

**UNIVERSIDAD NACIONAL DEL SANTA**

**ESCUELA DE POSTGRADO**

**MAESTRÍA EN INGENIERÍA DE SISTEMAS E INFORMÁTICA  
MENCIÓN EN GESTIÓN DE TECNOLOGÍAS DE INFORMACIÓN**



**Base de datos de difonos para la síntesis del habla del  
español peruano**

Tesis de investigación para optar el grado de Maestro en Ingeniería  
de Sistemas e Informática

Autor:

Bach. Yim Isaias Apestegui Florentino

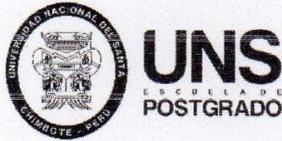
Asesora:

Dra. Diana Cecilia Muñoz Casanova

Nuevo Chimbote, Perú

2016

Registro Nro. \_\_\_\_\_



## CONSTANCIA DE ASESORAMIENTO DE LA TESIS DE MAESTRÍA

Yo, **Diana Cecilia Muñoz Casanova** mediante la presente certifico mi asesoramiento de la Tesis de Maestría titulada: **Base de datos de difonos para la síntesis del habla del español peruano**, elaborada por el bachiller **Yim Isaias Apestegui Florentino** para obtener el Grado Académico de Maestro en **Ingeniería de Sistemas e Informática con mención en Gestión de Tecnologías de Información** en la Escuela de Postgrado de la Universidad Nacional del Santa.

Nuevo Chimbote, 11 de octubre del 2016

---

Dra. Diana Cecilia Muñoz Casanova  
Asesora



**UNS**  
ESCUELA DE  
POSTGRADO

## HOJA DE CONFORMIDAD DEL JURADO EVALUADOR

### Base de datos de difonos para la síntesis del habla del español peruano

TESIS PARA OPTAR EL GRADO DE MAESTRO EN INGENIERÍA DE SISTEMAS E  
INFORMÁTICA CON MENCIÓN EN GESTIÓN DE TECNOLOGÍAS DE INFORMACIÓN

Revisado y Aprobado por el Jurado Evaluador:

.....  
Ms. Hugo Esteban Caselli Gismondi  
PRESIDENTE

.....  
Dra. Celinda Elcira Romero Salinas  
SECRETARIA

.....  
Dra. Diana Cecilia Muñoz Casanova  
VOCAL

## Dedicatoria

*Dedico con mucho cariño el presente informe de tesis a mis queridos hijos Marián y Adiel.*

## **Agradecimiento**

Quiero agradecer a Dios, fuente inagotable de energía y vida, quien armoniza todo lo que podemos percibir y lo que no, también.

Agradecimiento especial a la Dra. Diana Cecilia Muñoz Casanova, quien, con sus sugerencias como asesora de la presente tesis, me encaminó en la noble labor de la investigación.

Quiero agradecer a cada uno de los docentes de la Maestría en Ingeniería de Sistemas e Informática de la Universidad Nacional del Santa, quienes compartieron con nosotros sus conocimientos y experiencias, en las aulas y laboratorios de la escuela de postgrado.

Asimismo, agradezco a los alumnos de Comunicación Social, promoción 2015 y alumnos de Sistemas e informática, promoción 2014, de la UNS, quienes han contribuido en la puesta a prueba de algunos instrumentos empleados en la presente tesis de investigación.

Finalmente, a todas aquellas personas que, de una u otra manera, con sus sugerencias y aportes hicieron posible la culminación de la presente tesis.

## Índice

INTRODUCCIÓN .....	15
CAPÍTULO I .....	16
1 PROBLEMA DE INVESTIGACIÓN.....	17
1.1 Planteamiento y fundamentación del problema de investigación .....	17
1.2 Antecedentes de la investigación .....	19
1.3 Formulación del problema de investigación.....	21
1.4 Delimitación del estudio .....	21
1.4.1 Delimitación espacial .....	21
1.4.2 Delimitación temporal.....	21
1.4.3 Delimitación técnica.....	21
1.5 Justificación e importancia de la investigación .....	22
1.5.1 Justificación teórica .....	22
1.5.2 Justificación metodológica .....	22
1.5.3 Justificación práctica .....	22
1.6 Objetivos de la investigación: General y específicos .....	23
1.6.1 Objetivo general .....	23
1.6.2 Objetivos específicos.....	23
CAPÍTULO II .....	24
2 MARCO TEÓRICO .....	25
2.1 Fundamentos teóricos de la investigación .....	25
2.1.1 La voz.....	25
2.1.2 El habla .....	25
2.1.3 El español.....	26
2.1.4 El alfabeto español fonético según SAMPA.....	26
2.1.5 El español en América.....	27
2.1.6 El español peruano .....	28
2.1.7 Síntesis de la Voz .....	29
2.1.8 Diseño de Corpus .....	32
2.1.9 Fonética y fonología.....	33
2.1.10 Los Fonemas .....	33
2.1.11 Los difonos o difonemas .....	33
2.1.12 El proceso de síntesis a partir de texto.....	34

---

2.1.13	La técnica PSOLA.....	35
2.1.14	La técnica MBROLA.....	37
2.1.15	Resumen de la adopción de una postura teórica.....	37
2.2	Marco conceptual.....	38
2.2.1	Base de Datos.....	38
2.2.2	Base de datos de frases portadoras.....	38
2.2.3	Difono.....	38
2.2.4	Fonemas.....	38
2.2.5	Corpus peruano.....	38
2.2.6	Naturalidad.....	38
2.2.7	Inteligibilidad.....	39
2.2.8	Español peruano.....	39
2.2.9	Concatenación de difonos.....	39
2.2.10	Síntesis del Habla a partir de texto.....	39
CAPÍTULO III.....		40
3	MARCO METODOLÓGICO.....	41
3.1	Hipótesis central de la investigación.....	41
3.2	Variables e indicadores de la investigación.....	42
3.2.1	Definición conceptual.....	42
3.2.2	Definición operacional.....	42
3.2.3	Indicadores.....	43
3.2.4	Operacionalización de variables.....	43
3.3	Método de la investigación.....	43
3.4	Diseño de la investigación.....	44
3.5	Población y muestra.....	44
3.6	Actividades del proceso investigativo.....	45
3.7	Técnicas e instrumentos de recolección de datos.....	46
3.8	Procedimiento para la recolección de datos.....	46
3.9	Técnicas de procedimiento y análisis de los resultados.....	47
CAPÍTULO IV.....		48
4	RESULTADOS Y DISCUSIÓN.....	49
4.1	La concatenación de unidades fonéticas basada en difonos.....	49
4.2	Inventario de fonemas y difonos del español peruano.....	52
4.2.1	Inventario de fonemas.....	52
4.2.2	Metodología para el Inventario de difonos.....	56

4.3	Inventario de audio digital de frases portadoras de unidades de difonos.....	58
4.4	Etiquetado y segmentado de los fonemas de las frases portadoras.....	59
4.4.1	Consideraciones para la grabación del corpus.....	60
4.5	Procesamiento y síntesis, mediante concatenación de difonos.....	67
4.6	Determinar los niveles de inteligibilidad y naturalidad de la voz artificial generada.....	69
4.6.1	Aplicación del test de Inteligibilidad.....	70
4.6.2	Registro de datos en SPSS.....	72
4.6.3	Cálculo del Coeficiente Alfa de Cronbach.....	73
4.6.4	Tablas de frecuencia de Palabras Sueltas.....	74
4.6.5	Tabla de frecuencias de oraciones con sentido.....	75
4.6.6	Tabla de frecuencias de frases sin sentido.....	77
4.6.7	Tabla de frecuencias de la Variable Calificativo.....	78
4.6.8	Aplicación del Test de Naturalidad.....	79
4.7	Contrastación de la hipótesis.....	81
4.7.1	Paso1: Planteamiento de la hipótesis Nula y la hipótesis alternativa.....	81
4.7.2	Paso2: Nivel de significancia.....	82
4.7.3	Paso3: Estadístico de prueba y reglas de decisión.....	82
4.7.4	Aceptación o rechazo de Ho, H1 y H2.....	89
CAPÍTULO V.....		91
5 CONCLUSIONES Y RECOMENDACIONES.....		92
5.1	Conclusiones.....	92
5.2	Recomendaciones.....	93
REFERENCIAS BIBLIOGRÁFICAS.....		94
ANEXOS.....		98

## ÍNDICE DE TABLAS

Tabla 1 Valoración de los métodos de síntesis.	51
Tabla 2 Grafía y fonemas de vocales y consonantes del español peruano	53
Tabla 3 Grafía y fonemas vocálicas del español peruano	54
Tabla 4 Grafía y fonemas líquidas del español peruano	54
Tabla 5 Fonemas básicos para la generación de difonos	55
Tabla 6 Inventario de difonos	57
Tabla 7 Tabla de palabra(s) portadoras de difonos	59
Tabla 8 Estadísticos descriptivos	73
Tabla 9 Estadística de fiabilidad Alfa de Cronbach obtenido con SPSS	74
Tabla 10 Tabla de frecuencia de la variable Palabras Sueltas	75
Tabla 11 Tabla de frecuencias de la variable oraciones con sentido	76
Tabla 12 Tabla de frecuencias de la variables frases sin sentido	77
Tabla 13 Tabla de frecuencias de la variable Calificativo	78
Tabla 14 Tabla adaptada de la escala Five-grade del ITU	81
Tabla 15 Tabla K-S, para inteligibilidad y naturalidad	83

## ÍNDICE DE ECUACIONES

Ecuación 1. Unidades de análisis .....	36
Ecuación 2. Ecuación de solapamiento y suma de unidades de síntesis .....	37
Ecuación 3. Fórmula para hallar el tamaño muestral de una población infinita.....	44
Ecuación 4 Tamaño muestral obtenido al 95% de nivel de confianza .....	45
Ecuación 5 Fórmula del Alfa de Cronbach.....	73

## ÍNDICE DE FIGURAS

Fig. 1 El Alfabeto español fonético .....	27
Fig. 2 Fórmula del número de difonemas reales.....	34
Fig. 3: Modelo de proceso de un sintetizador .....	34
Fig. 4 Valoración de los métodos de síntesis.....	51
Fig. 5 Especificaciones del audio, para frases portadoras.....	61
Fig. 6 Grabación de frases portadoras, con Sound Forge 11.0.....	62
Fig. 7 Archivo audio de la palabra “acto” .....	63
Fig. 8 Archivo de audio de la palabra “acto, con mayor detalle de frecuencias” .....	63
Fig. 9 Segmentación de límites de difonos y fonema .....	65
Fig. 10 Espectrograma, pitch, intensidad y formantes de la palabra “acto” .....	65
Fig. 11 Un ejemplo de un archivo de texto de límites de fonemas .....	66
Fig. 12 Espectrograma de voz natural de la frase “la nube de Amazon”, pronunciado 3 veces ..	68
Fig. 13 Espectrograma de voz artificial de la frase “la nube de Amazon” pronunciado 3 veces.	68
Fig. 14 Espectrograma de voz natural de la frase “la nube de Amazon” pronunciado una vez ..	68
Fig. 15 Espectrograma de voz artificial de la frase “la nube de Amazon” z .....	69
Fig. 16 Escala Five-Grade para la valoración de la calidad de la voz.....	69
Fig. 17 Escala de Valoración para Ítems de Inteligibilidad .....	70
Fig. 18 Escala de Variable calificativo .....	71
Fig. 19 Extracto de los datos registrados en SPSS.....	72
Fig. 20 Gráfico de barras, para la variable palabras sueltas .....	75
Fig. 21 Gráfico de barra de la variable oraciones con sentido .....	76
Fig. 22 Gráfico de barras de la variable frases sin sentido.....	77
Fig. 23 Gráfico de barras de la variable Calificativo(agrupado) .....	78
Fig. 24 Escala five-grade para la variable similitud de voz natural .....	80
Fig. 25 Configuración de la prueba de Wilcoxon con $H_0=1$ y $H_1 \neq 1$ .....	84
Fig. 26 Resultados de la prueba de Wilcoxon con $H_0=1$ y $H_1 \neq 1$ .....	84
Fig. 27 Configuración de la prueba de Wilcoxon con $H_0=3$ y $H_1 < 3$ .....	86
Fig. 28 Resultados de la prueba de Wilcoxon con $H_0=3$ y $H_1 < 3$ .....	86
Fig. 29 Configuración de la prueba de Wilcoxon con $H_0=3$ y $H_1 > 3$ .....	88
Fig. 30 Resultados de la prueba de Wilcoxon con $H_0=3$ y $H_1 > 3$ .....	88

Fig. 31 Cuadro resumen de conclusiones de inteligibilidad y naturalidad .....	90
Fig. 32 Hipótesis aceptadas y/o rechazadas .....	90

## RESUMEN

La síntesis del habla como elemento de interacción hombre máquina, es un tema que hace que los investigadores continúen en el trabajo del perfeccionamiento de la emulación de la voz humana. Esta investigación comprende una de las fases del proceso de síntesis del habla, que es obtener una base de datos de difonos del español peruano, la misma que pueda servir para emular con cierto nivel de similitud una voz natural. Se presenta la técnica empleada para la obtención de los difonos y su almacenamiento respectivo. Luego, mediante un instrumento de concatenación de los mismos, se generan los audios con voz sintetizada. Para determinar el nivel de similitud de la voz sintetizada con la voz natural, se han realizado las pruebas de naturalidad e inteligibilidad.

*Palabras claves:* base de datos de difonos, síntesis del habla, conversión de texto a voz.

## **ABSTRACT**

Speech synthesis as an element of man machine interaction, is a topic that makes researchers continue to work on improving of the human voice emulation. This research comprises one of the stages of speech synthesis, which is to obtain a Peruvian Spanish diphone database, the same that demonstrates emulate with certain level of similarity with a natural voice. Techniques used for obtaining the respective diphones and storage, is presented. Then, using an concatenation instrument, audios with synthesized speech is generated. To determine the level of similarity of voice synthesized with natural voice, it has been done proofs of naturalness and intelligibility.

*Keywords:* diphone database, speech synthesis, converting text to speech.

## INTRODUCCIÓN

Existen varios servicios para el usuario, basados en tecnologías de síntesis de voz; empresas como Nuance y Google, por ejemplo, trabajan en ofrecer cada vez nuevos productos relacionados al tema de la emulación de la voz humana. Se habla de biometría de voz, de sistemas de respuesta de voz interactiva, de dispositivos móviles que manejan el tema de accesibilidad, y de nuevas versiones con cierto nivel de prestaciones, buscando el equilibrio entre usabilidad y accesibilidad. El motor de conversión de texto a voz de Google para Android, en el año 2014 pesaba 9.87Mb; hoy la versión 2016 pesa 18.85Mb (ApkHere, 2016), sin considerar el banco de sonidos o corpus que se estima se encuentra cercano a los 200Mb. Asimismo, el trabajo de construir grandes corpus de voz continúa, a fin de obtener mejoras en las tecnologías del habla y lograr la emulación de la voz humana (Garrido, 2015). Esta tesis apuntó a construir un corpus de los fonemas del español peruano, y permitió crear una base de datos de difonos, a partir de frases portadoras, métodos de segmentación y la técnica de concatenación que propone The MBROLA Project (2006). Se realizó las pruebas de síntesis de voz correspondientes, y se logró niveles de naturalidad e inteligibilidad aceptables. Para tal efecto se expone en el **capítulo I**, el problema de investigación, principales antecedentes de investigación y se formula el problema de investigación. En el **capítulo II**, se describen el marco teórico y conceptual necesario para establecer las bases teóricas de la investigación, reforzando los fundamentos de difonos y el mecanismo de concatenación. En el **capítulo III**, se establece el marco metodológico, se plantea la hipótesis, se definen las variables de investigación, instrumentos utilizados en el proceso de investigación y las técnicas del procedimiento de recolección de datos. En el **capítulo IV**, se presentan los resultados y discusiones de la metodología aplicada en la creación de la base de datos de difonos. Y en el **capítulo V**, se presentan las conclusiones y recomendaciones de la presente investigación.

## **CAPÍTULO I**

# **Problema de investigación**

## **1 PROBLEMA DE INVESTIGACIÓN**

En el presente capítulo se describen y analizan las conclusiones a las que han llegado otros autores en el tema. Asimismo, se plantea el problema de investigación.

### **1.1 Planteamiento y fundamentación del problema de investigación**

La interacción hombre máquina, que, aunque a la fecha no tiene una definición concreta, como realidad está compuesta por elementos que lo enmarcan dentro de varias disciplinas de investigación que estudian el intercambio de información a través de software, hardware, personas, computadoras, dispositivos; logrando resultados que han servido de beneficio y disfrute al hombre y la sociedad.

Actualmente, en el campo de la síntesis del habla las empresas e investigadores han alcanzado notables logros en la sintetización del habla como son los proyectos para dispositivos móviles Siri, de Apple, Sherpa, de Google así lo indica Garrido (2015) en su publicación. Allí se mencionan proyectos de automoción como el proyecto SYNC de Ford, proyectos de asistencia personal como Renfe y proyectos de ayuda a la discapacidad como JAWS, de Freedom Scientific, proyectos de ayuda a la escritura como WordQ, de GoQ, proyectos de apoyo al estudio como Kurzweil 3000/ Firefly, de Kurzweil Educational Systems.

Los ingenieros inmersos en las tecnologías del habla, necesitan del conocimiento fonético de una lengua y de la aplicación de técnicas y variaciones de las mismas, para llegar a una aproximación lingüística de la lengua de una zona, región o país.

Se está trabajando en obtener aproximaciones de corpus lingüísticos de las diversas lenguas en el mundo, pero sintetizar el habla de cada región, país tiene muchas particularidades; si a ello se agrega costumbrismos y modismos propios, el tema se vuelve aún más complejo; dada la cantidad de idiomas, dialectos que existen en el mundo.

La razón de esta complejidad es porque hay entre 5,000 y 6,000 idiomas hablados en el planeta, inclusive se afirma según estadísticas que ya hay 7000 (Lewis, Paul, Simons, & Fenning, 2016).

Se cuenta con síntesis del habla del Idioma español de España, del español de México, del español de Argentina, del español de Venezuela, y del español de Colombia, pero específicamente del español de Perú, a la fecha no. No se tiene un corpus en el español peruano; no se cuenta con una publicación de síntesis del habla en español peruano al 2016.

Se observa que las aplicaciones actuales en dispositivos electrónicos como celulares, celulares inteligentes usan un español de otro país, y no un español peruano. Como parte de nuestra identidad, se debe tener un corpus de nuestro español, que permita entre muchas utilidades, por ejemplo, generar capacidad de voz a los mudos y capacidad de lectura a los ciegos en nuestro propio dialecto: El español peruano, el español chimbotano específicamente.

Garrido (2015), en sus conclusiones, menciona que se acerca una convergencia entre los métodos de trabajo de la fonética y las tecnologías del habla, basados en el uso de grandes corpus y de herramientas automáticas, para la anotación y el procesamiento automático

Y es allí, donde apuntó esta investigación: Llegar a implementar una base de datos con la prosodia del español peruano estándar y que permita generar voz artificial.

## **1.2 Antecedentes de la investigación**

Aquí se presentan algunos trabajos de investigación y publicaciones, efectuados en los últimos años; cada uno con aportes valiosos para el tema de estudio de la presente investigación.

Hannane & Hamid (2014), en su publicación: Un sistema experto para la lectura automática de texto escrito en árabe estándar nos explica el proceso de la concatenación de unidades extraídas de palabras portadoras, las cuales no deben tener significado alguno para obtener una total independencia del difono creado y su contexto de pronunciación.

Como conclusión determina, que la concatenación aplicada al árabe estándar se torna difícil por la complejidad del lenguaje, por ello plantean que se utilice un sistema experto que ayude en la síntesis de fonemas.

Violante (2012) en su tesis: Construcción y evaluación del back-end de un sistema de síntesis de habla en español argentino, presentado en la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires, utiliza dos técnicas de síntesis distintas: síntesis por concatenación de unidades y síntesis basada en modelos ocultos de Markov. La autora trabajó con un inventario de sonidos provisto por el Laboratorio de Investigaciones Sensoriales, INIGEM, sobre el cual realizó modificaciones para mejorar la calidad de la voz sintetizada. Utiliza tests de inteligibilidad y naturalidad, contrastando los resultados con los obtenidos en trabajos previos.

La base de sonidos tomada de un laboratorio, permite deducir que es necesario un banco de sonidos del español peruano.

Sadowski & Salamanca (2011) en su artículo: El inventario fonético del español de Chile. ambos autores de la Universidad de la Frontera de Chile y de la Universidad de Concepción de Chile, presentan un inventario fonético de consonantes, el cual tiene como punto de referencia el castellano hablado en la provincia de Concepción (Chile).

Se presenta un sistema de representación fonética (International Phonetic Association), y propone un programa de investigación cuyo objetivo es la elaboración del inventario fonético definitivo y de consenso del español de Chile contemporáneo.

Dada la realidad de la fonética de dicho país (Chile), se tiene un referente de su fonética; el español chileno y de alófonos presentados en aquel tipo de dialecto, el estilismo, la fuerza vocal, incluso hasta orientación y segmentación de clase social, punto importante, dado que el dialecto en determinada región difiere de otra en un mismo país. Por ejemplo, en nuestro país, la fonética de la región de la selva difiere de la fonética de los hablantes de la costa, por lo que ya se tiene un referente de la limitación y alcance de la presente investigación.

Cave, Rodriguez, Mora, Sandrine, & Hirst (2006) en su publicación: Un sistema de síntesis de habla en español de Venezuela, de la universidad Los Andes de Mérida, Venezuela y la Université de Provence, Francia, presentan un sistema de habla a partir de texto en español venezolano.

El sistema utiliza la concatenación de 794 difonos, mediante el cual el sistema logra manejar la entonación y la velocidad. En dicho sistema recomienda un buen ingreso de texto es decir un texto ortográficamente correcto, dado que no se cuenta con un verificador lingüístico.

El trabajo de Cave et Al (2006), indica además que es necesario la creación de un corpus, a través de frases portadoras.

Flores (2005), en su tesis de la Pontificia Universidad Católica del Perú, desarrolla un modelo de entonación de un sintetizador de voz, para un sistema de información vía telefónica en cines. Allí describe que la entonación de la computadora juega un papel importante, pues en tanto que la entonación esté muy bien modelada y/o modulada, dotará de naturalidad al sistema.

Asimismo, dicha investigación se orienta a la concatenación por selección de unidades, dado que el sistema requiere emitir entonaciones con alto nivel de naturalidad, pues hay interacción con el cliente.

En relación a esta investigación, los conjuntos de palabras formados por difonos son elevados respecto a las palabras formadas por selección de unidades, asimismo no se tiene frases predefinidas como las que usa el sistema propuesto por Flores (2005), sino que los difonos, en teoría, permiten formar cualquier frase.

### **1.3 Formulación del problema de investigación**

¿Cómo desarrollar una base de datos de difonos para la síntesis del habla del español peruano con un nivel aceptable de inteligibilidad y naturalidad?

### **1.4 Delimitación del estudio**

#### **1.4.1 Delimitación espacial**

La base de datos generada corresponde al español peruano hablado en la Región Ancash, específicamente en la ciudad Chimbote, Nuevo Chimbote que cuenta con 151, 127 habitantes (INEI, 2015).

#### **1.4.2 Delimitación temporal**

Los métodos y técnicas están sujetos a experimentación. Se pueden lograr resultados similares o diferentes en el tiempo, dado que muchas de las técnicas se realizaron de manera manual, manteniendo un margen de error diferente en cada suceso de tiempo.

#### **1.4.3 Delimitación técnica**

Dado que el financiamiento de esta investigación se realizó con recursos propios, los equipos de audio y cómputo se ubicaron dentro de la categoría de semi profesionales, con características necesarias para los fines de la presente tesis de investigación.

Los resultados obtenidos, están sujetos a las técnicas aplicadas y su limitación técnica correspondiente. Por ende, se esperaría tener otro tipo

de resultados en la investigación, si se aplican instrumentos con calidad diferente al utilizado en esta tesis.

## **1.5 Justificación e importancia de la investigación**

### **1.5.1 Justificación teórica**

Con la presente tesis de investigación se quiere demostrar que la concatenación de difonos permitió llegar a niveles aceptables de emisión de voz artificial, sin llegar a utilizar demasiado consumo de recursos de hardware. Además, frente a otras teorías de concatenación, como las que utilizan las empresas comerciales, concatenación de difonos se espera que sea una buena alternativa para la emisión de voz en dialecto peruano. Para ello, se necesitó la comprensión teórica sólida de la emisión del habla natural y la artificial.

### **1.5.2 Justificación metodológica**

La presente tesis de investigación es importante porque en base a la metodología planteada se obtuvo una base de datos de difonos del español peruano, específicamente del español ribereño costeño, metodología que requiere pruebas de calidad y precisión en la captura de sonidos, para luego obtener la base de difonos mediante técnicas de segmentación, los mismos que mediante algoritmos de concatenación de aquellos difonos se pueda simular la voz humana.

Este instrumento podrá servir a otros proyectos de investigación futuros relacionados a tema.

### **1.5.3 Justificación práctica**

La justificación práctica se refiere a que teniendo una base de datos de difonos, éstos se puedan implementar en otros tipos de proyectos de desarrollo de software.

Dicha base de difonos se puede utilizar en dispositivos domóticos, en la emisión de sonidos artificiales de ayuda al usuario y/o reconocimiento de voz.

Puede ayudar a la mejor calidad de vida de las personas ciegas que necesiten un lector de texto con voz artificial, pero en nuestro propio dialecto.

Esta investigación, puede contribuir como un componente en la implementación de software de ayuda para el tratamiento de terapias del lenguaje o problemas de articulación de palabras.

## **1.6 Objetivos de la investigación: General y específicos**

### **1.6.1 Objetivo general**

Desarrollar una base de datos de difonos que, mediante la concatenación de unidades, permita la síntesis del habla del español peruano.

### **1.6.2 Objetivos específicos**

- Realizar un análisis de la literatura de la concatenación de unidades fonéticas basada en difonos.
- Elaborar un inventario de fonemas y difonos del español peruano, para determinar los difonos realizables o utilizables.
- Elaborar un inventario de audio digital de frases portadoras de unidades de difonos.
- Etiquetar y segmentar los fonemas de las frases portadoras, para obtener difonos.
- Procesar y sintetizar, la base de datos de difonos, mediante el método de concatenación de difonos
- Determinar los niveles de naturalidad e inteligibilidad de la voz artificial generada.

## **CAPÍTULO II**

# **Marco Teórico**

## **2 MARCO TEÓRICO**

En el presente capítulo se desarrolla un marco teórico necesario para la comprensión del tema de estudio, haciendo una revisión de literatura de investigaciones relacionadas, profundizar en el tema de investigación y establecer las pautas de nuestra postura teórica del mismo.

### **2.1 Fundamentos teóricos de la investigación**

Se debe comprender, los elementos principales que intervienen en la presente investigación, a partir del problema planteado: La base de datos de difonos, realizar la síntesis del habla del español peruano y consideraciones de naturalidad e inteligibilidad.

#### **2.1.1 La voz**

La Real Academia Española (2014), en su última edición 23<sup>o</sup> publicada en octubre del mismo año, define a la voz (proveniente de lat. *vox, vocis*), como “*Sonido que el aire expelido de los pulmones produce al salir de la laringe, haciendo que vibren las cuerdas vocales.*”

De lo anterior, se deduce que pronunciar palabras de manera oral incluye un proceso físico de emisión de sonido, que incluye la vibración de las cuerdas vocales, las mismas que producen el efecto sonoro deseado.

#### **2.1.2 El habla**

La Real Academia Española (RAE), presenta al habla en varias definiciones, de las cuales las que se relacionan con el tema de investigación son:

- “Manera especial de hablar”
- “Acto individual del ejercicio del lenguaje, producido al elegir determinados signos, entre los que ofrece la lengua, mediante su realización oral o escrita”.

- “Sistema lingüístico de una comarca, localidad o colectividad, con rasgos propios dentro de otro sistema más extenso”.

De lo anterior, se considera que el habla es esa manera especial de hablar que tiene una localidad o colectividad con rasgos y características propias y que se produce en el ejercicio de la realización oral o escrita.

### 2.1.3 El español

La RAE, define al español como: “Lengua común de España y de muchas naciones de América, hablada también como propia en otras partes del mundo.”

El español es una lengua común en muchas naciones de América, y el Perú en su proceso histórico, social, cultural y político lo adopta como lengua hablada con sus rasgos y características propios; de ahí que se hace mención al hecho de tener un español peruano.

### 2.1.4 El alfabeto español fonético según SAMPA

El alfabeto español según SAMPA (1996), presenta los siguientes elementos entre consonantes

<b>Consonantes</b>			
<b>Oclusivas</b>			
<b>Símbolo</b>	<b>Palabra</b>	<b>Transcripción</b>	
p	padre	"paDre	
b	vino	"bino	
t	tomo	"tomo	
d	donde	"donde	
k	casa	"kasa	
g	gata	"gata	
<b>Africada</b>			
tS	mucho	"mutSo	
jj	hielo	"jjelo	
<b>Fricativas</b>			
f	fácil	"faTil	
B	cabra	"kaBra	(= /b/)

T	cinco	"Tinko	
D	nada	"naDa	(= /d/)
s	sala	"sala	
x	mujer	mu"xer	
G	luego	"lweGo	(= /g/)
<b>Nasales</b>			
m	mismo	"mismo	
n	nunca	"nunca	
J	año	"aJo	
<b>Líquidas</b>			
l	lejos	"lexos	
L	caballo	ka"baLo	(or as jj)
r	puro	"puro	
rr	torre	"torre	
<b>Semivocales</b>			
j	rei	rrej	
	pie	pje	
w	deuda	"dewDa	
	muy	mwi	
<b>Vocales:</b>			
i	pico	"piko	
e	pero	"pero	
a	valle	"baLe	
o	toro	"toro	
u	duro	"duro	

Fig. 1 El Alfabeto español fonético

Fuente: (SAMPA, 1996)

### 2.1.5 El español en América

Según Juttsjö L. (2009) en su publicación: "Peruanismos: Origen, historia, cultura y lengua" menciona, que de manera general se puede distinguir a los hablantes del español europeo y hablantes del español americano ya que existen variantes en el español americano y textualmente dice:

"Para describir el `español americano', `español de América' o `español hablado en América' hay que tener en cuenta del conjunto de numerosas variantes nacionales y

dialectales, en lugar de pensar en una lengua homogénea hablada por todos los hispanohablantes americanos” (Justjö, 2009, p5).

Saralegui (1997) citado por Justjö (2009, p.6) menciona que:

Fue un dominicano, Pedro Henríquez Ureña, el primero que determinó de manera relevante, en 1921, las zonas dialectales del español en América. Henríquez Ureña da especial relevancia al sustrato prehispánico, según el cual, pueden distinguirse cinco zonas.

El autor refiere que estas cinco zonas comprender a los cubanismos, colombianismos, argentinismos, peruanismos, y semejantes.

O sea, al referirnos al español de América, nos referimos a los diversos dialectos que se han formado en el transcurso de la historia de América y por ende del Perú con su propio español americano, al cual Henríquez Ureña lo tipifica como peruanismo, zona dialectal que se encuentra en Perú, por ello, en ese proceso de formación, adaptación, mezcla y acoplamiento del español para generar el español de Perú, el español sufre una serie de alteraciones que obedecen a aspectos sociales, culturales, históricos inclusive hasta políticos (Cabrera, 2006).

#### **2.1.6 El español peruano**

Respecto al castellano que se habla en el Perú, Escobar (1978), planteó una distribución según zonas geográficas:

Variedad del Tipo1:

- A.- El andino: Comprende los valles andinos e interandinos de Norte y Sur, incluyendo Madre de Dios.
- B.- El Altiplánico: Comprende a Puno.
- C.- El de Litoral y Andes Occidentales sureños: Moquegua y Tacna.

Variedad del tipo 2.

D.- El de litoral norteño y central.

E.- El amazónico: Comprende Loreto, San Martín, Amazonas y Huánuco.

Además de esta propuesta de esquema de zonificación Escobar (1978) dice: "A estas alturas del trabajo solo tenemos seguridad de una cosa, aquí se propone una tarea que merece ser perfeccionada y proseguida". En efecto como todo investigador, Escobar deja abierta la posibilidad de que su propuesta sea perfeccionada y de hecho que los lingüistas deben haber continuado la investigación de Escobar y perfeccionado su propuesta.

Para los motivos de la presente tesis de investigación, se deduce que existe un español peruano con todas sus variedades dialectales y la variedad que nos compete, dada las delimitaciones de investigación presentadas anteriormente, es la del español costeño, o español ribereño como lo denomina Escobar.

Esto refuerza que se puede utilizar el español de nuestra localidad como el español representativo peruano, por ejemplo: México tiene ciudades norteñas y sureñas; cuando se escucha hablar a un mexicano no se distingue si es mejicano del norte o del sur, solo se reconoce su acento y se dice que el acento es mejicano.

Lo mismo si se escucha a un chileno, no se distingue si es del norte o del sur de Chile, solo se reconoce su acento genérico. Así, se puede decir lo mismo del español argentino, del español venezolano y otros.

### **2.1.7 Síntesis de la Voz**

Violante (2012), utiliza dos técnicas de síntesis distintas: síntesis por concatenación de unidades y síntesis basada en modelos ocultos de Markov.

### **2.1.7.1 La Síntesis por Concatenación de Unidades**

Según Fernández (2006) en su estudio de la síntesis de voz y su aplicación al software educativo indica que la síntesis concatenativa se basa en la concatenación de segmentos de voz grabados. Generalmente, la síntesis concatenativa produce los resultados naturales, pero la variación natural del habla y las técnicas de segmentación pueden llevar a una pérdida de naturalidad.

#### **Síntesis por selección de unidades**

La síntesis por selección de unidades debe tener base de datos de voz grabada. Durante la creación de la base de datos, el habla se segmenta en unidades de palabras, frases y oraciones.

Por ejemplo, un sistema de síntesis de voz para dar informaciones de vuelos tendrá naturalidad si la base de datos tiene grabaciones de informaciones de vuelos. Por ello la máxima naturalidad requiere que la base de datos sea muy grande por el orden de los Gb. de datos grabados (Fernández, 2006).

#### **Síntesis de difonos**

La síntesis de difonos necesita de una base de datos que contiene a todos los difonos de un lenguaje. El español, tiene en teoría unos 900 difonos, el alemán unos 2500. Se debe cuidar de que la prosodia de una oración se mantenga o no se pierda de manera significativa, para ello investigaciones como la del Español Venezolano han utilizado procesado digital de la señal, como codificación lineal predictiva, PSOLA o MBROLA. Se debe mencionar, que para fines comerciales síntesis de difonos se encuentra en investigación (Fernández, 2006).

Se plantea la postura de que la investigación continúa porque a pesar de lo que Fernández, pueda comentar hay investigaciones que se han desarrollado como la de Rojas, Blondet, & Mora (2012), donde se obtuvieron buenos resultados al dialecto venezolano.

Todo apunta a que el campo o dominio de aplicación en síntesis de difonos es mayor frente a un dominio específico.

### **Síntesis específica para un dominio**

La síntesis específica para un dominio une palabras y frases grabadas. Se usa en aplicaciones de un dominio particular y específico, como anuncios de viajes, salidas de bus, etc. Es la tecnología usada por diversos dispositivos parlantes. La naturalidad de esta es muy alta, porque la variedad de oraciones tiene límite. Este límite de frases y palabras de la base de datos, hace que la aplicación no sea de propósito general (Fernández, 2006).

Gracias a la teoría de Fernández (2006), se puede comprender a Violante (2012), el por qué la autora trabajó sobre un inventario de sonidos del laboratorio INIGEM, sobre el cual realizó modificaciones para mejorar la calidad de la voz sintetizada. Violante (2012), muestra los resultados que fueron evaluados, utilizando las pruebas de inteligibilidad y naturalidad. La autora utilizó técnicas de procesamiento de señales para lograr habla más expresiva y variaciones en la prosodia.

Hasta este punto se tuvo ya más claro la necesidad de tener un inventario de sonidos, lo cual constituyó una actividad de suma importancia en el desarrollo de la presente tesis, pues se tuvo que crearlo.

#### **2.1.7.2 Síntesis a partir de texto**

Cave, Rodriguez, Mora, Sandrine, & Hirst (2006), en el sistema que desarrollaron utilizan la concatenación de difonos, mediante el cual, el sistema logra manejar la entonación y la velocidad.

En dicho sistema recomiendan un buen ingreso de texto, es decir; un texto ortográficamente correcto, si es que no se cuenta con un verificador ortográfico lingüístico.

Cave et Al. (2006), sostienen que a través de las pruebas que realizaron, las señales de audio obtenidas con su sintetizador eran similares a las producidas por un hablante natural, y se puede cualificar, mas no muestran datos cuantitativos; asimismo, se notaron fallas en la coarticulación entre palabras, afectando su calidad respecto a su naturalidad.

Los tiempos de respuesta de dicho trabajo se encuentran entre 0.7 y 0.9 segundos, y recomienda que es muy importante tener en cuenta medidas cuantitativas respecto a la calidad de un sintetizador. Un análisis del dominio de la frecuencia podría indicarnos en términos de la similitud de las señales y saber qué criterios se deben mejorar.

Hasta este punto, los resultados de Cave et Al. (2006), nos indicaron que en la síntesis del habla se presentan problemas de coarticulación

### **2.1.8 Diseño de Corpus**

Torruella & Llisterri (1999, p6), en su publicación: Diseño de corpus textuales y orales, define ¿qué es un corpus?, quien menciona que un corpus es “Una colección de piezas de un lenguaje que son seleccionados y ordenados de acuerdo a un criterio lingüístico explícito, ordenados para ser usados como muestra del lenguaje”

La informática o computación no está directamente relacionada al concepto de “corpus”, pero hoy en día la informática facilita los procedimientos, eso es innegable, de hecho, facilita el procesamiento de grandes cantidades de datos y “Sería impensable crear un corpus prescindiendo de este medio o herramienta” (Torruella & Llisterri, 1999).

Entonces un corpus se debe manejar con una herramienta informática respectiva de procesamiento. Torruella & Llisterri (1999), hacen referencia a un corpus lingüístico informatizado, por ello, uno los objetivos de

presente investigación fue organizar y documentar cada de una de las piezas conformantes de un corpus.

### **2.1.9 Fonética y fonología**

La Fonética, estudia las variaciones articulatorias y acústicas de los sonidos del habla y cómo estos se perciben; la fonología, se ocupa de las unidades lingüísticas codificadas en las ondas sonoras que se producen y perciben (Listerri, 2016).

Así se tiene a:

La Fonética: Ciencia que estudia los diversos sonidos del habla.

La Fonología: Ciencia que estudia la organización de los sonidos en términos lingüísticos, de las diversas lenguas (Gil, 2007, p540).

Se hace necesaria la verificación del correcto pronunciamiento de las palabras, a través de un experto en fonología y/o fonética, debido a las variantes que pueda tener el español peruano. Dado que la pronunciación artificial a través de la síntesis, depende de los sonidos emitidos en las frases portadoras de los difonos.

### **2.1.10 Los Fonemas**

Los fonemas es el conjunto de abstracciones mentales de los sonidos del habla, para relacionar cada uno de ellos con una letra del alfabeto, sin embargo, algunas de ellas “son mudas” no producen sonidos como la “h”, y en ocasiones, varias letras se asocian a un mismo fonema (Correa, Rueda, & Argüello, 2010).

### **2.1.11 Los difonos o difonemas**

Un difono es el segmento que va de la mitad de un fonema hasta la mitad del fonema vecino, y se constituye como la unidad de síntesis. Esta unidad de síntesis se utiliza en la concatenación, con sus transiciones y coarticulaciones (Rojas, Blondet, & Mora, 2012).

El número de difonemas existentes, viene dado por el número total de fonemas elevado al cuadrado, menos los difonemas que nunca se presentan.

$$Difonemas_{totales} = (fonemas)^2 - difonos_{inexistentes}$$

Fig. 2 Fórmula del número de difonemas reales

Fuente: (Correa, Rueda, & Argüello, 2010)

### 2.1.12 El proceso de síntesis a partir de texto

El proceso de síntesis se realiza a través de un sintetizador. La función de este, es convertir un texto de entrada en una salida de audio. Para lograr esto, es necesario que dicho texto pase por una serie de procesadores encargados de realizar diferentes tareas hasta llegar al objetivo, que es una voz emulada (Correa, Rueda, & Argüello, 2010).



Fig. 3: Modelo de proceso de un sintetizador

Como se observa en la Figura 3, se presenta un esquema en la que el sintetizador se ejecuta secuencialmente, es decir, que la salida de un procesador se convierte en la entrada del siguiente.

Asimismo, Correa muestra un fragmento de los difonos utilizados para el español de Colombia en la Figura 3.1 Fragmento de difonos de español de Colombia, solo una porción de todo el conjunto real de difonos.

	pau	a	b	ch	d	e	f
pau	_ -	_ -a	_ -b	_ -ch	_ -d	_ -e	_ -f
a	a -	a-a	a-b	a-ch	a-d	a-e	a-f
b	b -	b-a	b-b	b-ch	b-d	b-e	b-f
ch	-	ch-a	ch-b	-	-	ch-e	-
d	d -	d-a	d-b	-	-	d-e	-
e	e -	e-a	e-b	e-ch	e-d	e-e	e-f
f	f -	f-a	-	-	-	f-e	-

Figura 3.1 Fragmento de difonos de español de Colombia,

Fuente: (Correa, Rueda, & Argüello, 2010)

### 2.1.13 La técnica PSOLA

En esta sección se considera los aspectos y fundamentos principales en la comprensión de los mecanismos de concatenación y por ello será necesario dos aspectos; la descripción del algoritmo y el análisis y la síntesis que utiliza PSOLA.

#### 2.1.13.1 Descripción del Algoritmo

De la Vega (2007), nos menciona que la Técnica PSOLA “Pitch Synchronous Overlap and Add” TD-PSOLA es un algoritmo que actúa en el tiempo y el funcionamiento del PSOLA se puede resumir en tres etapas:

Etapa 1.-Un análisis de la onda original, que se descompone en una serie de unidades de corta duración superpuestas denominadas “bloques” o “unidades” y en inglés se conoce como Short-term signal, ST.

Etapa 2.-La modificación de la prosodia, a partir de esta representación. Estas señales modificadas se convierten en bloques de síntesis.

Etapa 3.- La producción de la señal sintética.

De la Vega (2007), menciona que, mediante la superposición y suma de estas últimas, se genera la onda sintetizada, siendo este proceso el que

da lugar a PSOLA (Pitch Synchronous Overlap and Add), Solapamiento y suma sincronizada con la frecuencia fundamental.

### 2.1.13.2 Análisis y Síntesis Psola

De la Vega (2007), además explica que la señal de voz digitalizada  $s(n)$  se descompone en una serie de unidades superpuestas  $s_m(n)$  denominada **unidades de análisis**. Estas se obtienen multiplicando la señal por una secuencia de “ventanas”  $h_m(n)$  según la expresión:

$$s_m(n) = h_m(t_m - n)s(n)$$

*Ecuación 1. Unidades de análisis*

*Fuente: (De la Vega, 2007)*

La ecuación 1,  $h_m(n)$  representa una ventana simétrica y centrada en  $n=0$ . Los instantes  $t_m$  se seleccionan de manera síncrona con la frecuencia fundamental de la señal.

La secuencia de unidades es procesada para producir otro conjunto de unidades,  $s'_q(n)$ , que se denomina **unidades de síntesis**.

Se establecen marcas temporales  $t_q$ . La relación entre las unidades de análisis y las de síntesis viene de este modo determinada por una función de alineamiento temporal:  $t_q \rightarrow t_m$  que expresa la frecuencia de la onda sintetizada respecto a la de la original (De la Vega, 2007).

La señal de voz sintética  $s'(n)$  se obtiene mediante un proceso de solapamiento y suma de unidades de síntesis, según la ecuación 2.

$$s'(n) = \frac{\sum_q \alpha_q s'_q(n) h'_q(t'_q - n)}{\sum_q h_q'^2(t'_q - n)}$$

*Ecuación2. Ecuación de solapamiento y suma de unidades de síntesis*

*Fuente: (De la Vega, 2007)*

#### **2.1.14 La técnica MBROLA**

El objetivo del proyecto MBROLA, de la Facultad politécnica de Mons, Bélgica, es obtener un conjunto de sintetizadores de voz para la mayor cantidad de idiomas posibles y que puedan disponerse gratuitamente para aplicaciones no comerciales. La última meta es mejorar la investigación académica en síntesis de voz, particularmente en la generación de la prosodia, conocido por ser uno de los mayores retos de la síntesis de voz (Fernández, 2006).

Para efectos de esta investigación, se utilizó el mecanismo de concatenación y el software que proporciona el The MBROLA Project (2006); si bien es cierto, su algoritmo no está liberado, el uso del software con fines de investigación es libre.

#### **2.1.15 Resumen de la adopción de una postura teórica**

De lo mencionado en el presente marco teórico, desde la comprensión elemental de lo que es la voz y habla, hasta la profundidad teórica matemática del proceso de síntesis, se obtiene un fuerte sustento teórico, para mantener una postura respecto a que la fase que compete a la presente tesis se ha desarrollado, bajo criterios teóricos, lo que contribuyó a la implementación de una base de datos de difonos. Los mismos que mediante métodos de concatenación de unidades de difonos permitieron sintetizarse, emulando el español peruano con niveles sujetos a medición, de naturalidad e inteligibilidad.

Una base de difonos, elaborado bajo criterios teóricos debe incidir en el trabajo posterior de otras fases de síntesis del habla. En la presente tesis no entró en discusión el algoritmo de solapamiento, por cuanto MBROLA no ha liberado dicho algoritmo. Se conoce parcialmente su funcionamiento en base a las investigaciones realizadas a la fecha por otros autores. De manera que la investigación se ha centrado en la construcción de la base de difonos.

## **2.2 Marco conceptual**

### **2.2.1 Base de Datos**

Una base de datos es un conjunto de registros de datos, debidamente organizados y almacenados en formato digital.

### **2.2.2 Base de datos de frases portadoras**

Es un conjunto de registros de frases y textos que contienen las palabras seleccionadas con el difono en estudio.

### **2.2.3 Difono**

Segmento que va de la mitad de un fonema inicial hasta la mitad del fonema vecino.

### **2.2.4 Fonemas**

Los fonemas es la caracterización mental de los sonidos del idioma o habla, relacionado con una letra del alfabeto correspondiente.

### **2.2.5 Corpus peruano**

Base de datos de fonemas, difonemas, semisílabas, sílabas, palabras y frases, que forman parte del español peruano.

### **2.2.6 Naturalidad**

Grado o nivel de semejanza del sonido al emitir una palabra sintetizada respecto al emitido por una persona.

### **2.2.7 Inteligibilidad**

Grado o nivel de comprensión del sonido de la palabra sintetizada respecto al emitido por una persona.

### **2.2.8 Español peruano**

Español del distrito de Nuevo Chimbote, Perú, como el español representativo peruano (sustentado en el Marco Teórico).

### **2.2.9 Concatenación de difonos**

Técnica de procesamiento para solapar, enlazar difonos creados, mediante software.

### **2.2.10 Síntesis del Habla a partir de texto**

Generación automática por medio de software o hardware de una señal de audio, que emula la voz humana, a partir de un texto de entrada de un idioma.

### **CAPÍTULO III**

# **Marco metodológico**

### **3 MARCO METODOLÓGICO**

En esta sección, correspondiente al marco metodológico se determinó la metodología de desarrollo de la presente investigación, desde el planteamiento de nuestra hipótesis, definición de los objetivos y determinación de actividades del proceso de investigación, lo cual permitió centrarnos y no desviarnos en el proceso investigativo.

#### **3.1 Hipótesis central de la investigación**

##### **Hipótesis nula: Ho**

La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel no aceptable de inteligibilidad y naturalidad.

##### **Hipótesis alterna: H1**

La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel aceptable de inteligibilidad y naturalidad.

##### **Hipótesis alterna: H2**

La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel aceptable de inteligibilidad y un nivel poco aceptable de naturalidad.

##### **Hipótesis alterna: H3**

La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel poco aceptable de inteligibilidad y un nivel aceptable de naturalidad.

## **3.2 Variables e indicadores de la investigación**

### **3.2.1 Definición conceptual**

#### **Variable independiente: Base de datos de difonos**

Se entiende por base de datos de difonos al conjunto de registros de audio seleccionados mediante mecanismos de segmentación y que corresponde a cada uno de los difonos que contiene el alfabeto de un habla.

#### **Variable dependiente: Síntesis del habla del español peruano**

La síntesis del habla del español peruano es el procedimiento de generar una voz artificial no pregrabada, generada mediante un proceso de sintetización.

La síntesis de habla del español peruano es la producción artificial de habla humana.

### **3.2.2 Definición operacional**

#### **Variable independiente: Base de datos de difonos**

La base de datos de difonos es conjunto de registros de audio seleccionados mediante mecanismos de segmentación de una fuente de datos bibliográficos o de uso común para luego ser reproducidos y grabados en formato de audio digital y que corresponde a cada uno de los difonos que contiene los fonemas del alfabeto del habla.

#### **Variable dependiente: Síntesis del habla del español peruano**

La síntesis del habla del español peruano es un sistema sintetizador de habla llevado a cabo mediante software y produce emisiones de sonido a través un periférico de audio o parlante que tiene la capacidad de convertir texto en habla, mediante ciertos algoritmos de tratamiento de difonos. El audio emitido tendrá niveles de inteligibilidad y naturalidad.

### 3.2.3 Indicadores

#### Para la Base de datos de difonos

- Número de Unidades de difonos (Unidad de Audio wav)
- Número de Frases Portadoras (Unidad de Audio wav)

#### Para la Síntesis del habla del español peruano

- Palabras Sueltas
- Oraciones
- Frases sin sentido
- Similitud respecto a la voz humana

### 3.2.4 Operacionalización de variables

Variable	Dimensiones	Indicadores	Técnica	Instrumento	Validación
Variable independiente: Base de Datos de difonos	Fonemas del alfabeto del español	Frases portadoras.	Selección de frases portadoras	Inventario de frases portadoras	Cumplimiento de Requisitos de Software MBrola
		Difonos	Segmentación de difonos	Inventario de difonos	Cumplimiento de Requisitos de Software MBrola
Variable dependiente: Síntesis del habla del español peruano	Inteligibilidad	Palabras sueltas	Concatenación de difonos	Test de Inteligibilidad	Prueba alfa de Cronbach. Test de Wilcoxon
		Oraciones			
	Naturalidad	Similitud respecto a la voz humana		Test de naturalidad	Prueba alfa de Cronbach. Test de la Mediana

### 3.3 Método de la investigación

La presente investigación es de tipo: Tecnológico aplicada.

Aplicada por cuanto se utiliza conocimiento existente a la fecha y servirá de base para la generación nuevo productos. Tecnológica porque se utilizará conocimiento y tecnologías cuya finalidad es solucionar problemas o situaciones que el conocimiento científico consolidado como tecnología demanda.

Teniendo lo anterior, se empleó el método **inductivo-deductivo** de los aspectos, particulares de las teorías o normas.

### 3.4 Diseño de la investigación

La determinación del tipo de diseño de la investigación se apoya en que: Un estudio de investigación en el que se manipulan deliberadamente una o más variables independientes para medir sus efectos en una variable dependiente (supuestos efectos), dentro de una situación de control para el investigador. Hernández, R (2003, p 220). Respecto a ¿qué es un experimento?:

De acuerdo al enfoque de Hernández y a la definición de nuestras variable independiente y dependiente. Se tiene que:

La variable que comprende la Base de Datos de difonos, es una variable sujeta a la cantidad de difonos, nivel de precisión de segmentación, nivel de insonorización y ruido. Y la variable dependiente está sujeta a la variable independiente, pero sin opción a manipularse. Obviamente, lo que se desea es observar los resultados de la variable dependiente.

M → O

M: Base de Datos de Difonos

O: Síntesis del habla

### 3.5 Población y muestra

Al tener una población muy grande más de 100 000 habitantes en nuestra localidad, según el INEI (2015), se tiene en Nuevo Chimbote 151,127 habitantes, se usó las fórmulas estadísticas para poblaciones infinitas.

$$n = \frac{Z\alpha^2 \cdot p \cdot q}{i^2}$$

*Ecuación 3. Fórmula para hallar el tamaño muestral de una población infinita*

Donde:

*n: tamaño muestral.*

*Z*: valor correspondiente a la distribución de Gauss. Para un  $\alpha=0.05$ , se tiene un  $Z_{0.05} = 1.96$ . O sea para un nivel de significancia de 0.05 se tiene un nivel de confianza de 95%.

*p*: prevalencia esperada del parámetro a evaluar. En este caso, como se desconoce *p* se asigna ( $p=0.5=50\%$ ), entonces ( $q= 1-p=50\%$ )

*i*= error máximo admisible, se considera para la presente tesis un 11.4%  $\leq 20\%$  ( $i=0.114$ )

Reemplazando valores en la ecuación 3, se tiene:

$$n = \frac{1.96^2 \cdot 0.5 \cdot 0.5}{0.114^2} = 73.89$$

*Ecuación 4* Tamaño muestral obtenido al 95% de nivel de confianza

De la ecuación 4, se obtuvo el tamaño muestral de 74 usuarios.

### **3.6 Actividades del proceso investigativo**

- Realizar un análisis de la literatura de la concatenación de unidades fonéticas basada en difonos.
- Elaborar el inventario de fonemas y difonos del español peruano para determinar los difonos realizables o utilizables.
- Elaborar un inventario de audio digital de frases portadoras de unidades de difonos.
- Etiquetar y segmentar los fonemas de las frases portadoras
- Procesar y sintetizar la base de datos, mediante concatenación de difonos
- Determinar los niveles de naturalidad e inteligibilidad de la voz artificial generada.
- Análisis de datos para llegar a conclusiones que corrobore o no nuestra hipótesis.

### 3.7 Técnicas e instrumentos de recolección de datos

<b>TÉCNICA</b>	<b>INSTRUMENTO</b>
Observación.	Registro Visual Registro Auditivo
Edición de Audio	Portadoras wav
Encuestas.	Test de Naturalidad Test de inteligibilidad
Codificación	Audición (grabación)

#### **Software**

Microsoft Office	Java
Sound Forge	C
Gold Wave	Oracle Mysql
Audacy	SPSS, Minitab

### 3.8 Procedimiento para la recolección de datos. (validación y confiabilidad de los instrumentos).

Teniendo en consideración las variables de investigación, la definición operacional de las variables, la muestra, los recursos disponibles.

- Utilizar recursos disponibles para la recolección de datos como recursos económicos y tiempo programado.
- Generar la base de datos de sonidos con equipos de grabación.
- Medir de las fluctuaciones de las variables contenidas en las hipótesis previamente cuantificadas con valores numéricos, mediante instrumentos de medición diseñados.
- Segmentar la base de sonidos para obtener la base de difonos, mediante técnicas de segmentación de sonido con software, midiendo el inicio de corte y término de corte.
- Preparar data para el análisis de los resultados.

### **3.9 Técnicas de procedimiento y análisis de los resultados**

- Validar el instrumento de recolección de datos mediante la prueba del coeficiente alfa de Cronbach.
- Realizar análisis comparativos en base al Test de Naturalidad, a juicio de experto y encuestas a usuarios.
- Realizar análisis comparativos en base al Test de Inteligibilidad, a juicio de experto y encuestas a usuarios.
- Verificar la normalidad de la muestra mediante pruebas estadísticas paramétricas y no paramétricas, según sea el caso.

## **CAPÍTULO IV**

# **Resultados y discusión**

## **4 RESULTADOS Y DISCUSIÓN**

En el presente apartado se muestran los resultados que caracterizan a esta investigación, asimismo, el aporte producto de la experimentación de las metodologías y técnicas aplicadas.

### **4.1 La concatenación de unidades fonéticas basada en difonos.**

En base a la revisión de literatura para elaborar el marco teórico y los antecedentes del presente informe se resume que la síntesis concatenativa se basa en la concatenación de segmentos de voz grabados. Los resultados pueden tener un buen nivel de naturalidad, sin embargo, la variación natural del habla y las técnicas automatizadas de segmentación de formas de onda que se tienen a la fecha pueden dar como resultado defectos audibles, que conllevan a la pérdida de naturalidad del mismo. Se tienen tres tipos básicos de síntesis concatenativa:

**(1) Síntesis por selección de unidades:** Describe que la síntesis por selección de unidades utiliza una base de datos de voz grabada de fonemas, sílabas, palabras, frases y oraciones.

Para lograr la máxima naturalidad a menudo requiere que la base de datos sea muy amplia, llegando en algunos sistemas a los Gb. de datos grabados, por lo que se aplica generalmente a un dominio específico.

**(2) La síntesis de difonos:** Usa una base de datos mínima conteniendo todos los difonos que pueden aparecer en un lenguaje dado. El número de difonos depende de la fonotáctica del lenguaje. La calidad del habla resultante es generalmente peor que la obtenida mediante selección de unidades, pero más natural que la obtenida mediante sintetización de formantes. La síntesis de difonos adolece de los defectos de la síntesis concatenativa en cuanto que puede sonar robótica, pero tiene ventajas respecto a estas técnicas respecto al tamaño de la base de datos. Se continúa usando en investigación y en la mejora de la naturalidad.

**(3) Síntesis específica para un dominio:** Concatena palabras y frases grabadas para crear salidas completas. Se usa en aplicaciones donde la variedad de textos que el sistema puede producir está limitada a un particular

dominio. La naturalidad de estos sistemas puede ser muy alta porque la variedad de oraciones está limitada y corresponde a la entonación y la prosodia de las grabaciones originales, sin embargo, al estar limitados a unas ciertas frases y palabras de la base de datos, no son de propósito general.

Por otro lado, se tienen otros métodos de síntesis: (4) La síntesis de formantes que no es concatenativa y usa modelos acústicos y parámetros de frecuencia para crear formas de onda de habla artificial, el habla es totalmente robótica, pero muy inteligible a altas velocidades evitando defectos acústicos de los métodos concatenativas. (5) Síntesis híbrida según trata de utilizar las síntesis concatenativa con la de formantes, para minimizar efectos acústicos. (6) Síntesis basada en Modelos Ocultos de Markov (HMM) trata de modelar el tracto vocal, la fuente vocal y la prosodia, por modelos ocultos de Markov.

Se estableció una valoración (0) Bajo, (1) Medio, 2(Alto), para los criterios de naturalidad, inteligibilidad, múltiples dominios de aplicación y si utiliza técnicas concatenativas.

Lo valores de la tabla 1, son producto de una calificación del investigador en base al análisis bibliográfico de síntesis de voz.

De acuerdo a la Tabla 1: Valoración de los métodos de síntesis, se observa que la síntesis de mayor valoración para la presente tesis es la de síntesis de difonos, por cuanto tiene un mayor dominio de aplicación frente a las demás técnicas concatenativas, tal como se puede observar en la Fig.1

Cabe indicar, que los métodos 4, 5, y 6 correspondiente a la síntesis de formantes, síntesis híbrida y síntesis basada en HMM, no son concatenativas y que combinan el método concatenativa con otros para obtener mejores resultados de naturalidad.

Tabla 1 Valoración de los métodos de síntesis.

	<b>Método</b>	<b>Naturalidad</b>	<b>Inteligibilidad</b>	<b>Múltiples Dominios de Aplicación</b>	<b>Valoración</b>
1	Síntesis por selección de unidades	0	1	0	1
2	La síntesis de difonos	1	2	2	5
3	Síntesis específica para un dominio	1	2	0	3
4	Síntesis de formantes	0	2	0	2
5	Síntesis Híbrida	1	2	1	4
6	Síntesis HHM	1	2	1	4

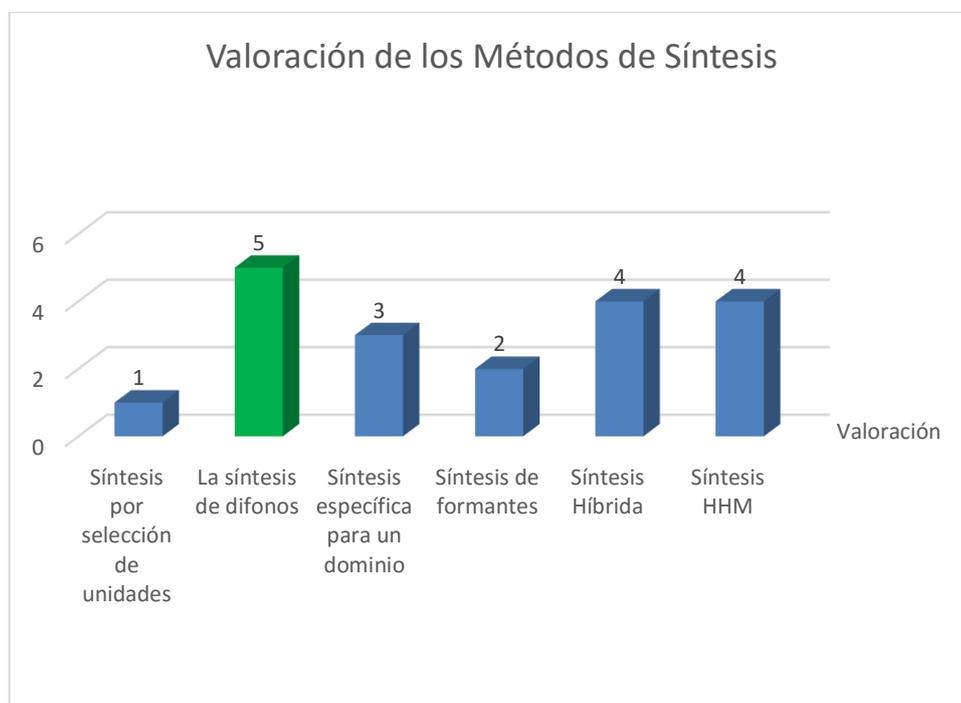


Fig. 4 Valoración de los métodos de síntesis

## Discusión

Se observa que de los tres primeros métodos concatenativos, síntesis de difonos tiene una mayor valoración (5), por cuanto el dominio de aplicación es alto, pero además se observa que los métodos 5 y 6 tienen una alta valoración (4), cercana a la de difonos, por cuanto tienen alta naturalidad.

Se debe indicar que de acuerdo al estado del arte las técnicas híbridas y HHM, tienen buenas posibilidades y potencialidad de extenderse a múltiples dominios de aplicación, conforme se mejoren dichos métodos, por ello esta tabla tiene una valoración cualitativa y no cuantitativa.

## 4.2 Inventario de fonemas y difonos del español peruano.

En esta sección de resultados se tienen dos partes: Una correspondiente a los fonemas del español peruano y otra que corresponde a los difonos que se van a poder utilizar.

### 4.2.1 Inventario de fonemas

El inventario de fonemas sirve para visualizar qué fonemas se tienen en el español peruano, basado en el español que establece la RAE-Asociación de Academias de la lengua española (2005) y también en base a Correa, Rueda, & Argüello (2010). En la Tabla 2, se muestran las grafías (en letra cursiva) y su correspondiente sonido o representación fonética (en letra normal y entre barras), en ella se observa que en el español peruano se produce el seseo, cuando se pronuncia la "z" y no se distingue de la pronunciación de la "S" y tampoco de la "C" cuando su letra vecina es la "e" o "i", por ello en la tabla se muestra que el sonido de la representación fonética de la "Z" y "S" y "Ce", "Ci" es el mismo.

Situación similar se presenta con la representación de las letras "y" y "ll", dado que en algunas zonas se produce el yeísmo que

consiste en pronunciar el sonido de “ll” como el sonido de “y”. Así las letras “b” y “v” producen un mismo sonido.

De lo mencionado en el párrafo anterior se observa en la columna “sonido” de la tabla 2, que existe letras que tienen una misma emisión de sonido, lo cual nos permite simplificar la representación sonora del fonema.

Tabla 2 *Grafía y fonemas de vocales y consonantes del español peruano*

Grafía	fonema	Aplicaciones	Sonido
<i>a</i>	/a/		a
<i>b</i>	/b/		b, v
<i>c</i>	/k/	/k/ seguida por a, o, u /z/ /s/ seguida e, i utilizado en zonas de seseo	k+
<i>d</i>	/d/		d
<i>e</i>	/e/		e
<i>f</i>	/f/		f
<i>g</i>	/g/	/g/ seguida por a, o, u, üe, üi /x/ seguida por e, i	g+
<i>h</i>		La letra <i>h</i> no representa ningún sonido en el español estándar actual, salvo en casos de aspiración	Es muda
<i>i</i>	/i/		i
<i>j</i>	/x/		j, g+
<i>k</i>	/k/		c+, k, q
<i>l</i>	/l/		l
<i>m</i>	/m/		m
<i>n</i>	/n/		n
<i>ñ</i>	/ñ/		Ñ
<i>o</i>	/o/		o
<i>p</i>	/p/		p
<i>q</i>	/k/		
<i>r</i>	/R/ o /r/		r+
<i>s</i>	/s/	/s/ /z/ en zonas de ceceo	c+, s, z
<i>t</i>	/t/		T
<i>u</i>	/u/		u+
<i>v</i>	/b/		v

w	/hu/	/hu/ seguido por una vocal Usada en palabras de origen extranjero	u+
x	/x/, /s/, /k-/s/		x, c+
y	/L/	/hi/, seguido por una vocal	y, ll
z	/s/	/s/ en zonas de seseo	

Tabla 3 Grafía y fonemas vocálicas del español peruano

Grafía	fonema	Letras que lo producen	Fonema similar
á	/A/	á	
é	/E/	é	
í	/I/	í	
ó	/O/	ó	
ú	/U/	ú	

Tabla 4 Grafía y fonemas líquidas del español peruano

Grafía	fonema	Letras que lo producen	Fonema similar
ch	/ch/	Combinación de c-h	c-h
gu	/g/	/g/ seguida por e, i	g+
ll	/l/	/l/ al final de una palabra /L/ en cualquier otro caso /ll/ (/y/ en zonas de yeísmo)	l+
rr	/R/		r+ con otra r+
sh	/sh/	Sonido palatal fricativo	s+
ü	/u/		ü

+ El fonema tendrá una representación sonora diferente dependiendo de la letra vecina.

Fuente: Tabla basada en (RAE- Asociación de Academias de la lengua española, 2005) y adaptada de (Correa, Rueda, & Argüello, 2010)

De la Tabla 2, 3 y 4 simplificando los fonemas que tienen el mismo sonido, se obtuvo la Tabla 5, en la cual se muestra los fonemas que se van a necesitar, para la generación de difonos.

*Tabla 5 Fonemas básicos para la generación de difonos*

Nro.	DIFONO	USO
1	_	Pausa
2	a	Vocal inacentuada
3	a*	Vocal acentuada
4	e	Vocal inacentuada
5	e*	Vocal acentuada
6	i	Vocal inacentuada
7	i*	Vocal acentuada
8	j	Glide: para "i" en diptongos
9	o	Vocal inacentuada
10	o*	Vocal acentuada
11	u	Vocal inacentuada
12	u*	Vocal acentuada
13	w	Glide: para "u" en diptongos
14	b	Oclusiva sonora: Al inicio de sílaba
15	B	Fricativa sonora: Al inicio de sílaba
16	k	Oclusiva sorda: Al inicio de sílaba
17	tʃ	Africada: para el sonido de la digrafía "ch"
18	d	Oclusiva sonora: Al inicio de sílaba
19	D	Fricativa sonora: Al inicio de sílaba
20	f	Fricativa sorda
21	g	Oclusiva sonora: Al inicio de sílaba
22	G	Fricativa sonora: Para "gr", "gl", "gā", "ge", "go", "gu", "gui", "gue"
23	h	Fricativa sorda: (no se encuentra en SAMPA)
24	l	Líquida
25	L	Lateral, para el fricativo palatal "ll" y la "y",
26	m	Nasal
27	n	Nasal
28	J	Nasal: para el sonido de la "ñ"
29	p	Oclusiva sorda: Al inicio de sílaba
30	r	Líquida
31	rr	Líquida
32	s	Fricativa sorda
33	s2	Fricativa sorda: alófono en distensión de la "s"
34	t	Oclusiva sorda: Al inicio de sílaba

Tabla de fonemas, para la generación de difonos basados en SAMPA, y en (Rojas, Blondet, & Mora, 2012)

#### 4.2.2 Metodología para el Inventario de difonos

Una vez que se tiene la tabla de fonemas (Tabla 3) con las consideraciones de simplificar aquellas con el mismo sonido, se procede a determinar qué difonos se van a formar como resultado de la relación (tipo producto cartesiano) entre fonemas.

De manera teórica, si se tienen 33 fonemas más un “silencio” tendrías 34 símbolos (que contienen a las 27 letras y 5 dígrafos del español, según la última edición de la Real Academia Española (2010). Luego, aplicando la fórmula de la Fig.2 multiplicando  $34 \times 34$  se tiene un total de 1156 difonemas entre difonos existentes y no existentes. Se entiende por difonos no existentes, aquellos difonos que no se encuentran en palabra alguna para pronunciarse de acuerdo a la RAE.

Como resultado se obtiene un inventario de difonos del español hablado en el Perú (**EHP**), tal como muestra la Tabla 6; cabe indicar que solo se está mostrando una parte de la tabla completa, y solo se observa las vocales acentuadas y no acentuadas.

#### Discusión

Dado que las Tablas 2, 3 y 4 están fuertemente basadas en lo que propone SAMPA (1996), se debe encontrar algún mecanismo para situaciones de difonos que SAMPA no contempla, sobre todo para el uso de la “h”.

Rojas, Blondet, & Mora (2012), soluciona el problema de SAMPA, respecto al casos de las “h”, sin embargo, en dicho trabajo aún no muestra una solución al problema de los dígrafos “sh”, para palabras como champú, y palabras extranjerizadas como “short”, “Shalom”, que forman parte de nuestra lengua local.

Tabla 6 Inventario de difonos

fxf		a	a*	e	e*	I	i*	j	o	o*	u	u*	w	b	B	k
	+	ta	ta*	te	te*	ti	ti*	tj	to	to*	tu	tu*	tw	tb	tB	tk
<b>a</b>	a+	ata	ata*	ate	ate*	ati	ati*	aj	ao	ao*	au	au*	aw	ab	aB	ak
<b>a*</b>	a*+	a*ta	a*ta*	a*te	a*te*	a*ti	a*ti*	a*tj	a*to	a*to*	a*tu	a*tu*	a*tw	a*tb	a*tB	a*tk
<b>e</b>	e+	eta	eta*	ete	ete*	eti	eti*	ej	eo	eo*	eu	eu*	ew	eb	eB	ek
<b>e*</b>	e*+	e*ta	e*ta*	e*te	e*te*	e*ti	e*ti*	e*tj	e*to	e*to*	e*tu	e*tu*	e*tw	e*tb	e*tB	e*tk
<b>i</b>	i+	ita	ita*	ite	ite*	iti	iti*	ij	io	io*	iu	iu*	iw	ib	iB	ik
<b>i*</b>	i*+	i*ta	i*ta*	i*te	i*te*	i*ti	i*ti*	i*tj	i*to	i*to*	i*tu	i*tu*	i*tw	i*tb	i*tB	i*tk
<b>j</b>	j+	jta	jta*	jte	jte*	ji	ji*	jj	jo	jo*	ju	ju*	jw	jb	jB	jk
<b>o</b>	o+	ota	ota*	ote	ote*	oti	oti*	oj	oo	oo*	ou	ou*	ow	ob	oB	ok
<b>o*</b>	o*+	o*ta	o*ta*	o*te	o*te*	o*ti	o*ti*	o*tj	o*to	o*to*	o*tu	o*tu*	o*tw	o*tb	o*tB	o*tk
<b>u</b>	u+	uta	uta*	ute	ute*	ui	ui*	uj	uo	uo*	uu	uu*	uw	ub	uB	uk
<b>u*</b>	u*+	u*ta	u*ta*	u*te	u*te*	u*ti	u*ti*	u*tj	u*to	u*to*	u*tu	u*tu*	u*tw	u*tb	u*tB	u*tk
<b>w</b>	w+	wta	wta*	wte	wte*	wi	wi*	wj	wo	wo*	wu	wu*	ww	wb	wB	wk
<b>b</b>	b+	bta	bta*	bte	bte*	bi	bi*	bj	bo	bo*	bu	bu*	bw	bb	bB	bk
<b>B</b>	B+	Bta	Bta*	Bte	Bte*	Bi	Bi*	Bj	Bo	Bo*	Bu	Bu*	Bw	Bb	BB	Bk
<b>k</b>	k+	kta	kta*	kte	kte*	ki	ki*	kj	ko	ko*	ku	ku*	kw	kb	kB	kk

fxf es el producto fonema por fonema para generar un difono, cada difono que exista debe estar dentro de una frase portadora.

### 4.3 Inventario de audio digital de frases portadoras de unidades de difonos.

En esta sección, se construyó una base de datos de frases portadoras, donde cada frase portadora contiene un difono existente. Si existe una palabra del español peruano que contiene al difono, entonces se confirma la existencia del difono (Correa, Rueda, & Argüello, 2010).

Al no encontrar una palabra en el español que contenga al difono (dxd), se descarta el difono y se considera como difono inexistente.

#### Método

- Para encontrar una frase portadora, se buscó palabras en el diccionario de la Real Academia Española (2014) y en bibliografía de la localidad de Chimbote.
- Se trata que el difono se encuentre en la parte central de la palabra portadora, salvo los casos de difonos que están al inicio y fin de palabra.
- Se elaboró una tabla con los atributos: [Relación fxf], [palabra(s)\_portadora].
- Una frase portadora puede estar compuesta de una palabra o la unión de dos o más cadenas de palabras.

Por ejemplo, para el siguiente texto “la nube de ámazon”,

- l a - n u b e - d e - Á m a z o n -

Se requieren los siguientes difonos:

-l la a- -n nu ub be e- -d de e- -á ám ma az zo on n-

Los difonos requeridos se encuentran en la Tabla 7, donde el difono fxf se encuentre en cada palabra. El texto “la nube de ámazon” requiere de 18 difonos, como existen dos difonos similares (“e-”) entonces es necesario 17 difonos. Con estos 17 difonos, teóricamente se debe construir el texto ejemplo. Se debe mencionar que estos 17 difonos con sus respectivas palabras portadoras son

solo un fragmento de la tabla de palabra(s) portadoras de difonos del español peruano, que se muestra en el anexo; pues el objetivo fue elaborar la tabla completa de palabras portadoras de difonos, lo cual no fue una tarea trivial, sino que ha requerido de inversión de muchas horas y fineza en la selección de palabras portadoras, dado que, en cada instante, se tiene que verificar si la palabra portadora, además de contener al difono, cubre los aspectos fonéticos del difono de acuerdo a SAMPA.

Tabla 7 Tabla de palabra(s) portadoras de difonos

Nro.	f xf		Palabra(s) portadora
1	_	a*	<i>acto</i>
2	a*	m	<i>chamba</i>
3	a	s	<i>empastado</i>
4	a	_	<i>cola sintética</i>
5	e	_	<i>abastece poco</i>
6	o	n	<i>moneda</i>
7	u	B	<i>nublado</i>
8	B	e	<i>abeja</i>
9	_	d	<i>dolor</i>
10	d	e	<i>adecuado</i>
11	_	l	<i>limpieza</i>
12	l	a	<i>pala</i>
13	m	a	<i>camaradería</i>
14	_	n	<i>nativo</i>
15	n	u	<i>anulado</i>
16	s	o	<i>absoluto</i>
17	n	-	<i>adán hombre</i>

#### 4.4 Etiquetado y segmentado de los fonemas de las frases portadoras

Para mostrar resultados en esta sección se ha seguido los siguientes pasos, las condiciones de la grabación, etiquetado y extracción, tal como recomienda Correa, Rueda, & Argüello (2010).

#### **4.4.1 Consideraciones para la grabación del corpus**

En esta sección se han utilizado los siguientes equipos en una sala insonorizada de los sonidos exteriores, con materiales reductores de ruido.

- Un auricular Estéreo de frecuencia: 20-16,000Hz, Impedancia de 32 Ohm, Sensibilidad  $120 \pm 3$ dB.
- Un micrófono de respuesta de frecuencia 50 a 15,000 Hz., no se recomienda micrófonos de condensador por cuanto puede captar sonidos externos, utilizar micrófonos dinámicos.
- Una laptop con tarjeta de sonido incorporada, core i5, 12 Gb de RAM.

En cuanto a software de audio, se ha utilizado:

- Sound forge 11.0, para la captura del audio.
- Gold Wave, para visualizar los niveles de audio obtenidos en el proceso de captura.

##### **4.4.1.1 Respecto al formato del archivo de audio en Sound Forge 11.0**

Se tienen las siguientes especificaciones para el formato de audio que de las frases portadoras: Extensión de archivo de tipo WAV, Sample rate a 32,000 Hz, Bit depth: 16 bits y un canal Mono. La herramienta sound forge nos da la posibilidad de grabar a dos canales, pero para los fines de la presente investigación y por sugerencia de Rodríguez, Mora, & Cavé (2006), se propone a un solo canal, obviamente, dos canales no justifican, dado que la voz es única y no se utilizará otras técnicas de grabación de sonido como el paneo por ejemplo, en la que se requiere separación de canales. Así, se tienen las especificaciones tal como se puede observar en la Fig. 5

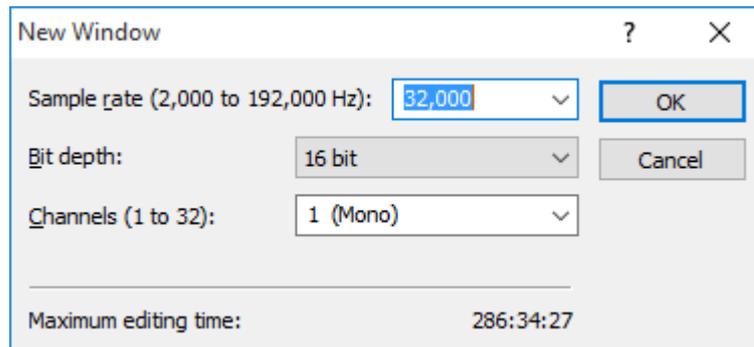


Fig. 5 Especificaciones del audio, para frases portadoras

#### 4.4.1.2 Respecto a la grabación de frases portadoras.

Respecto a la grabación de frases portadoras, se tuvieron que grabar cada uno de las palabras portadoras de la Tabla 7, las mismas que deben ir dentro una frase portadora más extensa, para este caso se ha utilizado la frase:

*“unos dicen \_\_\_\_\_, pero no todos dicen \_\_\_\_\_”.*

Esta frase en realidad no tiene sentido, pues al completar el espacio vacío, se tiene: *“unos dicen **acto**, pero no todos dicen **acto**”*. La idea de aplicar esta técnica es cumplir con lo que concluye Hannane & Hamid (2014): “Que el proceso de concatenación de unidades extraídas de palabras portadoras no debe tener significado alguno, para obtener una total independencia del difono creado y su contexto de pronunciación”; y tiene razón, por ello la frase no debe tener sentido o al menos no es el objetivo de la técnica.

En la Fig. 6, se observa la grabación de la frase *“unos dicen **acto**, pero no todos dicen **acto**”* en la que las frecuencias del rectángulo menor corresponden a *“unos dicen”*, las frecuencias de rectángulo mayor corresponden a *“pero no todos dicen”*, y las frecuencias dentro de las elipses corresponde a la palabra *“acto”*.

Se graba dos veces la palabra *“acto”*, dado que nos permitirá:

- Verificar la estructura de la palabra portadora, y la similitud de sus frecuencias capturadas en tiempos diferentes.
- Nos permitirá escoger la mejor captura de audio para dicha palabra portadora.

Luego se graba en el formato especificado y con el nombre de acuerdo al id de difono, por ejemplo D0002.wav o PORTADORA-002.wav.

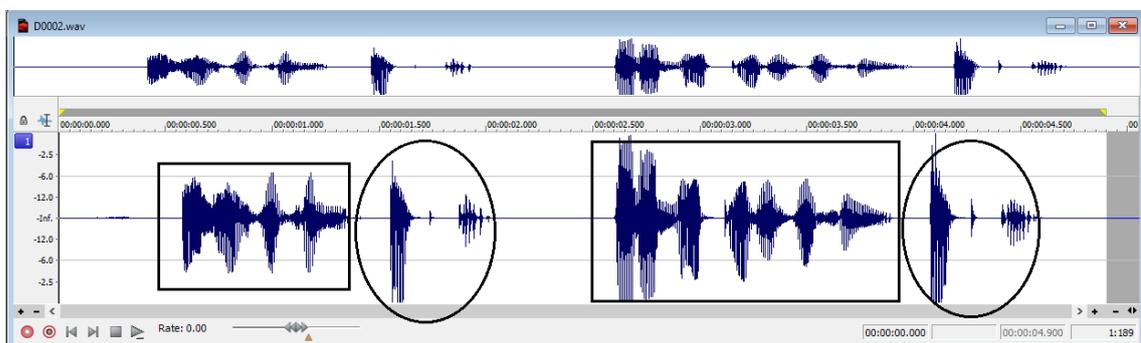


Fig. 6 Grabación de frases portadoras, con Sound Forge 11.0

#### 4.4.1.3 Respecto a la palabra portadora del difono

Hasta este punto de la investigación, se obtuvo el conjunto de frases, de acuerdo al inventario de frases portadoras y se procede al recorte del archivo Dxxxx.wav o PORTADORA-xxx.wav, en el que quedó solo la palabra que contiene al difono, para el caso de ejemplo que se muestra en este informe, se obtuvo solo el audio de la palabra “acto”, que se muestra en la Fig.7.

En la figura 8, se observa la misma palabra, pero con mayor detalle de frecuencias, que tiene tres partes notorias; la primera a la vocálica y consonante “act”, la segunda parte a la sílaba “to” y la tercera a la continuidad del sonido “o”, en decadencia.

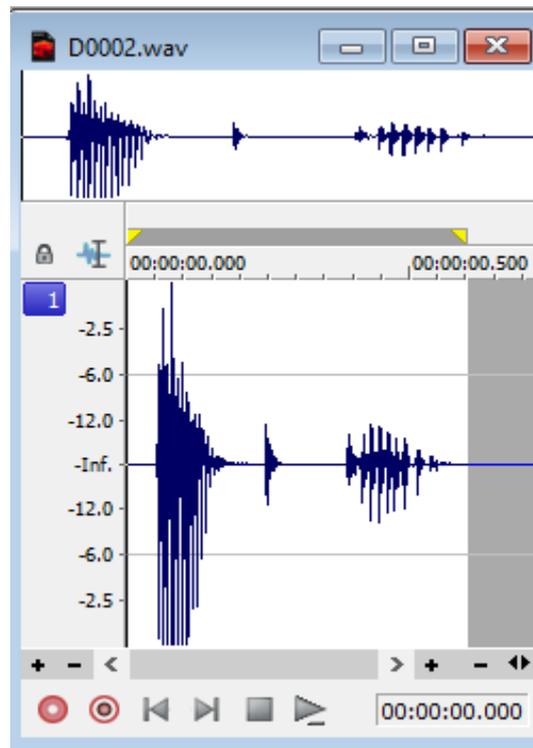


Fig. 7 Archivo audio de la palabra "acto"

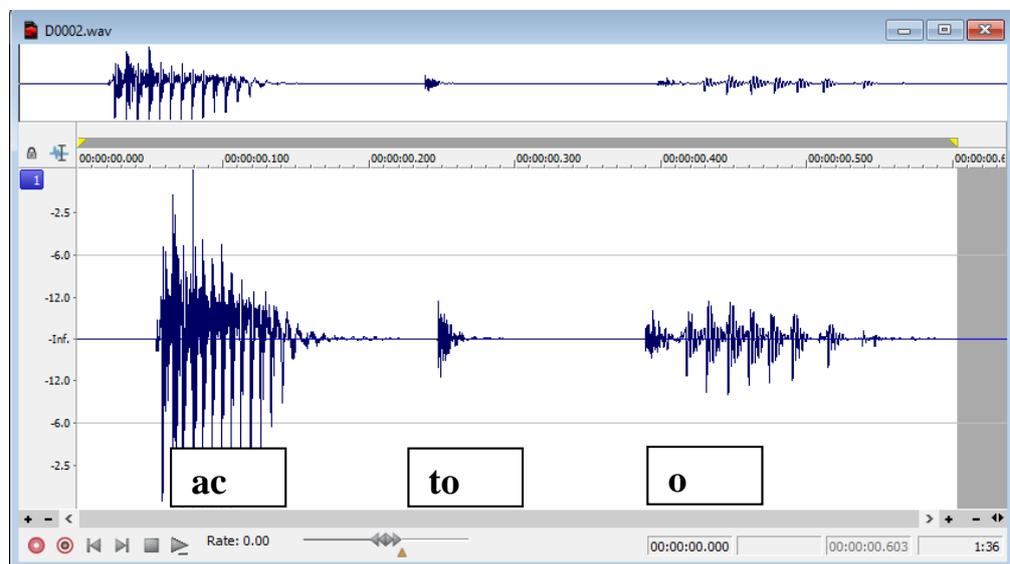


Fig. 8 Archivo de audio de la palabra "acto, con mayor detalle de frecuencias"

#### 4.4.1.4 Respecto al etiquetado y segmentación de difonos.

En esta sección se tiene que comprender dónde empieza y culmina el difono, además, dónde es el límite del fonema.

Por definición para este caso se habla de límite izquierdo del difono, límite derecho del difono y el límite del fonema.

De acuerdo al marco teórico, cada difono está comprendido por la mitad de un fonema y la mitad del fonema vecino. De acuerdo a la Tabla 7: se tiene por ejemplo que el difono “\_a”, o sea un silencio con el fonema “a”. Para obtener el difono correspondiente se establece “medio silencio” más “media vocal a”.

#### **Método**

Mediante el software Diphone Studio, propuesto por The MBROLA Project (2006), se selecciona el audio de la palabra portadora de difono.

Se escucha el audio obtenido.

Se verifica dónde inicia y culmina el sonido del fonema izquierdo y se marca en la mitad del fonema de la izquierda. Ver Fig. 9

Se verifica dónde inicia y termina el sonido del fonema vecino derecho y se marca en la mitad del fonema de la derecha. Ver Fig. 9.

Para las dos situaciones anteriores se utilizó el programa Praat para visualizar el espectrograma de frecuencias de los sonidos correspondientes a la palabra portadora. Ver Fig. 10.

Luego se verifica dónde es el límite entre fonema y fonema, para establecer el marcado del mismo.

De esta manera quedó segmentado, el fonema, estableciendo el fonema objetivo motivo del estudio.

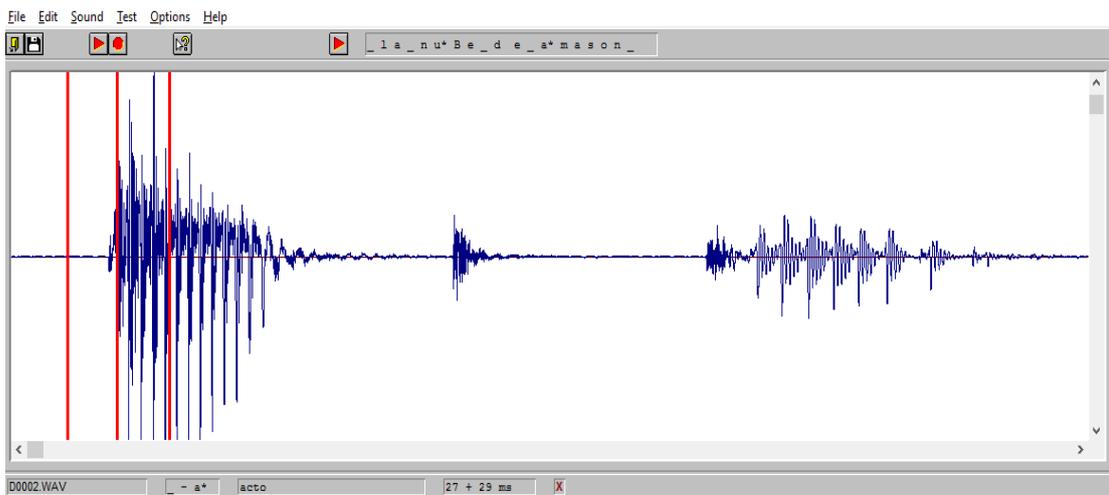


Fig. 9 Segmentación de límites de difonos y fonema

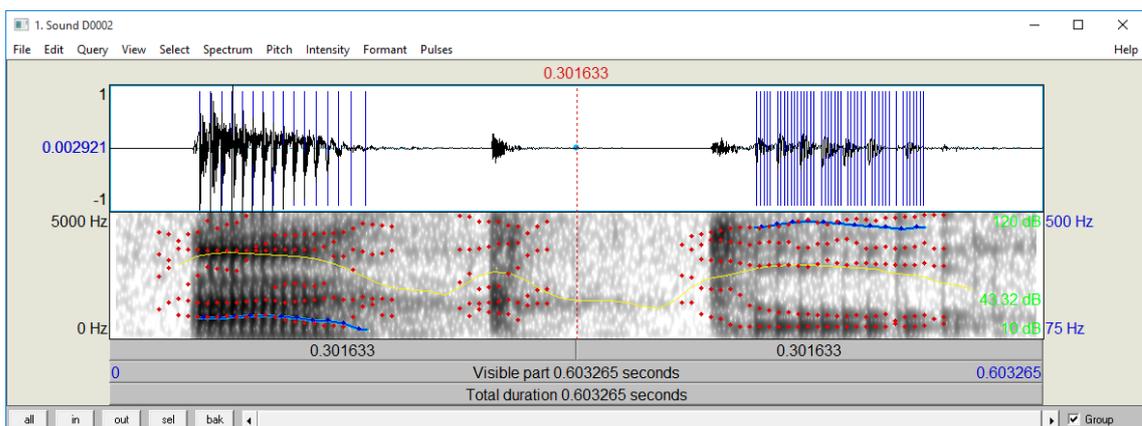


Fig. 10 Espectrograma, pitch, intensidad y formantes de la palabra "acto"

Finalmente al guardar en Diphone Studio, las tres marcas de segmentación del fonema, se guarda en un archivo de texto, previamente preparado con todos los difonos del inventario, los nombres del archivo de audio y la palabra portadora, con delimitadores de campo con las siguientes características: El audio simple rate, el fonema izquierdo, el fonema derecho, el nombre del archivo de audio, la palabra portadora, el límite izquierdo del fonema izquierdo, el límite derecho del fonema vecino derecho y el límite de fonemas.

Tal como se muestra en la Fig. 11, se tiene una porción de lo que es el archivo de límites de fonemas del inventario de fonemas, cuya delimitación de fronteras establece el difono correspondiente. Por ejemplo, en la línea 1 se tiene el simple rate de 32000, a partir de la línea dos, se tienen los fonemas del inventario.

En el registro de la línea 2 se tienen:

\_ : fonema izquierdo

a : fonema derecho

D0002.wav: nombre de archivo de audio

acto: palabra portadora

1389: límite del fonema izquierdo

3312: límite de fonemas

5268: límite del fonema derecho

De manera que los límites de fonema establecidos representan el sonido del difono deseado.

De esta manera se realiza el etiquetado y segmentado de todo el inventario de difonos

1	!	32000					
2	_	a*	D0002.WAV	acto	1389	3312	5899
3	a*	m	D0042.WAV	chamba	10347	12495	15116
4	a	s	D0053.WAV	empastado	16325	20149	24043
5	a	_	D0059.WAV	cola sintetica	12833	14287	15970
6	e	_	D0119.WAV	abastece poco	31486	33833	35766
7	o	n	D0222.WAV	moneda	6761	8812	12196
8	u	B	D0259.WAV	nublado	8270	9532	12915
9	u*	B	D0260.WAV	tuba	5358	8828	12480
10	B	e	D0306.WAV	abeja	10379	12088	13877
11	_	d	D0358.WAV	dolor	2353	5095	7592
12	d	e	D0364.WAV	adecuado	6805	9603	14652
13	_	l	D0469.WAV	limpieza	2390	4596	6655
14	l	a	D0470.WAV	pala	9441	11589	14177
15	m	a	D0508.WAV	camaraderia	9600	11596	13674
16	_	n	D0527.WAV	nativo	596	2394	6240
17	n	u	D0550.WAV	anulado	5876	8976	11855
18	n	u*	D0551.WAV	nunca	3749	4619	6360
19	s	o	D0640.WAV	absoluto	10991	13755	19616
20	n	s	D0734.WAV	adan hombre	18203	22361	29199

Fig. 11 Un ejemplo de un archivo de texto de límites de fonemas

## Discusión

Como se ha descrito, en la presente sección en cuanto al etiquetado y segmentado de las fronteras de fonemas para generar un difono, el proceso es manual.

El investigador solo en base a las recomendaciones de los antecedentes de investigación, en cuanto a segmentación, procede a determinar la posición de las fronteras, lo cual conlleva a realizar prueba y error tratando de obtener un difono audible; sin embargo, al ser un proceso manual, el tiempo de segmentación se vuelve relativo, dependiendo de la calidad de captura del audio de frase portadora original y desconociendo el efecto de la coarticulación del medio difono con los demás difonos resultantes.

Por otro lado, existen difonos creados que necesitan una fuerte agudeza auditiva, y ello también está sujeto a incertidumbre en cuanto a su coarticulación.

### **4.5 Procesamiento y síntesis, mediante concatenación de difonos.**

En esta sección, se procedió a generar la síntesis del habla, en base a los difonos creados y en base a la técnica MBROLA. La técnica MBROLA, es una de las mejores técnicas para concatenación de difonos frente a la técnica PSOLA, tal como nos explican los antecedentes de investigación.

En la Fig. 12 se observa el espectrograma de la frase “la nube de ámazon” pronunciado tres veces y en contraste con la imagen de la Fig. 13, que es la misma frase, pero emitida con voz artificial. Se nota que la estructura y forma son parecidas, lo cual nos lleva a predecir que las frecuencias de sonidos pueden ser similares, cualitativamente hablando, pero la voz natural tiene mayor sonoridad, amplitud de frecuencia en todo el recorrido. La Fig. 14 nos ayuda a entender el comportamiento de las articulaciones y vocalización de una voz natural de la misma frase, y que en contraste con la voz artificial representada en la Fig. 15, tiene mayor sonoridad y mayor amplitud para la mayoría de frecuencias; sin embargo, se considera aceptable dado que, aunque existen cortes o intervalos de silencio en la concatenación de difonos, se logra entender la frase dada la similitud de frecuencias.

Los puntos de color rojo en las figuras 14 y 15 nos representan respectivamente los formantes de las palabras emitidas tanto natural como artificialmente. En la que se nota que hay mayor continuidad en la voz natural, pero en la voz artificial, el concatenador trata de alguna manera de no perder esa continuidad, de manera que se tenga algún margen aceptable de prosodia que incide en la naturalidad del mismo.

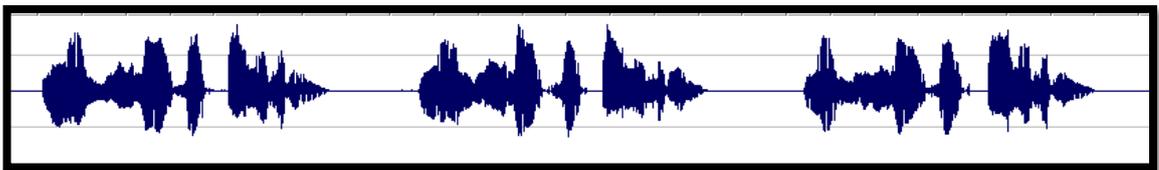


Fig. 12 Espectrograma de voz natural de la frase "la nube de Amazon", pronunciado 3 veces

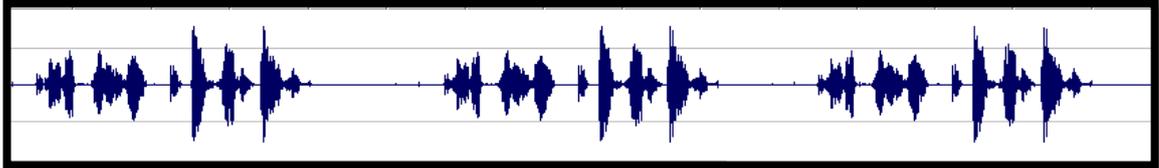


Fig. 13 Espectrograma de voz artificial de la frase "la nube de Amazon" pronunciado 3 veces

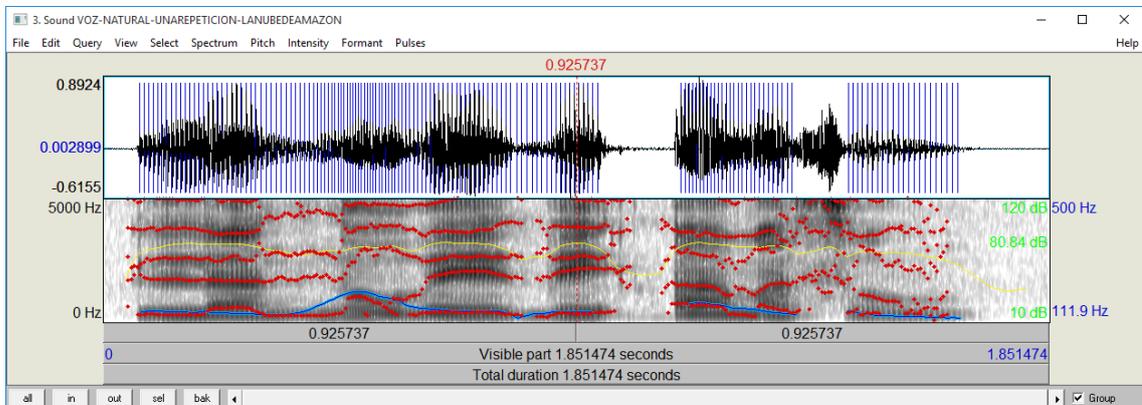


Fig. 14 Espectrograma de voz natural de la frase "la nube de Amazon" pronunciado una vez

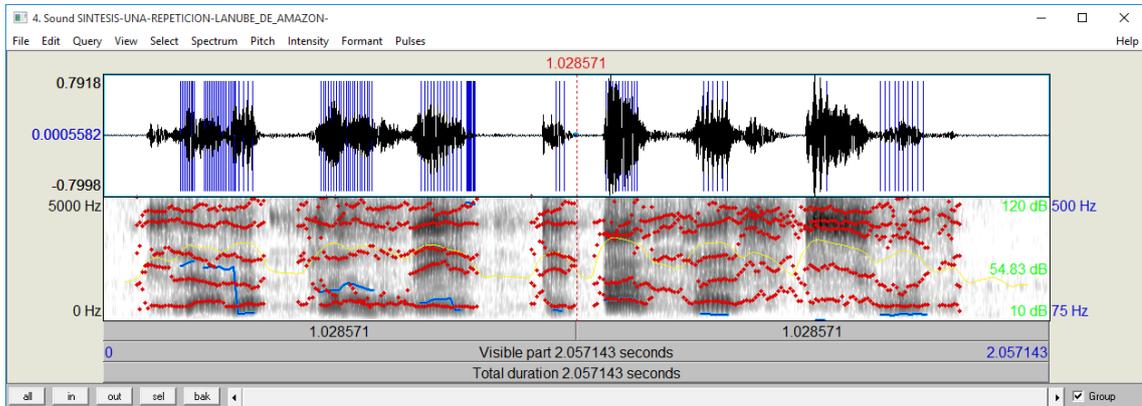


Fig. 15 Espectrograma de voz artificial de la frase “la nube de Amazon” pronunciado una vez

#### 4.6 Determinar los niveles de inteligibilidad y naturalidad de la voz artificial generada.

Para poder determinar los niveles de naturalidad e inteligibilidad, se ha tomado una muestra de 75 usuarios a quienes se ha aplicado dos Test de audición para determinar niveles de inteligibilidad y naturalidad. La escala de valoración utilizada en esta investigación, se basa en la escala Five-Grade, propuesta para medir la calidad de la voz en la Recomendación ITU-R bs.1284-1 de la Unión Internacional de Telecomunicaciones, (ITU, 2003). Como se observa en la Fig. 16, la escala presenta 5 grados de calidad: 1,2,3,4 y 5 para Malo, Pobre, Regular, Bueno y Excelente respectivamente.

Quality		Impairment	
5	Excellent	5	Imperceptible
4	Good	4	Perceptible, but not annoying
3	Fair	3	Slightly annoying
2	Poor	2	Annoying
1	Bad	1	Very annoying

Fig. 16 Escala Five-Grade para la valoración de la calidad de la voz.

Fuente: imagen de la recomendación de la Unión Internacional de Telecomunicaciones (ITU, 2003)

Basado en la escala Five-Grade, se ha realizado una adaptación de la valoración para el caso del Test de inteligibilidad, ver Fig. 17. La misma escala se ha utilizado para el Item1, Item2 e Item3, Item 4 del test de inteligibilidad.

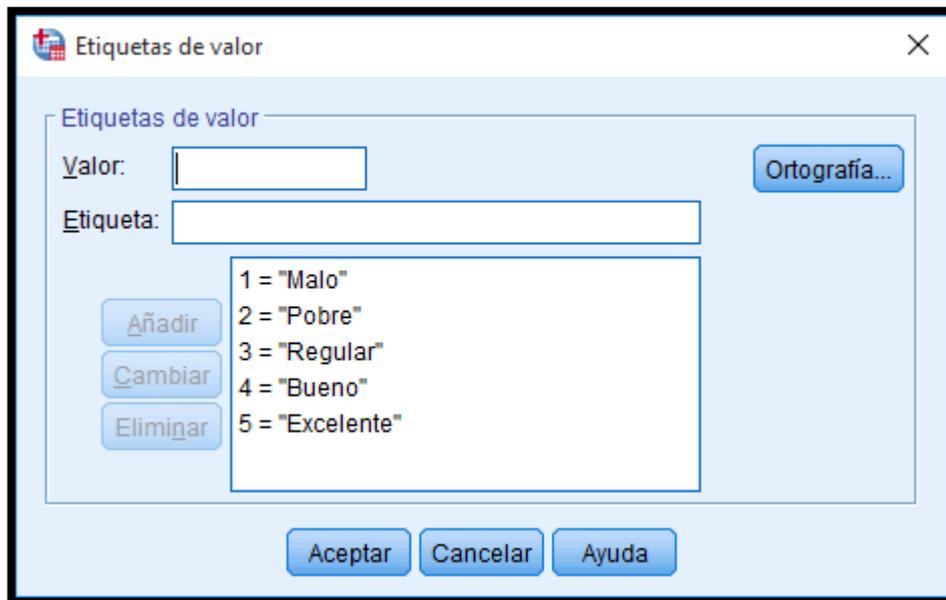


Fig. 17 Escala de Valoración para Ítems de Inteligibilidad

#### 4.6.1 Aplicación del test de Inteligibilidad.

Para el test de inteligibilidad se ha determinado 4 ítems, 3 de ellos corresponden al nivel de comprensión de los sonidos emitidos y el ítem 4 corresponde a una valoración resultado de la experiencia de los tres primeros ítems.

**Ítem 1: Palabras sueltas (Variable Palabras sueltas)**, que son palabras tomadas deliberadamente con contengan, vocal consonante vocal, inicio de consonante y término en consonante, inicio en vocal y término en vocal, inicio en vocal y término en consonante e inicio en consonante y término en vocal. Las palabras elegidas fueron: *Tranquilidad, direccionalidad, discriminación, imágenes, lectura, caracol, cocina, araña, educación y responsabilidad*. Esta variable (palabras sueltas) se aplicó a inteligibilidad.

**Ítem 2: Oraciones (Variable Oraciones con Sentido)**, que fueron frases tomadas del uso de vida cotidiana de los estudiantes universitarios, que tienen sentido completo y requiere un mayor procesamiento para el concatenador y mayor atención de parte de los usuarios, así, se tienen las siguientes oraciones:

*Cumple con tus tareas universitarias.*

*Ayuda en las tareas de la casa.*

*Trabaje con orden y limpieza.*

*Tus padres y profesores valoran tus logros.*

*El cebiche salpreso es muy agradable.*

**Ítem 3: Frases sin sentido (Frases sin sentido)**, que fueron frases de en las que el concatenador las reproduzca y el cerebro del usuario tome mayor tiempo en procesarlas. La idea no es crear una oración con sentido, sino crear una independencia en la audición aumentando la complejidad de audición del usuario. Así, se tienen las siguientes frases:

*Mejor vez cada ser por esfuerzos te.*

*Colegio tu en feliz sientes te.*

*Preferencia de juego tu elige.*

*Vuelan los cerdos en la cebichería.*

**Ítem 4: Calificativo:** que es una valoración resultado de la sumatoria de los ítem1, ítem2, e ítem3, bajo la siguiente escala, tal como se muestra en la Fig. 18.



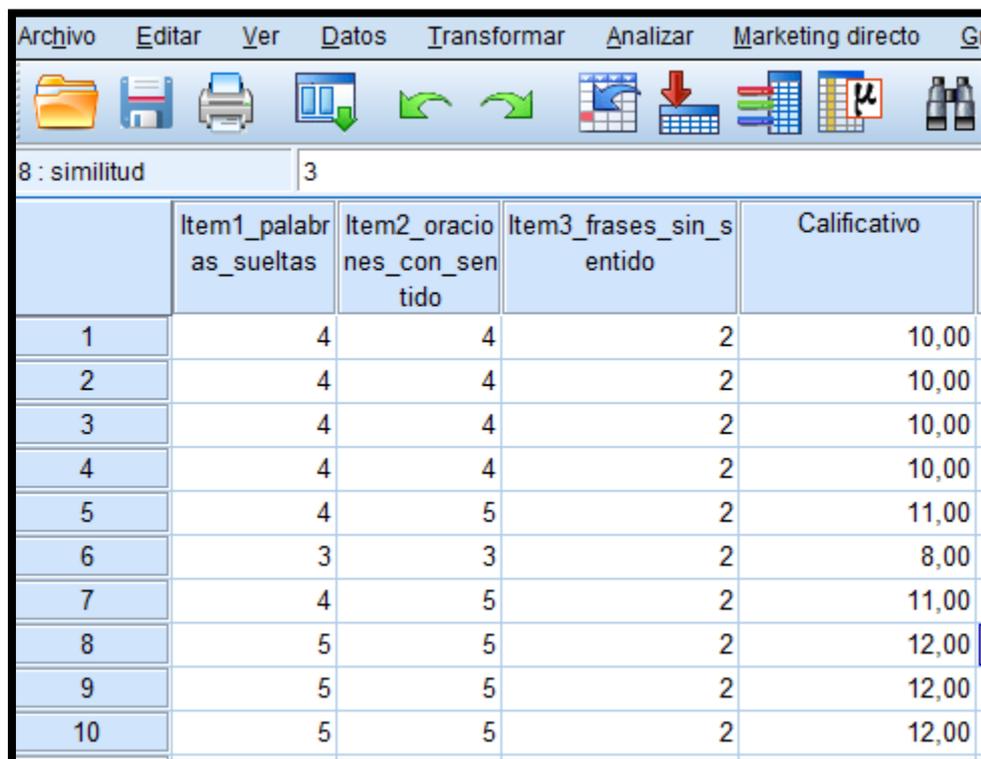
3,00	= "Malo"
4,00	= "Malo"
5,00	= "Malo"
6,00	= "Pobre"
7,00	= "Pobre"
8,00	= "Pobre"
9,00	= "Regular"
10,00	= "Regular"
11,00	= "Regular"
12,00	= "Bueno"
13,00	= "Bueno"
14,00	= "Excelente"
15,00	= "Excelente"

Fig. 18 Escala de Variable calificativo

Luego se configura en SPSS, de manera que la escala de 3 a 15, se ajuste a la escala five-grade.

#### 4.6.2 Registro de datos en SPSS

Una vez realizado el Test de Inteligibilidad a 75 personas (usuarios), se procedió a registrar los datos obtenidos en el instrumento de evaluación, en el programa de software estadístico SPSS. En la Fig. 19 se puede observar un extracto de los datos ingresados, luego de realizar el test de inteligibilidad, en los anexos se puede observar los datos completos de los 75 usuarios.



	Item1_palabras_sueltas	Item2_oraciones_con_sen_tido	Item3_frases_sin_s entido	Calificativo
1	4	4	2	10,00
2	4	4	2	10,00
3	4	4	2	10,00
4	4	4	2	10,00
5	4	5	2	11,00
6	3	3	2	8,00
7	4	5	2	11,00
8	5	5	2	12,00
9	5	5	2	12,00
10	5	5	2	12,00

Fig. 19 Extracto de los datos registrados en SPSS

### 4.6.3 Cálculo del Coeficiente Alfa de Cronbach

Apoyados en el software SPSS se obtuvo el coeficiente alfa de Cronbach, agregando un campo más (Sub\_Total), donde Sub\_Total es la suma de las variables (ítem1+ítem2+ítem3+Calificativo), con el propósito de obtener varianzas de totales de cada ítem, tal como se observa en Tabla 8 de Estadísticos descriptivos.

Tabla 8 Estadísticos descriptivos

Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desviación estándar
Palabras Sueltas	75	2	5	4,41	,699
Oraciones con sentido	75	2	5	4,59	,639
Frases sin sentido	75	1	3	1,83	,476
Calificativo (agrupado)	75	2	5	3,91	,374
N válido (por lista)	75				

A continuación, se procede a obtener el coeficiente alfa de Cronbach, para determinar la validez del instrumento aplicado

$$\alpha = \frac{K}{K-1} \left[ 1 - \frac{\sum S_i^2}{S_T^2} \right]$$

Ecuación 5 Fórmula del Alfa de Cronbach

Donde:

K= número de ítems

S<sup>2</sup><sub>i</sub>= varianza del ítem i.

∑ S<sup>2</sup><sub>i</sub>= sumatoria de la varianza de ítems.

S<sup>2</sup><sub>T</sub>= total de la varianza

$$\alpha = (4/(4-1)) * (1 - (.489 + .408 + .226 + 1.640)/6.261)$$

$$\alpha = 0.744928$$

Tabla 9 Estadística de fiabilidad Alfa de Cronbach obtenido con SPSS

Alfa de Cronbach	N de elementos
.745	4

En este caso se obtuvo un coeficiente aprox. de .745 lo que indica que este instrumento tiene un alto grado de confiabilidad, validando su uso para la recolección de datos. Dado que por convención el Coeficiente Alfa de Cronbach debe ser mayor a 0.5, para determinar la validez del instrumento.

#### **DISCUSIÓN:**

Observar en la Tabla 8 de estadísticos descriptivos que la media de palabras sueltas tiene un valor de 4.41, la media de oraciones con sentido tiene un valor de 4.59, la media de frases sin sentido tiene un valor de 1.83 y la media de Calificativo es 10.81, este calificativo se ve afectado por la media de frases sin sentido que es baja con respecto al de palabras sueltas y oraciones con sentido. Ello se explica por cuanto se presentó mayor dificultad para comprender frases sin sentido, emitidas por la concatenación de difonos y aún si una voz natural produjera frases sin sentido, un usuario oyente tendría problemas de inteligibilidad.

En cuanto al coeficiente alfa de Cronbach, como se puede analizar en la fórmula, se está expresando la correlación entre ítems, para medir la validez del instrumento, este coeficiente .745 para los motivos de la presente investigación se considera aceptable, por cuanto está acercándose al valor 1; y garantiza que si se repite el experimento la validez está garantizada a .745. aprox. 0.8.

#### **4.6.4 Tablas de frecuencia de Palabras Sueltas**

A continuación, se muestran los resultados de los estadísticos descriptivos de frecuencia en la Tabla 10, para la variable Palabras

Sueltas, donde se observa un 1.3% de los usuarios tuvo una comprensión pobre, un 8% tuvo una comprensión regular, un 38.7% tuvo una buena comprensión y un 52% tuvo una excelente comprensión de palabras sueltas, lo cual indica, que la base de difonos creado está permitiendo niveles aceptables de inteligibilidad, donde un 90.7% de usuarios tuvieron una buena y excelente comprensión de palabras sueltas emitidas por concatenación de difonos, ver Fig. 20.

Tabla 10 Tabla de frecuencia de la variable Palabras Sueltas

		Palabras Sueltas			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Pobre	1	1,3	1,3	1,3
	Regular	6	8,0	8,0	9,3
	Bueno	29	38,7	38,7	48,0
	Excelente	39	52,0	52,0	100,0
	Total	75	100,0	100,0	

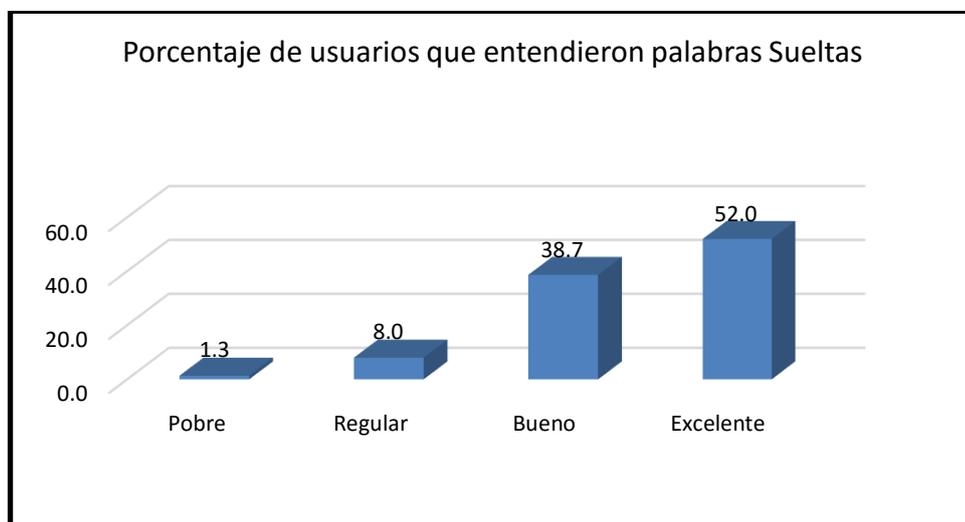


Fig. 20 Gráfico de barras, para la variable palabras sueltas

#### 4.6.5 Tabla de frecuencias de oraciones con sentido

A continuación, se muestran los resultados de los estadísticos descriptivos de frecuencia en la Tabla 11, para la variable

Oraciones con Sentido, donde se observa un 1.3% de los usuarios tuvo una comprensión pobre, un 4% tuvo una comprensión regular, un 29.3% tuvo una buena comprensión y un 65,3% tuvo una excelente comprensión de oraciones con sentido, lo cual indica, que la base de difonos creado está permitiendo niveles aceptables de inteligibilidad, donde un 94.6% de usuarios tuvieron una buena y excelente comprensión de oraciones con sentido emitidas por concatenación de difonos, ver Fig. 21.

Tabla 11 Tabla de frecuencias de la variable oraciones con sentido

		Oraciones con sentido			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Pobre	1	1,3	1,3	1,3
	Regular	3	4,0	4,0	5,3
	Bueno	22	29,3	29,3	34,7
	Excelente	49	65,3	65,3	100,0
	Total	75	100,0	100,0	

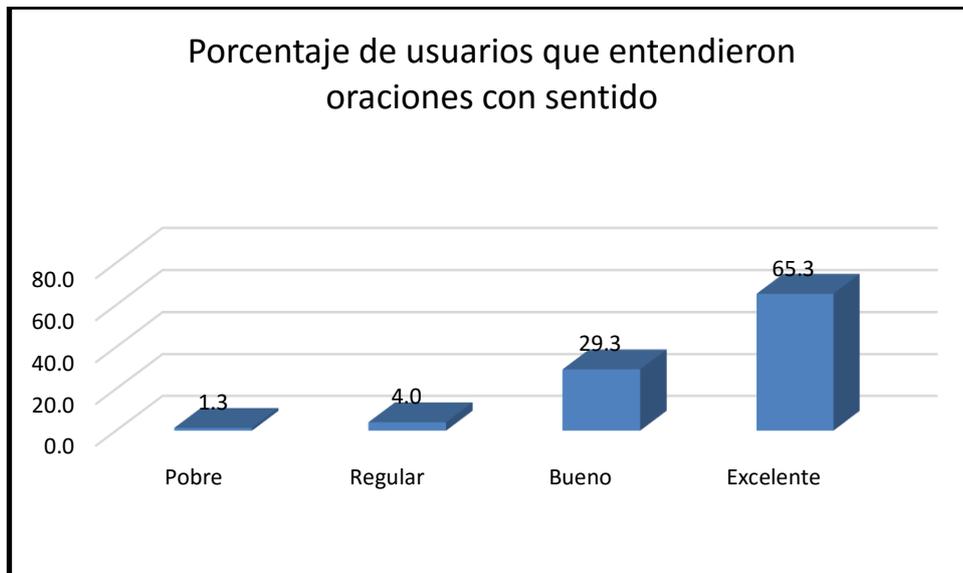


Fig. 21 Gráfico de barra de la variable oraciones con sentido

#### 4.6.6 Tabla de frecuencias de frases sin sentido

Tal como se observa los resultados en la Tabla 11 frecuencias de la variable Frases sin sentido, un 21.3% de los usuarios tuvo una mala comprensión, un 74,3% tuvo una comprensión pobre y un 4% tuvo una regular comprensión, lo cual indica, que esta variable en vez de sumar la inteligibilidad la disminuye, donde el 96% de usuarios tuvieron una mala y pobre comprensión de frases sin sentido emitidas por concatenación de difonos, ver Fig. 21.

Tabla 12 Tabla de frecuencias de la variables frases sin sentido

		Frases sin sentido			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Malo	16	21,3	21,3	21,3
	Pobre	56	74,7	74,7	96,0
	Regular	3	4,0	4,0	100,0
	Total	75	100,0	100,0	

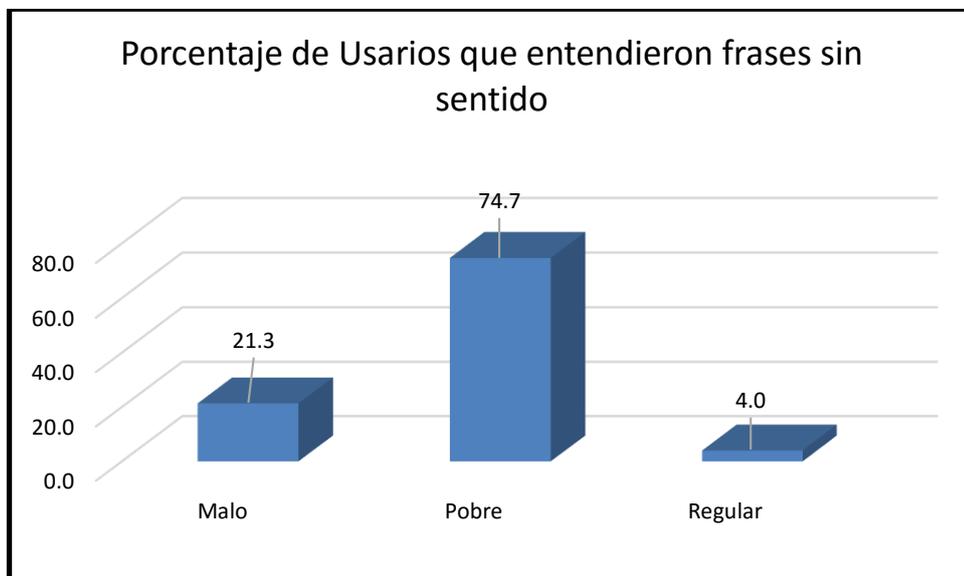


Fig. 22 Gráfico de barras de la variable frases sin sentido

#### 4.6.7 Tabla de frecuencias de la Variable Calificativo

Tal como se observa los resultados en la Tabla 12 tabla de frecuencias de la variable Calificativo, un 1.3% de los usuarios tuvo una pobre comprensión, un 8% tuvo una comprensión regular, un 89.3% tuvo una buena comprensión y un 1.3 tuvo una excelente comprensión, lo cual indica que la valoración de Calificativo, resultado de los tres ítem1, 2 y 3 presenta un buen calificativo en la inteligibilidad por concatenación de difonos, ver Fig. 23

Tabla 13 Tabla de frecuencias de la variable Calificativo

		Calificativo (agrupado)			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Pobre	1	1,3	1,3	1,3
	Regular	6	8,0	8,0	9,3
	Bueno	67	89,3	89,3	98,7
	Excelente	1	1,3	1,3	100,0
	Total	75	100,0	100,0	

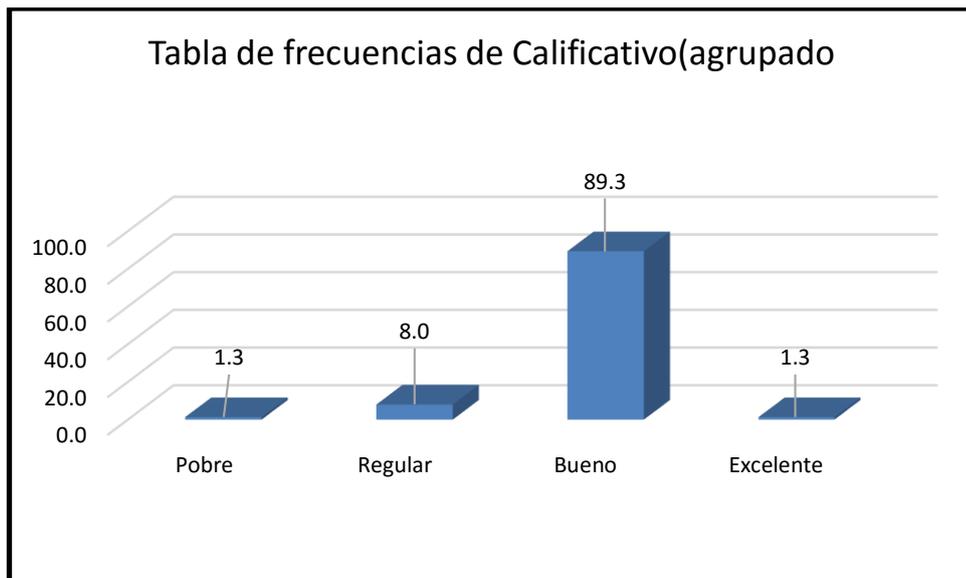


Fig. 23 Gráfico de barras de la variable Calificativo (agrupado)

### **Discusión de los resultados de las tablas de frecuencias.**

Las tablas de frecuencias nos muestran que la media de porcentaje de personas que comprendieron frases con sentido (4.59) es ligeramente mayor respecto al porcentaje de personas que entendieron palabras sueltas (4.41), lo cual indica que en ambos ítem se obtuvieron resultados aceptables, frente a la media de 1.83 de frases sin sentido.

El manejo de esta variable (frases sin sentido) es importante para representar situaciones reales de emisión de voz natural como en ambientes de mucho ruido o problemas de audición. En este caso dada la velocidad de pronunciación de las frases sin sentido, no se da mucho tiempo al cerebro humano poder procesar el sonido emitido por el concatenador de difonos. Es importante recalcar que esta situación también se presenta en un ambiente natural de pronunciación de frases sin sentido, por lo cual se considera válido obtener un 96% de usuarios que no pudieron comprender las frases sin sentido emitidas por el concatenador de difonos, tal como se puede observar en la Fig. 22

#### **4.6.8 Aplicación del Test de Naturalidad**

Para determinar la naturalidad de la voz generada por concatenación de difonos se procedió a pedir que los 75 usuarios, den una valoración five-grade, tal como se observa en la Fig. 24 donde, 1,2,3,4,5 corresponden a Nada, Poco, A medias, Mucho, Muchísimo.

**Similitud a la voz humana**, variable que se utilizó para recibir la valoración de los usuarios en cuanto a su apreciación de similitud de las frases escuchadas respecto a la voz emitida por un ser humano. La pregunta fue: *¿Crees que esta voz se parece a la de un ser humano?*

1 = "Nada"
2 = "Poco"
3 = "A medias"
4 = "Mucho"
5 = "Muchísimo"

Fig. 24 Escala five-grade para la variable similitud de voz natural

Los resultados obtenidos, luego de aplicar el test de naturalidad, fueron los siguientes: Un 5.3% de usuarios percibe que nada se parece a una voz humana, un 25% percibe que se parece poco, un 53.3% percibe que se parece A medias y 16% percibe que se parece mucho, por lo que se puede afirmar que un 69.3% de usuarios se encuentran en el Rango de A medias y Mucho.

Similitud a la Voz Humana					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Nada	4	5,3	5,3	5,3
	Poco	19	25,3	25,3	30,7
	A medias	40	53,3	53,3	84,0
	Mucho	12	16,0	16,0	100,0
	Total	75	100,0	100,0	

## DISCUSIÓN

El test de Naturalidad, muestra un 53.3% de personas que perciben que la voz por concatenación de difonos respecto a una voz natural se parece "A medias", por ser una apreciación subjetiva es aceptable para los fines de la presente tesis, sin embargo, para una investigación posterior se debe mejorar la denominación de esta escala o perfeccionarla.

#### 4.7 Contrastación de la hipótesis

Para la presente tesis, basados en la escala five-grade para inteligibilidad y naturalidad, se tiene que un nivel aceptable se encontraría en el rango de 3 a 5.

Tabla 14 Tabla adaptada de la escala Five-grade del ITU, para inteligibilidad y naturalidad

Puntuación	Según ITU	Naturalidad	Aplicado a Niveles de Calidad Aceptable en (NCA)- Inteligibilidad
1	Malo	Nada	No aceptable
2	Pobre	Poco	Poco aceptable
3	Regular	A medias	Aceptable
4	Bueno	Mucho	Muy aceptable
5	Excelente	Muchísimo	Totalmente aceptable

Tabla basada en la escala five grade según (ITU, 2003)

##### 4.7.1 Paso1: Planteamiento de la hipótesis Nula y la hipótesis alternativa

Se tiene:

###### **Hipótesis Nula: Ho**

La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel no aceptable de inteligibilidad y naturalidad.

###### **Hipótesis Alterna: H1**

La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel aceptable de inteligibilidad y naturalidad.

###### **Hipótesis Alterna: H2**

La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel aceptable de inteligibilidad y un nivel poco aceptable de naturalidad.

### **Hipótesis Alternativa: H3**

La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel poco aceptable de inteligibilidad y un nivel aceptable de naturalidad.

#### **4.7.2 Paso2: Nivel de significancia**

$$\alpha \leq 0.05,$$

$$\text{Nivel de confianza} = 1 - \alpha = 95\%.$$

#### **4.7.3 Paso3: Estadístico de prueba y reglas de decisión**

A continuación, verifican la distribución de las muestras y los estadísticos respectivos.

##### **4.7.3.1 Verificación de distribución de muestras en SPSS**

La Tabla 15 nos muestra el número de casos válidos que son 75 y los parámetros de la distribución seleccionada, es decir, de la distribución normal (Media y Desviación típica). A continuación, ofrece las diferencias más extremas entre las frecuencias acumuladas empíricas y teóricas (la más grande de las positivas, la más pequeña de las negativas y la más grande de las dos en valor absoluto). Por último, ofrece el estadístico de K-S ( $Z = .505$  y  $.296$ ) para inteligibilidad y naturalidad respectivamente y su nivel crítico (Significación asintótica bilateral = 0,000). **Puesto que el valor del nivel crítico es muy pequeño (menor que 0,05), se rechaza la hipótesis de normalidad y se concluye que las puntuaciones de inteligibilidad y naturalidad no se ajustan a una distribución normal.** Por tanto, no se puede usar una prueba paramétrica como la prueba Z o la prueba T-Student, para contrastar la hipótesis, sino una prueba estadística no paramétrica.

Tabla 15 Tabla K-S, para inteligibilidad y naturalidad

Prueba de Kolmogorov-Smirnov para una muestra			Inteligibilidad	Naturalidad
N			75	75
Parámetros normales <sup>a,b</sup>	Media		3,91	2,80
	Desviación estándar		,374	,771
Máximas diferencias extremas	Absoluta		,505	,296
	Positivo		,388	,238
	Negativo		-,505	-,296
Estadístico de prueba			,505	,296
Sig. asintótica (bilateral)			,000 <sup>c</sup>	,000 <sup>c</sup>

- a. La distribución de prueba es normal.
- b. Se calcula a partir de datos.
- c. Corrección de significación de Lilliefors.

#### 4.7.3.2 Prueba estadística No paramétrica de Wilcoxon con Minitab

La prueba de Wilcoxon, para contrastar hipótesis nos permitirá decidir, si se acepta o rechaza las hipótesis nulas y/o alternativas, basados en la mediana de la muestra obtenida.

#### 4.7.3.3 Análisis cuando $H_0=1$ y $H_1 \neq 0$

La Fig.25 muestra la configuración de la Prueba de Wilcoxon para las variables Inteligibilidad y Naturalidad a un 95% de nivel de confianza, y como mediana de la prueba el valor 1, de acuerdo a la escala el indicador de “No aceptable”, y la Hipótesis alterna sería si la mediana es diferente del valor 1.

Entonces, la prueba sería verificar el contraste de la prueba de la mediana para el valor 1 contra la mediana diferente de 1. Ver Fig.26.

**Donde se observa:**

**Para el caso de la variable de Inteligibilidad los siguiente:**

- N=75 casos,
- Nro. de prueba= 75,
- Estadística de Wilcoxon=2850.0,
- Valor P= 0.000 y

La media estimada=4.00

**Para el caso de la variable de Naturalidad los siguiente:**

N=75 casos,

Nro. de prueba= 71, se ha encontrado 5 casos con valor igual a 1

Estadística de Wilcoxon=2556.0,

Valor P= 0.000 y

La media estimada=3.00

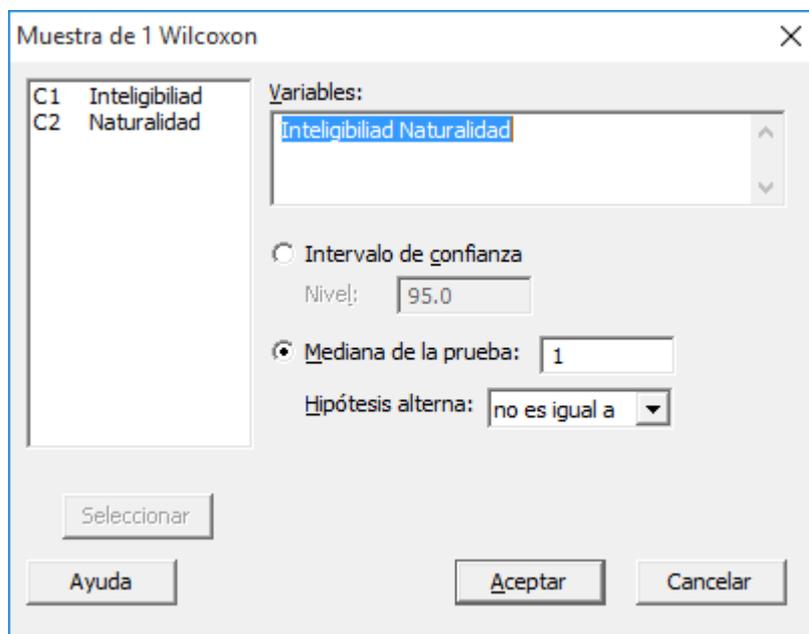


Fig. 25 Configuración de la prueba de Wilcoxon con  $H_0=1$  y  $H_1 \neq 1$

<b>Prueba de clasificación con signos de Wilcoxon: Inteligibilidad, Naturalidad</b>					
Prueba de la mediana = 1.000 vs. la mediana $\neq$ 1.000					
	N	Número de prueba	Estadística de Wilcoxon	P	Mediana estimada
Inteligibilidad	75	75	2850.0	0.000	4.000
Naturalidad	75	71	2556.0	0.000	3.000

Fig. 26 Resultados de la prueba de Wilcoxon con  $H_0=1$  y  $H_1 \neq 1$

En ambas situaciones de Inteligibilidad y Naturalidad nos interesa poner atención al valor P, que es el grado de significancia cuando el valor de prueba de la mediana es 1.

**En el caso de inteligibilidad:** como se observa en la Fig. 26,  $P=0.000$  es menor que 0.05 nos indica que  $H_0$ , se rechaza y se acepta  $H_1$ .

**En el caso de naturalidad:** como se observa en la Fig. 26,  $P=0,000$  es menor que 0.05, nos indica que  $H_0$ , se rechaza y se acepta  $H_1$ .

***Conclusión 0: Se rechaza que la mediana de inteligibilidad y naturalidad sea 1 y se acepta que la mediana de inteligibilidad y naturalidad es diferente de 1.***

#### **4.7.3.4 Análisis cuando $H_0=3$ y $H_1<3$**

La Fig.27 muestra la configuración de la Prueba de Wilcoxon para las variables Inteligibilidad y Naturalidad aun 95% de nivel de confianza, y como mediana de la prueba el valor 3, de acuerdo a la escala el indicador de "Aceptable", y la Hipótesis alterna sería si la mediana se encuentra por debajo del valor 3.

Entonces, la prueba sería verificar el contraste de la prueba de la mediana para el valor 3 contra la mediana menor a 3. Ver Fig.28.

**Donde se observa:**

**Para el caso de la variable de Inteligibilidad los siguiente:**

N=75 casos,

Nro. de prueba= 69, se ha encontrado 6 casos con valor igual a 3

Estadística de Wilcoxon=2380.5

Valor P= 1.000 y

La media estimada=4.00

**Para el caso de la variable de Naturalidad los siguiente:**

N=75 casos,

Nro. de prueba= 35, se ha encontrado 40 casos con valor igual a 3

Estadística de Wilcoxon=192.0

Valor P= 0.022 y

La media estimada=3.00

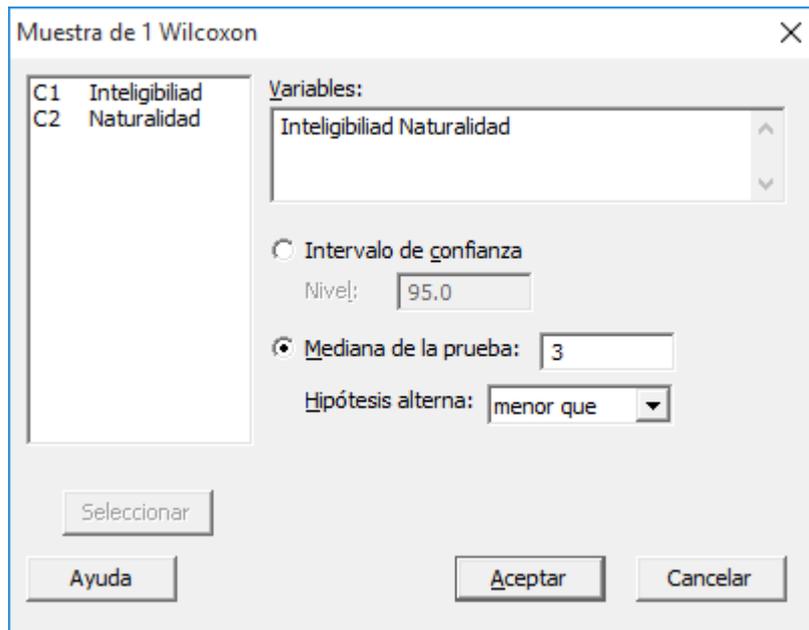


Fig. 27 Configuración de la prueba de Wilcoxon con  $H_0=3$  y  $H_1 <3$

Prueba de clasificación con signos de Wilcoxon: Inteligibilidad, Naturalidad					
Prueba de la mediana = 3.000 vs. la mediana < 3.000					
	N	Número de prueba	Estadística de Wilcoxon	P	Mediana estimada
Inteligibilidad	75	69	2380.5	1.000	4.000
Naturalidad	75	35	192.0	0.022	3.000

Fig. 28 Resultados de la prueba de Wilcoxon con  $H_0=3$  y  $H_1 <3$

En ambas situaciones de Inteligibilidad y Naturalidad nos interesa ponerle atención al valor P, que es el grado de significancia cuando el valor de prueba de la mediana es 3.

**En el caso de inteligibilidad:** como se observa en la Fig. 28,  $P=1.000$  es mayor que  $0.05$  nos indica que  $H_0$ , se acepta y se rechaza  $H_1$ .

**Conclusión 1:** *Se acepta que la mediana de inteligibilidad sea 3 y se rechaza que la mediana de inteligibilidad sea menor que 3.*

**En el caso de naturalidad:** como se observa en la Fig. 28,  $P=0,022$  es menor que  $0.05$ , nos indica que  $H_0$ , se rechaza y se acepta  $H_1$ .

**Conclusión 2:** *Se rechaza que la mediana de naturalidad sea 3 y se acepta que la mediana de naturalidad sea menor que 3.*

#### 4.7.3.5 Análisis cuando $H_0=3$ y $H_1>3$

La Fig.29 muestra la configuración de la Prueba de Wilcoxon para las variables Inteligibilidad y Naturalidad aun 95% de nivel de confianza, y como mediana de la prueba el valor 3, de acuerdo a la escala el indicador de "Aceptable", y la Hipótesis alterna sería si la mediana se encuentra por encima del valor 3.

Entonces, la prueba sería verificar el contraste de la prueba de la mediana para el valor 3 contra la mediana mayor a 3. Ver Fig.30.

**Donde se observa:**

**Para el caso de la variable de Inteligibilidad los siguiente:**

$N=75$  casos,

Nro. de prueba= 69, se ha encontrado 6 casos con valor igual a 3

Estadística de Wilcoxon=2380.5

Valor  $P= 0.000$  y

La media estimada=4.00

**Para el caso de la variable de Naturalidad los siguiente:**

$N=75$  casos,

Nro. de prueba= 35, se ha encontrado 40 casos con valor igual a 3  
 Estadística de Wilcoxon=192.0  
 Valor P= 0.978 y  
 La media estimada=3.00

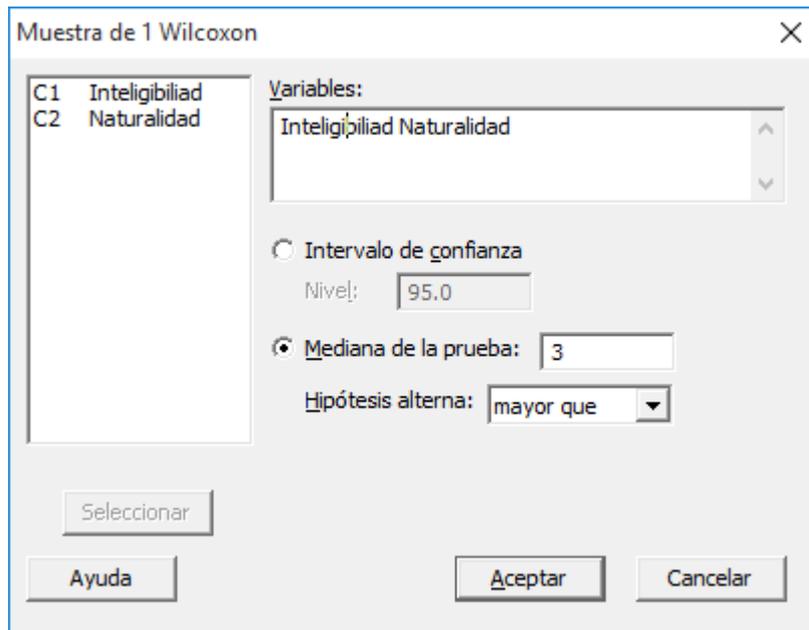


Fig. 29 Configuración de la prueba de Wilcoxon con  $H_0=3$  y  $H_1 >3$

**Prueba de clasificación con signos de Wilcoxon: Inteligibilidad, Naturalidad**

Prueba de la mediana = 3.000 vs. la mediana > 3.000

	N	Número de prueba	Estadística de Wilcoxon	P	Mediana estimada
Inteligibilidad	75	69	2380.5	0.000	4.000
Naturalidad	75	35	192.0	0.978	3.000

Fig. 30 Resultados de la prueba de Wilcoxon con  $H_0=3$  y  $H_1 >3$

En ambas situaciones de Inteligibilidad y Naturalidad similar al caso anterior nos interesa ponerle atención al valor P, que es el grado de significancia cuando el valor de prueba de la mediana es 3.

**En el caso de inteligibilidad:** como se observa en la Fig. 30,  $P=0.000$  es menor que 0.05, nos indica que  $H_0$ , se rechaza y se acepta  $H_1$ .

**Conclusión 3:** *Se rechaza que la mediana de inteligibilidad sea 3 y se acepta que la mediana de inteligibilidad sea mayor que 3.*

**En el caso de naturalidad:** como se observa en la Fig. 28,  $P=0,978$  es mayor que 0.05, nos indica que  $H_0$ , se acepta y se rechaza  $H_1$ .

**Conclusión 4:** *Se acepta que la mediana de naturalidad sea 3 y se rechaza que la mediana de naturalidad sea mayor que 3.*

#### 4.7.4 Aceptación o rechazo de $H_0$ , $H_1$ y $H_2$

**Conclusión 0:** Nos indica que se rechaza el nivel de “No Aceptable” para naturalidad e inteligibilidad y se admite un nivel diferente a “No aceptable”.

**Conclusión 1:** Nos indica que el nivel de inteligibilidad es “Aceptable” y se rechaza que el nivel sea menor.

**Conclusión 2:** Nos indica que se rechaza que la naturalidad sea “Aceptable” y admite que su nivel sea menor o sea “Poco Aceptable”.

**Conclusión 3:** Nos indica que se rechaza que la inteligibilidad sea “Aceptable” y se acepta que su nivel sea mayor o sea “Muy Aceptable”.

**Conclusión 4:** Nos indica que se acepta que la naturalidad, sea “Aceptable” y se rechaza que el nivel sea mayor.

En la Fig. 31 se presenta un cuadro resumen obtenido en base a las conclusiones 0,1,2,3 y 4, donde se puede observar, que Inteligibilidad admite “Aceptable” y “Muy Aceptable”, mientras naturalidad admite “Poco Aceptable” y “Aceptable”, por lo que se puede concluir que, de acuerdo a los resultados obtenidos, las Hipótesis nula  $H_0$  se rechaza, la Hipótesis

H1 se acepta, la hipótesis H2 se acepta y la hipótesis H3 se rechaza, ver Fig.32, todo ello bajo la ponderación de la escala five-grade, a un nivel de confianza de 95% y bajo el estadístico no paramétrico de Wilcoxon, con un nivel de significancia de 0.05.

	Inteligibilidad				Naturalidad			
	No Aceptable	Poco Aceptable	Aceptable	Muy Aceptable	No Aceptable	Poco Aceptable	Aceptable	Muy Aceptable
<b>Conclusión 0</b>								
<b>Conclusión 1</b>								
<b>Conclusión 2</b>								
<b>Conclusión 3</b>								
<b>Conclusión 4</b>								
<b>Resultado</b>								

Fig. 31 Cuadro resumen de conclusiones de inteligibilidad y naturalidad

HIPOTESIS	ACEPTACION /RECHAZO
<b>Hipótesis Nula: Ho</b> La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel no aceptable de inteligibilidad y naturalidad.	
<b>Hipótesis Alterna: H1</b> La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel aceptable de inteligibilidad y naturalidad.	
<b>Hipótesis Alterna: H2</b> La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel aceptable de inteligibilidad y un nivel poco aceptable de naturalidad.	
<b>Hipótesis Alterna: H3</b> La concatenación de unidades de una base de datos de difonos permite la síntesis del habla del español peruano con un nivel poco aceptable de inteligibilidad y un nivel aceptable de naturalidad.	

Fig. 32 Hipótesis aceptadas y/o rechazadas

## **CAPÍTULO V**

# **Conclusiones y recomendaciones**

## **5 CONCLUSIONES Y RECOMENDACIONES**

### **5.1 Conclusiones**

Se realizó un análisis de los métodos de concatenación de unidades fonéticas, determinándose, que a la fecha la concatenación basada en difonos, es una buena solución para la síntesis del habla, por cuanto su aplicación no se limita a un dominio específico, sino que tiene potencialidad para extenderse a muchos dominios, debido a que no tiene las restricciones de los demás métodos de concatenación.

Se elaboró un inventario de fonemas y difonos y se consideró 802 difonos realizables, concluyendo que para los fines de la presente tesis de investigación 802 difonos es una cantidad aceptable para cubrir las posibles palabras a emular.

Se elaboró un inventario de audio digital de 802 frases portadoras de unidades de difonos, todas en formato wav, para mantener la calidad sin compresión, concluyendo que la independencia y neutralidad de la frase portadora de palabras que contienen difonos a grabarse, permiten capturar un mejor grupo de fonemas para la generación de difonos.

Se etiquetó y segmentó los fonemas de las frases portadoras, obteniendo 802 difonos, debidamente segmentados e inventariados, con sus respectivos límites de frontera de fonema y su respectivo archivo de audio, lo que permitió el desarrollo de la base datos de difonos que, mediante la concatenación de los mismos, permitió la síntesis del habla del español peruano, bajo el alcance y límites de la presente tesis de investigación

Se procesó y sintetizó la base de datos de difonos mediante el método de concatenación de difonos, realizando las pruebas necesarias para palabras sueltas, oraciones y frases sin sentido.

Mediante la contrastación de las hipótesis se determinó los niveles de naturalidad e inteligibilidad de la voz artificial generada por concatenación de difonos,

concluyendo que, si existe un nivel aceptable de inteligibilidad y naturalidad, comprobado mediante la prueba de Wilcoxon, dando validez a nuestra hipótesis alternativa H1 que, si existe un nivel aceptable de inteligibilidad y un nivel poco aceptable de naturalidad, dando validez a nuestra hipótesis alternativa H2.

Se determinó, además, que existe un nivel “muy aceptable” de inteligibilidad, situación hipotética no contemplada al inicio del estudio, pero que bajo la prueba de Wilcoxon, se obtuvo la validez necesaria.

## **5.2 Recomendaciones**

Continuar la investigación en términos de aplicar HMM juntamente con difonos, que se estima debe dar mejores resultados, que una concatenación pura de difonos.

Se debe mejorar la base de datos de difonos, con palabras extranjerizadas que ya forman parte de nuestra habla, dado que para la presente investigación se trabajó con una test de 20 frases portadoras sin palabras asimiladas de otros idiomas.

Se deben desarrollar algoritmos para hacer segmentación y etiquetado de manera automática o artificial, para mejorar el proceso y obtener resultados más cuantificables.

Se recomienda mejorar la técnica de unión de fronteras de difonos de MBROLA, para mejorar la prosodia, aun cuando en la inteligibilidad se tienen niveles aceptables, investigaciones posteriores deben incidir en obtener mayor naturalidad.

Se recomienda desarrollar un sintetizador de voz, para la concatenación de difonos, que contemple mejoras en la prosodia y unión de fronteras, del español peruano.

## REFERENCIAS BIBLIOGRÁFICAS

- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6ta ed.). Washington, DC.
- ApkHere. (2016). *ApkHere.com*. Obtenido de ApkHere.com:  
<http://es.apkhere.com/app/com.google.android.tts>
- Balcázar, P., González, N., Gurrola, G., & Moysén, A. (2013). *Investigación cualitativa*. México: Universidad Autónoma del Estado de México. Obtenido de <http://ri.uaemex.mx/handle/20.500.11799/21589>
- Branza, M. (2003). *Los fonemas del español*. Obtenido de Los fonemas del español:  
<http://ebooks.unibuc.ro/filologie/spaniola/2.htm>
- Cabrera, J. (2006). *El dialecto, la voz de los pueblos por excelencia*. Venezuela: Papeles de Padrón.
- Cave, C., Rodriguez, M., Mora, E., Sandrine, C., & Hirst, D. (2006). Un sistema de síntesis de habla en español de Venezuela. *Laboratoire Parole & Langage*. Obtenido de <http://www.lpl-aix.fr/~fulltext/2476.pdf>
- Correa, C., Rueda, H., & Argüello, H. (2010). Síntesis de voz por concatenación de difonemas, para el español de Colombia. *Revista Iberoamericana en Sistemas, Cibernética e Informática*, 7(1), 19-24. Obtenido de [http://www.iiis.org/CDs2010/CD2010CSC/CISCI\\_2010/PapersPdf/CA371ML.pdf](http://www.iiis.org/CDs2010/CD2010CSC/CISCI_2010/PapersPdf/CA371ML.pdf)
- De la Vega, L. (2007). *Diseño de un sintetizador de voz del idioma español hablado en México*. México: Universidad Nacional Autónoma de México. Obtenido de <http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/1915/delavegasegura.pdf?sequence=1>
- Escobar, J. (1978). *Variaciones Lingüísticas del castellano del Perú* (1 ed.). Lima, Perú: Talleres Industrial Gráfica.
- Fernández, F. (2006). *Estudio de la Síntesis de voz y su Aplicación al Software Educativo*. Valladolid, España: GrupodeInvestigación en Reutilización y Orientación a Objeto. Obtenido de <http://www.giro.infor.uva.es/proyectos/memAgenteMS.pdf>
- Flores, C. (2005). *Modelo de entonación para un sintetizador de voz con aplicación a un sistema de información vía telefónica en cines*. Tesis para optar el título de

- Ingeniero Electrónico, Pontificia Universidad Católica del Perú, Lima, Perú.  
Obtenido de  
[http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/285/FLORES\\_ALVAREZ\\_CRISAIDA\\_MODELO\\_ENTONACION\\_SINTETIZADOR\\_VOZ.pdf?sequence=1&isAllowed=y](http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/285/FLORES_ALVAREZ_CRISAIDA_MODELO_ENTONACION_SINTETIZADOR_VOZ.pdf?sequence=1&isAllowed=y)
- García Ocampo, R. A. (2005). *Síntesis de voz en español*. México: Instituto Politécnico Nacional. Obtenido de  
<http://itzamna.bnct.ipn.mx/dspace/bitstream/123456789/15844/1/Sintesis%20de%20voz%20en%20espa%C3%B1ol.pdf>
- Garrido, J. M. (2015). Fonética experimental y tecnologías del habla. *Revista de estudios lingüísticos hispánicos: NORMAS*, 67-79. Obtenido de  
<https://ojs.uv.es/index.php/normas/article/view/6822/6614>
- Gil, J. (2007). *Fonética para profesores del español: De la Teoría a la práctica*. Madrid: Arco/Lbros. Obtenido de  
<https://books.google.es/books?id=WhblMAAACAAJ>
- Hannane, T., & Hamid, A. (2014). Un sistema experto para la lectura automática de texto escrito en árabe estándar. *International Journal on Natural Language Computing (IJNLC)*, 3(2). doi:10.5121/ijnlc.2014.3201
- INEI. (2015). *Perú Instituto Nacional de Estadística e Informática INEI*. Obtenido de Perú Instituto Nacional de Estadística e Informática INEI:  
<https://www.inei.gob.pe/estadisticas/indice-tematico/poblacion-y-vivienda/>
- ITU. (2003). *International Telecommunication Union*. Obtenido de  
[https://www.itu.int/dms\\_pubrec/itu-r/rec/bs/R-REC-BS.1284-1-200312-I!!PDF-E.pdf](https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1284-1-200312-I!!PDF-E.pdf)
- Jusjö, L. (2009). *Peruanismo: Origen, historia, cultura y lengua*. Gothenburg University . Suecia: Gothenburg University Publications Electronic Archive. Obtenido de  
[https://gupea.ub.gu.se/bitstream/2077/20835/1/gupea\\_2077\\_20835\\_1.pdf](https://gupea.ub.gu.se/bitstream/2077/20835/1/gupea_2077_20835_1.pdf)
- Lewis, Paul, M., Simons, G. F., & Fenning, C. D. (2016). Ethnologue: Languages of the world. *SIL International. Online version*(19). Dallas, Texas, Estados Unidos. Obtenido de Ethnologue: Languages of the world:  
<https://www.ethnologue.com/statistics>

- Llisterri, J. (12 de Setiembre de 2016). *Grupode Fonética*. Obtenido de Grupode Fonética, Universidad de Barcelona, :  
[http://liceu.uab.es/~joaquim/phonetics/fon\\_def\\_ambits/fonetica\\_fonologia.html#Fon\\_tica\\_y\\_fonolog\\_a](http://liceu.uab.es/~joaquim/phonetics/fon_def_ambits/fonetica_fonologia.html#Fon_tica_y_fonolog_a)
- RAE- Asociación de Academias de la lengua española. (2005). *Diccionario Panhispánico de dudas*. Colombia: Santillana.
- Real Academia Española. (2010). *Principales novedades de la última edición de la Ortografía de la lengua española*. Obtenido de  
[http://www.rae.es/sites/default/files/Principales\\_novedades\\_de\\_la\\_Ortografia\\_de\\_la\\_lengua\\_espanola.pdf](http://www.rae.es/sites/default/files/Principales_novedades_de_la_Ortografia_de_la_lengua_espanola.pdf)
- Real Academia Española. (2014). *Diccionario de la Lengua Española*. Obtenido de Diccionario de la Lengua Española: <http://dle.rae.es/?w=diccionario>
- Rodriguez H., M. (s.f.). *Manuel Rodriguez Hourcadette*. Obtenido de Manuel Rodriguez Hourcadette: <http://webdelprofesor.ula.ve/ingenieria/hourcade/>
- Rodríguez, M., Mora, E., & Cavé, C. (2006). Síntesis de voz en el dialecto venezolano por medio de la concatenación de difonos. *Revista Ciencia e Ingeniería*, 27(1). Obtenido de  
<http://ecotropicos.saber.ula.ve/db/ssaber/Edocs/papers/humanidades/mora-elsa/dialecto-venezolano.pdf>
- Rojas, N., Blondet, M. A., & Mora, E. (2012). Diccionario de difonos del sintetizador de voz SEVEN. Manual de instalación y uso. *Revista del Centro de Investigación y Atención Lingüística C.I.A.L., Enero-Diciembre*(16), 172-185. Obtenido de  
<http://erevistas.saber.ula.ve/index.php/lenguayhabla/article/view/4162/3946>
- Sadowski, S., & Salamanca, G. (2011). El inventario fonético del español de Chile. *Onomázein*(24), 61-84. Obtenido de  
[https://www.researchgate.net/publication/258514485\\_El\\_inventario\\_fonetico\\_de\\_l\\_espanol\\_de\\_Chile\\_principios\\_orientadores\\_inventario\\_provisorio\\_de\\_consonantes\\_y\\_sistema\\_de\\_representacion\\_AFI-CL](https://www.researchgate.net/publication/258514485_El_inventario_fonetico_de_l_espanol_de_Chile_principios_orientadores_inventario_provisorio_de_consonantes_y_sistema_de_representacion_AFI-CL)
- SAMPA. (1996). Recuperado el 10 de enero de 2016, de SAMPA computer readable phonetic alphabet: <http://www.phon.ucl.ac.uk/home/sampa/spanish.htm>

Saralegui, C. (1997). *El Español Americano: Teoría y textos*. Pamplona, España:  
Ediciones Universidad de Navarra.

*The MBROLA Project*. (2006). Recuperado el 30 de marzo de 2016, de  
<http://tcts.fpms.ac.be/synthesis/mbrola.html>

Torruella, J., & Llisterri, J. (1999). *Diseño de corpus textuales y orales*. Universidad  
Autónoma de Barcelona, Departamento de Filología e Informática. Obtenido de  
[http://latel.upf.edu/traductica/lc/material/torruella\\_llisterri\\_99.pdf](http://latel.upf.edu/traductica/lc/material/torruella_llisterri_99.pdf)

Violante, L. (2012). *Construcción y evaluación del back-end de un sistema de síntesis  
de habla del español argentino*. Tesis de Licenciatura, Universidad de Buenos  
Aires, Buenos Aires. Obtenido de  
<http://www.dc.uba.ar/inv/tesis/licenciatura/2012/violante.pdf>

## **ANEXOS**

# **Anexos**

**Anexo 1: Test – Audición de voz \*      fecha:**

Hola, ..... a continuación, escucharás una serie de sonidos, se te pide que escribas lo que el audio dice. Muchas Gracias. Comunicación Social III.

1	
2	
3	
4	
5	

6	
7	
8	
9	
10	

11	
12	
13	
14	
15	

16	
17	
18	
19	
20	

CALIFICATIVO
--------------

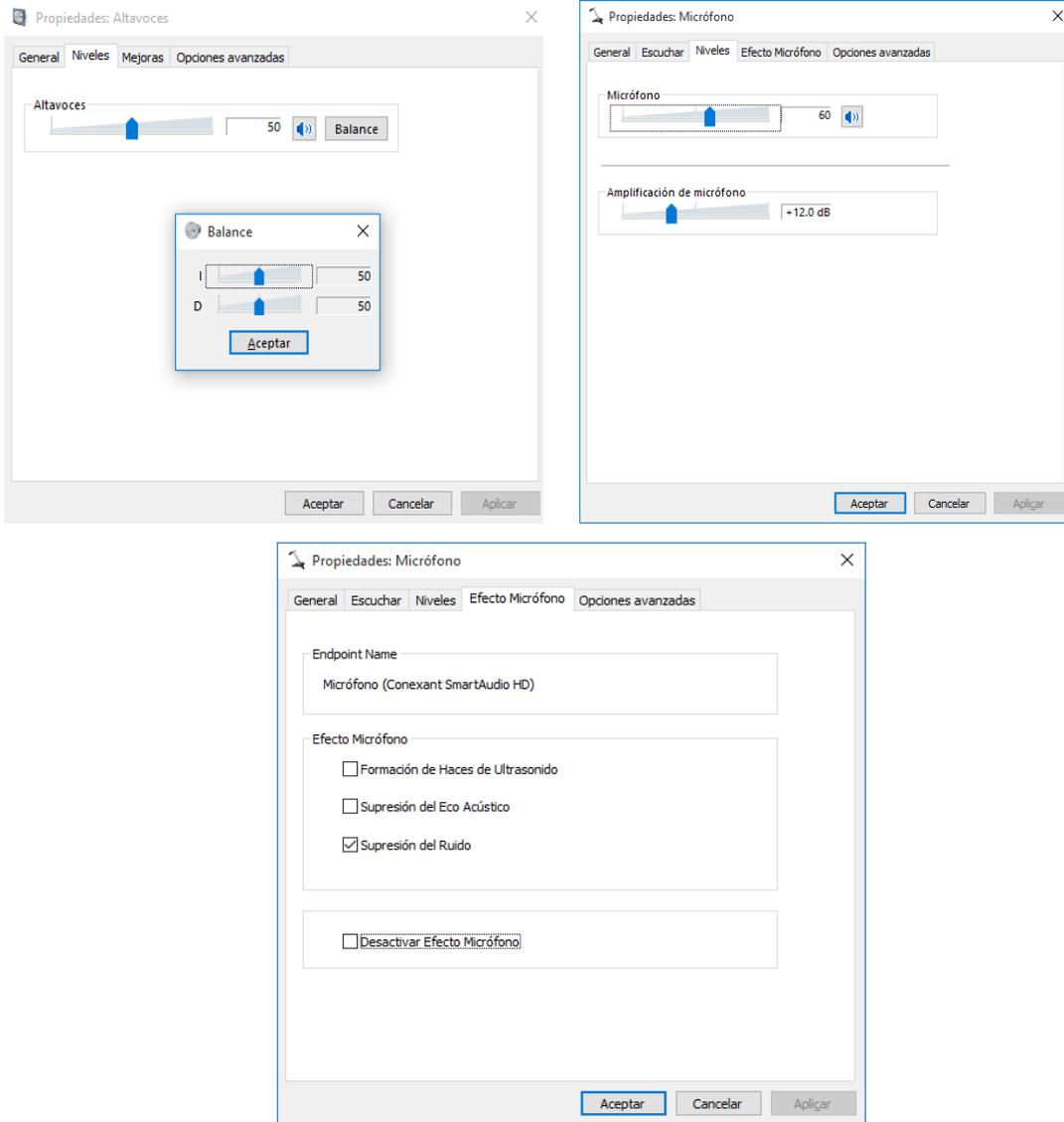
\*Solamente el aplicador del debe saber que se está midiendo la inteligibilidad de la voz sintetizada, lo usuarios no.

**Anexo 2: Test – audición de voz \* fecha:**

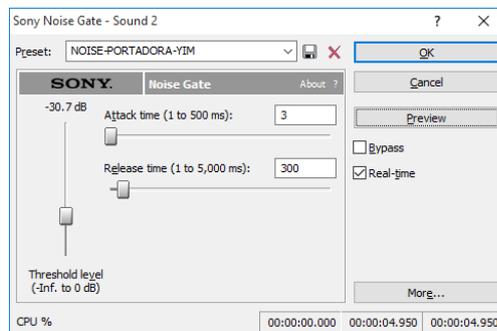
<i>¿Crees que esta voz se parece a la de un humano?</i>		MARCA CON UNA X SOLO UNA OPCION
Muchísimo	5	
Mucho	4	
A medias	3	
Poco	2	
En nada	1	

\*Solamente el evaluador debe saber que está midiendo la naturalidad de la voz sintetizada, los usuarios no.

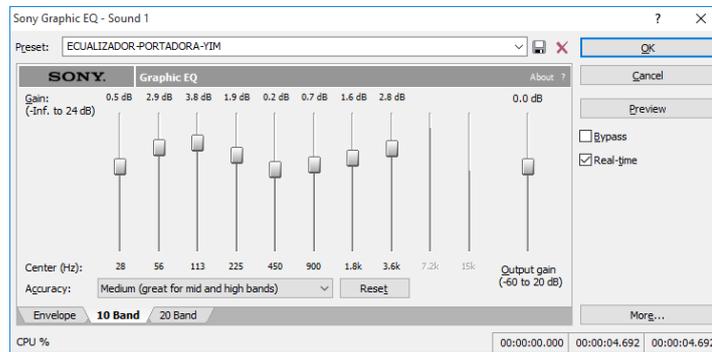
### Anexo 3: Configuración de los dispositivos de reproducción.



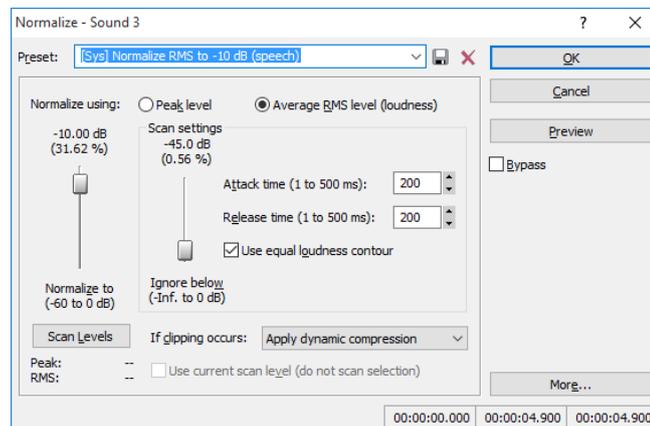
### Configuración de reducción de ruido noise gate.



### Configuración de ecualización para mejora de las frecuencias medias.



### Configuración de normalización.



#### Anexo 4.- Tabla de difonos y palabras portadoras

La tabla de difonos y palabras portadoras, de este anexo, usa los símbolos propuestos por SAMPA (1996) y símbolos utilizados por Rodríguez et al (2006).

nro	f	f	palabra portadora de difono	nro	f	f	palabra portadora de difono
1	_	a	aliento	51	a	rr	abarrotar
2	_	a*	acto	52	a*	rr	espárrago
3	a	a	laantena	53	a	s	empastado
4	a*	a*	tedránimo	54	a*	s	casa
5	a	a*	Abraham	55	a	s2	castrado
6	a*	a	comerá aceituna	56	a*	s2	padrastró
7	a	e	amaestrado	57	a	t	batea
8	a*	e	contrae matrimonio	58	a*	t	cinemática
9	a	e*	antiaéreo	59	a	_	cola sintética
10	a*	e*	usaráépsilon	60	a*	_	cantará muy bien
11	a	i*	recaída	61	_	e	el elefante
12	a*	i*	seráídolo	62	_	e*	el épico
13	a*	j	tráiler	63	e	e	teentiendo
14	a	j	laimágen	64	e*	e*	leeréexodo
15	a	o	laoladebesique	65	e	e*	deéter
16	a*	o	cacao de la selva	66	e*	e	tendréexito
17	a	o*	faraón	67	e	a	plazavea
18	a*	o*	produciráóxido	68	e*	a	patea fuerte
19	a	u*	baúl	69	e	a*	teatro
20	a*	u*	tendráúlceras	70	e*	a*	tendréeánimo
21	a*	w	caucáu	71	e	i*	egreído
22	a	w	mauricio	72	e*	i*	abréido
23	a	B	labeca	73	e	j	afeite
24	a*	B	jugaba	74	e*	j	pleito
25	a	k	aquello	75	e	o	aleonado
26	a*	k	alpaca	76	e*	o	aréola
27	a	tS	agachada	77	e	o*	apoteósico
28	a*	tS	máchica	78	e*	o*	cantaréopera
29	a	D	madera	79	e	w	rehundir
30	a*	D	acobardado	80	e	u*	transeúnte
31	a	f	azafata	81	e*	u*	seréútil
32	a*	f	geográfico	82	e*	w	terapéutico
33	a	G	estómago	83	e	B	ceviche
34	a*	G	decágono	84	e*	B	débil
35	a	h	laguna	85	e	k	recordar
36	a*	h	trágico	86	e*	k	récord
37	a	l	balada	87	e	tS	desechado
38	a*	l	festiváljuvenil	88	e*	tS	mecha
39	a	L	lallama	89	e	D	acueducto
40	a*	L	valle	90	e*	D	adréde
41	a	m	afirmación	91	e	f	adefesio
42	a*	m	chamba	92	e*	f	encéfalo
43	a	n	cansado	93	e	G	alegoría
44	a*	n	canto	94	e*	G	orégano
45	a	J	acompañada	95	e	h	ejecuta
46	a*	J	empaña	96	e*	h	oveja
47	a	p	acapela	97	e	l	meloso
48	a*	p	acápite	98	e*	l	bélico
49	a	r	abarcar	99	e	L	Bellido
50	a*	r	páramo	100	e*	L	pléyade

nro	f	f	palabra portadora de difono	nro	f	f	palabra portadora de difono
101	e	m	membrete	151	i	f	simplifica
102	e*	m	endémico	152	i*	f	científico
103	e	n	encender	153	i	G	figureti
104	e*	n	biogénesis	154	i*	G	dígito
105	e	J	señalizar	155	i	h	fijar
106	e*	J	santeño	156	i*	h	díjole
107	e	p	aceptar	157	i	l	acémila
108	e*	p	estrépito	158	i*	l	bisílaba
109	e	r	mermelada	159	i	L	billar
110	e*	r	estera	160	i*	L	pasillo
111	e	rr	ferrol	161	i	m	ánimo
112	e*	rr	paupérrimo	162	i*	m	centímetro
113	e	s	anestesia	163	i	n	afinador
114	e*	s	apéstegui	164	i*	n	caféina
115	e	s2	luces tiza	165	i	J	escudriñar
116	e*	s2	doméstico	166	i*	J	piña
117	e	t	aletargador	167	i	p	anticipado
118	e*	t	energético	168	i*	p	bípedo
119	e	_	abastece poco	169	i	r	admirador
120	e*	_	conté poco	170	i*	r	círculo
121	_	i	_idiota	171	i	rr	antirrábico
122	_	i*	_íntimo	172	i*	rr	pírrico e írrito
123	i	a	abreviado	173	i	s	visión
124	i*	a*	comiantes	174	i*	s	lindísimo
125	i	a*	diálogo	175	i	s2	islam
126	i*	a	decíase	176	i*	s2	aprisma
127	i	e	tiernamente	177	i	t	pleitista
128	i*	e	hematíes	178	i*	t	ítalo
129	i	e*	piénsalo	179	i	_	papi poco
130	i*	e*	permitiéso	180	i*	_	ají poco
131	i	i	antihistimina	181	_	o	orate
132	i*	i*	chíismo	182	_	o*	ópera
133	i*	i	partímpetuoso	183	o	o	coordina
134	i	o	estudioso	184	o*	o*	aprendióóracol
135	i*	o	tardío poco	185	o	o*	oósfera
136	j	o	audiometría	186	o*	o	plantóórquídeas
137	j	o*	angiólogo	187	o	a	loable
138	i*	o*	partíotro	188	o*	a	coautor
139	j	u	Piura	189	o	a*	aminoácidos
140	j	u*	ventiún	190	o*	a*	cantóálgo
141	i*	u	víuno	191	o	e	autoestima
142	i*	u*	partíuno	192	o*	e	corroe
143	i	B	altibajo	193	o*	e*	metióéso
144	i*	B	bíblico	194	o	e*	poético
145	i	k	mediático	195	o	j	agroindustrial
146	i*	k	pico	196	o*	i*	cantóítalo
147	i	tS	cevichero	197	o	i*	arcoíris
148	i*	tS	chichamorada	198	o*	j	hoy poco
149	i*	D	antídoto	199	o	w	power
150	i	D	ácido	200	o	u	estadounidense

nro	f	f	palabra portadora de difono	nro	f	f	palabra portadora de difono
201	o	u*	finoúgrío	251	u*	j	comecuy poco
202	o*	u*	compróúltimo	252	w	i*	acúícola
203	o	w	show	253	u*	o	dúo
204	o	B	adobera	254	w	o	ingenuo
205	o*	B	aeróbico	255	w	o*	licuó poco
206	o	k	chocar	256	u*	o*	champúóptimo
207	o*	k	enfócate	257	u	u	aylluunido
208	o	tS	anochecidas	258	u	u*	ichuúltimo
209	o*	tS	querroche	259	u*	u	bambúutilizado
210	o	D	codificado	260	u*	u*	Perúúnico
211	o*	D	código	261	u	B	nublado
212	o	f	choferazo	262	u*	B	tuba
213	o*	f	catastrófico	263	u	k	azucarado
214	o	G	abogado	264	u*	k	azúcar
215	o*	G	sismógrafo	265	u	tS	elcucharon
216	o	h	acongojado	266	u*	tS	púchica
217	o*	h	prójimo	267	u*	D	púdrete
218	o	l	polar	268	u	D	adjudicar
219	o*	l	católico	269	u	f	manufactura
220	o	L	pollada	270	u*	f	búfalos
221	o*	L	trambollo	271	u	G	cónyuge
222	o	m	bombaza	272	u*	G	púgil
223	o*	m	cómplice	273	u	h	brujería
224	o	n	moneda	274	u*	h	brújula
225	o*	n	admisión	275	u	l	dulcería
226	o	J	carroñero	276	u*	l	úlceras
227	o*	J	Toño	277	u	L	cuyero
228	o	p	adoptado	278	u*	L	aúlla
229	o*	p	utópico	279	u	m	acumulado
230	o	r	dormitorio	280	u*	m	número
231	o*	r	calórico	281	u	n	punkurí
232	o	rr	borrador	282	u*	n	túnica
233	o*	rr	prórroga	283	u	J	puñalada
234	o*	s	proposito	284	u*	J	pezuña
235	o	s2	acoso	285	u	p	chupamedias
236	o*	s2	próstata	286	u*	p	cúpula
237	o	t	chimbotano	287	u	r	durable
238	o*	t	sótano	288	u*	r	siderúrgico
239	o	_	palo poco	289	u	rr	burrada
240	o*	_	plantó poco	290	u*	rr	burro
241	_	u	ubicado	291	u	s	acusador
242	_	u*	útil	292	u*	s	músico
243	u*	a	garúa poco	293	u	s2	degustar
244	w	a*	cuádruple	294	u*	s2	mayúscula
245	w	a	aguado	295	u	t	computación
246	u*	a*	VirúÁncash	296	u*	t	escúter
247	u*	e	actúe poco	297	u	_	chau poco
248	w	e	desahuevado	298	u*	_	Chimú poco
249	w	e*	huérfano	299	_	b	baño
250	u*	e*	PerúÉxito	300	b	a	bahído

nro	f	f	palabra portadora de difono	nro	f	f	palabra portadora de difono
301	b	a*	básico	351	tS	a*	champa
302	B	a	comprueBA	352	tS	e	anochecer
303	B	a*	arribar	353	tS	e*	escuchépoco
304	B	B	subvención	354	tS	i	unchinguirito
305	B	k	subcontrata	355	tS	i*	unachichadejora
306	B	D	abdicar	356	tS	o	achorado
307	b	e	suversión	357	tS	o*	mechón
308	B	e	jarabe	358	tS	u	esChumpitaz
309	B	e*	alfabético	359	tS	u*	eschúcaro
310	B	G	subgerente	360	_	d	dolor
311	b	i	abreviador	361	d	a	lasdalinás
312	B	i	ámBito	362	d	a*	cadáver
313	b	i*	tabique	363	D	a	podaelgrass
314	B	i*	bígamo	364	D	a*	adánico
315	B	h	inobjetable	365	D	B	adventista
316	b	l	ablandado	366	d	e	adecuado
317	B	l	amoblado	367	D	e	acordebello
318	B	L	subyugar	368	d	e*	andén
319	B	m	submaxilar	369	D	e*	académico
320	B	n	abnegado	370	D	f	Standford
321	b	o	muybonito	371	D	G	landgrave
322	b	o*	adobón	372	d	i	eldiario
323	B	o	aborto	373	D	i	desidia
324	B	o*	carbónico	374	d	i*	audífono
325	B	p	subprefectura	375	D	i*	budín
326	b	r	abrefácil	376	D	h	adjunto
327	B	r	abril	377	D	l	cuodlibeto
328	B	s	abstenerse	378	D	L	ciudadllano
329	B	s2	absoluto	379	D	m	admirable
330	B	t	subteniente	380	D	n	aDnominal
331	b	u	ambulante	381	d	J	comedñuña
332	B	u	estábueno	382	d	o	adornado
333	b	u*	minibúsuevo	383	D	o	abrigadopoco
334	B	u*	eratabú	384	d	o*	bidóngrande
335	_	k	cactus	385	D	o*	bordópoco
336	k	a	lacasona	386	D	p	ciudadprimera
337	k	a*	campo	387	D	k	vodka
338	k	D	anecdótico	388	d	r	ajedrecista
339	k	l	aclamado	389	D	r	Adrián
340	k	n	biotecnólogo	390	D	rr	verdadríca
341	k	o	acorazonada	391	D	s	adscrito
342	k	o*	AncónLima	392	D	s2	adsorbente
343	k	r	acreditación	393	D	t	verdadttotal
344	k	s	accesorio	394	d	u	aducir
345	k	t	infectado	395	D	u	aduana
346	k	u	acuicultura	396	d	u*	chocadura
347	k	u*	acústica	397	D	u*	pedúnculo
348	k	_	kayak poco	398	D	_	ciudad poco
349	_	tS	Chota	399	_	f	foca
350	tS	a	pachamanca	400	f	a	acéfalo

nro	f	f	palabra portadora de difono	nro	f	f	palabra portadora de difono
401	f	a*	abrefácil	451	h	e	adjetivo
402	f	k	kafkiano	452	h	e*	ejército
403	f	e	confesionario	453	h	f	relojfino
404	f	e*	esférico	454	h	g	relojgrone
405	f	G	afgano	455	h	i	eljinete
406	f	i	afinado	456	h	i*	ajípanca
407	f	i*	astrofísico	457	h	h	relojgrande
408	f	l	chifles	458	h	l	cambujlindo
409	f	o	afortunado	459	h	L	cambujlleno
410	f	o*	metafórico	460	h	m	cambujmarino
411	f	r	afrodisiaco	461	h	n	relojnegro
412	f	t	difteria	462	h	o	ajonjolí
413	f	u	afuera	463	h	o*	callejónoscuro
414	f	u*	sulfúrico	464	h	p	relojplano
415	_	g	ganso	465	h	rr	relojruso
416	g	a	adelgazar	466	h	s	relonSanyo
417	g	a*	inorgánico	467	h	t	cambujteñido
418	G	a	abrigadamente	468	h	u	Junín
419	G	a*	agárrate	469	h	u*	conjúbilo
420	G	d	amígdala	470	h	_	reloj poco
421	g	e	alberguepequeño	471	_	l	limpieza
422	g	e*	portugués	472	l	a	pala
423	G	e	Aguedita	473	l	a*	chalán
424	G	e*	manguera	474	l	B	albaca
425	g	i	aguinaldo	475	l	k	alcachofa
426	g	i*	borceguíperuano	476	l	d	alcalde
427	G	i	águila	477	l	e	aletazos
428	G	i*	sanguíneo	478	l	e*	LaCaleta
429	g	l	esglotón	479	l	f	alfalfa
430	G	l	aglutinado	480	l	G	álgebra
431	G	m	dogma	481	l	i	felicidad
432	G	n	cognitivo	482	l	i*	película
433	g	o	mongodb	483	l	h	aljaba
434	g	o*	alegórico	484	l	l	eI-loro
435	G	o	agolpar	485	l	L	ágillanero
436	G	o*	cagón	486	l	m	filmoteca
437	g	r	agradecer	487	l	n	vulnerable
438	G	r	agreste	488	l	J	eIñato
439	G	s	tungsteno	489	l	o	adolorido
440	G	t	washingtoniano	490	l	o*	analógico
441	g	u	agujero	491	l	p	alpaca
442	g	u*	según	492	l	rr	milrayas
443	G	u	agua	493	l	s	propulsa
444	G	u*	fulgúreo	494	l	t	adulto
445	_	h	Jaén	495	l	u	aluvión
446	h	a	Cajamarca	496	l	u*	metalúrgico
447	h	a*	unajáquima	497	l	_	canal poco
448	h	B	relojbueno	498	_	L	llanta
449	h	k	relojCalavinKlein	499	L	a*	Huallanca
450	h	D	relojDamasko	500	L	a	Ayacucho

nro	f	f	palabra portadora de difono	nro	f	f	palabra portadora de difono
501	L	e	Lambayecano	551	n	t	abatimiento
502	L	e*	Lambayequé	552	n	u	anulado
503	L	i	chanchamayino	553	n	u*	nunca
504	L	i*	ensayístico	554	_	J	ñato
505	L	o	Pasamayopeligroso	555	J	a	mañanero
506	L	o*	Sayón	556	J	a*	piñata
507	L	u	ayudantía	557	J	e	puñetea
508	L	u*	mayúsculo	558	J	e*	Cañete
509	_	m	mate	559	J	i	albañilería
510	m	a	camaradería	560	J	i*	alfeñique
511	m	a*	máchika	561	J	o	pañodelagrimas
512	m	b	Chimbote	562	J	o*	acuñomonedas
513	m	e	Diómedes	563	J	u	pañuelo
514	m	e*	cosmético	564	J	u*	lañuña
515	m	i	comisaría	565	_	p	pelota
516	m	i*	Antamina	566	p	a	porPativilca
517	m	l	kremlinología	567	p	a*	Sipán
518	m	L	Abrahamllama	568	p	e	esperuano
519	m	m	ommiada	569	p	e*	terapéutico
520	m	n	calumnia	570	p	i	antipirético
521	m	o	acomodado	571	p	i*	espíritu
522	m	o*	atmósfera	572	p	l	explanada
523	m	p	acampar	573	p	n	hipnotismo
524	m	s	jamster	574	p	o	elpopular
525	m	t	femtogramo	575	p	o*	hipótesis
526	m	u	unamulisa	576	p	r	comprende
527	m	u*	ChimúGrande	577	p	s	autopsia
528	m	_	Abraham poco	578	p	t	acepta
529	_	n	nativo	579	p	u	apuntar
530	n	a	ganadero	580	p	u*	crepúsculo
531	n	a*	CamanáCiudad	581	r	a	paradero
532	n	b	buenvivir	582	r	a*	Huaráz
533	n	k	PunkuríUns	583	r	B	arbusto
534	n	d	altoandino	584	r	k	cercado
535	n	e	carbonería	585	r	D	abordaje
536	n	e*	venéreo	586	r	e	parecido
537	n	f	anfetamina	587	r	e*	diurético
538	n	g	Conga	588	r	f	amorfo
539	n	i	arsénico	589	r	G	argollero
540	n	i*	alienígena	590	r	i	polisacárido
541	n	h	conjugador	591	r	i*	algorítmico
542	n	l	desenlodar	592	r	h	alforja
543	n	L	panllevar	593	r	l	aperlado
544	n	m	inmuneidad	594	r	L	superyó
545	n	n	connotación	595	r	m	dormir
546	n	J	unñato	596	r	n	carnero
547	n	o	afanosamente	597	r	J	cogerñuña
548	n	rr	enramado	598	r	o	Pamparomás
549	n	s	censurado	599	r	o*	cabrón
550	n	S2	amansado	600	r	p	arponear

nro	f	f	palabra portadora de difono	nro	f	f	palabra portadora de difono
601	r	rr	cogerroca	651	s2	u*	unSublime
602	r	s	apoyarse	652	s2	_	yafuimos pocos
603	r	t	artista	653	_	t	tauri
604	r	u	aruñazo	654	t	a	patacino
605	r	u*	Perú	655	t	a*	bitácora
606	r	_	pedir poco	656	t	B	basquetbol
607	_	rr	rabanito	657	t	D	postdoctoral
608	rr	a	esrrabioso	658	t	e	picantería
609	rr	a*	subterráneo	659	t	e*	sodaestéreo
610	rr	e	arremete	660	t	G	postgrado
611	rr	e*	contrarréplica	661	t	i	astillero
612	rr	i	arrimado	662	t	i*	botínperuano
613	rr	i*	aserrín	663	t	l	atletismo
614	rr	o	abarrotado	664	t	m	algoritmo
615	rr	o*	pizarrón	665	t	n	etnocacerista
616	rr	u	arrugado	666	t	o	abotonado
617	rr	u*	verruja	667	t	o*	anatómico
618	_	s	sopa	668	t	p	postproduccion
619	s	a	pasamanos	669	t	r	EITrome
620	s	a*	bursátil	670	t	u	atunero
621	s2	a	lasamigas	671	t	u*	betúnguinda
622	s2	a*	dasasco	672	w	k	embaucado
623	s2	B	asbesto	673	p	j	apropiado
624	s2	k	escalafón	674	p	w	Chimbotepuerto
625	s2	D	desdicha	675	p	_	microchip poco
626	s	e	casetera	676	b	e*	diabético
627	s	e*	desértico	677	b	j	fobia
628	s2	e	vaselina	678	b	w	abuelita
629	s2	e*	estáshecho	679	B	f	subfebril
630	s2	f	estratósfera	680	B	j	cobija
631	s2	G	rasga	681	B	w	esBueno
632	s	i	universidad	682	t	j	tiene
633	s	i*	antitusígeno	683	t	w	atuendo
634	s2	i	pasión	684	t	_	Sunat poco
635	s2	i*	vasija	685	d	j	andinismo
636	s2	h	desjoder	686	d	w	duele
637	s2	l	aislante	687	D	d	Reddedatos
638	s2	L	disyuntivo	688	D	tS	amistadChimbotana
639	s2	m	abismo	689	D	j	acedia
640	s2	n	limosna	690	D	w	unduede
641	s2	J	lasñatas	691	k	e	BesiquePlaya
642	s	o	absoluto	692	k	e*	catequesis
643	s2	o*	asómate	693	k	i	maquillaje
644	s2	p	vispera	694	k	i*	bioquímica
645	s2	rr	desratizar	695	k	G	bistecGrande
646	s2	s	lassandalias	696	k	tS	bistecChalaco
647	s2	t	abastece	697	k	f	bistecfrío
648	s	u	asustado	698	k	h	bistecJimbeño
649	s	u*	ConoSur	699	k	m	lucma o lúcumá
650	s2	u	lasuvas	700	k	L	bistecllano

nro	f	f	palabra portadora de difono	nro	f	f	palabra portadora de difono
701	k	rr	bistecraro	751	j	t	taitaDios
702	k	j	adquiere	752	j	D	paracaidista
703	k	w	acueducto	753	j	k	incaico
704	g	j	guiado	754	j	G	raigambre
705	g	w	guajira	755	j	s	paisaje
706	G	j	aguisado	756	j	s2	aislante
707	G	w	aguaje	757	j	h	arraján
708	tS	j	teleguiado	758	j	m	elAimara
709	tS	w	riachuelo	759	j	n	chanfainita
710	tS	_	sandwich poco	760	j	l	frailecito
711	f	j	confiado	761	j	L	unailu
712	f	w	refuerzo	762	j	r	chaira
713	s	o*	bolsóngrande	763	j	_	Rai poco
714	s	j	grasiento	764	w	p	pauperismo
715	s	w	absuelto	765	w	B	carnauba
716	s2	o	hacesobras	766	w	t	aeronauta
717	s2	tS	deschabe	767	w	D	caudillo
718	s2	j	deshierbas	768	w	G	inaugura
719	s2	w	lashuelgas	769	w	tS	unagauchada
720	h	tS	relojChino	770	w	s	causalidad
721	h	w	jueves	771	w	s2	auspicio
722	m	d	forumdocente	772	w	h	Jauja
723	m	k	forumcalidad	773	w	m	enAumento
724	m	g	forumgenial	774	w	n	fauna
725	m	tS	forumChinecas	775	w	l	faulero
726	m	f	forumfinal	776	w	L	elaullido
727	m	h	forumjuvenil	777	w	f	naufragio
728	m	rr	forumReal	778	w	r	laurencio
729	m	j	abdominales	779	w	rr	aurragado
730	m	w	amueblado	780	w	_	chau poco
731	n	o*	económico	781	p	p	desktopprincipal
732	n	p	input	782	p	B	desktopbásico
733	n	tS	anchoveta	783	i*	i*	ajíinca
734	n	j	ingeniero	784	w	i	saharai
735	n	w	renuevo	785	r	j	variedad
736	n	_	adan hombre	786	r	w	trueque
737	k	p	bistecpobre	787	i*	w	aquíhuincha
738	k	B	bistecBueno	788	i	w	bicihuca
739	k	k	bistecCocido	789	u	j	ichuhierba
740	_	w	huesos	790	s	t	estado
741	_	j	hielo	791	_	S	shock
742	l	j	aliento	792	S	a	Shalom
743	l	w	afluyente	793	S	e	sheriff
744	L	w	polluelo	794	S	i	ancashino
745	rr	j	arriendo	795	S	o	show
746	rr	w	marrueco	796	S	u	Ashu
747	l	tS	elcharqui	797	a	S	flash
748	r	tS	archivador	798	e	S	cueshte
749	j	p	guaipe	799	i	S	espanglish
750	j	B	Laive	800	o	S	Cotosh
				801	u	S	usha
				802	S	_	ídish poco

### Anexo 5. Datos obtenidos en la aplicación del test de inteligibilidad

USUARIO	ITEM1	ITEM2	ITEM3	ITEM4	USUARIO	ITEM1	ITEM2	ITEM3	ITEM4
Usuario1	4	4	2	10,00	Usuario39	2	4	2	8,00
Usuario2	4	4	2	10,00	Usuario40	4	4	1	9,00
Usuario3	4	4	2	10,00	Usuario41	4	5	1	10,00
Usuario4	4	4	2	10,00	Usuario42	5	5	2	12,00
Usuario5	4	5	2	11,00	Usuario43	5	5	1	11,00
Usuario6	3	3	2	8,00	Usuario44	3	2	1	5,00
Usuario7	4	5	2	11,00	Usuario45	4	5	2	11,00
Usuario8	5	5	2	12,00	Usuario46	5	5	2	12,00
Usuario9	5	5	2	12,00	Usuario47	5	5	2	12,00
Usuario10	5	5	2	12,00	Usuario48	3	4	2	9,00
Usuario11	4	5	2	11,00	Usuario49	5	4	1	10,00
Usuario12	5	5	2	12,00	Usuario50	5	5	2	12,00
Usuario13	5	5	2	12,00	Usuario51	5	4	2	11,00
Usuario14	4	5	2	11,00	Usuario52	5	3	2	10,00
Usuario15	4	4	2	10,00	Usuario53	5	5	2	12,00
Usuario16	4	4	3	11,00	Usuario54	5	4	2	11,00
Usuario17	5	5	2	12,00	Usuario55	5	3	2	10,00
Usuario18	5	4	2	11,00	Usuario56	5	5	2	12,00
Usuario19	4	5	2	11,00	Usuario57	5	5	2	12,00
Usuario20	4	4	2	10,00	Usuario58	5	5	2	12,00
Usuario21	5	5	2	12,00	Usuario59	5	5	2	12,00
Usuario22	4	4	2	10,00	Usuario60	5	5	2	12,00
Usuario23	4	5	2	11,00	Usuario61	5	5	2	12,00
Usuario24	3	5	3	11,00	Usuario62	3	5	2	10,00
Usuario25	4	4	1	9,00	Usuario63	5	5	3	13,00
Usuario26	4	5	2	11,00	Usuario64	5	5	2	12,00
Usuario27	5	5	2	12,00	Usuario65	5	5	2	12,00
Usuario28	5	4	1	10,00	Usuario66	5	5	2	12,00
Usuario29	5	5	1	11,00	Usuario67	4	5	2	11,00
Usuario30	5	4	2	11,00	Usuario68	4	5	2	11,00
Usuario31	4	5	1	10,00	Usuario69	5	5	2	12,00
Usuario32	4	5	2	11,00	Usuario70	5	4	1	10,00
Usuario33	3	4	1	8,00	Usuario71	4	5	1	10,00
Usuario34	4	5	1	10,00	Usuario72	5	5	2	12,00
Usuario35	4	5	1	10,00	Usuario73	5	4	1	10,00
Usuario36	4	5	1	10,00	Usuario74	5	4	2	11,00
Usuario37	4	5	2	11,00	Usuario75	4	5	2	11,00
Usuario38	5	5	2	12,00					