultado 3 tipos de audio parte de un texto ingresado, estos son: primero, la generación de audio por concatenación PSOLA; segundo, la generación de audio por concatenación PSOLA en combinación del algoritmo sentimental, y tercero, la generación de audio parte de un texto evaluado por autómatas adaptativos. Además, se presenta una evaluación en donde se compara el rendimiento de los autómatas adaptativos y el algoritmo de síntesis de concatenación PSOLA en combinación de un algoritmo sentimental, toma en cuenta sus características de tono y duración de unidades fonéticas.

4. MARCO TEÓRICO

4.1. Sintetizador de voz

Con el pasar del tiempo la relación hombre-máquina se ha incrementado, lleva al hombre al uso diario de dispositivos tecnológicos, esto con el fin de apoyar a nuestras actividades diarias y proporcionarnos facilidad para realizar algunas tareas. Por dichos motivos ha surgido la necesidad de hacer más amigable nuestra interacción con la tecnología, un ejemplo de esto el sintetizador de voz. El cual permite generar de forma artificial la voz humana a partir de un texto, en donde se entiende por sintetizador a la acción de colocar las partes o elemento en un resultado completo. Al unir los conceptos de sintetizador con la voz, se entiende por sintetizador de voz a la generación artificial de la voz humana desde un texto, también conocido TTS (Text to Speech por sus siglas en inglés). En general, el sintetizador de voz parte de una entrada de texto, el cual es procesado y se obtienen las unidades fonéticas que son comparadas, con una base de datos de voz, la cual concatena los resultados y genera una salida de voz artificial (Varghese y Hande, 2015).

Por otra parte, el sintetizador de voz es una herramienta muy útil para las personas con discapacidades visuales o de lectura para comprender de mejor forma el lenguaje. Dentro de las actividades que se realizan para producir un sintetizador de voz se tienen que tomar en cuenta el análisis de texto, análisis de fonemas, entonación, ritmo, acento, emoción, síntesis articulatoria como factores principales para lograr una voz de calidad que permita la comprensión del lenguaje (Alande, Sharma y Chavan, 2015).

4.1.1. Técnicas de síntesis de voz

Para lograr un mejor acierto en la interacción entre humanos y computadoras se presentan diferentes técnicas por las cuales se puede construir un sintetizador de voz, toma en cuenta que cada una de las técnicas puede resaltar alguna característica especial, dentro de las técnicas se encuentran las siguientes:

- Síntesis articulatoria
- Síntesis de formantes
- Síntesis por concatenación
- HMM (Hidden markov model, por sus siglas en inglés)

4.1.1.1. Síntesis articulatoria

La síntesis articulatoria parametriza la producción de voz humana por medio de un sistema directo, es decir, modela los órganos vocales para que la producción vocal sea lo más parecida a la voz natural. Esta técnica utiliza los siguientes 5 parámetros:

- Área de la abertura de los labios
- Constricción formada por el depresor lingual
- Apertura de las cavidades nasales
- Área global promedio
- Expansión del tracto vocal

Este tipo de técnica tiene potencial para aumentar la calidad de la voz, pero sus estudios no han sido exitosos, dado que desde el punto de vista teórico cuenta con factores que pueden aumentar y mejorar la síntesis de la voz, pero es difícil aplicarlos en la práctica y sus resultados no son determinantes (Adeyemo y Idowu, 2015).

4.1.1.2. Síntesis de formantes

Esta técnica describe la frecuencia en la resonancia del tracto vocal, obtiene resultados altos en lo que a inteligibilidad se refiere, pero resultados bajos en la naturalidad de la voz. Esta síntesis consiste en la reconstrucción artificial de las características con las que cuentan los formantes que van a ser producidos (Adeyemo y Idowu, 2015).

De tal forma, este método aplica parámetros de la frecuencia fundamental y expresa los niveles de ruido en el tiempo para formar una onda de voz artificial (Kayte, Waghmare y Gawali, 2015).

4.1.1.3. Síntesis por concatenación

La técnica de síntesis por concatenación utiliza segmentos de voz grabada que selecciona de una base de datos y los une para formar una expresión deseada, presenta alta calidad en el resultado. Uno de los inconvenientes de esta técnica es que se ve limitada por la capacidad de memoria que requiere (Adeyemo & Idowu, 2015). Dentro de las subcategorías que se encuentran en la síntesis por voz, se puede mencionar:

- Síntesis de unidades de selección o síntesis basado en corpus
- Síntesis de difonos
- Síntesis de dominio específico

4.1.1.4. PSOLA (Pitch Synchronous Overlap and Add)

Uno de los principales métodos utilizados para la técnica de síntesis por concatenación es PSOLA, el cual es un método que se basa en la descomposición de una serie de ondas elementales en donde cada uno de las ondas representa períodos de tonos sucesivos de señal y la suma total

reconstituye la entrada inicial. Una característica importante de mencionar es que PSOLA no trabaja con modelos de ordenamiento por lo cual no pierde ningún detalle de la señal inicial (Mousa, 2010). Existen diferentes tipos de PSOLA, dentro de las cuales se pueden mencionar:

- Time Domain TD-PSOLA
- Frequency Domain FD-PSOLA
- Linear Predictive LP-PSOLA

En donde, el más utilizado es el Time Domain TD-PSOLA por su eficiencia, pero los otros modelos permiten realizar más configuraciones con respecto al tono y permiten mayor control sobre el espectro de la síntesis de la señal.

4.1.1.5. Difonema (par de fonemas)

Este se define como un par contiguo de fonemas, los cuales se utilizan para grabar difonemas para la síntesis de voz haciéndola más natural, dado que la pronunciación de cada uno de los fonemas depende de los que acompañan su entorno (Jewalikar, 2013).

Por otro lado, se puede decir que un difonema consiste básicamente en dos fonemas que se encuentran en una variedad de lenguajes (DÂRDALĂ, 2008), en donde dicha unión se realiza por la frecuencia más estable entre cada fonema, generalmente por el medio del fonema. La dificultad de esto radica en que no todos los fonemas cuentan con sonidos extensos y no se presenta la oportunidad de encontrar su sonido estable, por lo cual es de gran ayuda el contar con los trifonemas para encontrar dicha estabilidad (Trujillo y Roig, 2008).

4.1.1.6. Hidden markov model (HMM)

Modelo estadístico que se caracteriza por la secuencia del espectro de voz y su uso es común en el reconocimiento de voz y el sistema de síntesis de voz proporciona un sonido natural del lenguaje. HMM crea modelos estocásticos que realiza una comparación probabilística para las expresiones desconocidas que pueden ser generadas por cada modelo. Esta técnica a diferencia de la síntesis por concatenación no consume mucha memoria, por el contrario si necesita gran cantidad de los recursos de CPU (Kayte, Waghmare y Gawali, 2015).

4.1.2. Prosodia

El lenguaje cuenta con un conjunto de características que permiten determinar la calidad y aspectos que detallan el sentido del mismo. Dentro de dichos aspectos se encuentra la expresión oral, el tono, la entonación, entre otros aspectos. A esto se le conoce como prosodia, la cual fue descrita por primera ocasión en 1785, y se define como la habilidad que se tiene con el trato de acentos y de la cuantificación de sílabas, en donde el conteo de sílabas se refiere a la longitud de la misma y el acento se refiere al tono y la intensidad con la que son pronunciadas dichas sílabas (Muñoz, 2014).

Otros aspecto a resaltar dentro de la prosodia es la singularidad con la que resalta los segmentos del lenguaje oral y escrito, en lo que expresividad, fraseo y entonación se refiere, permite aportar valor al reconocimiento de las palabras y la comprensión del texto, enfatiza palabras, frases de segmentos y en su carácter interrogativo y exclamativo del texto (Calero, 2014).

4.1.2.1. Tono y fonética

Dentro de las características más importantes de la voz se consideran el tono y la fonética, las cuales nos permiten identificar y analizar la calidad de la voz. Se puede definir al tono como el nivel o altura acústica del sonido, que cumple con la función de valorar la estructura melódica (Barroso, 2012). Las principales características fonéticas del tono se encuentran en el dominio, son diferentes en los distintos lenguajes. El término de tono se refiere particularmente a la forma en la que se escucha el sonido, se clasifica en una escala que va de un tono bajo a un tono alto deja a un lado las propiedades físicas del sonido, en donde su principal correlativo es nombrado como Frecuencia Fundamental (Fromkin, 2014).

Por otra parte, la fonética se enfoca en los sonidos que se emite y se percibe la generación de voz y la transición de ésta al oyente en ondas del sonido, es de importancia resaltar que la fonética es uno de los campos más amplios del lenguaje (Rogers, 2014). También se puede decir que la fonética es objetiva y no subjetiva, por lo que se describen los sonidos del lenguaje en un sentido imparcial, dado que no se puede imponer pensamientos sobre los demás, con respecto a la pronunciación (Brown, 2014).

4.1.3. Corpus

Corpus se define como un conjunto de textos ensamblados en un formato establecido, dentro de la rama de la lingüística, el corpus lingüístico es definido como un medio de almacenamiento informático y analizable de forma automática o semiautomática. Dicho almacenamiento se compone de colecciones de texto o de muestras orales (Kenny, 2014). Otras definiciones establecen que el corpus es una colección finita de texto que es entendible por el computador, también se puede decir que el corpus es un conjunto de texto

escrito o voz que puede servir como una base de análisis y de la cual se puede obtener descripciones lingüísticas.

Dentro de los principales tipos que se tienen se encuentran los corpus escritos y los corpus orales. Los corpus escritos pueden dividirse entre generales o especializados depende de su finalidad. Dos de los corpus más extensos en español son los desarrollados por la RAE (Real Academia Española), que tienen por nombre CREA (Corpus Real Academia Española) y el CORDE (Corpus Diacrónico del Español). Al igual que el corpus escrito el corpus oral también cuenta con CORLEC (Corpus Oral de Referencia de la Lengua Española Contemporánea) de la Universidad Autónoma de Madrid (Rodríguez, Hurtado y Beeby, 2015).

Luego de conocer la definición de un corpus se puede establecer ciertos criterios que se deben de tomar en cuenta para definir un corpus que se acople a las necesidades que se tengan. Dentro de los criterios se puede mencionar:

- Finalidad con la que se quiere utilizar: definir las condiciones y finalidad que se le dará al corpus.
- Límites del corpus: tomar en cuenta el lenguaje, variaciones geográficas y el tiempo de creación.
- Tipo de corpus: conjunto de textos, vocales, cantidad, codificación.
- Población: tomar en cuenta que sector poblacional se utilizará para el corpus.
- Número y longitud del texto:
- Crecimiento del corpus: Tomar en cuenta que la información del lenguaje puede ocupar cantidades grandes de espacio.

- Software y Hardware: Listar y definir las necesidades y los resultados que se quieren alcanzar para tomar en cuenta las diferentes herramientas de software y hardware existentes.
- Aspectos legales: tomar en cuenta que algunas herramientas se encuentran bajo algún tipo de licencia privativa que niega el uso de terceros, por lo cual se tiene que estar informado de cada componente a utilizar para formar un corpus.
- Presupuesto: tener presente los gastos que se darán con respecto a herramientas, personal humano y cualquier otro factor que altere nuestro presupuesto.

En conclusión, se puede afirmar que el corpus es un conjunto de datos finitos, éste se encuentra en un formato informático, el cual cuenta con criterios de diseño definidos y que funciona como representación de una o más lenguas (Rodríguez, Hurtado y Beeby, 2015).

4.2. Semántica

Para comprender mejor el lenguaje se debe de tener claro un aspecto muy importante como lo es la semántica, la cual representa una dimensión del lenguaje que ayuda en la comprensión de los elementos lingüísticos, oraciones, acciones, propiedades y todo hecho del mundo (Gutiérrez, 2006). También se puede decir que la semántica se caracteriza por la relación entre entidades del mundo y la forma en la que se utilizan los símbolos del lenguaje.

Por otra parte, en términos de nivel de lenguaje, la semántica constituye la relación correcta entre el significado y las ideas de las palabras, las cuales se identifican como un conjunto de fonemas (Rodríguez, Vaquero, Saz y Lleida, 2008). Dentro de la semántica se pueden mencionar diferentes aspectos como:

- Semántica léxica: esta se encuentra relacionada con el significado de las palabras.
- Semántica gramatical: esta se encuentra relacionada con el significado de los elementos, categorías, estructuras y los procesos gramaticales.
- Semántica pragmática: esta trata de las relaciones lógicas como lo pueden ser la ampliación, presuposición, vinculación entre otros aspectos.

En relación directa entre la semántica y el lenguaje se presentan múltiples acepciones, de acuerdo al significado connotativo que describe las asociaciones que se realizan entre las palabras y los motivos lingüísticos, conceptuales, culturales, de escolaridad, entre otros. Con respecto a las diferentes lenguas se puede encontrar aspectos como doble sentido del lenguaje, metáforas, significados distintos de una misma palabra, etc. (Liberal y Nazaret, 2015).

4.2.1. Análisis de sentimientos

Debido a la creciente interacción de personas en el internet surge a gran escala la producción de opiniones y comentarios en las diferentes plataformas como lo son: redes sociales, blogs, wikis, tiendas on-line, entre otros sitios web. Por lo que se ha mostrado una alta tendencia en conocer cuál es la aceptación, rechazo o sentimientos que refleja con su escritura un usuario al responder un comentario en una red social o al evaluar la experiencia que ha tenido con un producto o servicio. A esto se le conoce como el análisis de sentimientos (Miranda, Guzmán & Santamaría, 2017).

En el campo de procesamiento de lenguaje natural y de la recuperación de información, el análisis de sentimientos aplica técnicas semánticas para lograr un análisis particular de una representación estructurada de opiniones y su

relación de datos, que presentan mayor precisión en clasificar la polaridad del texto y resaltar los modelos emocionales (Poggi, 2016).

4.2.2. Variables lingüísticas

Estas interpretan las normas que conforman el uso correcto del lenguaje, así como el conjunto de signos y reglas condicionadas por factores como el tiempo, la posición geográfica, nivel sociocultural y el contexto. Derivado a que los hablantes no utilizan la lengua de forma uniforme, sino que esta depende de muchos factores. La primera variable identificada fue la social, que cumple con el objetivo de analizar la influencia de la lengua en situación como la edad, el sexo, origen étnico, clase social y el lugar en que se produce la comunicación. Dicho término aparece en el año 1952 por H. C. Currie y en 1964 en Estados Unidos.

Por otra parte, W. Labov define la importancia de la variación lingüística y describe el nivel interno de la lengua como competencia y como actuación el nivel externo y los factores sociales (Marcos F. G., 1999). En 1981, Coseriu distingue 3 tipos de variables internas en la lengua (diatópica, diastrática y diafásica), donde se resalta el latín vulgar, español de la Península y el español de América (Penny R., 2006). Luego, la Real Academia Española reconoce las variables de un idioma, indica la forma correcta de expresarse e intenta evitar expresiones inadecuadas. La variable sociolingüística se desarrolla mediante 3 campos: la lingüística, antropología y sociología. Desde los aportes mencionados se ha determinado que no existe una forma común de clasificar el lenguaje ni su expresión, por dicha variedad se diversificó en las características personales de quien emplea la lengua y por otro lado, las condiciones por factores de contexto. Dentro de la variable por factores de contexto se encuentra la variable diafásica o estilística. Y dentro de las variables que

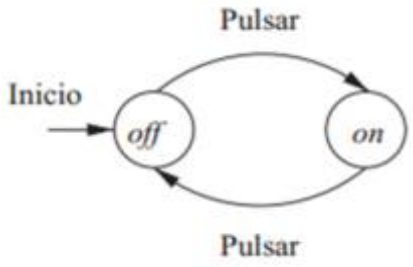
presentan características personales las variables geográficas, diastráticas y diacrónicas.

4.3. Autómatas

Los autómatas se denomina como el estudio de los dispositivos que realizan cálculos abstractos, esto quiere decir, que realizan cálculos de las máquinas. Estos fueron denominados como autómatas finitos en la década de los años cuarenta y cincuenta. Inicialmente dichos autómatas fueron considerados para modelar el funcionamiento del cerebro, posteriormente se tomaron en cuenta para realizar otras acciones (Hopcroft, Motwani y Ullman, 2008).

En un contexto más tecnológico los autómatas finitos (Ver Figura 1) se construyen como un modelo el cual es de utilidad para diferentes tipos de Hardware y Software. Dentro de la rama del software este se utiliza para diseñar y probar el comportamiento de circuitos digitales. En términos del lenguaje, se puede encontrar en el analizador léxico de un compilador, el cual se encarga de separar el texto en unidades lógicas para obtener información de la comprensión del mismo, identifica términos como identificador, palabras claves y signos de puntuación. También es utilizado para explorar textos largos dentro de páginas web y permite determinar la frecuencia de las palabras, frases o permite identificar patrones (Hopcroft, Motwani y Ullman, 2008).

Figura 1. **Modelo de autómata finito de un interruptor de apagado/encendido**



Fuente: Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2008). Teoría de autómatas, lenguajes y computación. *Pearson Prentice-Hall*. (p. 3)

4.4. Base de datos nosql

El término NoSQL (No Structure Query Language or Not only Structure Query Language, por sus siglas en inglés) fue acuñado a finales de los años 90, cuenta con características como el manejo de grandes volúmenes de información, en donde su mayor diferencia es que no cumple con los esquemas tradicionales de tipo relacional. Junto con este cambio de perspectivas han surgido diferentes características de interés en las que se incluyen el rendimiento, la escalabilidad, replicación, distribución y la necesidad de satisfacer la gran demanda de datos de los sistemas como comunidades, buscadores, blog, redes sociales y muchos otros (del Busto y Enríquez, 2013). Una definición más concreta de las bases de datos NoSQL es que son sistemas que almacenan información, las cuales mantienen un esquema diferente al ER (Entidad-Relación), no se impone ninguna estructura de datos formales, en otras palabras, no se impone una estructura de tablas y relaciones que tengan un sentido referencial, sino que son de tipo más flexible como lo son formatos clave-valor, Mapeo de columnas, documentos o Grafos (Valenzo, Valencia y Castro, 2013).

Otros aspecto importante a resaltar es que aunque el término de NoSQL ha existido por muchos años no se le había dado la relevancia, sino hasta el 2009 que un empleado de Last.fm Johan Oskarsson organizó un evento con el tema de base de datos distribuidas de código abierto no relacionales (Valenzo, Valencia & Castro, 2013).

NoSQL cuenta con características que permiten adaptar la necesidad de un sistema para brindar una solución más eficiente, dentro de las que se puede mencionar se encuentran:

- Consistencia eventual
    - Por medio de comunicación de nodos.
    - Flexibilidad en la consistencia
    - Se realiza cada cierto período.
    - En contraposición de propiedades ACID (Atomicity, Consistency, Isolation, Durability, por sus siglas en inglés) de base de datos relacionales, se aplica el concepto de BASE (Basically Available Soft-state Eventual Consistency, por sus siglas en inglés)
- Estructura distribuida
    - Típicamente por distribución de datos por mecanismos de tablas hash
- Escalabilidad horizontal
    - Incremento de equipos (nodos)
    - capacidad de proceso limitado por nodo
- Tolerancia a fallos y redundancia

Las plataformas con las que cuenta nosql son las siguientes:

- Llave-valor
    - Se almacenan valores asociados a una llave
    - Estructuras sencillas

    - Mejor rendimiento
    - Ejemplos:
        - Cassandra, Apache
        - BigTable, de Google
        - Dynamo, de Amazon, etc.
- Documentos
    - Particularización de llave-valor pero en documentos
    - Permite consultas complejas
    - Ejemplos
        - CouchDB, de Apache
        - MongoDB, de 10gen MongoDB
        - RavenDB, de HibernatingRhinos, etc.
- Columnas
    - Los valores se almacenan en columnas, no en filas
    - Utilidad al manejar datos agregados
- Grafos
    - Unidad básica el nodo, entidades en modelo relacional
    - Aristas, relaciones en modelo relacional
    - Ejemplos:
        - Neo4j
        - DEX
        - AllegroGraph
        - OrientDB
        - InfiniteGraph
- Objetos
    - Datos son objetos
    - Punteros son relaciones
    - Bajo rendimiento, con opción a operaciones de complejidad superior

- Ejemplos:
  - db4o
  - GemStone S
  - Objectivity/DB

Basadas en tuplas, multivaluadas, jerárquicas, entre otras.

4.4.1. Mongodb

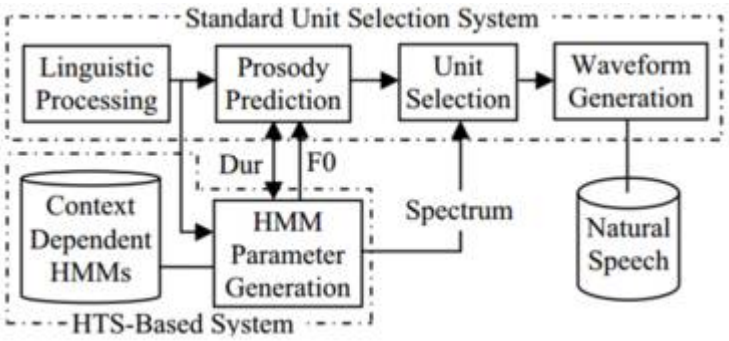
Es una base de datos NoSQL muy robusta que presenta características como la flexibilidad, escalabilidad, replicación, indexación, balanceo de carga entre otras más. Es una base de datos orientada a documentos, la cual reemplaza el concepto de filas a un modelo más flexible como lo son los documentos. Dicha orientación se enfoca en hacer posible la representación de relaciones jerárquicas complejas con un campo simple, hace más sencillo el desarrollo de sistemas con los lenguajes orientados a objetos. Otra característica que es importante mencionar es que no utiliza esquemas predefinidos, las llaves y valores no contienen tipos o tamaños y hace el agregar o remover campos sea una tarea relativamente sencilla (Chodorow, 2013).

MongoDB fue diseñado para ser escalable y realiza de forma eficiente la separación de datos en múltiples servidores toma en cuenta el balanceo de carga y distribuyendo de forma automática los documentos y enruta las peticiones de los usuarios a las máquinas correctas. En conclusión, se puede decir, MongoDB es una base de datos NoSQL en donde la unidad básica son los documentos, maneja esquemas dinámicos, una instancia puede hospedar diferentes bases de datos, cada documento tiene una llave única dentro de la colección de base de datos y permite mejorar en gran medida la consulta de gran cantidad de información (Chodorow, 2013).

4.5. Arquitectura de software

La arquitectura de un programa de software es una parte muy importante para resaltar y cumplir con ciertas características como acceso a los datos, escalabilidad, desempeño, conjunto de componentes, distribución física, entre otros. Por lo tanto, se puede definir a la arquitectura como la estructura global que proporciona la integridad conceptual de un sistema (Capobianco y en Ciencias, 2014), la cual contiene los detalles dinámicos esenciales (Hernández, Escandón, Acosta y Rivera, 2015) y que se representa en componentes, conexiones y restricciones. Por otra parte la arquitectura de un sistema nos puede brindar una idea de la calidad del sistema y también nos permite resaltar las posibles debilidades y riesgos que presente el sistema (Pons, Rodríguez y Maribona, 2012). Un ejemplo de arquitectura es la que se muestra en la Figura 2.

Figura 2. **Hibryd TTS architecture**



Fuente: Sainz, I., Erro, D., Navas, E., & Hernáez, I. (2011). A Hybrid TTS Approach for Prosody and Acoustic Modules. In *INTERSPEECH* (pp. 333-336).

4.5.1. Arquitectura modular

La arquitectura modular se encuentra basada en componentes, que pueden ser interconectados y que funcionan en conjunto (Pérez, Valdiviezo, Pérez Otero, Liberatori, Rexachs del Rosario, Luque y Lassarre, 2010). Dicha

arquitectura facilita la identificación y solución de errores, modificación por partes del sistema y la adición de nuevos componentes sin afectar al sistema completo (Benito, Gaspar, Rivas, Martínez, Rodriguez y Ramírez, 2011), dentro de sus principales ventajas se puede mencionar, compatibilidad con estándares, compuesto por componentes, integración de sistemas, adaptabilidad, acoplamiento, independencia funcional y escalabilidad como las más importantes a destacar (Alvarado, 2013).

4.6. Modelo-vista-controlador (MVC)

A principio de los 70's se establecieron conceptos como objetos, clases, encapsulación, herencia y polimorfismo, los cuales siguen vigentes hasta nuestros tiempos en lenguajes como C++, JAVA entre otros. Junto a estos conceptos se presentó una interfaz provista por SmallTalk, la cual se basó en un patrón de diseño que tiene como base la separación de 3 módulos identificados y con tareas definidas, el Modelo, la Vista y el Controlador (MVC) (Bascón Pantoja, 2011).

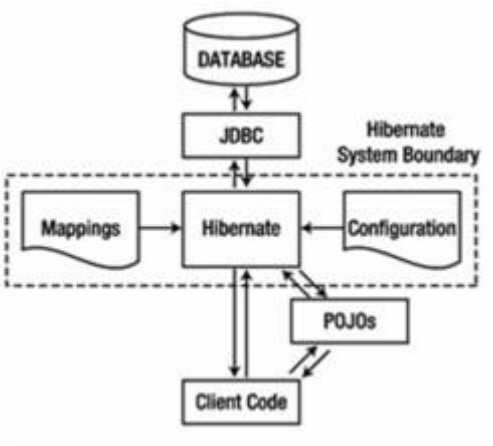
Para comprender mejor los módulos se presenta la siguiente explicación:

- Modelo: Dentro de este módulo se representan los datos del programa y sus transformaciones. El modelo no se relaciona directamente con los otros módulos, es el sistema el cual se encarga de la comunicación de cambios.
- Vista: presenta los objetos visuales que vienen del modelo y que interactúa de forma preferente con el controlador.
- Controlador: brindar significado a las órdenes del usuario, esto por medio del Modelo y se encarga de orquestar la interacción entre la Vista y el Modelo, mantiene el sistema sincronizado (González y Romero, 2012).

4.7. Hibernate

Es una herramienta que simplifica la conexión con la base de datos y las aplicaciones realizadas en Java (Ver Figura 3), con el fin de mostrar los datos como objetos simples. El acceso se realiza por medio de gestor de sesiones, otro aspecto importante a mencionar es que Hibernate es un ORM (Object Relational Mapper, por sus siglas en inglés) el cual provee dos tipos de interfaces, el primero es un Hibernate nativo y la segunda interfaz es Java EE-standar Java Persistence API. Las cuales influyen en la forma con la que se conectan con el desarrollo Java (Ottinger, Minter y Linwood, 2014).

Figura 3. T**he role of Hibernate in a Java Application**



Fuente: Ottinger, J. B., Minter, D., &Linwood, J. (2014). An Introduction to Hibernate 4.2.In *Beginning Hibernate* (pp. 1-7).Apress.

5. MARCO METODOLÓGICO

5.1. Tipo de investigación

El presente trabajo de graduación muestra una investigación experimental y cualitativa, la cual contribuye en la línea de investigación de la tecnología de la información y la comunicación para apoyo a la educación. Como objetivo principal generó un prototipo de software que permite el ingreso de texto plano en idioma español para generar una voz artificial como resultado, en donde se implementaron los algoritmos de síntesis por concatenación PSOLA y un algoritmo semántico sentimental de evaluación de texto.

5.2. Diseño de investigación

El presente trabajo se realizó mediante investigación experimental y cualitativa de los aspectos necesarios para implementar un algoritmo de síntesis por concatenación PSOLA en combinación de un algoritmo sentimental que mejore las condiciones de obtener un tono y duración de unidades fonéticas de mayor calidad en idioma español, para Guatemala en lo que a voz artificial se refiere.

Como parte de la investigación cualitativa, más específicos un estudio de caso, se aplicaron elementos y características que apoyan a mejorar el resultado de la voz artificial, presenta un estudio que detalla la influencia de las variables lingüísticas sobre la comprensión de un mensaje, en otras palabras, qué tipo de variables lingüísticas apoyan en mayor y menor medida al tono y duración de unidades fonéticas de la voz en Guatemala.

Para alcanzar los objetivos descritos se desarrolló un prototipo que facilita los siguientes experimentos:

- Evaluación y clasificación de texto
- Generación de audio artificial
- Optimización del tono y duración de unidades fonéticas
- Comparación de resultados.

Para la investigación experimental se realizó una investigación cuantitativa, donde se analizaron los resultados de los experimentos, parte de una técnica de recolección de datos, a través de encuestas por correo electrónico, en las cuales se les presentaron audios generados por el prototipo a un grupo de personas y se les hicieron ciertas preguntas para determinar si se cumple con los objetivos de la investigación.

5.3. Método de investigación

La presente investigación se llevó a cabo en seis fases, las cuales presentan tres fases para el desarrollo del prototipo: una de experimentación, una de recolección y evaluación de resultados y por último la de redacción de resultados.

5.3.1. Fase 1: investigación documental

En la fase de investigación documental, se realizó la búsqueda y recolección de información que brinda una visión general del funcionamiento y de los tipos de sintetizador de voz que se pueden generar, asimismo de las diferentes técnicas y métodos que se utilizan para manipular la conversión de texto a voz. También se obtuvo información relacionada a los factores que

apoyan al incremento de calidad de la voz artificial con respecto a las características del tono y duración fonética de la voz. Dentro de dicha información se presentan los siguientes puntos:

a. Investigación variables lingüísticas

i. **Investigación de tipos de variables lingüísticas:** se determinaron las variables lingüísticas que se pueden utilizar para mejorar la calidad del tono y la duración de voz generada por un sintetizador.

ii. Determinar las características de las variables lingüísticas: se investigaron las características de las variables lingüísticas que afectan la calidad del tono y fonética.

iii. Medir el grado de influencia de las variables lingüísticas respecto a características de la voz: se determinó el grado de influencia que tiene cada una de las variables lingüísticas en relación a la calidad del tono y fonética.

b. Valor que aporta un algoritmo semántico al algoritmo PSOLA

i. **Investigación de algoritmo PSOLA:** investigación de características de algoritmo PSOLA y su aplicación de la generación de voz artificial.

ii. **Investigación de algoritmo semántico**: investigación de características de algoritmo semántico para definir emociones a la generación de voz artificial.

iii. **Influencia del sentido del texto para mejorar la voz artificial**: se determinó la forma de aplicar sentido al texto mediante un servicio de análisis de sentimientos el cual influye en la concatenación de unidades que generan la voz artificial.

c. Características de un autómata adaptativo que influyen en el resultado de un sintetizador de voz

i. **Investigación de las características de un autómata adaptativo:** se investigaron las características y estados que definen el funcionamiento de un autómata adaptativo

ii. **Factores del autómata adaptativo que se utilizan para la categorización de texto en un sintetizador de voz:** se investigaron los factores que permiten la clasificación del texto para la selección de unidades de voz en la generación de voz artificial.

iii. **Determinar las propiedades de comparación entre autómatas adaptativos y el método de síntesis por concatenación PSOLA en combinación de un algoritmo semántico:** se determinaron las propiedades para validar que resultado de voz artificial presenta una mejora en lo que a tono y duración de unidades fonéticas se refiere.

5.3.2. Fase 2: diseño de prototipo

En esta fase, se realizó un plan para estructurar un prototipo funcional que permite soportar los objetivos principales.

a. Variables lingüísticas

i. Diseño de selección de variables lingüísticas que se tomaron en cuenta para la creación de las unidades utilizadas en el corpus de voz, en otras palabras es la selección del sujeto de prueba para realizar las grabaciones de voz a utilizar que cumpla con las características investigadas.

ii. Diseño de la lógica de como guardan las unidades en el corpus de voz.
iii. Diseño de la estructura final de la base de datos para almacenar los textos y sonidos que se utilizan en el corpus de voz.

b. Valor que aporta un algoritmo semántico al algoritmo PSOLA

i. Diseño de evaluación y clasificación de texto de entrada para la generación de audio artificial.
ii. Diseño de consulta de algoritmo semántico sentimental para determinar a qué partes del texto se aplican los estados, dentro de los que se puede mencionar aspecto feliz, triste y neutro.
iii. Diseño de aplicación de estados semánticos para generar el resultado final de voz artificial.

c. Comparar características de un autómata adaptativo que influye en el resultado de un sintetizador de voz

i. Diseño de evaluación y clasificación de texto por estados de autómata.
ii. Diseño de selección de unidades del corpus de voz para la generación de voz artificial.
iii. Diseño de presentación de resultado de voz detalla el número de iteraciones que realiza por estado para generar una respuesta final.

5.3.3. Fase 3: desarrollo de prototipo

En esta fase, se realizó el desarrollo del código fuente del prototipo que contiene las librerías que convierten el texto de entrada a estructuras de datos, utiliza principalmente tecnologías de desarrollo como JAVA y MongoDB.

Asimismo, se aplicó el desarrollo para realizar la generación de voz artificial, mediante síntesis de voz por concatenación PSOLA con la opción de implementar un algoritmo semántico sentimental para apoyar al sentido del texto. Por último, se desarrolló la opción de generar voz por el método de autómata adaptativo para su posterior uso en la experimentación del prototipo al ser comparado con la voz generada por la síntesis por concatenación PSOLA en combinación de un algoritmo semántico.

a. Variables lingüísticas

i. Generación de fragmentos de voz por sujeto de pruebas toma en cuenta las variables lingüísticas.
ii. Implementación de carga de unidades de texto en la base de datos para interpretar los sonidos del corpus.
iii. Implementación de carga de sonidos en el corpus de voz.

b. Valor que aporta un algoritmo semántico al algoritmo PSOLA

i. Implementación de la evaluación y clasificación de texto de entrada para la generación de audio.
ii. Implementación de transformación de texto de entrada que identifica los estados sentimentales.
iii. Implementación de estados semánticos al resultado final de voz artificial.

c. Comparar características de un autómata adaptativo que influyen en el resultado de un sintetizador de voz

i. Implementación de la evaluación y clasificación de texto por autómata.
ii. Desarrollo de selección de unidades del corpus de voz para generación de voz artificial

iii. Generación de resultados de voz indica el número de iteraciones que realiza por estado para generar la respuesta.

5.3.4. Fase 4: experimentación

Para la experimentación se realizó una breve comparación de audios generados por el prototipo, estos generados por clasificación de autómata adaptativo y por síntesis de concatenación PSOLA en combinación con algoritmo semántico. Dicho análisis se llevó a cabo con la herramienta de software MATLAB, toma en cuenta las características de *pitch* entre uno y otro. Además se generaron audios con cada técnica (síntesis por concatenación PSOLA con y sin algoritmo semántico y por autómatas adaptativos) y se proporcionaron a un grupo de personas para conocer si alguna técnica cumple con el objetivo de mejorar el tono y duración fonética, esta recolección de información se hizo a través de encuestas.

a. Variables lingüísticas

i. Se estableció un caso real, el cual apoya al reconocimiento de los aspectos principales de los tipos de variables lingüísticas (Geográficas, diacrónicas, sociales y diastráticas).

ii. Se redactó una encuesta que en sus preguntas apoya a los diferentes tipos de variables lingüísticas.

iii. Se realizaron diferentes cargas de la información y de consulta de los audios por separado dentro del corpus de voz para comprobar su conexión y la correcta generación de los mismos.

b. Valor que aporta un algoritmo semántico al algoritmo PSOLA

i. Generación de voz artificial con síntesis por concatenación PSOLA y validación del grado de aceptación en lo que a calidad del tono y duración de unidades fonéticas se refiere.

ii. Generación de voz artificial con síntesis por concatenación PSOLA en combinación con algoritmo semántico sentimental y validación del grado de aceptación en lo que a calidad del tono y duración de unidades fonéticas se refiere.

iii. Evaluación mediante la herramienta de MATLAB, las características de PITCH y de frecuencia de onda los resultados de voz con síntesis por concatenación PSOLA y síntesis por concatenación PSOLA en combinación con un algoritmo semántico sentimental.

c. Comparar características de un autómata adaptativo que influyen en el resultado de un sintetizador de voz

i. Generación de voz artificial por la técnica de autómatas adaptativos y validación del grado de aceptación en lo que a calidad del tono y duración de unidades fonéticas se refiere.

ii. Generación de voz con síntesis por concatenación PSOLA en combinación con algoritmo semántico y de voz artificial por la técnica de autómatas adaptativos para evaluar cuál de las dos presenta más características favorecedoras en relación al tono y la duración de unidades fonéticas de la voz artificial.

iii. Generación de voz con síntesis por concatenación PSOLA con y sin combinación con algoritmo semántico y de voz artificial por la técnica de autómatas adaptativos para evaluar cuál de las tres presenta más características favorecedoras en relación al tono y la duración de unidades fonéticas de la voz artificial.

5.3.5. Fase 5: recolección y evaluación de resultados

Dentro de esta fase se realizaron encuestas para determinar mediante la opinión de una muestra de la población cual o cuales técnicas brindan un incremento en la calidad del tono y duración de las unidades fonéticas de la voz artificial. Dicha recolección de información se realizó mediante encuestas enviadas por medios electrónicos (correo electrónico, formularios de google), luego se procedió a realizar un conteo de los datos obtenidos para brindar y explicar el resultado final, esto para la parte experimental del trabajo de investigación.

Por otra parte, para la investigación cualitativa sobre la influencia de las variables lingüísticas en la mejora del tono y la duración de las unidades fonéticas se realizaron encuestas presenciales y un estudio de un caso propuesto, las cuales nos apoyaron a determinar qué tipo de variable lingüísticas influye en mayor y menor medida a la calidad de la voz.

a. Variables lingüísticas

i. Evaluación de datos obtenidos y generación de tabla de resultados que determinan que tipo de variables lingüísticas influyen en mayor y cuales en menor medida al tono y a la duración de unidades fonéticas de la voz artificial.

ii. Evaluación de encuestas realizadas, simplificación de datos obtenidos.

iii. Especificaciones de las bases de datos utilizadas, en cuanto a tamaño, número de elementos, entre otros aspectos.

b. Valor que aporta un algoritmo semántico al algoritmo PSOLA

i. Conteo y evaluación de datos obtenidos por las encuestas realizadas para determinar la influencia en las características de

tono y duración de unidades fonéticas en la voz artificial del algoritmo PSOLA y algoritmo PSOLA en combinación de algoritmo semántico.

ii. Recolección y evaluación de datos obtenidos de la herramienta MATLAB al comparar las características de PITCH entre algoritmo PSOLA y algoritmo PSOLA en combinación de algoritmo semántico.

c. Comparar características de un autómata adaptativo que influyen el resultado de un sintetizador de voz

i. Conteo y evaluación de datos obtenidos por las encuestas realizadas para determinar la influencia en las características de tono y duración de unidades fonéticas en la voz artificial del algoritmo PSOLA, algoritmo PSOLA en combinación de algoritmo semántico y autómata adaptativo.

5.3.6. Fase 6: redacción de informe final

Como conclusión del experimento, se redactó el presente informe final que muestra todas las bases utilizadas que apoyaron a construir un prototipo y su posterior evaluación, para tomar en cuenta desde que punto de vista se llegaron a las conclusiones descritas al final. De la misma forma se presentan los datos utilizados en las encuestas y todos los resultados obtenidos con las técnicas de análisis de información.

a. Variables lingüísticas

i. Detalle que muestra las variables que apoyan a mejorar las características de la voz artificial como el tono y la duración de las unidades fonéticas en idioma español para Guatemala.

b. Valor que aporta un algoritmo semántico al algoritmo PSOLA

i. Detalle de las características que mejoran la calidad de la voz artificial al combinar un algoritmo semántico al algoritmo PSOLA.

c. Comparar características de un autómata adaptativo que influyen en el resultado de un sintetizador de voz

i. Detalle de las mejoras y/o desventajas que muestra un autómata adaptativo en comparación de un algoritmo semántico, considera las características de tono y duración de unidades fonéticas de la voz artificial.

5.4. Instrumentos de recolección de información

Para llevar a cabo la investigación se utilizaron los siguientes instrumentos de recolección:

- Fuentes secundarias:
  - Artículos científicos.
  - Tesis de maestría/doctorado.
  - Documentación de herramientas de desarrollo.
- Fuentes primarias:
  - Recolección de datos por medios electrónicos:
    - Correo electrónico
    - Formularios de Google
  - Encuestas presenciales

5.5. Variables e indicadores

En la Tabla I, se puede apreciar las diferentes variables que están sujetas a la experimentación dentro de la investigación y la generación del prototipo,

toma en cuenta sus elementos de medición como los son los indicadores y las sub variables que las describen.

Tabla I. **Variables e indicadores**

Variables	Definición	Sub variables	Indicadores	Dimensiones
Técnica de síntesis por concatenación PSOLA	Método que descompone una serie de ondas elementales en donde las ondas se representan en periodos de tonos sucesivos.	1. Corpus de voz 2. Separación por difonemas. 3. Concatenación de sonidos.	1.Número de audios en la base de datos 2. Tiempo de generación de resultado. 3. Naturalidad del sonido generado.	Cuantitativos(rangos de aceptación)
Técnica de autómatas adaptativos	Método que aplica a los autómatas, que se encargan de realizar cierto número de operaciones en donde progresivamente se ejecuta hasta que cumple con las condiciones a cumplir.	1. Autómata 2. Formatos de reconocimiento.	1. Número de iteraciones a realizar. 2. Detección de estados. 3. Facilidad para comprender la señal de voz. 4. Naturalidad del sonido generado.	Cuantitativos(rangos de aceptación)
Técnica de algoritmo semántico	Algoritmo que comprende el sentido del texto y le permita descifrar la tonalidad y tiempo de duración de voz	1. Optimización del sonido. 2. Características del sentido del texto.	1. Grado de comprensión de texto. 2. Número de optimizaciones de sonido. 3. Naturalidad del sonido generado.	Cuantitativos (rangos de aceptación)
Características del tono y duración de unidades fonéticas, respecto a las variables lingüísticas	Aporta valor al incremento de la calidad de la voz generada por sintetizador de voz.	1. Intensidad de frecuencia de voz. 2. Duración de frecuencia de voz	1. Variables que influyen en mayor y en menor medida al tono de la voz generada. 2. Variables que influyen en mayor y menor medida a la duración de la voz generada.	Cualitativos

Fuente: elaboración propia.

5.6. Técnicas de análisis de información

Con el fin de analizar los resultados obtenidos por el experimento, se utilizó estadística descriptiva para medir los indicadores que expresan los logros obtenidos por cada uno de los objetivos descritos.

5.6.1. Técnica descriptiva

Dicha técnica se utilizó para identificar los elementos y características que influyen en el resultado de la voz artificial, analiza mediante un estudio de caso las variables lingüísticas que afectan al idioma español para Guatemala. Como resultado del estudio de caso, se realizó una serie de encuestas a personas de una oficina para identificar el grado de influencia de las variables lingüísticas sobre el entendimiento del lenguaje. Luego de esto se categorizaron los resultados de las encuestas realizadas y el grado de aceptación para las características del tono y duración de unidades fonéticas de la voz. El análisis para brindar una respuesta final sobre que categoría influye más y cuál menos sobre el español para Guatemala, se obtuvo del conteo de incidencias.

5.6.2. Técnica de encuesta

Como parte del estudio se realizaron encuestas que fueron enviadas por medios electrónicos, dichas encuestas tienen el propósito de determinar cuál o cuáles técnicas utilizadas para generar voz artificial a partir de texto, presentan mejoras en las características de tono y duración de unidades fonéticas de la voz artificial. El proceso que se realizó para dicho estudio, mediante la técnica de encuestas se encuentra descrito en la Figura 4.

Figura 4. **Proceso de técnica de encuesta**

Fuente: elaboración propia.

Se tomó en cuenta el análisis de las variables cualitativas y las variables cuantitativas. Para las variables cuantitativas las de carácter discreto, valores específicos, y para las variables cualitativas las variables de tipo ordinal que presentan una escala establecida de valores y que finalmente serán evaluadas cuantitativamente.

Para determinar un conjunto de datos validos se optó por seleccionar el tipo de muestreo no probabilístico, donde se desconoce la muestra y es un procedimiento informal y más natural, el cual consiste en seleccionar sujetos por tener mayor accesibilidad a ellos.

Con respecto a la técnica de recolección de datos, se realizó mediante el estudio de encuestas, las cuales fueron enviadas por correo electrónico y formularios de Google.

El diseño que se utilizó para realizar la encuesta (entrevista) que estudia las variables lingüísticas que apoyan al tono y duración de unidades fonéticas de la voz artificial es el que se encuentra en la Tabla II:

Tabla II. **Encuesta de estudio de caso variables lingüísticas**

No.	Pregunta
1	Cuando usted escucha hablar a una persona de uno de los departamentos de Guatemala, ¿le es fácil entenderle?
2	¿Qué diferencias cree usted que hay en la forma de hablar de las personas de diferentes departamentos?
3	¿Cree usted que la rapidez con la que hablen las personas afecta entender sus palabras? ¿Podría explicarme?
4	¿Es usted capaz de entender lo que le dice una persona cuando le habla rápido?
5	¿Le es fácil entender lo que dicen los jóvenes cuando tiene alguna conversación con ellos? ¿Puede darme detalles?
6	¿Qué palabras le parecen desconocidas cuando tiene alguna conversación con otra persona?
7	¿Por lo general usted habla con personas en ambientes libre de distracciones? ¿Me puede ampliar?
8	¿Considera que las personas, por lo general, hablan muy fuerte o muy callados? ¿Puede darme algunos ejemplos?
9	¿En las conversaciones que usted sostiene con otras personas, usted nota que las palabras se pronuncian completas o no?
10	¿Cree usted que las personas pronuncian las palabras adecuadamente? ¿Por qué?
11	¿Considera usted que el desconocimiento de la ortografía o la gramática influye en que las personas pronuncien bien o mal las palabras?
12	¿Cree usted que la diferencias sociales afectan en la forma como hablan las personas?
13	¿Estima usted que las personas se expresan de manera diferente en situaciones diferente? ¿Puede darme algún ejemplo?
14	¿El uso de palabras extranjeras es común en conversaciones que usted sostiene con otras personas?
15	¿Usted modifica su lenguaje y/o características de su voz al hablar con diferentes tipos de personas? ¿De qué forma?
16	¿Cuáles cree que sean las características necesarias en la voz para comprender mejor un mensaje?

Fuente: elaboración propia.

Para la recolección de datos sobre el valor que genera cada una de las técnicas del prototipo se utilizaron 4 encuestas que están divididas en 2 grupos,

el primer grupo engloba las primeras 3 encuestas que se enfocan en evaluar cada técnica por separado; para el segundo grupo se evaluaron las técnicas en conjunto para determinar que aporte es más significativo en cuestiones del tono y duración de unidades fonéticas de la voz artificial. Por último, se analizaron los resultados de cada uno de los grupos para brindar las conclusiones finales.

El primer grupo cuenta con la siguiente encuesta (Ver Tabla III), la cual se aplica para las 3 técnicas (algoritmo PSOLA, algoritmo PSOLA en combinación con algoritmo semántico y autómatas adaptativos) por separado, como resultado 3 evaluaciones.

Tabla III. **Encuesta por técnica del prototipo**

No.	Pregunta	Respuestas
1	En su opinión, ¿qué tan entendible es la voz artificial presentada? Donde 5 es entendible y 0 no se entiende el audio.	0, 1, 2, 3, 4, 5
2	En su opinión, ¿Cuál es la calidad del tono de la voz artificial presentada? Donde 5 es muy buena calidad y 0 es muy mala calidad.	0, 1, 2, 3, 4, 5
3	¿Cómo considera la calidad de pronunciación de las palabras de la voz artificial presentada? Donde 5 es la pronunciación es entendible y 0 no se entienden las palabras pronunciadas.	0, 1, 2, 3, 4, 5
4	¿En su opinión, cuál de estas características es aceptable en la voz artificial presentada? Donde el tono es la calidad del volumen de la voz, duración de la voz es si se entiende el mensaje y ninguna si en verdad no cumplió con dichas características.	• TONO • DURACIÓN DE LA VOZ • AMBOS • NINGUNA
5	En su opinión, ¿cree usted que la voz artificial presentada mejoraría si fuera de un sexo o de edad diferente? ¿Sería significativo el cambio?	• SI • NO
6	En su opinión, ¿cree usted que la voz artificial presentada mejoraría o se entendería mejor si utilizará palabras distintas a las que presenta el audio?	• SI • NO
7	Identificaría como una característica de la voz artificial presentada un tono formal, es decir, que la voz se expresa con propiedad.	• SI • NO
8	Identificaría como una característica de la voz artificial presentada palabras de otras regiones o que utilicen solo en otras partes del país.	• SI • NO

Fuente: elaboración propia.

El segundo grupo, utilizó la encuesta que se presenta en la Tabla IV, con la que se realizó la comparación de las técnicas utilizadas.

Tabla IV. **Encuesta comparación de técnicas del prototipo**

No.	Pregunta	Respuestas
1	Cree usted que existe alguna mejora del audio **DELTA** con respecto al audio **GAMA** donde 5 es si mejora significativamente el audio **DELTA** y 0 no le encuentro mejora alguna al audio **DELTA**.	0, 1, 2, 3, 4, 5
2	Podría indicar si usted reconoce algún tipo de emoción en el audio **DELTA**, donde 5 es si totalmente y 0 es no reconozco nada todo suena igual.	0, 1, 2, 3, 4, 5
3	En su opinión, ¿Qué audio (**DELTA y BETA**) presenta mejoras en el tono y pronunciación de la voz artificial?	• Audio **DELTA** • Audio **BETA**
4	En su opinión, ¿Cuál de los audios escuchados presenta mejora en alguna de las características: calidad en tono y/o duración de la voz?	• Audio **GAMA** • Audio **DELTA** • Audio **BETA** • Audio **GAMA y DELTA** • Audio **DELTA y BETA** • Audio **GAMA y BETA** • Todos • Ninguno
5	¿Cuál de los audios según su criterio suena mejor y más entendible?	• Audio **GAMA** • Audio **DELTA** • Audio **BETA**

Fuente: elaboración propia.

Para resumir se utilizaron las técnicas siguientes:

- **Técnica de análisis cualitativo:** se categorizaron los resultados, donde los datos se reducen a unidades llamadas categorías, las cuales fueron recolectados por estudio de caso y analizados de forma cuantitativa.

- **Técnica de análisis cuantitativo:** se utilizaron técnicas de frecuencias para determinar los valores con mayor incidencia y presenta los resultados en diagramas o gráficas de frecuencias como la distribución.

6. PRESENTACIÓN DE RESULTADOS

6.1. Diseño del prototipo sintetizador de voz

El diseño propuesto fue utilizado para llevar a cabo la implementación de un Sintetizador de Voz de idioma español para Guatemala, el cual permite satisfacer los objetivos descritos, por tal razón en los siguientes puntos se detallan las principales características que lo conforman.

6.1.1. Funcionalidad general

Como parte práctica del estudió se propuso un algoritmo de síntesis por concatenación PSOLA en combinación con un algoritmo semántico para mejorar el tono y la duración fonéticas de síntesis de voz en el idioma español para Guatemala. En términos generales, el prototipo se encarga de obtener una entrada de texto la cual mejorará en aspectos de tono y duración de unidades fonéticas determina mejor el sentido con el que se debe expresar la voz.

Por lo antes descrito, se diseñó un prototipo que permite el ingreso de texto por carga de un archivo con extensión TXT en español o por un formulario desarrollado en el lenguaje de programación JAVA, luego de un proceso de conversión de texto a voz se genera como resultado un archivo con extensión WAV que permite apreciar un audio de voz artificial, el proceso descrito se presenta en la Figura 5.

Figura 5. **Concepto general del prototipo de síntesis de voz**

Fuente: elaboración propia.

6.1.2. Arquitectura

Como esquema general de la arquitectura se presenta en la Figura 6 una distribución de servidores que permiten relacionar cada uno de los componentes del sistema. Dentro de dicho esquema se encuentran los siguientes:

- **Servidor de base de datos**: dentro de este servidor se maneja el componente que se relaciona con la base de datos NoSQL, para obtener los fragmentos de audio que se utilizan para presentar las soluciones, dicho componente es: MongoDB.

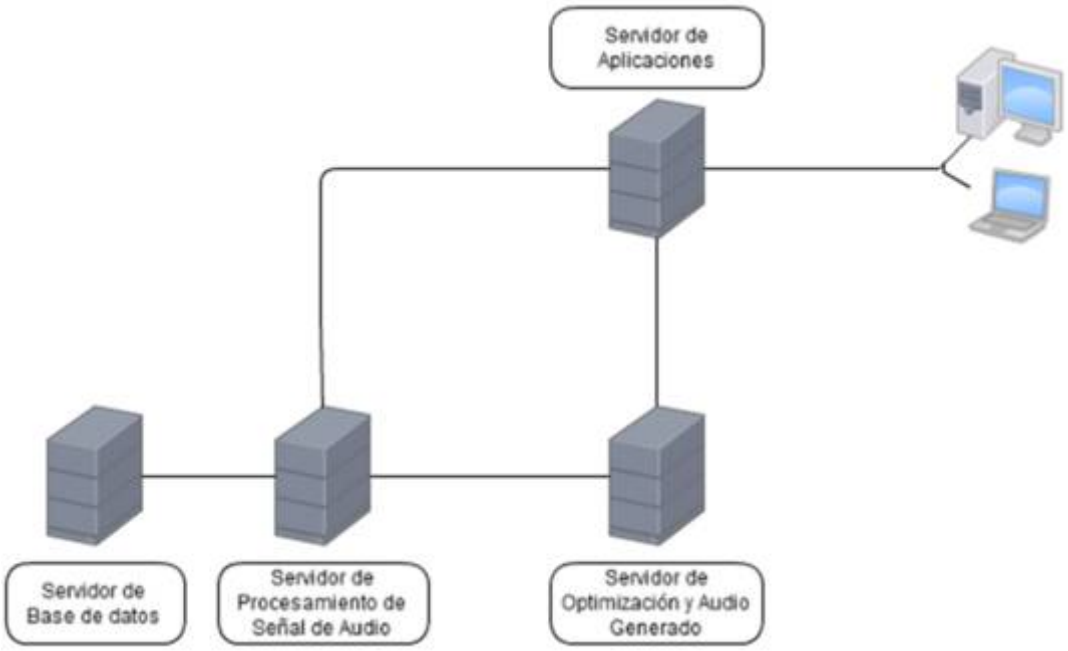
- **Servidor de procesamiento de señal de audio**: este servidor se utiliza para realizar la concatenación de los audios que son el resultado del texto de entrada, los componentes que lo conforman son: Generación de audio y relación con el corpus de voz.

- **Servidor de aplicaciones**: este servidor contiene toda la lógica de la capa de negocio (Controlador) para manejar las entradas de texto, los

componentes que la conforman son: procesamiento y clasificación de texto, método de concatenación y la aplicación del algoritmo de sentimientos.

- **Servidor de optimización y audio generado:** servidor que permite aportar optimizaciones en el audio generado (optimización de tono y duración de unidades fonéticas), permite generar la voz artificial y guardar los audios producidos.

Figura 6. **Esquema general de la arquitectura**



Fuente: elaboración propia.

Para la generación del prototipo, se tomó en cuenta la reducción del tiempo de respuesta y el almacenamiento del corpus de voz, por lo que se utilizó una base de datos no relacional NoSQL, la cual maneja variables no

estructuradas para almacenar el audio. De la misma forma se establece que los servidores anteriormente descritos serán simulados por componentes generados por JAVA, estas bibliotecas conocidas como JAR's.

6.1.3. Modelo vista controlador

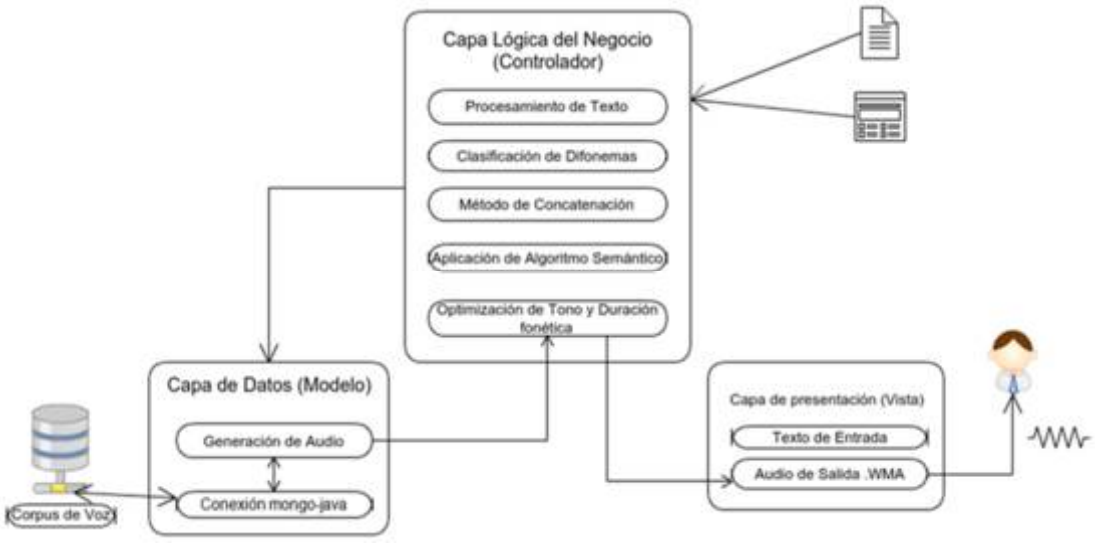
Es importante mencionar que dentro de las solución se propuso un patrón MVC (ver Figura 7) aplicada a una arquitectura modular que permite la facilidad de cambios sin afectar al diseño principal. Dentro de dichas capas se puede mencionar:

- **Capa lógica del negocio (controlador)**: El controlador es la capa que permite la gestión del modelo y la vista. Asimismo se enfoca en la lógica del negocio.
  - **Procesamiento de texto**: se encarga de leer el texto de entrada y lo clasifica en texto simple.
  - **Clasificación de texto**: clasifica las unidades fonológicas que representan la descripción de los sonidos de la lengua en español.
  - **Método de concatenación**: aplicación de método PSOLA para la generación de audio.
  - **Aplicación de algoritmo semántico**: aplicación de método que permite representar el sentido del texto, a partir de su clasificación.
  - **Optimizar tono y duración fonética**: Optimizar la salida recibida del modelo de datos para mejorar los aspectos del tono y la duración fonética mediante la semántica del texto.
- **Capa de datos (modelo)**: El modelo es la capa que representa y controla el acceso de la información del sistema.
  - **Generación de audio**: permite construir un audio obtiene información de texto de entrada y la base de datos, luego de la clasificación y procesamiento.

- **Corpus de voz**: base de datos que contiene la información de tipo audio.

- **Capa de presentación (vista)**: La vista es la capa que interactúa directamente con el usuario en lo que entradas de texto y salidas de audio del sistema se refiere.
  - **Texto de entrada**: texto inicial que se desea convertir a voz artificial
  - **Audio de salida .WAV**: generación de resultado en formato de audio .WAV.

Figura 7. **Arquitectura MVC (modelo vista controlador) general**



Fuente: elaboración propia.

6.1.4. Corpus de voz

El conjunto de grabaciones con formato de audio denominado corpus de voz, se realizó en una base de datos no relacional (NoSQL), la cual se utiliza en el prototipo de sintetizador de voz para la concatenación de los audios para generar la voz artificial a partir de la evaluación y clasificación de texto. En los

siguientes puntos se detallará como se conforma dicho corpus de voz en su estructura y contenido de información.

6.1.4.1. Diseño de corpus

El diseño del corpus de voz, toma en cuenta la creación de dos bases de datos no relacionales, las cuales son:

- **Corpus**: esta base de datos detalla toda la información relacionada a los archivos de audio a guardar en la base de datos **Sonidosdb**, es decir, contiene información sobre el identificador, nombre del audio, fonema, palabra relacionada, el tamaño y un ejemplo de uso. En la Figura 8, se puede visualizar un ejemplo de la estructura utilizada en formato de intercambio de información JSON (JavaScript Object Notation por sus siglas en inglés) para guardar la información de los archivos de audio que están presentes dentro del prototipo.

Figura 8. **Ejemplo JSON utilizado para la base de datos corpus**

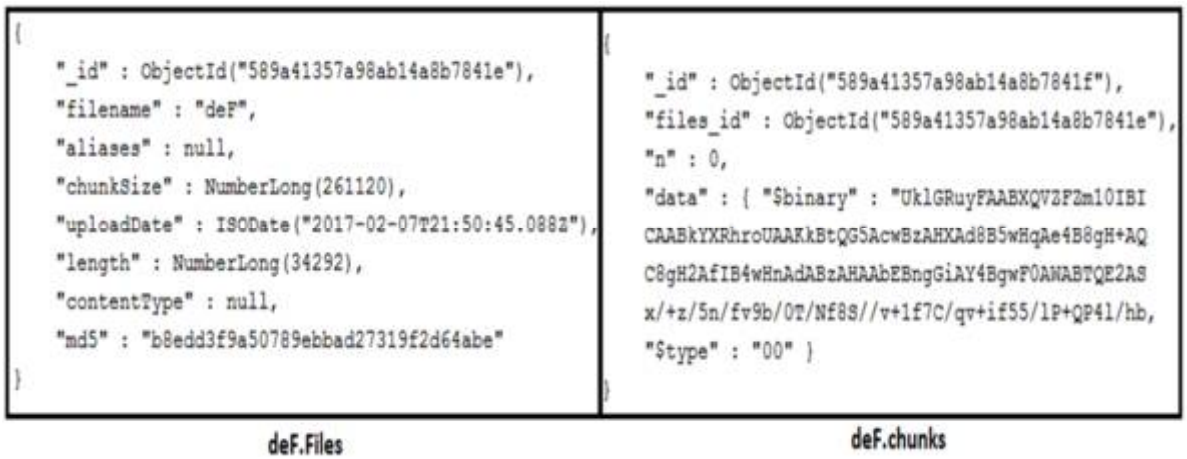
```
{
    "_id" : ObjectId("589a41267a98ab14a8b780da"),
    "Id" : "63",
    "nombre" : "al",
    "sonido" : {
        "audio" : "",
        "sexo" : ""
    },
    "fonema" : "",
    "palabra" : "",
    "tamano" : "",
    "ejemplo" : ""
}
```

Fuente: elaboración propia.

- **Sonidos db**: esta base de datos contiene almacenados los audios en concreto que se utilizan dentro del prototipo. La distribución que utiliza Mongo db para manejar el tipo de archivos de audio es crear dos colecciones por archivo, las cuales se presentan a continuación:
    - **Sección de Files (metadata)**: maneja una sección de archivo por audio, con las propiedades de Id, tamaño, ChunkSize, fecha de carga, hash md5 y nombre del archivo.
    - **N secciones de *chunks***: estas tienen un tamaño de 255KB, por lo que depende del tamaño total del audio genera uno o más *chunks*, guarda información de Id, identificador del archivo padre y un número que determina su secuencia.

Como se puede ver en la figura 9, se presenta un ejemplo JSON de cómo se almacena la información de un archivo de audio en la base de datos de Mongo DB.

Figura 9. **Ejemplo JSON para archivos y chunks**



Fuente: elaboración propia.

6.1.4.2. Características de base de datos

Dentro de las características de la base de datos, luego de cargarlas de información se presentan las propiedades descritas en la Tabla V.

Tabla V. **Características de base de datos**

Propiedades	**Corpus**	**Sonidos db**
Colecciones	3	1,188
Objetos	569	7,117
Tamaño promedio de cada documento	238.762b = 0.00023MB	7628.27b = 0.0076MB
Tamaño total de data	135,856.00b =0.1358MB	54,290,464.00b = 54.290 MB
Tamaño total de espacio asignado a las colecciones	184,320.00b =0.184MB	867,446,784.00b = 867.44 MB
Número de extensiones o conjuntos de índices	5	1,787
Número total de índices en todas las colecciones	1	2,372
Tamaño total de todos los índices creados en la base de datos	32,704.00b = 0.0327MB	19,393,472.00b = 19.393 MB

Continúa Tabla V.

Tamaño total de archivos de data	67,108,864.00b =67.10 MB	1,543,503,872.00b = 1.54 GB
Tamaño total de Colecciones	16MB	16MB

Fuente: elaboración propia.

6.1.4.3. Condiciones de grabación

Para obtener los archivos de audio se realizaron 569 grabaciones, las características para llevar a cabo dicha tarea se presentan en la Tabla VI a continuación.

Tabla VI. **Condiciones de grabación**

Lugar	Casa
Frecuencia de Muestreo	16 KHZ
Resolución	16 bits
Número de Canales	1 mono
Formato de grabación	WAV
Tiempo promedio	0 -1 segundo
Micrófono	Dinámico

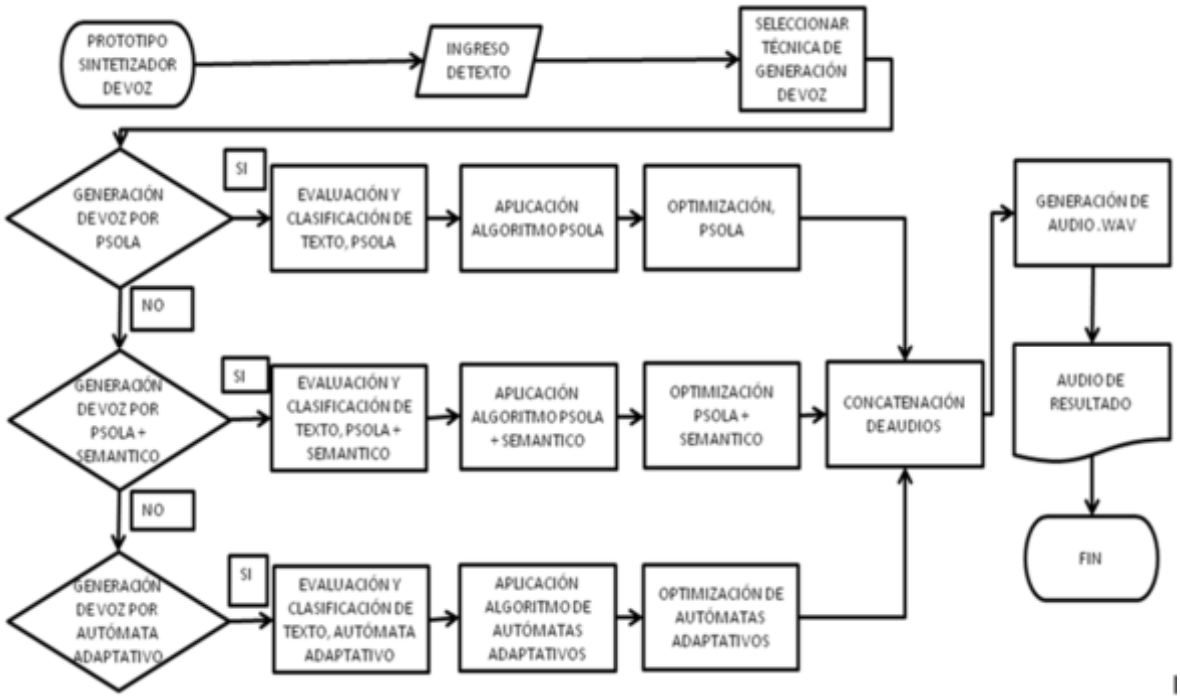
Fuente: elaboración propia.

6.2. Prototipo sintetizador de voz

Para cubrir los objetivos de esta investigación y los aspectos de diseño establecidos, se generó un prototipo como resultado, el cual permite concretamente la generación de voz artificial por tres medios; el primero generar voz artificial utiliza la síntesis por concatenación PSOLA; segundo, la generación de voz artificial utiliza la síntesis por concatenación PSOLA en

combinación de un algoritmo semántico, y por último, la generación de voz artificial utiliza autómatas adaptativos para la clasificación de texto. El flujo de las tres técnicas presentadas se puede apreciar en la Figura 10, la cual indica los procesos que apoyan a la generación del resultado.

Figura 10. **Diagrama de flujo generación de voz artificial 1 iteración**



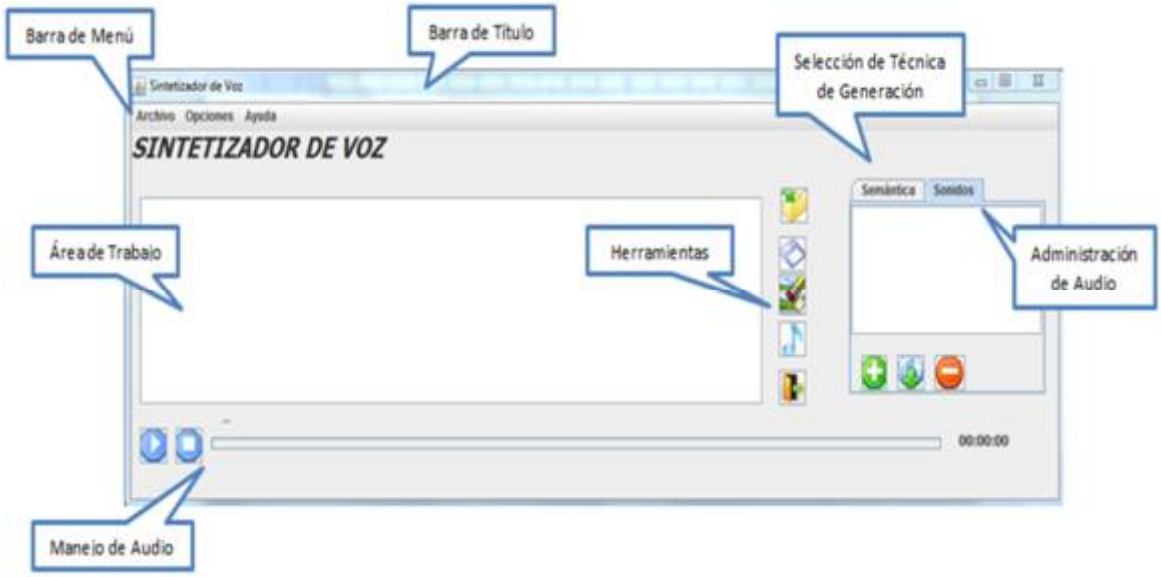
Fuente: elaboración propia.

6.2.1. Funcionalidad del prototipo

El prototipo cuenta con distintas funcionalidades que apoyan a generar el resultado esperado, funciones primarias como la evaluación, clasificación de texto y generación de voz artificial y funciones secundarias como cargar archivos de entrada, descargar y cargar audios, reproducir audios y

configuraciones del prototipo. En la Figura 11 se puede apreciar dichas funcionalidades en general.

Figura 11. **Ventana principal de prototipo sintetizador de voz**



Fuente: elaboración propia.

Es de importancia mencionar que el prototipo cuenta con configuraciones que determinan la dirección local donde se almacenarán todos los audios generados por defecto, la carpeta local que indica donde se encuentran las voces masculinas y femeninas para alimentar el corpus de voz, y por último, el género con el que se desea generar la voz resultante (Ver Figura 12).

Figura 12. **Ventana de configuraciones de prototipo de sintetizador de voz**

Fuente: elaboración propia.

6.2.2. Proceso de sintetizado de voz

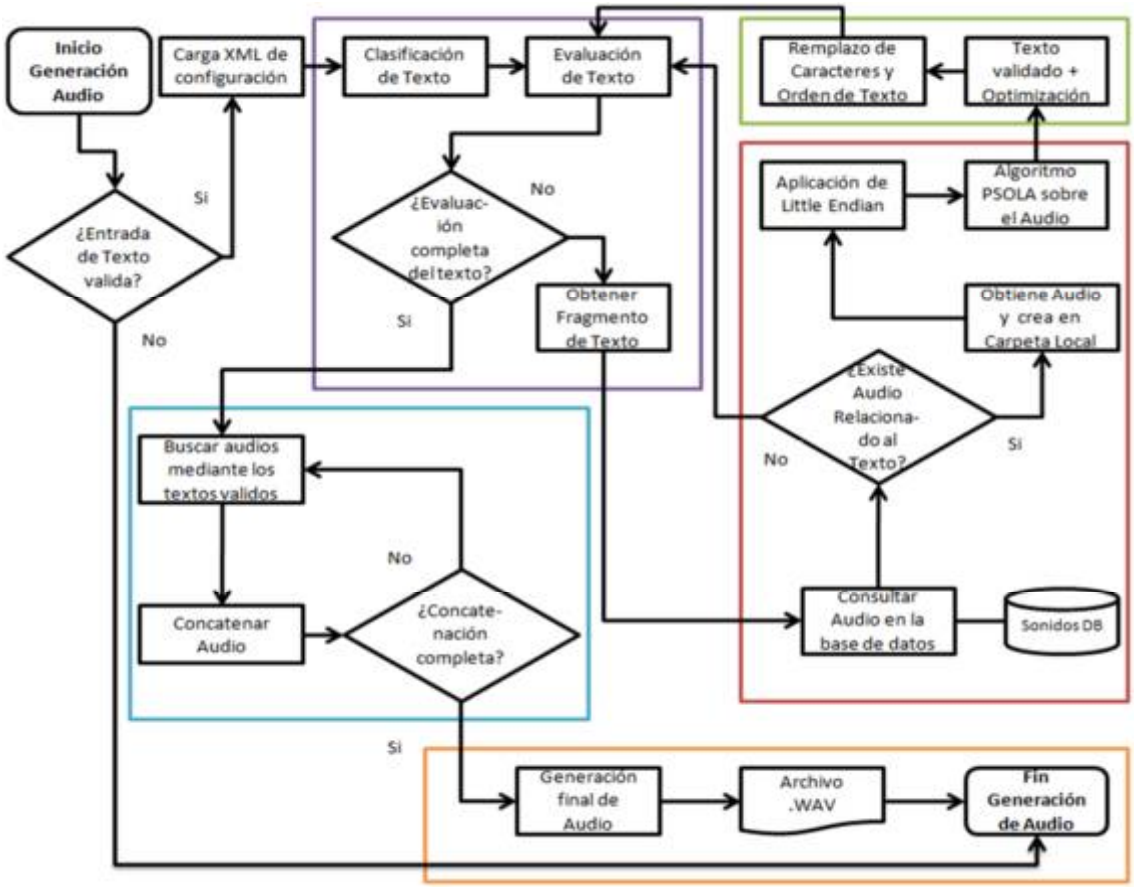
Dentro de esta sección se muestran con mayor detalle las técnicas utilizadas para la generación de voz artificial en el prototipo de sintetizador de voz, resalta la lógica y componentes utilizados para llevar a cabo el resultado esperado.

6.2.2.1. Síntesis por concatenación PSOLA

Este método se encarga de descomponer una serie de ondas elementales, para esta generación se estableció la propiedad *Little Endian,* la cual establece que los bytes de menor peso se almacenen en la dirección más baja de memoria y el byte de mayor peso en la dirección más alta, esto con la intención de hacerlo más sencillo de implementar.

En la Figura 13 se puede ver el diagrama de flujo que muestra los procesos que se llevan a cabo para generar un archivo de audio con la técnica de síntesis por concatenación PSOLA. Donde el recuadro de color morado identifica el proceso de clasificación y evaluación de texto; el recuadro de color verde, la optimización, el recuadro rojo, el algoritmo PSOLA; el recuadro celeste, concatenación de audios, y por último, el recuadro naranja muestra el proceso de generación de audio.

Figura 13. **Diagrama de flujo síntesis por concatenación PSOLA**



Fuente: elaboración propia.

6.2.2.2. Síntesis por concatenación PSOLA en combinación de algoritmo semántico

Como parte de la investigación y con el objetivo de obtener mejor calidad en la generación de voz artificial, aumentar las características de tono y duración de unidades fonéticas se realizó la implementación de un algoritmo semántico que en este caso es un algoritmo sentimental, el cual se consume de un servicio en internet, dicho servicio es prestado por la empresa *Meaning Cloud* LLC. Esta es una empresa estadounidense con base en New York, que se especializa en software de análisis semántico, con 20 años de experiencia tecnológica.

¿Cuál es el proceso para evaluar texto con el algoritmo de análisis de sentimientos de *Meaning Cloud*?, este servicio como se mencionó se encuentra en la nube, para poder consultarlo se tiene que crear una cuenta en el sitio www.meaningcloud.com, la cual permite un plan gratis con 40,000 peticiones mensuales. Dicho plan cuenta con una llave que se utiliza para realizar las consultas vía http, también permite la configuración del lenguaje del texto ingresado y el formato con el que se desea obtener la respuesta (XML o JSON). En la Figura 14 se puede ver un ejemplo de la respuesta en formato XML que se utiliza para determinar el análisis de sentimientos de la palabra “feliz”. Donde la etiqueta <score_tag> contiene el sentimiento del texto, que en este caso es P, el cual se cataloga como positivo (feliz).

Figura 14. **Respuesta XML análisis de sentimientos, palabra feliz**


<sentence_list>
<sentence>
<text>
<![CDATA[feliz]]>
</text>
<inip>0</inip>
<endp>4</endp>
<bop>y</bop>
<confidence>100</confidence>
<score_tag>P</score_tag>
<agreement>AGREEMENT</agreement>
<segment_list>

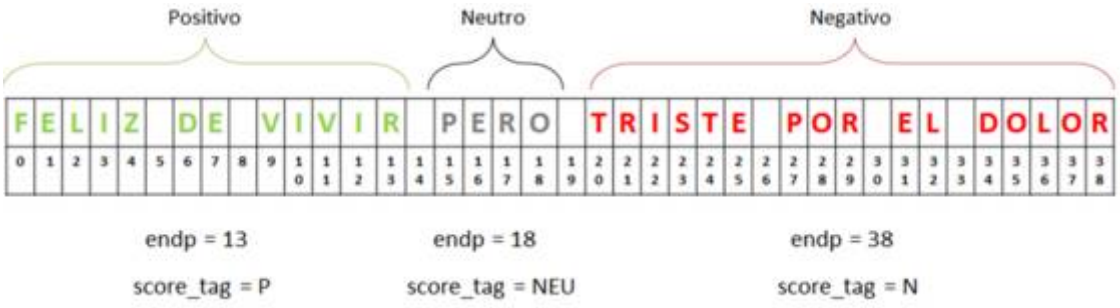
<text>
<![CDATA[feliz]]>
</text>
<inip>0</inip>
<endp>4</endp>
<confidence>100</confidence>
<score_tag>P</score_tag>
<agreement>AGREEMENT</agreement>
<polarity_term_list>
<polarity_term>
<text>
<![CDATA[feliz]]>
</text>
<inip>0</inip>
<endp>4</endp>
<confidence>100</confidence>
<score_tag>P</score_tag>
</polarity_term>
</polarity_term_list>

</segment_list>
</sentence>
</sentence_list>
</response>

Fuente: Sentiment Analysis – Console https://www.*meaningcloud*.com/developer/sentiment-analysis/console Consulta: 10 de febrero de 2017.

Toma en cuenta lo anterior, el prototipo toma el resultado del análisis de sentimientos y se enfoca en las etiquetas , <endp> y <score_tag>. Recorre todos los segmentos (fragmentos de texto), los cuales indican donde terminan en la etiqueta de <endp>, esto permite determinar donde inicia y termina el texto que tenga algún sentimiento para realizar su respectiva clasificación. Por ejemplo, se analizó la frase "**Feliz de vivir, pero triste por el dolor**" y se obtuvo un XML como el de la Figura 14 pero con más segmentos, dado que el texto presenta más de un sentimiento. En la Figura 15, se puede ver el resultado de la clasificación del texto al aplicarle el algoritmo de sentimiento, así como los valores de las etiquetas que permiten identificar los distintos límites.

Figura 15. **Clasificación de texto por método semántico del prototipo**



Fuente: elaboración propia.

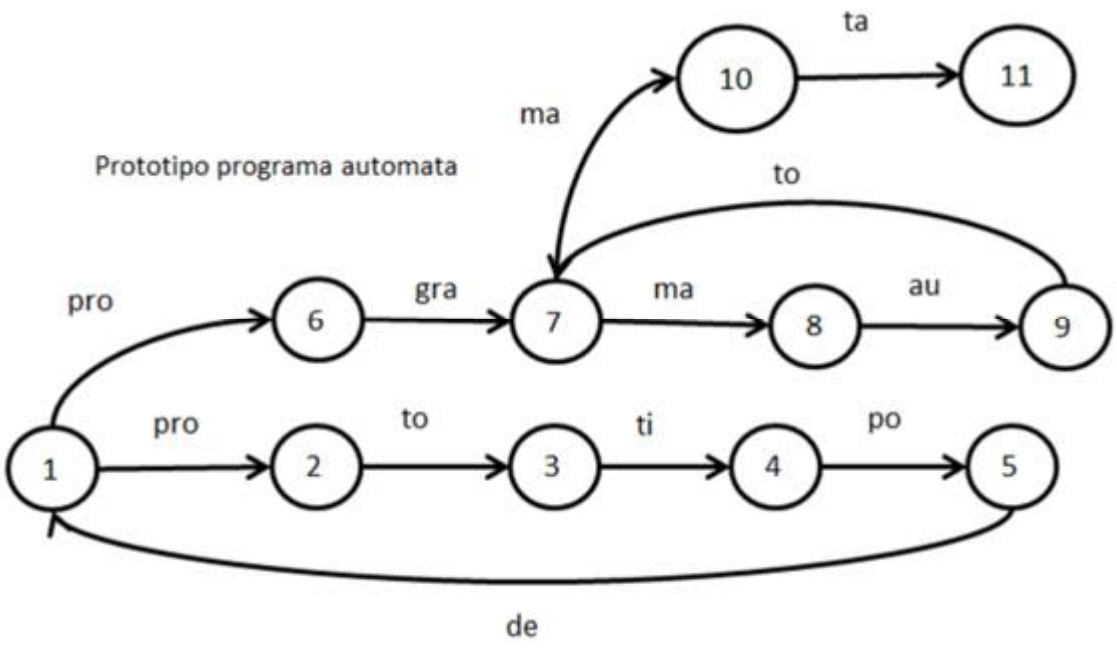
En resumen, la aplicación de síntesis por concatenación PSOLA en combinación de un algoritmo semántico, se realiza con el mismo proceso de concatenación PSOLA de la técnica anterior con dos únicos cambios, las cuales son las entradas, dado que se identifica el sentido del texto y luego el cambio sobre la generación y selección de audios a concatenar.

6.2.2.3. Autómatas adaptativos

La tercera técnica con la que cuenta el prototipo es la generación de voz artificial, mediante estados de autómatas adaptativos, estos estados se representan por silabas, es decir, la clasificación de texto que se realiza al generar la voz es por silabas del texto ingresado y contiene estados finitos.

En la Figura 16 se puede apreciar un ejemplo de la evaluación de estados que se realiza dentro del prototipo, dichos estados son parte de la clasificación de texto que se realiza.

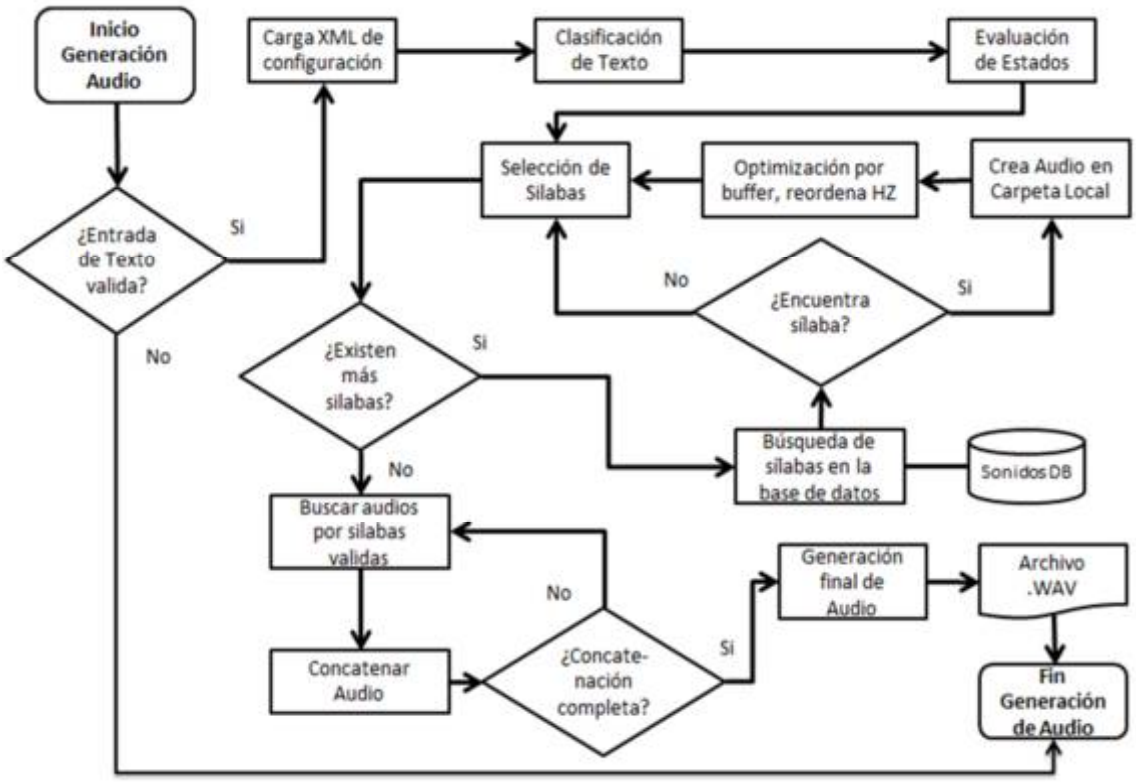
Figura 16. **Evaluación de estados del prototipo**



Fuente: elaboración propia.

Luego de la clasificación del texto y evaluación se realiza el proceso que se presenta en la Figura 17, el cual muestra la relación sobre la aplicación de estados del autómata en la evaluación del texto y la recuperación de audio en la base de datos de Mongo DB.

Figura 17. **Proceso de autómata adaptativo**



Fuente: elaboración propia.

6.3. Síntesis de presentación de resultados

Como parte de la investigación y el cumplimiento de los objetivos propuestos, se realizaron dos tipos de análisis: el primero fue un estudio de caso sobre las variables lingüísticas que afectan al entendimiento de la voz

artificial en lo que a características del tono y la duración de las unidades fonéticas de la voz se refieren, y el segundo, la evaluación de encuestas que fueron analizadas mediante estadística descriptiva.

6.3.1. Ambiente de experimentos

El proceso de experimentación se llevó a cabo en dos tipos de ambientes: uno de forma electrónica (análisis cuantitativo), y el otro, de forma física (análisis cualitativo).

6.3.1.1. Análisis cualitativo

El experimento que se interpretó por análisis cualitativo se enfocó en el cumplimiento del objetivo específico que determina el tipo de variables lingüísticas que se pueden utilizar para mejorar la calidad del tono y la duración de voz generada por un sintetizador de voz. Dicho experimento se realizó en 6 fases, estas son las siguientes:

- **Conceptualización del tema**: para entablar una buena comunicación se deben de cumplir ciertas características que apoyen al entendimiento de un mensaje, se puede mencionar por ejemplo, el tono de la voz, medirlo adecuadamente para que no sea tan agresivo o tímido, encontrar un equilibrio. Por otro lado, la duración de las unidades de la voz también influye en la correcta pronunciación de las palabras, en la forma de cómo se combinan las mismas y en la duración que le agrega. También existen otro tipo de características como las variantes lingüísticas que determinan factores externos como el nivel sociocultural, el territorio geográfico, el cambio del vocabulario en función del tiempo y la diferencia del tipo de persona con quien realiza la comunicación.

- **Selección y énfasis en un fenómeno en particular**: en la presente investigación, se determinó qué tipo de variables o variantes lingüísticas afecta a la calidad del tono y la duración de voz en la comunicación oral en el idioma español en el territorio guatemalteco. Por esto, se plantearon las siguientes preguntas:
    - ¿Cuáles son las variables o variantes lingüísticas?
    - ¿Cómo afecta el tono y la duración fonética de la voz a comprender un mensaje de comunicación oral?
    - ¿Qué variables o variantes lingüísticas afectan al tono y duración de la voz, para mejorar la comprensión de un mensaje cuando se sostiene una conversación en idioma español?
- **Recolección de información**: recolección de datos en bruto de las entrevistas y documentos como archivos, reportes, artículos y propuestas que permitan obtener hechos o experiencias. En la Tabla II se puede apreciar la estructura de la entrevista, que realizó a 16 personas.
- **Organización, clasificación y edición de los datos**: organización de las ideas relacionadas a las entrevistas, clasificación de ideas y las relaciones encontradas. Esto se puede ver en el Anexo1, la cual contiene los datos de las entrevistas en una matriz de resultados.
- **Triangulación de datos:** permite apoyar los argumentos que se obtuvieron con respecto al trabajo de campo (entrevistas). Por lo que en la Tabla VII se puede ver la relación y resultados obtenidos de las entrevistas.

Tabla VII. **Triangulación, resultados de entrevista**

Resultados de entrevista			
No.	**Pregunta**	**Resultado**	**Relación con preguntas principales**
1	Cuando usted escucha hablar a una persona de uno de los departamentos de Guatemala, ¿le es fácil entenderle?	El 75% de las personas indicó que sí le es fácil entender	¿Cómo afecta el tono y la duración fonética de la voz a comprender un mensaje de comunicación oral?
2	¿Qué diferencias cree usted que hay en la forma de hablar de las personas de diferentes departamentos?	Acento, tono de voz, modismos y lengua materna	
3	¿Cree usted que la rapidez con la que hablen las personas afecta entender sus palabras? ¿Podría explicarme?	Sí, depende de la costumbre	
4	¿Es usted capaz de entender lo que le dice una persona cuando le habla rápido?	La mayoría de las veces sí, pero depende de la pronunciación	
5	¿Le es fácil entender lo que dicen los jóvenes cuando tiene alguna conversación con ellos? ¿Puede darme detalles?	En un 90% de las veces sí	¿Cuáles son las variables o variantes lingüísticas?
6	¿Qué palabras le parecen desconocidas cuando tiene alguna conversación con otra persona?	Yolo, abreviaturas, modismos, LOL, apear	
7	¿Por lo general, usted habla con personas en ambientes libre de distracciones? ¿Me puede ampliar?	Sí, la mayoría de las veces, el sonido distrae, pero no afecta totalmente	¿Qué variables o variantes lingüísticas afectan al tono y duración de la voz para mejorar la comprensión de un mensaje cuando se sostiene una conversación en idioma español?
8	¿Considera que las personas, por lo general, hablan muy fuerte o muy callados? ¿Puede darme algunos ejemplos?	Depende de las circunstancias	¿Cómo afecta el tono y la duración fonética de la voz a comprender un mensaje de comunicación oral?
9	¿En las conversaciones que usted sostiene con otras personas, usted nota que las palabras se pronuncian completas o no?	La mayoría no, utilizan abreviaturas por moda, utilizan muletas, otras sí hablan bien	
10	¿Cree usted que las personas pronuncian las palabras adecuadamente? ¿Por qué?	Depende de la educación o la lectura de la persona	

Continúa Tabla VII.

11	¿Considera usted que el desconocimiento de la ortografía o la gramática influye en que las personas pronuncien bien o mal las palabras?	Sí, porque si las escriben mal las pronuncian mal	¿Cuáles son las variables o variantes lingüísticas?
12	¿Cree usted que la diferencias sociales afectan en la forma como hablan las personas?	Depende más del nivel educativo y que tan culta sea la persona	
13	¿Estima usted que las personas se expresan de manera diferente en situaciones diferente? ¿Puede darme algún ejemplo?	Sí, depende de las emociones o de con quien se hable	
14	¿El uso de palabras extranjeras es común en conversaciones que usted sostiene con otras personas?	Sí el uso de inglés	
15	¿Usted modifica su lenguaje y/o características de su voz al hablar con diferentes tipos de personas? ¿De qué forma?	Depende del entorno y la persona	
16	¿Cuáles cree que sean las características necesarias en la voz para comprender mejor un mensaje?	Tono moderado, voz clara, fluida, pronunciación, duración de la voz y palabras claras	¿Qué variables o variantes lingüísticas afectan al tono y duración de la voz para mejorar la comprensión de un mensaje cuando se sostiene una conversación en idioma español?

Fuente: elaboración propia.

- **Interpretación y generación de reporte**: Luego se identificaron los siguientes puntos:
    - Las variables lingüísticas aportan distintas características al lenguaje las cuales pueden mejorar o perjudicar el entendimiento del mensaje, dentro de estas se encuentran la variable geográfica, estilística (lenguaje formal e informal), sociocultural (sexo, cultura,

edad, etc.) y las diacrónicas (vocabulario que cambia con el paso del tiempo).

- En las entrevistas realizadas se brindó un punto de vista más general sobre lo que se piensa del entendimiento de la comunicación oral, en donde el factor de pronunciación y el tono de la voz marcan una diferencia significativa para comprender un mensaje. Parte de esto, se puede definir que existen influencias externas que pueden marcar el entendimiento de un mensaje como lo son las geográficas, dado que cada región tiene distintas formas de hablar y comunicarse.
- Para determinar el grado de relación de las variables lingüísticas con las características de la voz como: el tono y la duración fonética se evaluaron las diferentes fuentes, que permitieron generar una matriz de medición (ver Tabla VIII), las cuales muestran el grado de relación que tienen cada una con el propósito de definir qué combinación de variables y características de la voz apoya mejor a comprender mejor un mensaje.

La clasificación se realizó con base a tres niveles: alto, medio y bajo. Los cuales interpretan rangos de aceptación sobre la comprensión de un mensaje; el nivel bajo, una comprensión escasa e insuficiente; el nivel medio, una comprensión parcial, y el nivel alto, una comprensión considerable o total. Estos definen la importancia de las variables lingüísticas en relación a las características de la voz (tono y duración de unidades fonéticas), dicha importancia se detalla en la Tabla VIII, la cual identifica que variables presentan mayor influencia.

Tabla VIII. **Nivel de relación variables lingüísticas y características de la voz**

VARIABLES		Nivel de relación para apoyar la comprensión de un mensaje	
LINGÜÍSTICAS	**Geográfica**	Media	Alta
	Estilística	Alta	Alta
	Sociocultural	Media	Baja
	Diacrónicas	Baja	Baja
		TONO DE VOZ	**DURACIÓN FONÉTICA**
		CARACTERÍSTICAS DE LA VOZ	

Fuente: elaboración propia.

6.3.1.2. Análisis cuantitativo

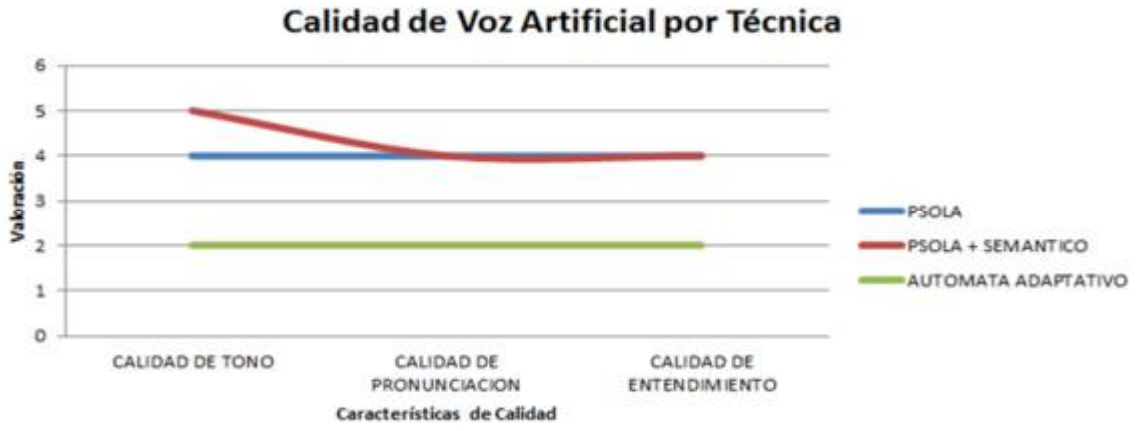
El experimento que se interpretó por análisis cuantitativo, se enfocó en el cumplimiento de los objetivos específicos de evaluar el valor que un algoritmo semántico aporta en el algoritmo PSOLA al entendimiento de la voz artificial producida por un sintetizador de voz y comparar si los autómatas adaptativos pueden mejorar el tono y la duración de la fonética del sintetizador de voz. Para esto se obtuvieron los resultados de las encuestas realizadas por formularios de Google, esta se conformó de 4 secciones, en donde las primeras 3 son de evaluación de cada técnica por separado, y la última sección compara las tres técnicas en conjunto. Es importante mencionar que dentro de las encuestas que se realizaron no se dio a conocer el nombre de la técnica que generó cada

resultado de audio artificial, para no influir en las respuestas de los encuestados.

Los nombres que se utilizaron para esto fueron gama (técnica síntesis por concatenación PSOLA), delta (técnica síntesis por concatenación PSOLA en combinación de algoritmo semántico) y beta (técnica de autómata adaptativo).

Las encuestas fueron respondidas por 20 personas en donde se evaluaron las tres técnicas de generación de voz por separado y como un todo. Al evaluar los resultados por análisis de frecuencias encontramos que, toma en cuenta las características de la voz, más específicamente en la calidad, las técnicas que utilizan PSOLA fueron muy parecidas pero la que utilizó el análisis semántico obtuvo un 20% más de aceptación (Ver Figura 18) en la característica de calidad de tono, es decir, que es la característica más notoria de la voz artificial. La valoración indica los niveles de aceptación en donde 0 es muy malo, 1 es malo, 2 es debe mejorar, 3 es bien, 4 es muy bien y 5 es excelente.

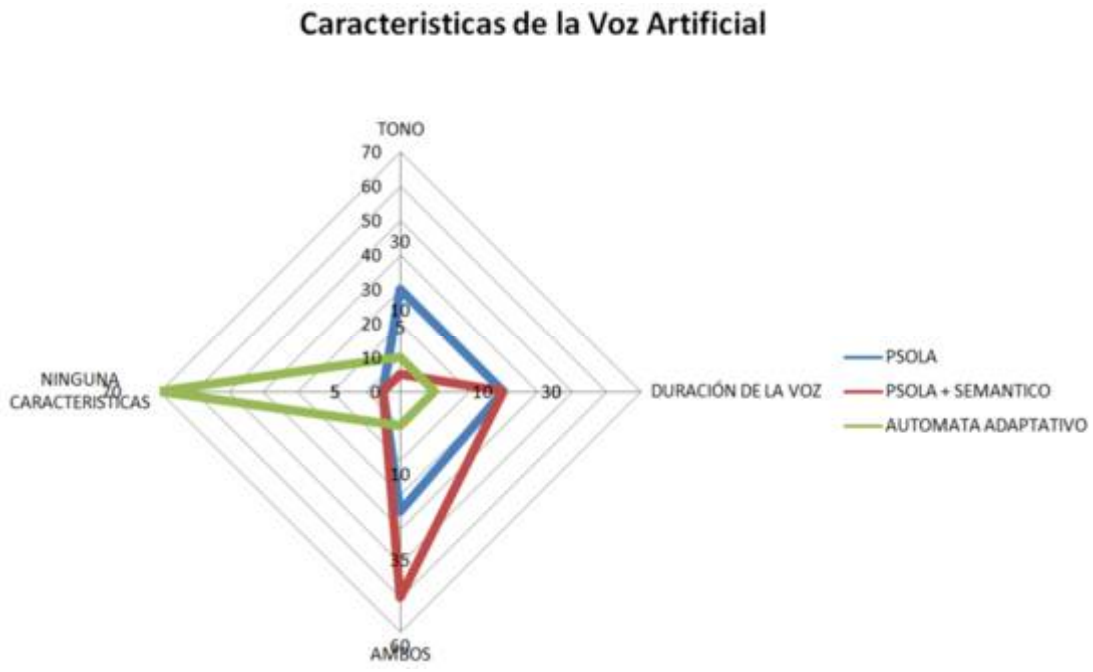
Figura 18. **Gráfica de calidad de la voz artificial**



Fuente: elaboración propia.

Luego al evaluar la presencia y el grado de aceptación de las características de la voz, dentro de las técnicas de generación de voz artificial y no cual se identifica más como se hizo en la Figura 18, se presenta que la técnica de autómata adaptativo se orienta más a no contar con la presencia de ninguna de las características, el algoritmo de síntesis por concatenación PSOLA cuenta con las dos características, pero en menor proporción; por último, se aprecia que la técnica PSOLA en combinación de un algoritmo semántico destaca las dos propiedades de la voz (Tono y Duración de la voz) más que contar con cada una por separado (Ver Figura 19).

Figura 19. **Gráfica de características de la voz artificial**

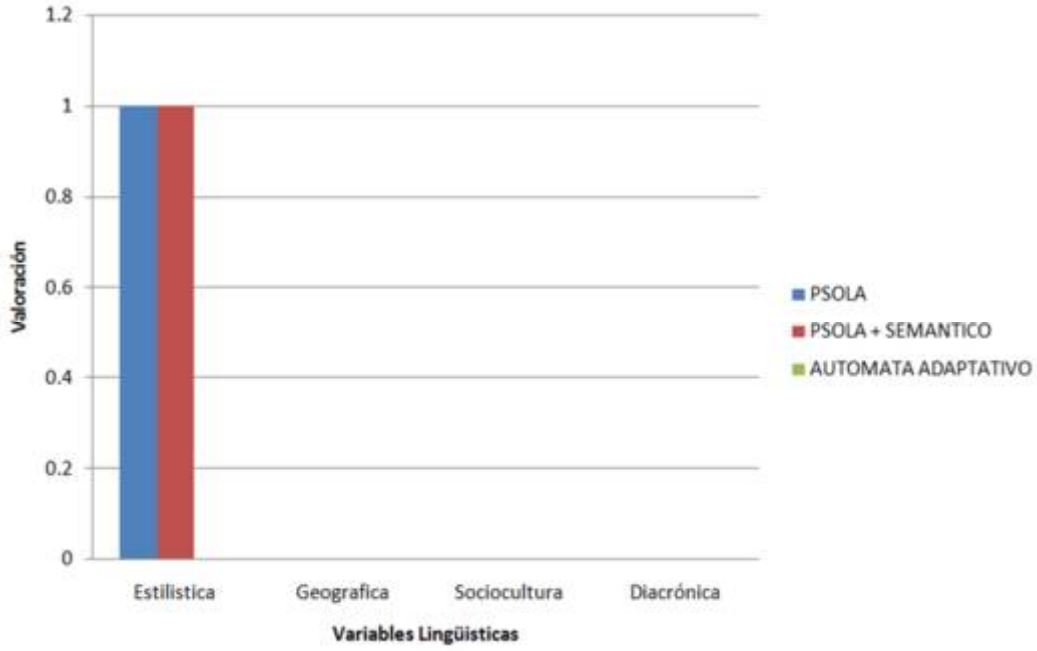


Fuente: elaboración propia.

Por último, en la evaluación de cada una de las técnicas por separado se puede resaltar que la variable lingüística que más influye es la estilística, pero

solo en las técnicas que utilizan PSOLA; por otro lado, se obtiene poca o ninguna presencia de las otras variables lingüísticas, la cual no es significativa para la investigación. Es importante mencionar que la valoración se representa como: 0.2 el 20% hasta 1 el 100% de las respuestas obtenidas (Ver Figura 20).

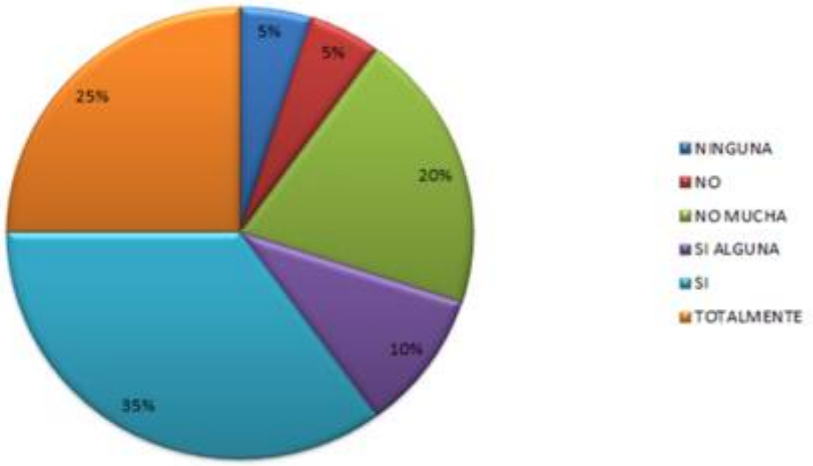
Figura 20. **Gráfica de presencia de variables lingüísticas por técnica**



Fuente: elaboración propia.

El algoritmo sentimental en combinación con la síntesis por concatenación PSOLA presenta que el 60% de los encuestados (Ver Figura 21) identificó algún tipo de emoción significativa en la voz artificial escuchada, toma en cuenta un rango de aceptación 4-5. Donde el número 4 significa que se encontró algún tipo de emoción en el audio presentado y el número 5 significa que totalmente se encuentro un tipo de emoción en el audio presentado.

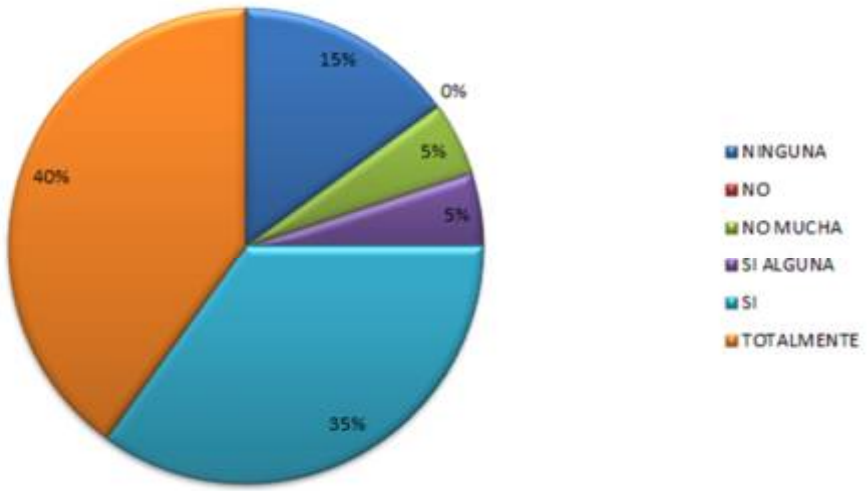
Figura 21. **Gráfica de presencia de emoción en el audio de síntesis por concatenación PSOLA en combinación de algoritmo semántico**



Fuente: elaboración propia

Por otra parte, en la evaluación de las encuestas que se enfocaron a tomar en cuenta las técnicas del prototipo en conjunto, se resalta en la Figura 22 que el 75% de los encuestados reconoce una mejora del audio artificial generado por la técnica síntesis por concatenación PSOLA en combinación de algoritmo semántico sobre la técnica de síntesis por concatenación PSOLA. Toma en cuenta un rango de aceptación de 4-5. Donde el número 4 significa que sí se reconoce una mejora del audio artificial, generado por la técnica de síntesis por concatenación PSOLA en combinación de un algoritmo semántico sobre la síntesis, por concatenación PSOLA, y el número 5 significa que totalmente es mejor la técnica de síntesis por concatenación PSOLA en combinación de un algoritmo semántico.

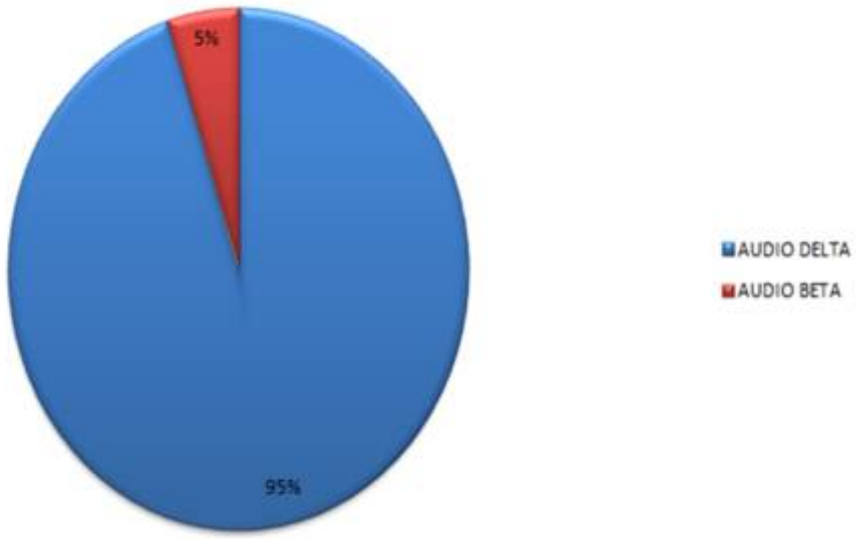
Figura 22. **Gráfica comparación de mejora de la síntesis por concatenación PSOLA en combinación de algoritmo semántico sobre síntesis por concatenación PSOLA**



Fuente: elaboración propia.

En una comparación directa entre la síntesis por concatenación PSOLA en combinación de un algoritmo semántico, y la técnica por Autómata Adaptativo obtiene que el 95% de las personas identifican que existen mejoras en el tono y la pronunciación de la voz artificial sobre la técnica que incluye el algoritmo semántico (Ver Figura 23).

Figura 23. **Gráfica comparación síntesis por concatenación PSOLA en combinación de un algoritmo semántico (audio delta) y autómatas adaptativos (audio beta)**

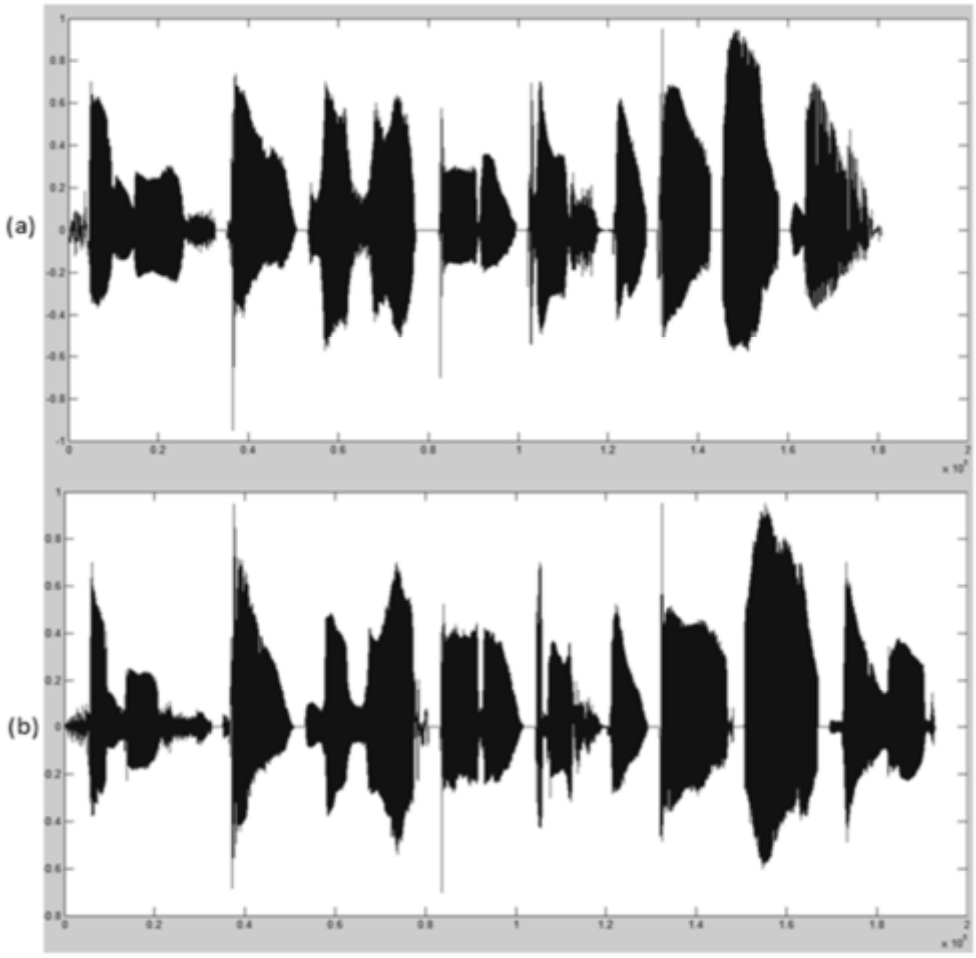


Fuente: elaboración propia.

6.3.2. Comparación de audios

Como parte del experimento, se realizó una comparación de la señal de audio “Feliz de vivir pero triste por el dolor”, en donde la Figura 24 (a) muestra la señal generada por la técnica de Síntesis por concatenación PSOLA y la Figura 24 (b) muestra la señal generada por la técnica de Síntesis por concatenación PSOLA en combinación de un algoritmo semántico. La principal diferencia que pueda notar son los sonidos altos y bajos, otro punto a resaltar es que los dos audios tienen un tiempo de duración de 4 segundos.

Figura 24. **Comparación de señal de audio**

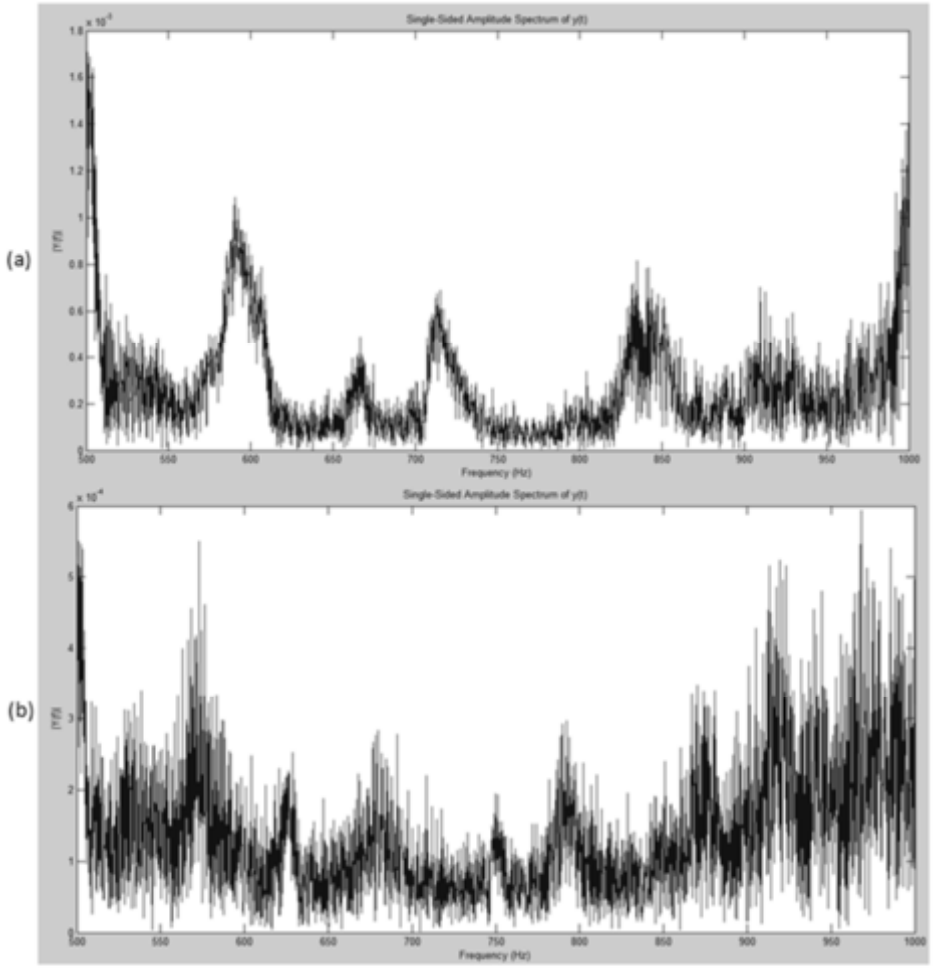


Fuente: elaboración propia, con base a Matlab.

Por otra parte, se realizó un análisis de espectro de amplitud de señal, donde se puede apreciar la frecuencia de los audios (Ver Figura 25). Al agregarle más emoción a la pronunciación de la segunda técnica se aprecia que su frecuencia se ve más pronunciada y abarca mayor área, en otras palabras su tono de voz se ve alterado por las diferentes emociones que se

encontraron con el algoritmo semántico y cuando la voz se torna más triste tiende a aumentar la frecuencia.

Figura 25. **Comparación de frecuencias de audio**



(a) Frecuencia Síntesis por Concatenación PSOLA

(b) Frecuencia Síntesis por concatenación PSOLA + algoritmo semántico

Fuente: elaboración propia, con base a Matlab.

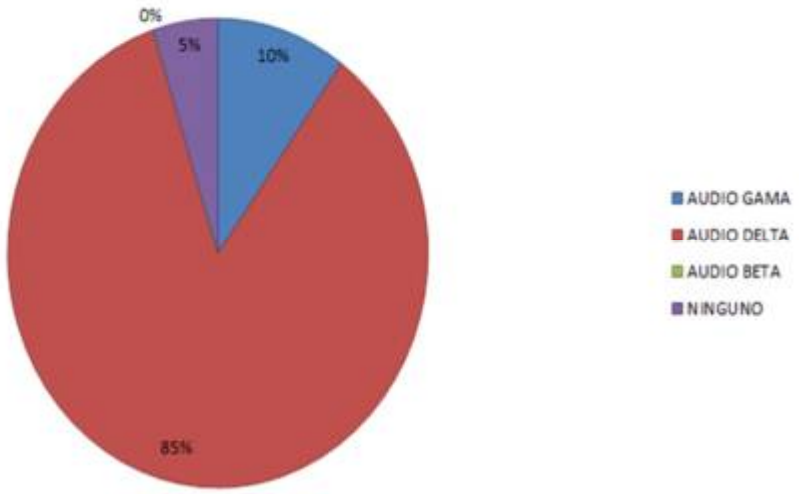
6.3.3. Resultados finales

En el presente trabajo se utilizaron las variables lingüísticas que representan factores de diversificación gramatical que aporta una serie de normas que permiten que la lengua adopte reglas de entendimiento uniforme. A diferencia de la variedad estándar como el vulgarismo que introduce extranjerismos al idioma español y no son indicadores concretos. De acuerdo a lo presentado en la Tabla VIII, la variable lingüística que junto al tono de la voz y duración fonética influye más en la comprensión y mejora de la calidad de la voz, es la variable lingüística Estilística (comunicación formal e informal), por otra parte el estudio refleja que el aspecto que menos toma en cuenta la calidad de voz, así como su influencia en la comprensión de un mensaje, es la variable diacrónica (vocabulario que cambia con el paso del tiempo). En resumen, se toma en cuenta la influencia sobre el tono y la duración de unidades fonéticas de las variables lingüísticas se presenta el siguiente listado de mayor a menor grado de influencia:

- Estilística
- Geográfica
- Sociocultural
- Diacrónica

Como parte de los resultados del grupo 2 que analizó las tres técnicas en conjunto, se presenta que 17 personas de la muestra total de 20, identificaron que la técnica que genera el audio más entendible y que a su criterio suena mejor es la técnica de síntesis por concatenación PSOLA en combinación de un algoritmo semántico (Ver Figura 26).

Figura 26. **Gráfico comparación de técnicas en conjunto, PSOLA (audio gama), PSOLA + semántico (audio delta) y autómata adaptativo (audio beta)**

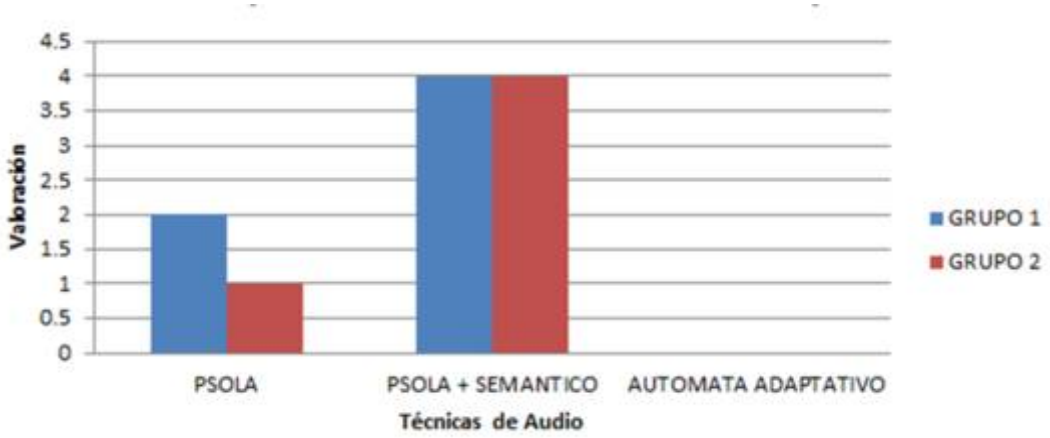


Fuente: elaboración propia.

Luego del análisis de los dos grupos: evaluación de las técnicas de generación de voz artificial por separado y en conjunto, se identificó que la técnica que más resalta es la de síntesis por concatenación PSOLA en combinación con un algoritmo semántico, considera las características de la voz como el tono y la duración de unidades fonéticas, las cuales benefician el entendimiento de la voz artificial producida por un sintetizador de voz. Asimismo, se comparó la técnica de autómata adaptativo y la síntesis por concatenación PSOLA en combinación de un algoritmo semántico y se determinó que los autómatas adaptativos no aportan mayor beneficio al entendimiento de un mensaje toma en cuenta las características del tono y la duración fonética de un sintetizador de voz del idioma español para Guatemala.

En la Figura 27, se puede apreciar el nivel de valoración para cada técnica de audio en los grupos propuestos en las técnicas de análisis de información dentro del marco metodológico. Dicho nivel de valoración se refiere al número de veces que una técnica resalta significativamente sobre las demás, en las diferentes evaluaciones realizadas. La técnica que implementa un algoritmo semántico en combinación de un algoritmo PSOLA la que más características positivas resalta dentro del grupo 1 y 2.

Figura 27. **Gráfica comparación de resultado**



Fuente: elaboración propia.

En resumen, se evaluaron las técnicas de generación de voz artificial del prototipo que fue desarrollado. Luego se evidencio que el aporte que brinda un algoritmo semántico a un algoritmo PSOLA es significativo en el entendimiento de la voz artificial, dado que al realizar las comparaciones contra las otras técnicas individualmente y en conjunto, resalta que en los dos grupos el algoritmo semántico mejora las características del tono y la duración de unidades fonéticas que mejoran el entendimiento de la voz artificial. Por otra parte, al realizar la comparación directa entre la técnica de síntesis por

concatenación PSOLA en combinación de un algoritmo semántico, y el autómata adaptativo se demostró que los autómatas no aportan valor para mejorar el tono ni la duración de la fonética de un sintetizador de voz, esto en las condiciones presentadas dentro de la investigación.

7. DISCUSIÓN DE RESULTADOS

7.1. Discusión de rendimiento de sintetizador de voz

La principal fuente de información del sintetizador de voz desarrollado es el corpus de voz, por lo que el desempeño del mismo depende de la forma en la que se agregaron y se realizaron las grabaciones de voz. Parte de esto se toman en cuenta las variantes o variables lingüísticas que más afectan a las características de la voz, como el tono y duración de unidades fonéticas. Toma en cuenta los resultados obtenidos en la investigación sobre la influencia de las variables lingüísticas, nos enfocamos en la Tabla VIII en donde la variable lingüística con mayor presencia en las características de la voz mencionadas es la estilística. Por tal hecho, se identifica que grabar la voz del corpus con características más formales en el tono y la pronunciación (duración de las unidades fonéticas) mejora el entendimiento del lenguaje y sugiere un incremento en la naturalidad.

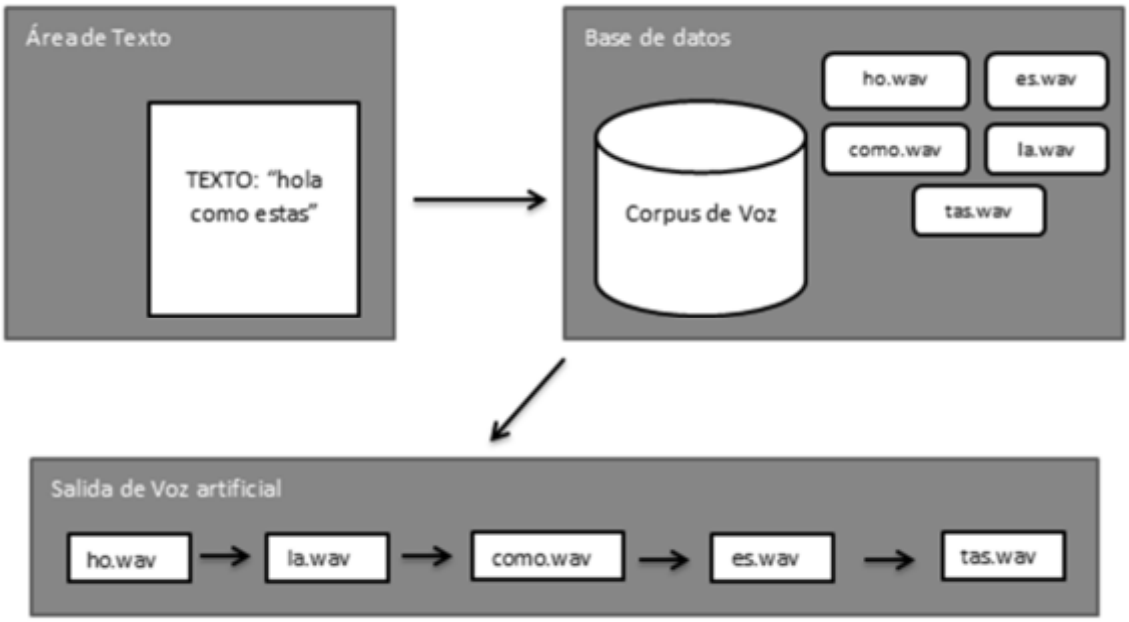
Como función principal del prototipo desarrollado, se encuentra la opción de realizar un análisis semántico del texto, más específicamente un análisis de sentimientos que se consume de *meanin gcloud*. Con este algoritmo se permite identificar el sentido del texto y así determinar el aspecto ideal para la generación de voz artificial, además permite el análisis de datos para encontrar el significado de contenido no estructurado. Un estudio del 2010 realizado por Claudia Correa, Hoover F. y Henry Arguello el cual tiene por nombre: “Síntesis por concatenación de difonemas para el español de Colombia” se enfocó en crear voz artificial mediante pares de fonemas, para esto utilizaron una base de datos transaccional y su voz resultante se escucha robótica y con problemas de

naturalidad en la concatenación. Aunque es una técnica que resalta los acentos y pronunciación de Colombia. El prototipo generado por la presente investigación permite una clasificación diferente del texto de entrada, además de una base de datos no sql y la implementación de un algoritmo sentimental. Continua con la base de datos, esta fue seleccionada para reducir el tiempo de respuesta y tener acceso más rápido a los datos al contar con el audio almacenado directamente. Por otra parte, aproximadamente para encontrar un audio dentro de la base de datos se toma un tiempo máximo de 0.040 segundos y un tiempo mínimo de 0.01 segundos, parte de esto se puede calcular que el tiempo total de generación de audio final es afectado directamente por el número de archivos de audio que se deseen consultar. Por ejemplo, si se tienen 15 audios el tiempo de respuesta máximo seria de 15 * 0.04 = 0.6 segundos y el tiempo mínimo de respuesta de 15 * 0.01 = 0.15 segundos, entonces en promedio se tendrá ((0.04+0.01)/2) * 15 = 0.375 segundos de respuesta.

Otro estudio, es el que propone Alexandre Trilla en el 2009, titulado “Natural Language Processing techniques in Text-to-Speech Synthesis and Automatic Speech Recognition” el cual se enfoca en las técnicas de procesamiento del lenguaje natural, este muestra más atención a técnicas de clasificación y resumen de texto pero esto presenta problemas de lentitud y son muy tediosos, por este antecedente la investigación se enfoca en los valores de entrada del prototipo que pueden brindar mejor flexibilidad para identificar diferentes combinaciones de textos, es decir, el clasificador de texto que se desarrolló permite interpretar una gran cantidad de combinaciones de audios del corpus de voz. Dado que la clasificación permite obtener audios ya sea que estos se encuentren en diferentes estructuras como: frases, trío de palabras, pares de palabras, palabras, silabas o letras en concreto. Por ejemplo, se buscó la frase “hola como estas” como texto de entrada, y por otro lado tenemos en el corpus de voz los audios: ho.wav, es.wav, como.wav, la.wav y tas.wav. Con

esto se obtuvo por silabas la palabra “hola”, por palabra completa “como” y por silabas de nuevo la palabra “estas” para formar la frase (Ver Figura 28).

Figura 28. **Clasificación y generación de audio**



Fuente: elaboración propia.

Como se puede apreciar a lo largo de la investigación se han desarrollado n cantidad de enfoques respecto a la generación de voz artificial, esto se refleja en los antecedentes descritos, con la presente investigación se tomó el enfoque sobre 3 puntos que a nuestra percepción apoyan a la construcción de un sintetizador de voz, los cuales son:

- La clasificación del texto de entrada.
- Añadir una base de datos No SQL para el manejo de información no estructurada como los archivos de audio.
- La aplicación de un algoritmo semántico para brindar mayor naturalidad a la voz artificial generada, mejora las características del tono y la duración de unidades fonéticas.

7.2. Avances logrados

Al evaluar los resultados obtenidos de la investigación y el prototipo realizado, se resaltan ciertos avances que permiten influir positivamente en el entendimiento de la voz artificial generada por un sintetizador de voz, toma en cuenta las características del tono y duración de unidades fonéticas de la voz.

7.2.1. Algoritmo semántico aplicado a síntesis por concatenación PSOLA

Tomando en cuenta la voz artificial que se genera mediante el prototipo por las 3 diferentes técnicas desarrolladas, se estableció que la síntesis por concatenación PSOLA en combinación de un algoritmo semántico mejora la característica del tono de la voz artificial en un 20% con respecto a la síntesis por concatenación PSOLA y en un 60% respecto al Autómata Adaptativo, como se puede apreciar en la Figura 18. Otra forma de visualizar dicho evento es en la Figura 25, dado que se puede identificar una variación significativa sobre la frecuencia de audio y la intensidad del sonido. Parte de la técnica de síntesis por concatenación PSOLA en combinación de un algoritmo semántico resalta algún tipo de emoción (como se puede apreciar en la gráfica de la Figura 21), y como resultado grafico un mayor espectro de área que la técnica de síntesis por concatenación PSOLA.

Un hecho curioso que se puede resaltar del análisis de frecuencias de audio, es que cuando se presenta alguna emoción en el audio artificial de tipo triste según el clasificador de sentimientos (*Meaning cloud*), este presenta mayor espectro de área aunque a nuestros oídos suene el tono de voz más bajo.

Otro aspecto a destacar es que cuando se evaluaron todas las técnicas en lo que a características de la voz se refiere, se aprecia en la Figura 19 que la técnica de síntesis por concatenación PSOLA en combinación de un algoritmo semántico tiene más presencia cuando se toman en cuenta 2 características en conjunto (tono y duración de las unidades fonéticas) más que cuando se evalúan cada una por separado. Analiza este punto de vista se obtiene una mejora que aporta un valor significativo sobre la técnica de síntesis por concatenación PSOLA normal, la cual aunque se muestra estable no resalta alguna característica en concreto.

Por lo que la presencia de un algoritmo semántico puede mejorar significativamente el resultado en términos de tono y duración de unidades fonéticas de la voz artificial, dado que brinda un nivel de aceptación más familiar y nos ayuda a identificar una comunicación más natural, resalta hechos más humanos esto dentro de los términos en los que se enfoca la presente investigación como lo son los sentimientos de tristeza, felicidad y neutro.

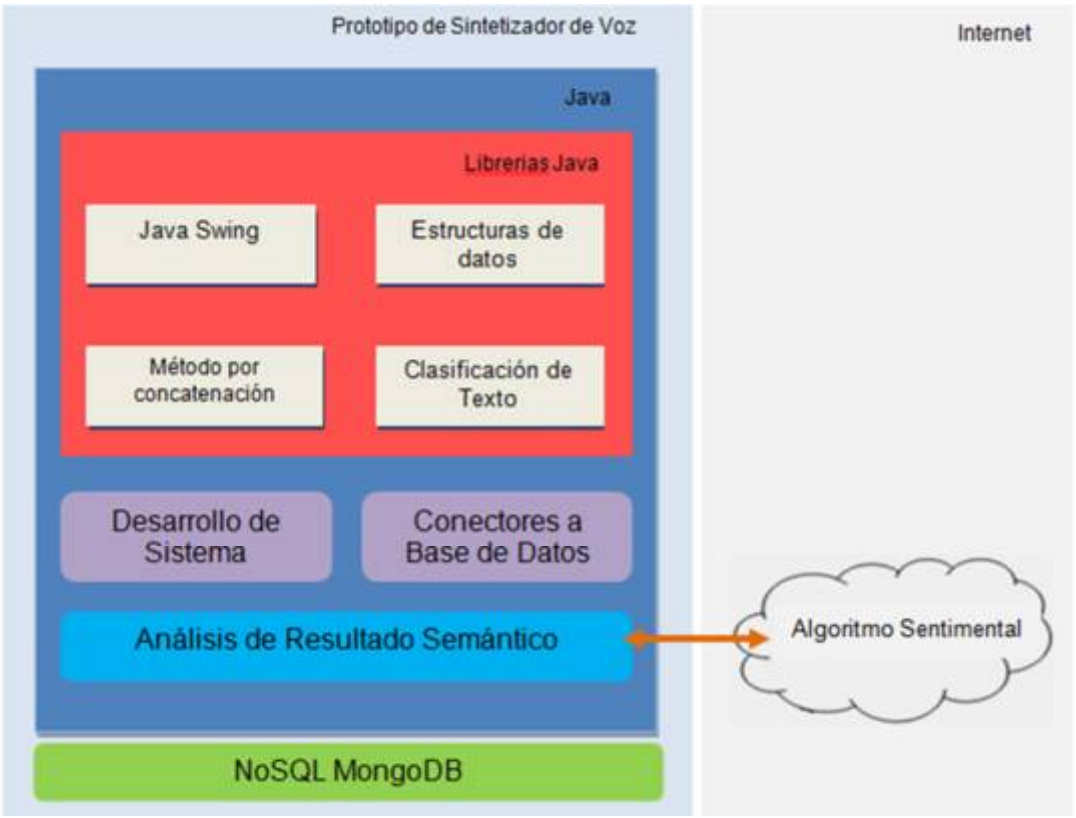
Por último, al revisar el análisis de las técnicas de generación de voz del prototipo se puede destacar que el tipo de comparación, ya sea cada técnica por separado o en conjunto, la técnica de autómatas adaptativos no mejora las características de la voz como tono y duración de unidades fonéticas dado que en todas las evaluaciones no presenta ninguna mejora, como se aprecia en la Figura 23 y Figura 26. Esto comparándola con las otras técnicas y toma en

cuenta las condiciones propuestas en el desarrollo del prototipo de la presente investigación, dado que el resultado puede cambiar si los factores como el corpus de voz o la clasificación de texto se realizan de forma distinta.

7.2.2. Tecnología utilizada

Dentro de las tecnologías utilizadas como se puede apreciar en la Figura 29 se encuentran las librerías de JAVA, tales como: arrays, vectores, iterators, linkedList y herramientas de JAVAX SWING. Por otra parte, contiene métodos de evaluación y clasificación de texto, algoritmo semántico que determina el sentido del texto y clasificación de autómatas adaptativos por estados en silabas. La herramienta de desarrollo que se utilizó fue NetBeans IDE 8.1, debido a su facilidad de crear interfaces y forma de uso, dentro de esta se utilizaron conectores de mongo-java para facilitar la conexión con el corpus de voz desarrollado en la base de datos nosql Mongo db. Como parte del algoritmo semántico implementado se utilizó un servicio en la nube (*Meaning Cloud*) el cual determina mediante un texto la clasificación de emoción, estas son texto positivo (feliz), neutro y texto negativo (triste).

Figura 29. **Diagrama de Tecnologías**



Fuente: elaboración propia.

7.3. Impacto

Como objetivo implícito de la investigación, se quiere crear algo útil que aporte valor al conocimiento general, por lo cual se considera que se puede alcanzar algún impacto positivo en el ámbito educativo, social y tecnológico.

7.3.1. Educativo

La educación es un factor muy importante en el crecimiento personal y en el proceso de adaptación, dado que apoya a la comunicación de mejor manera

y aporta crecimiento nuestro acervo cultural. Como aporte educativo del prototipo generado por la investigación, se presenta una forma fácil de conocer la pronunciación de las palabras en idioma español para Guatemala, así como conocer modismos o expresiones que utilizamos los guatemaltecos, siempre y cuando sea parte de nuestro corpus de voz. También es de utilidad para los niños que se encuentran aprendiendo a leer porque les permite por un medio electrónico jugar con las palabras y replicar lo sonidos.

7.3.2. Social

Una sociedad cuenta con diversos problemas que se pueden apoyar con tecnología para ir mejorándolos paulatinamente. Con la intención de esto y a manera de apoyo social, el prototipo generado por la investigación aporta valor a las personas que tengan algún desorden del habla para que puedan practicar la pronunciación de las palabras al repetirlas muchas veces y practicar tanto como lo crean necesario o a personas con algún desorden de lectura (dislexia) para que el prototipo pueda ir leyendo texto junto a la persona y exista una mejor comprensión de la lectura. Asimismo, el prototipo puede ser útil para las personas no videntes o con poca capacidad visual, al brindarles acceso a la información cuando cuenten con un texto que este dentro de la computadora.

7.3.3. Tecnológico

El crecimiento tecnológico ha sido impresionante en los últimos años, existe tecnología casi para todo y para todos. En este caso, el prototipo que se generó como resultado de esta investigación, es uno de los primeros sintetizadores de voz desarrollados especialmente para Guatemala y toma en cuenta que se utilizó el apoyo de la tecnología de análisis de sentimientos de *Meaning cloud*. Además de utilizar esta tecnología y reconocer el sentido del

texto en diferentes emociones como un mensaje positivo (feliz), neutro o negativo (triste) es un aporte tecnológico al mezclarlo con un sintetizador de voz, el cual permite la interpretación de frases, trio de palabras, pares de palabras, palabras, silabas o letras individualmente. Asimismo, se integró un corpus de voz desarrollado en una base de datos nosql para que los audios se guardaran directamente en la base de datos con la intención de mejorar y reducir el tiempo de respuesta.

7.4. Propuesta a futuro

Actualmente, el prototipo se encuentra orientado al idioma español para Guatemala, pero existe la posibilidad que se pueda utilizar con otro corpus de voz, es decir, una base de datos alimentada con frases, palabras, silabas o letras con la fonética y modismos de otro país de Latinoamérica o que tengan como lengua el idioma español, para que con sus propias tonalidades y expresiones puedan generar voz, a partir de un texto escrito en español, esto toma en cuenta las limitaciones que se describen en la investigación.

Dado que el prototipo de esta investigación es una aplicación de escritorio, un paso futuro que se puede dar es subir un corpus de voz a la web y realizar las conexiones pertinentes con el prototipo para que este pueda ser consumido desde la tecnología Applet de JAVA. Y de esta forma hacerlo un programa web.

Con respecto al servidor donde se encuentra almacenado el corpus de voz se tienen las siguientes características:

- Procesador Intel Dual Core
- 4 GB RAM
- Disco de 500 GB
- Windows 7, 64 bits

Al mejorar dichas características se brinda una mejora significativa al rendimiento de respuesta de los sonidos almacenados en la base de datos. Por ejemplo, se podría implementar mejores especificaciones como:

- Procesador Intel Core I7
- 16 GB RAM
- Disco de 1T
- Windows Server 2016

7.5. Limitaciones

- La aplicación de sintetizador de voz en idioma español tiene las siguientes limitaciones:
  - El analizador de texto de entrada no valida el texto con signos de puntuación, tildes, caracteres extraños o lenguaje que no pertenezca al idioma español.
  - La generación de voz artificial se encuentra limitada por la cantidad de elementos en el corpus de voz, es decir, que el prototipo será capaz de generar la voz a partir de un texto pero solo de las voces que se encuentren grabadas en la base de datos.
  - El reproductor de audio con el que cuenta el prototipo solo permite hacer las funciones básicas como reproducir, pausar, detener y seleccionar audio.
  - El tipo de audio que utiliza el prototipo es .WAV para incluir otro tipo de audio se deben de realizar las pruebas pertinentes.

CONCLUSIONES

1. Se desarrolló el prototipo sintetizador de voz para Guatemala, el cual implementa un algoritmo semántico al algoritmo de síntesis por concatenación PSOLA, con ello se obtuvo una mejora del 60% en el tono y la duración de unidades fonéticas de la voz y un 20% sobre la calidad y aceptación del tono de la síntesis de voz artificial en español. La aplicación de un algoritmo semántico presenta mayor influencia sobre la clasificación y evaluación del texto de entrada del sintetizador de voz.

2. Se determinaron las variables lingüísticas que tienen influencia sobre las características de la voz, la variable estilística con mayor presencia sobre la mejora de la calidad del tono, la duración de unidades fonéticas de la voz artificial de un sintetizador de voz, y la variable diacrónica con menor influencia.

3. Se evaluó el valor que aporta un algoritmo semántico al combinarlo con un algoritmo de síntesis por concatenación PSOLA, se determinó que presenta una mejora significativa al entendimiento de la voz artificial producida por un sintetizador de voz, al mejorar en un 60% la presencia de algún tipo de emoción en el tono y la pronunciación de las palabras, para obtener una voz artificial más natural. Además resalta un incremento del 75% de entendimiento con respecto a la técnica de generación de voz de un algoritmo PSOLA convencional.

4. Se comparó la técnica de autómatas adaptativos contra la técnica de síntesis por concatenación PSOLA en combinación de un algoritmo semántico, y se determinó que los autómatas adaptativos no aportan mayor beneficio al entendimiento de un mensaje y no mejoran el tono y la duración de la fonética de la voz artificial de un sintetizador de voz. Toma en cuenta que el audio artificial generado por la técnica de autómatas adaptativos produce sonidos muy robóticos y poco entendibles.

RECOMENDACIONES

1. Para mejorar la calidad del prototipo de Sintetizador de Voz, se sugiere la implementación de nuevos algoritmos en la clasificación y evaluación del texto de entrada que validen la representación de todos los elementos del idioma español, tales como: signos de puntuación, exclamación, admiración, tildes y todos los signos ortográficos que brindan una mejora en la comprensión de la voz artificial en el idioma español, para Guatemala.

2. Se sugiere la utilización de un ambiente más hermético para realizar las grabaciones de voz que pertenecen al corpus, con esto se mejoraría y amplificaría la claridad de los sonidos de voz generados por el prototipo.

3. Se puede implementar una base de datos distribuida del concepto nosql, para incrementar su velocidad de respuesta, incrementar el número de audios almacenados en la base de datos, y permitir que se consulte desde diferentes puntos.

4. Es recomendable implementar una arquitectura REST (Representational State Transfer, por sus siglas en inglés) que permita el consumo de servicios HTTP para crear una aplicación web que utilice todos los algoritmos del sintetizador de voz y que sea accesible desde cualquier dispositivo conectado a internet.

REFERENCIAS BIBLIOGRÁFÍCAS

1. Adeyemo, O. O., &Idowu, A. (2015). Development and integration of Text to Speech Usability Interface for Visually Impaired Users in Yoruba language. *African Journal of Computing &ICT* , 8(1), 87-94.

2. Agüero, P. D. (2012). Sintetizador de voz aplicada a la traducción voz a voz (Doctoral dissertation, Tesis Doctoral. Universidad Politécnica de Cataluña. http://hdl.handle.net/10803/97035).

3. Alande, S. S., Sharma, S. S., & Chavan, A. A. (2015). Text to Speech Converter. *International Journal Of Computer Science And Applications*, *8*(2).

4. Alonso, A., Sainz, I., Erro, D., Navas, E., &Hernaez, I. (2013). Sistema de conversión texto a voz de código abierto para lenguas ibéricas. Procesamiento del lenguaje natural, 51, 169-175.

5. Alvarado, G. (2013) Desarrollo de juegos educativos para niños con enfermedades prolongadas en centros hospitalarios development of educational games for children with long terms diseases in hospital Universidad tecnológica de Panamá.

6. Amar Amar, J. J. (2011). Educación infantil y desarrollo social. Investigación & Desarrollo, (7).

7. Barroso, L. G. (2012). Análisis descriptivo de la entonación andaluza en oraciones interrogativas y exclamativas. *LL Journal*, *7*(2).

8. Bascón Pantoja, E. (2011). El patrón de diseño Modelo-Vista-Controlador (MVC) y su implementación en Java Swing. *Revista Acta Nova*, *2*(4).

9. Benito Oterino, B., Gaspar Escribano, J. M., Rivas Medina, A., Martínez Díaz, J. J., Rodríguez, O., & Ramírez, M. S. (2011). Evaluación de la peligrosidad sísmica en España para aplicaciones relacionadas con la seguridad nuclear. Resultados preliminares.

10. Brown, A. (2014). Pronunciation and Phonetics. New York: Routledge, 1(1), 3-6

11. Calero, A. (2014). Fluidez lectora y evaluación formativa. Investigación sobre Lectura, España: Asociación española de compresión lectora, 1, 33-48.

12. Capobianco, M., & en Ciencias, L. (2014). Módulo 3: Arquitectura. Universidad nacional del sur, Argentina

13. Chodorow, K. (2013). *MongoDB: the definitive guide*. " O'Reilly Media, Inc.".

14. Correa, P., Rueda, H., & Arguello, H. (2010). Síntesis de voz por concatenación de difonemas para el español de Colombia. Revista Iberoamericana en Sistemas, Cibernéticos e Informática, 7(1), 19-24.

15. DÂRDALĂ, M. (2008). Software System for Vocal Rendering of Printed Documents. Revista Informatica Economicănr, 2(46), 90.

16. de la Vega Segura, L.E., & Camacho, J. H Diseño de un sintetizador de voz del idioma español hablado en México.

17. del Busto, H. G., & Enríquez, O. Y. (2013). Bases de datos NoSQL. *Revista Telem@ tica*, *11*(3), 21-33.

18. Fasold, R. (1990). The Sociolinguistic of Language. Oxford.

19. Fong, T., Thorpe, C., &Baur, C. (2003). Collaboration, dialogue, human-robot interaction. In Robotics Research (pp. 255-266). Springer Berlin Heidelberg.

20. Fresneda, M. D., & Mendoza, E. (2005). Trastorno especifico del lenguaje: Concepto, clasificaciones y criterios de identificación. Revista de Neurología, 41(1).

21. Fromkin, V. A. (Ed.). (2014). *Tone: A linguistic survey*. New York: Academic Press.

22. González, Y. D., & Romero, Y. F. (2012). Patrón Modelo-Vista-Controlador. *Revista Telem@ tica*, *11*(1), 47-57.

23. Gutiérrez, C. M. (2006). Semántica cognitiva: modelos cognitivos y espacios mentales. *A Parte Rei: revista de filosofía*, (43), 5.

24. HERNÁNDEZ, R. V. R., ESCANDÓN, J. M. S., ACOSTA, C. A. Á., & RIVERA, R. A. Comparación empírica entre el proceso unificado y

el desarrollo de software por prototipos. In *19th Annual Western Hemispheric Trade Conference April 15-17, 2015*| *Laredo, TX, USA Conference Proceedings* (p. 235).

25. Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2008). Teoría de autómatas, lenguajes y computación. *Pearson Prentice-Hall*. (p. 3)

26. Jewalikar, V. (2009). Improving automatic phonetic segmentation for creating singing voice synthesizer corpora (Tesis de maestría). Universitat Pompeu Fabra, Barcelona.

27. Jun, S. A. (Ed.). (2014). Prosodic typology II: the phonology of intonation and phrasing (Vol. 2). Oxford University Press.

28. Kayte, S., Waghmare, K., & Gawali, B. (2015). Marathi Speech Synthesis: A review. International Journal on Recent and Innovation Trends in Computing and Communication, 3, 3708-3711.

29. Kenny, D. (2014). *Lexis and Creativity in Translation: A Corpus Based Approach*. New York: Routledge.

30. King, S. (2014). Measuring a decade of progress in text-to-speech. Loquens,1(1), e006.

31. Liberal, M., & Nazaret, M. (2015). Análisis de la competencia básica comunicativo-lingüística y de la comprensión lectora en alumnos/as marroquíes. (Tesis doctoral). Universidad de Extremadura, España.

32. Marcos, F. G. (1999). Fundamentos críticos de sociolingüística (Vol. 9). Universidad Almería.

33. Muñoz Builes, D. M. (2013). Análisis prosódico de un informante de Medellín en el marco de la metodología del proyecto amper (Atlas Multimedia de Prosodia del Espacio Románico (Tesis de Maestría). Universidad de Antioquia, Medellín.

34. Ottinger, J. B., Minter, D., & Linwood, J. (2014). An Introduction to Hibernate 4.2. In *Beginning Hibernate* (pp. 1-7). Apress.

35. Penny, R. (2006). What did sociolinguistics ever do for language history?: The contribution of sociolinguistic theory to the diachronic study of Spanish. Spanish in Context, 3(1), 49-62.

36. Pérez Ibarra, M., Valdiviezo, L. M., Pérez Otero, N., Liberatori, H., Rexachs del Rosario, D., Luque Fadón, E., & Lasserre, C. M. (2010). CLUSIM: simulador de clusters para aplicaciones de cómputo de altas prestaciones basado en OMNeT++. In *XVI Congreso Argentino de Ciencias de la Computación*.

37. Pons, Y. M., Rodríguez, A. E. L., &Maribona, M. G. (2012). Arquitectura de Software para la Plataforma de Gestión de Aprendizaje ZERA. *Serie Científica*,*5*(2).

38. Roca, J. M. (1990). Situación actual de la síntesis de voz. Estudios de fonética experimental, 4, 147-166.

39. Rodríguez Inés, P., Hurtado Albir, A., &Beeby, A. (2008). Uso de corpus electrónicos en la formación de traductores (inglés-español-inglés) (Tesis doctoral). Universitat Autónoma de Barcelona, Barcelona.

40. Rodríguez, W. R., Vaquero, C., Saz, O., & Lleida, E. (2008). Aplicación de las tecnologías del habla al desarrollo del pre lenguaje y el lenguaje. In *IFMBE PROCEEDINGS* (Vol. 18, No. 2, p. 1064). SPRINGER SCIENCE+ BUSINESS MEDIA.

41. Rogers, H. (2014). *The sounds of language: An introduction to phonetics*. New York: Routledge.

42. Sainz, I., Erro, D., Navas, E., &Hernáez, I. (2011). A Hybrid TTS Approach for Prosody and Acoustic Modules. In *INTERSPEECH* (pp. 333-336).

43. Trilla, A. (2009). Natural Language Processing techniques in Text-To-Speech synthesis and Automatic Speech Recognition. Department of Tecnologies Media Enginyeria I Arquitectura La Salle (Universitat Ramon Llull), Barcelona, Spain 2009.

44. Trujillo Sánchez, J., & Roig de Zárate, J. (2008). Motor de veu natural per dispositius mòbils. Universitat Autónoma de Barcelona, Barcelona.

45. Valenzo, M. R., Valencia, R. E. C., & Castro, J. M. M. (2013). Integración de búsquedas de texto completo en Bases de Datos NoSQL. *Vínculos*, *8*(1), 80-92.

46. Varghese, J. M., & Hande, S. (2015). Design of Gujarati Text-to-Speech System. *International Journal of Research*, *2*(5), 1017-1019.

47. Yamagishi, J., Veaux, C., King, S., &Renals, S. (2012). Speech synthesis technologies for individuals with vocal disabilities: Voice banking and reconstruction. Acoustical Science and Technology, 33(1), 1-5.

48. Miranda, C. H., Guzman, J., & Santamaria, R. (2017). A review of Sentiment Analysis in Spanish. TECCIENCIA, 12(22), 6.

49. Olivares Poggi, C. A. (2016). Revisión sistemática sobre la aplicación de ontologías de dominio en el análisis de sentimiento.

ANEXOS

ANEXO 1: Tablas de matriz de resultados

La matriz de resultados es la simplificación de las entrevistas analizadas, esta se presenta en las siguientes 4 tablas.

Tabla 1. **Matriz de resultados 1**

Resultados (aspectos a resaltar)					
	No.	**Entrevistado 1**	**Entrevistado 2**	**Entrevistado 3**	**Entrevistado 4**
PREGUNTAS	**1**	A la mayoría Los acentos cuestan	No por Modismos Por acento	Es fácil	No siempre
	2	Tono de voz Alta voz Rápido Acento	Forma de hablar Por cultura	Modismos	Diferentes etnias Lengua materna
	3	Si pronuncian bien no dificulta	Si afecta	No afecta mientras sea español	Si afecta
	4	Si se complica	Cuando conozco a la persona si	Cuando sea en español.	La mayoría de veces no
	5	Si por lo general No definición exacta pero si el sentido	Algunas veces si Por palabras de otras culturas	No tengo problema Modismos muy rebuscados	Si lo entiendo
	6	Se comprenden según el contexto	Yolo	No recuerdo Palabra buscada en el diccionario	No recuerdo
	7	Existe distracción Ruido Vehículos Personas hablando	Si factores como ruido Música Distractores	Dificulta la comunicación	Ruido (personas, trafico)
	8	Algunas hablan fuerte o muy callado	Depende el ambiente	Depende la persona	Hablan callado
	9	Reducen las palabras o frases Abreviaturas de otro lenguaje	Usan abreviaturas	Hablan bien	Pronuncian incompletas
	10	No siempre Adoptamos mal el lenguaje	Diferencia social y académica	No muchas veces	No siempre
	11	Si influye Al escribir y pronuncia mal	SI Afecta en pronunciación Acentos	Si es culta	Mas por el nivel de educación
	12	Si pero depende más si es culto y educado.	Por cultura Estado económico	No afecta	Diferente nivel educativo
	13	Si Hablan diferente en lugares diferentes	Si, depende con quien se hable	Si depende el lugar y persona	Según la emoción
	14	Si adoptamos palabras extranjeras	Uso de ingles	Si vocabulario en ingles	Palabras en ingles
	15	Si depende de la confianza con la persona	Si las modifica	Depende el entorno o medio	Se acopla al ambiente
	16	Fluidez, Tono moderado Pronunciación, Palabras del lenguaje	No distracciones Voz clara Fluida	Fuerza Tono	Voz clara Fuerte Frases completas

Fuente: elaboración propia.

Tabla 2. **Matriz de resultados 2**

	Entrevistados (aspectos a resaltar)				
	No.	**Entrevistado 5**	**Entrevistado 6**	**Entrevistado 7**	**Entrevistado 8**
PREGUNTAS	**1**	Si la mayoría de veces	Si es fácil	Si por lo general	La mayoría de veces
	2	Acento Modismos	Acento Modismo	Tono de voz Lugar de origen	Acento
	3	No me confunden	Si afecta Palabras a medias	Algunas veces	Si
	4	No a veces no	Si no entrecorta palabras si	Cuesta a veces	Si
	5	La mayoría de veces Si no hablan con palabras que no entienda	Si No son complicadas	Es fácil	Si por influencia externa
	6	Abreviaciones de otras culturas	Modismos	Jerga Locación	Si modismos palabras extranjeras
	7	Ruido Personas hablando	Si, en salas de reuniones	Si en general	Si
	8	Depende la emoción	Depende de la persona	Normal no fuerte ni callado	Callado
	9	No, se expresan con abreviaturas	Algunas veces	Abreviaturas	Uso de muletas
	10	No Agregan letras a las palabras	No siempre	Depende de la persona	Palabras chapinas
	11	Si escriben mal hablan mal	Si afecta la gramática	Si escuchan mal Escriben mal	Si influye
	12	Depende si son cultas	No afecta, es más si es culto	Si afecta	Educación
	13	Emociones Lugar	Emociones	Emociones	Depende la persona
	14	En raras ocasiones Ingles	A veces ingles	Ingles	Sí, ingles
	15	Si depende si lo conozco	No modifico	No	Depende las personas
	16	Claro Tono de voz claro Ordenado	Voz clara	Buen tono Palabras con significado	Vocalizar Gestionar

Fuente: elaboración propia.

Tabla 3. **Matriz de resultados 3**

Entrevistados (aspectos a resaltar)					
	No.	**Entrevistado 9**	**Entrevistado 10**	**Entrevistado 11**	**Entrevistado 12**
P R E G U N T A S	**1**	Si	En un 80%	Difícil	No, si se apoyan de gestos
	2	Acentos	Tono de voz Modismo Frases	Idioma materno	Acento
	3	La mayoría de veces	Si afecta Deben tener pausas	Se entiende mejor	Si
	4	Si	Pocas veces	Si	Si
	5	Si	90% si	Hay que aprender modismos de jóvenes	Si depende modismo
	6	Ninguna	LOL (frases en inglés)	Palabras en ingles	Referencia a objetos, comida
	7	Si	Ruido Celulares	Son ambientes de estudiante	Si en oficina o grupo cerrado
	8	Mas callado	Depende la situación	Depende de la familia	Diferente tipo de voz Tímidos
	9	Depende la persona	No, por moda	No	Algunas
	10		No porque toman vocabulario de otras culturas.	Grado académico Lectura	Estudiadas Lectura
	11	Si	Si porque se pronuncian mal.	Si	Lenguaje heredado
	12	Si	No es más nivel académico Cultura Aprendizaje	Si	Educación
	13	Si	Emociones Lugares Estado físico	Entrono	Llamar la atención Amigos
	14	No	Ingles	Si	Si
	15	Si	Depende la persona Edad Formal e informal.	Otro idioma	Lenguaje Ortográfica
	16	Tono de voz Pronunciación	Tono de voz Duración Vocabulario.	Tono de voz	Tono de voz Rapidez Explicar

Fuente: elaboración propia.

Tabla 4. **Matriz de resultados 4**

Entrevistados (aspectos a resaltar)					
	No.	**Entrevistado 13**	**Entrevistado 14**	**Entrevistado 15**	**Entrevistado 16**
PREGUNTAS	**1**	Si	Si	50%	Si
	2	Depende el lugar Acento	Acento Vocabulario	Pronunciación	Acento Pronunciación, Tono
	3	Costumbre	Lugar de origen	No	Si
	4	No	No	NO	No
	5	Modismos Otro idioma	Difícil entender	Si por léxico o dialogo	No
	6	Términos	Depende el entorno	Depende lugar de origen	Apear
	7	Cambia la conversación	Hora de comida	No	En ocasiones
	8	Depende Lugar Personas	Depende ambiente	Muy fuerte	Depende la persona y ocasión
	9	Algunos	No	En ocasiones	La mayoría de veces
	10	Poco estudio	Hablan correctamente	No todas	Si
	11	Estudio	Por moda o por App se cambia el lenguaje español	Si	Educación
	12	Educación	Sí, depende lugar de origen.	No depende de la educación	Educación
	13	Depende Lugar Personas	Si	Depende el estado o ambiente	Si
	14	No	No	Si	Si, Ingles
	15	Si tipo de persona	Depende la confianza	Si	Depende la persona
	16	Gestionar Pronunciación	Despacio Palabras claras.	Pronunciación Tono Fluidez	Fluidez Tono de voz

Fuente: elaboración propia.