

# **TRABAJO ESPECIAL DE GRADO**

## **PROTOCOLO DE TELECOMUNICACIONES PARA LA TRANSMISIÓN CODIFICADA DE FONEMAS ORIGINADOS A PARTIR DE SEÑALES DE VOZ.**

**Presentado ante la Ilustre  
Universidad Central de Venezuela  
por el Br. Ruiz B., Jesús M.  
para optar al Título de  
Ingeniero Electricista**

**Caracas, 2017**

# **TRABAJO ESPECIAL DE GRADO**

## **PROTOCOLO DE TELECOMUNICACIONES PARA LA TRANSMISIÓN CODIFICADA DE FONEMAS ORIGINADOS A PARTIR DE SEÑALES DE VOZ.**

TUTOR ACADÉMICO: Prof. Bruzual Zeldivar

Presentado ante la Ilustre  
Universidad Central de Venezuela  
por el Br. Ruiz B., Jesús M.  
para optar al Título de  
Ingeniero Electricista

Caracas, 2017

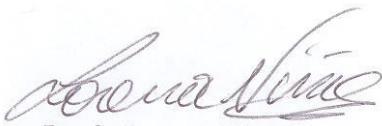
## CONSTANCIA DE APROBACIÓN

Caracas, 03 de noviembre de 2017

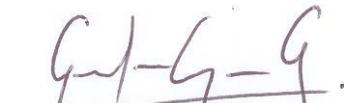
Los abajo firmantes, miembros del Jurado designado por el Consejo de Escuela de Ingeniería Eléctrica, para evaluar el Trabajo Especial de Grado presentado por el Bachiller Jesús M. Ruiz B., titulado:

### “PROTOCOLO DE TELECOMUNICACIONES PARA LA TRASMISIÓN CODIFICADA DE FONEMAS ORIGINADOS A PARTIR DE SEÑALES DE VOZ”

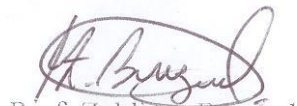
Consideran que el mismo cumple con los requisitos exigidos por el plan de estudios conducente al Título de Ingeniero Electricista en la mención de Comunicaciones, y sin que ello signifique que se hacen solidarios con las ideas expuestas por el autor, lo declaran APROBADO.



Prof. Lorena Núñez  
Jurado



Prof. Gerlis Caropresse  
Jurado



Prof. Zeldivar Bruzual  
Tutor Académico

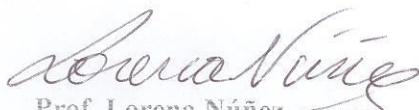


## ACTA


Quienes Suscriben, Miembros del Jurado designado por el Consejo de Escuela de Ingeniería Eléctrica de la Facultad de Ingeniería, de la Universidad Central de Venezuela para examinar el Trabajo Especial de Grado presentado por el Bachiller *JESÚS M. RUIZ B.*, Cédula de Identidad N° 18.829.214, bajo el título "**PROTOCOLO DE TELECOMUNICACIONES PARA LA TRASMISIÓN CODIFICADA DE FONEMAS ORIGINADOS A PARTIR DE SEÑALES DE VOZ**"; a los fines de cumplir con el requisito legal para optar al grado de Ingeniero Electricista, dejan constancia de lo siguiente:

Considerando que el Trabajo Especial de Grado presentado y defendido por el Bachiller en referencia, presenta un grado de originalidad y excelencia tanto en su contenido como en su presentación formal que lo convierte en una fuente referencial útil, acuerda por unanimidad solicitarle a las autoridades correspondientes de la Facultad de Ingeniería, tengan a bien otorgar **MENCIÓN HONORÍFICA**, al Trabajo Especial de Grado antes mencionado.

En Caracas, a los tres (03) días del mes de noviembre de 2017.

  
Prof. Lorena Núñez  
Jurado

  
Prof. Gerlis Caropresse  
Jurado

  
Prof. Zeldivar Bruzual  
Tutor Académico



## **AGRADECIMIENTOS**

A la Universidad Central de Venezuela.

Casa de estudio que me dio la oportunidad de alcanzar esta meta.

A Vero.

Por tu ayuda en edición y corrección de este trabajo, obviamente.

Pero más aún... por tu apoyo incondicional en todos los aspectos de la vida.

Gracias a ti, no solo este trabajo, sino esta carrera.

Por siempre en deuda... por siempre agradecido.

**Ruiz B. Jesús M.**

**PROTOCOLO DE TELECOMUNICACIONES  
PARA LA TRANSMISIÓN CODIFICADA DE FONEMAS  
ORIGINADOS A PARTIR DE SEÑALES DE VOZ.**

**Tutor Académico: Prof. Zeldivar Bruzual.**

**Tesis. Caracas, U.C.V. Facultad de Ingeniería. Escuela de Ingeniería Eléctrica. Opción:  
Comunicaciones.**

**2017, nº pág. 93.**

**Palabras Claves:** Fonemas, Español, Venezuela, Codificación, PCM, Telecomunicaciones.

**Resumen:** Este trabajo desarrolla la propuesta de una trama digital para la transmisión de audio, basada en fonemas de voz. En primera instancia se presenta la investigación teórica en cuanto a fonética y caracterización del sonido de habla, de manera de determinar las características a ser transmitidas. En el trabajo se incluye el estudio de las probabilidades de ocurrencia para los fonemas del español de Venezuela y se compara con trabajos previos, lo que permite proponer una tabla de codificación basada en un código Huffman. También se presentan los diagramas de flujo necesarios en el desarrollo de un algoritmo, para simular la generación de fonemas o realizar la identificación de fonemas en cadenas de texto y posteriormente producir las tramas del tamaño deseado para cualquier cantidad de fonemas obtenidos. Esto permite realizar pruebas comparativas y modificaciones en la trama. Finalmente, al comparar la trama propuesta contra la del estándar PCM de la recomendación G.711 de la ITU-T, se obtuvo un tamaño de trama aproximadamente 170 veces menor y una eficiencia de 99% de la trama propuesta en relación la PCM, a velocidades de habla promedio.

# ÍNDICE

<b>INTRODUCCIÓN.....</b>	<b>1</b>
<b>CAPÍTULO I.....</b>	<b>3</b>
<b>EL PROBLEMA.....</b>	<b>3</b>
1.1. Planteamiento del problema .....	3
1.2. Justificación .....	4
1.3. Alcance .....	5
1.4. Antecedentes .....	5
1.5. Objetivos .....	7
1.5.1. Objetivo General .....	7
1.5.2. Objetivos Específicos.....	7
<b>CAPÍTULO II.....</b>	<b>8</b>
<b>MARCO TEÓRICO.....</b>	<b>8</b>
2.1. Telecomunicaciones.....	8
2.1.1. Estructura de sistema de telecomunicaciones:.....	9
2.2. Modulación .....	11
2.3. Codificación .....	11
2.4. Digitalización .....	12
2.4.1. Muestreo:.....	12
2.4.2. Cuantificación:.....	13
2.4.3. Codificación:.....	13
2.4.4. Multiplexado: .....	15
2.5. Estándar o protocolo de comunicaciones.....	15
2.5.1. Protocolos digitales:.....	15
2.6. Estructura del lenguaje hablado .....	25
2.6.1. Fonema:.....	25
2.6.2. Grafemas:.....	26
2.6.3. Alófono:.....	26
2.6.4. Representación fonética: .....	26
6.5. Alfabeto Fonético Internacional:.....	26

2.7.	Español de Venezuela .....	28
2.7.1.	Seseo: .....	29
2.7.2.	Yeísmo: .....	29
2.7.3.	Betacismo venezolano: .....	29
2.8.	Características del sonido hablado .....	31
2.8.1.	Frecuencia fundamental: .....	31
2.8.2.	Armónicos: .....	31
2.8.3.	Formantes: .....	32
2.8.4.	Timbre: .....	32
2.8.5.	Velocidad de elocución: .....	33
2.8.6.	Intensidad del sonido: .....	33
2.8.7.	Umbral de frecuencia auditiva: .....	36
<b>CAPITULO III .....</b>		<b>38</b>
<b>MARCO METODOLÓGICO .....</b>		<b>38</b>
3.1.	Realización de la librería fonética .....	38
3.2.	Características fonéticas a transmitir .....	42
3.2.1.	Tiempo .....	42
3.2.2.	Intensidad: .....	43
3.2.3.	Tono o frecuencia: .....	44
3.2.4.	Elementos suprasegmentales: .....	45
3.2.5.	Código identificador del fonema: .....	46
3.2.6.	Construcción de la trama: .....	47
3.3.	Comparación con el estándar PCM .....	48
3.3.1.	Programa generador de fonemas: .....	48
3.3.2.	Pruebas de comparación: .....	53
3.4.	Retardo en transmisión de fonemas: .....	55
<b>CAPITULO IV .....</b>		<b>57</b>
<b>RESULTADOS .....</b>		<b>57</b>
4.1.	Verificación del programa generador de fonemas .....	57
4.2.	Comparación con PCM. ....	60
4.2.1.	Tamaño de la trama: .....	60



4.2.4. Eficiencia: .....	63
4.3. Cambios en los parámetros del esquema .....	65
4.3.1. Velocidad del habla: .....	65
4.4. Retardo en transmisión de fonemas: .....	67
<b>CAPITULO V .....</b>	<b>69</b>
<b>CONCLUSIONES Y RECOMENDACIONES .....</b>	<b>69</b>
5.1. Conclusiones.....	69
5.2. Recomendaciones .....	70
<b>MATERIALES DE REFERENCIA CONSULTADOS .....</b>	<b>71</b>
1. Referencias bibliográficas .....	71
2. Bibliografía .....	73
3. Glosario: .....	75

## LISTA DE TABLAS

Tabla 1: Fonemas consonánticos seleccionados.....	30
Tabla 2: Fonemas vocálicos seleccionados. ....	30
Tabla 3: Rangos de velocidades de elocución.....	33
Tabla 4: Frecuencia de aparición fonética en el programa de radio.....	39
Tabla 5 Frecuencia de aparición fonética por tipo de fonema. ....	40
Tabla 6: Fonemas emitidos por minuto en el programa de radio. ....	40
Tabla 7: Tabla comparativa de resultados de frecuencia fonética. ....	41
Tabla 8: Densidad fonética del programa de radio utilizado .....	43
Tabla 9: Codificación Huffman para la lista de fonemas.....	46
Tabla 10: Grupos fonéticos del programa generador. ....	49
Tabla 11: Comparación de tramas PCM y CPF. ....	60
Tabla 12: Resultados obtenidos a velocidad máxima. ....	66
Tabla 13: Resultados obtenidos a velocidad mínima.....	67

## LISTA DE FIGURAS

Figura 1: Diagrama de bloques simplificado de sistema de telecomunicaciones. (Tomasi, 2003)...	10
Figura 2: Ejemplos de códigos de línea. ....	14
Figura 3: Diagrama de puertos de entrada y salida MIDI. [12] .....	17
Figura 4: Diagrama de interconexiones MIDI. [12] .....	17
Figura 5: Esquema de la trama MIDI. [12].....	18
Figura 6: Lista de mensajes MIDI. [12] .....	18
Figura 7: Diagrama de bloques de comunicación MADI. [14].....	21
Figura 8: Trama MADI por canal. [14] .....	21
Figura 9: Codificación de expansión MADI. [14] .....	22
Figura 10: Esquema de puerto transmisor MADI. [14] .....	23
Figura 11: Esquema de puerto receptor MADI. [14].....	23
Figura 12: Fonemas consonánticos pulmonicos del Alfabeto Fonético Internacional. [16] .....	27
Figura 13: Símbolos de características suprasegmentales del IPA. [16] .....	28
Figura 14: Curva PUA vs Frecuencia del sonido (Moore). [24].....	35
Figura 15: Curva PUA contra frecuencia del sonido (Fastl y Zwicker). [25] .....	36
Figura 16: Mínima variación perceptible de frecuencia. [25] .....	37
Figura 17: Caracteres suprasegmentales IPA implementados. [16] .....	45
Figura 18: Trama detallada para un fonema a transmitir .....	47
Figura 19: Diagrama de flujo de programa generador en modo aleatorio.....	51
Figura 20: Diagrama de flujo de programa generador en modo directo. ....	52
Figura 21: Comparación entre señal analógica, trama PCM y trama CPF.....	58
Figura 22: Trama CPF detallada para fonemas /k/ y /a/. ....	59
Figura 23: Comparación de tamaños de trama PCM y CPF.....	61
Figura 24: Tamaños de trama CPF en generación aleatoria. ....	62
Figura 25: Eficiencia de trama CPF respecto a la trama PCM. ....	64
Figura 26 : Ventana de tiempo de transmisión en la trama CPF.....	65

# INTRODUCCIÓN

Un fonema es la representación mínima del lenguaje hablado que permite la distinción entre palabras de un idioma. Son construcciones abstractas diseñadas para el estudio formal del lenguaje a partir del sonido hablado y su significado dentro del idioma. Un fonema tiene significado lógico dentro del lenguaje y no debe ser confundido con un grafema (representaciones graficas de los sonidos o también llamados letras). [1]

En la estructura del lenguaje, producto de la unión de varios fonemas, y pasando por formas intermedias, se llega a la construcción de las palabras que, como bien se sabe, son portadoras de información lingüística concreta y que finalmente, son las unidades que se desea transmitir entre interlocutores.

Todo lenguaje (idioma) está conformado por un conjunto limitado de fonemas, que se mezclan entre sí, para formar palabras en una cantidad de combinaciones también limitada. Por razones culturales, fonemas iguales tienen diferentes probabilidades de ocurrencia en idiomas diferentes; incluso, en regiones diferentes donde se hable el mismo idioma, lo que crea el incentivo de ajustar la forma en la que se maneja los mensajes de cada idioma, según sus características estadístico-fonéticas.

“La frecuencia con que los sonidos se repiten en la cadena hablada caracteriza, sin duda, fonéticamente una lengua dada. Igualmente, la frecuencia con que los fonemas aparecen en la elocución con valor distintivo contribuye a caracterizar fonológicamente tal lengua, mostrando el rendimiento funcional de cada uno de ellos.” (Alarcos Llorach, 1965, pág. 197)

Un protocolo de comunicación define un conjunto concreto de normas y reglas de transmisión que permiten a los equipos de comunicación ponerse de acuerdo en cómo debe realizarse la comunicación a través de un canal determinado [2]. Estos suelen estar definidos por medio de un estándar, o recomendación, en el caso de las recomendaciones de la ITU.

Existen muchos estándares usados en codificación y transporte de audio, un ejemplo es el estándar MIDI, usado en la comunicación entre instrumentos musicales y con vigencia desde los años 80. Este protocolo se centra en transferencia de datos (notas, tiempos, tono, volumen, sincronización, etc.) y comandos de control entre instrumentos musicales, pero no incluye el manejo de voz. [3]. Existen muchos otros protocolos, tales como: MIDI y *Open Sound Control*, que serán descritos a lo largo de este proyecto. [4]

En este trabajo se pretende lograr el diseño de una trama eficiente para la transmisión codificada de fonemas de voz, aplicada especialmente para el idioma español de Venezuela. La hipótesis que se plantea es que una codificación diseñada especialmente para tal fin (para un tipo de información específica), generaría un aumento considerable en la eficiencia del uso del canal de comunicaciones en comparación con un procedimiento de codificación y muestreo actualmente usado en la transmisión de señales de audio, como lo es, el estándar PCM (*Pulse Code Modulation*).

Un protocolo que “entienda” mejor el contenido de la información que se desea enviar, es decir, que conozca el vocabulario o en este caso, los fonemas del Español de Venezuela, además características que ayuden a caracterizar inequívocamente cada uno de estos, es posible disminuir la cantidad de bits enviados por cada mensaje y por lo tanto, lograr un aumento en la eficiencia del canal de transmisión.

Cabe destacar que existe un proyecto macro, que forma parte de una línea de investigación, cuya estructura planteada consta de tres trabajos especiales de grado: Se necesitará la implementación de una primera etapa de recepción de audio, y reconocimiento de los fonemas (tema que no es objetivo del presente Trabajo de Grado). La segunda etapa, que consiste en el diseño de un protocolo de transmisión y codificación, temas que se abarcan en este trabajo. La tercera, una etapa de recepción, que generará el audio a partir de los fonemas transmitidos (tema que tampoco es objeto del presente trabajo).

En el capítulo I, se presentan las características del problema estudiado, los objetivos que se quieren lograr, las razones que motivan a la investigación y las limitaciones de la misma. El capítulo II consta del marco teórico, donde se hace un recorrido por la teoría que sustenta el trabajo, tanto desde el punto de vista de las telecomunicaciones; estándares y protocolos; como de los conceptos lingüísticos necesarios para la investigación, las características del idioma en estudio (Español venezolano) y las características del sonido del habla en general. El capítulo III abarca la metodología que se llevó a cabo, describiendo paso por paso cómo se atacó cada objetivo específico y finalmente, antes de pasar al siguiente capítulo, se explican las pruebas de comparación que se realizaron entre ambas tramas. El capítulo IV presenta los resultados de dichas pruebas y por último, en el capítulo V se describen las conclusiones y recomendaciones que resultaron de toda la investigación.

# **CAPÍTULO I**

## **EL PROBLEMA**

En este capítulo se abarcan los lineamientos que definen el problema de la investigación, los puntos que justifican y delimitan el trabajo, los antecedentes de importancia y finalmente se definen los objetivos que van ser cumplidos a lo largo del mismo.

### **1.1. Planteamiento del problema**

Si en un sistema de telecomunicaciones de cualquier naturaleza se pudiera disponer de un ancho de banda infinito y canales sin ruido (condiciones ideales), este proyecto tendría poco atractivo de investigación. Pero bien se sabe que esta nunca es la realidad y uno de los grandes problemas que siempre está presente en todo sistema de telecomunicaciones, es la limitación del canal de comunicaciones en cuanto a velocidad de transmisión (ancho de banda).

Ya sea por limitantes del propio canal o de los equipos de trabajo, siempre se tiene un ancho de banda restringido con el cual trabajar y por lo tanto limitantes en la velocidad que se puede usar. El aumento de las velocidades de transmisión, así como la reducción del ancho de banda utilizado por cada usuario sin el deterioro perceptivo de la calidad del mensaje, permite dar cabida a servicios más complejos y/o a mayor cantidad de usuarios al sistema.

El desarrollo tecnológico ha ayudado en gran medida a solventar este problema, creando mejores materiales, mejores equipos y nuevas herramientas que han permitido mejorar las características de transmisión. Pero las barreras físicas son difíciles de romper y muchos de los aumentos en velocidades de transmisión, y también los más comunes, han sido posibles gracias a mejoras en las técnicas de modulación, codificación y compresión de la información, lo que ha permitido aumentar de velocidad a pesar de mantener las mismas tecnologías y materiales.

Los sistemas de comunicaciones planteados de forma genérica pueden involucrar desperdicios en el ancho de banda para cierto tipo de mensajes (o idiomas en este caso), y un sistema de codificación dedicado y diseñado especialmente para los requerimientos del tipo de mensaje que manejará, debe aportar grandes mejoras en la eficiencia de la comunicación.

Una vez dicho esto surgen las siguientes preguntas, ¿Es posible elaborar un método de codificación eficiente para la transmisión de fonemas en español de Venezuela?, y si es así, ¿Cómo se puede implementar dicho procedimiento?, ¿Cuánto afecta éste en la calidad de audio, de manera que valga la pena su implementación? y ¿Cuánto más eficiente es, comparado con PCM?

## **1.2. Justificación**

De obtenerse una codificación que demuestre ser más eficiente que los métodos actuales, representará una gran mejora en las telecomunicaciones, no solo del país, específicamente hablando del Español de Venezuela, sino para cualquier idioma, ya que de ser el caso, este trabajo podría ser fácilmente escalable a otros idiomas, lenguajes o dialectos, e incluso dedicados específicamente para acentos regionales muy marcados.

De ser posible reducir la cantidad de datos necesaria para representar un mensaje, permitiría ganar “espacio” en las tramas enviadas y así es posible dar servicio a una mayor cantidad de usuarios. Esto involucra una mejora económica y de prestación de servicio considerable ya que permite el acceso a un mayor número de usuarios bajo la misma plataforma. Además se estaría aprovechando capacidad de almacenamiento y de procesamiento (recursos de hardware y software) en los equipos que actualmente se usan en comunicaciones en sus tareas de envío y recuperación de mensajes, así como la posibilidad de emplear equipos de menores capacidades (menos costosos) para realizar comunicación de manera exitosa.

De ser el caso, si se sabe de antemano que se trabajará sobre un canal problemático o con usuarios que requieran de una alta calidad de servicio, los espacios libres en las tramas se podrían emplear para incorporar bits de redundancia y seguridad a los mensajes o incluso para mejorar la calidad del audio transmitido.

### **1.3. Alcance**

Este trabajo forma parte de un proyecto de orden superior que constituye una línea de investigación planteada por el Departamento de Comunicaciones, que engloba al menos tres trabajos especiales de grado, cada uno abarcando una de tres etapas. La primera de estas etapas es la de elaborar un procedimiento que permita a partir del audio (hablado), la identificación de fonemas y sus características del habla necesarias para su transmisión y recepción a través de un sistema de telecomunicaciones. La segunda etapa, que es la que se abarca en el presente trabajo, involucra el diseño de un protocolo de comunicaciones eficiente para la transmisión de estos fonemas. Y la última etapa abarca el dispositivo receptor de estas tramas, que generaría señales de audio a partir de las mismas.

Por lo tanto, en este trabajo se llevará a cabo el diseño de la codificación de las tramas y los parámetros caracterizadores de fonemas generados en la etapa anterior. Además se debe hacer un estudio del tipo de fonemas que, con más probabilidad, serán procesados por el sistema y el respectivo estudio de eficiencia respecto a otro método de codificación, que en este caso será el de modulación por impulsos codificados o PCM.

### **1.4. Antecedentes**

- Laura Espinosa Camacho, en su trabajo de grado, “Compresión de voz para transmisión en redes de datos”, presenta una investigación detallada sobre métodos de codificación de voz usados en telecomunicaciones, tales como PCM, LPC, entre otros, información que será de gran ayuda teórica en el presente trabajo. [5]
- Mario Sánchez Monroy, presentó en su trabajo de Maestría “Codificación de voz en sub-bandas”, un proyecto parecido al presente en cuanto a la estructura del sistema total. En su trabajo, Monroy diseña un esquema de codificación y transporte de voz, también en la búsqueda de minimizar la cantidad de bits transmitidos. A pesar de que no aplica el concepto de fonemas a su investigación (si no, discriminación de las bandas frecuenciales de la señal, según su aporte de información o de energía), el parecido estructural del sistema y sus resultados al comparar diferentes métodos de codificación puede ser útil y será estudiado. [6]



- Rodrigo Gonzales y Azucena Mejorada presentaron en 2008 su trabajo para optar por el título de Ingeniero, “Técnicas de grabación y reproducción de sonido”, donde muestran especial interés en el estándar MIDI, explicando detalladamente su funcionamiento, componentes y las tramas que maneja. Tentativamente, el estándar MIDI será una guía importante para este trabajo, ya que implementa una codificación de música a partir de las notas musicales como tal, tomando en cuenta su contenido informático, y no solo del sonido que se desea transmitir. [7]
  
- En el trabajo de grado, “Robustez en reconocimiento fonético de voz para aplicaciones telefónicas”, José Ignacio Puertas Tera en 2000, describe un sistema de reconocimiento de voz usando técnicas matemáticas y probabilísticas mientras explica la generación, a partir de voz, de los elementos fonéticos que la componen. A pesar de que este trabajo podría ser de mayor utilidad en la primera etapa de la línea de investigación, se incorpora en la presente, ya que explica detalladamente la obtención de parámetros fonéticos en un sistema de reconocimiento de voz. [8]

## **1.5. Objetivos**

### **1.5.1. Objetivo General**

Plantear una alternativa para la transmisión de voz en español de la región central-norte de Venezuela, a partir de la codificación de fonemas.

### **1.5.2. Objetivos Específicos**

- Realizar una investigación sobre estándares usados para la transmisión de música y voz.
- Realizar una librería (o tabla) de los fonemas utilizados en el español estudiado para un número limitado de palabras, usadas en medios públicos y conteniendo probabilidad de ocurrencia de cada uno.
- Identificar las características o parámetros más importantes a transmitir del fonema.
- Diseñar un protocolo o método de codificación eficiente para la transmisión digital de los fonemas utilizados en español estudiado.
- Realizar una comparación entre el método de codificación diseñado y PCM en cuanto a número de bits utilizados, eficiencia del código y velocidad de transmisión para un conjunto de palabras específicas.

## CAPÍTULO II

### MARCO TEÓRICO

En este capítulo se presenta el estudio teórico necesario para desarrollar los objetivos de este trabajo. La revisión bibliográfica de múltiples estudios involucrados en el tema, así como la investigación sobre lenguaje, fonética, idiomas, señales de audio de habla y sobre los estándares y protocolos más usados para la transmisión de señales de audio. Al finalizar esta fase se contó con la información necesaria que hace posible el desarrollo de los objetivos posteriores.

#### 2.1. Telecomunicaciones

A lo largo de la historia, los seres humanos han tenido la necesidad de comunicarse entre sí, tratando de solventar las distancias que nos separan y siempre en busca de métodos más rápidos y eficientes para ello. Desde la invención del telegrama, que de alguna forma marcó el nacimiento de las telecomunicaciones, hasta nuestra época actual, el ser humano ha estado en la búsqueda de alternativas para el envío de mensajes a otras personas y cada vez con mayores exigencias, llegando a hacer de estos sistemas una parte de nuestra vida cotidiana, incluso haciéndonos dependientes de ellos para la mayoría de las actividades de la sociedad moderna.

El concepto de telecomunicaciones se refiere al uso de señales y sistemas eléctricos para el transporte de mensajes, ya sea a través del espacio libre o por medios guiados (como cables).

En este trabajo se prestará especial atención en las telecomunicaciones electrónicas. Las cuales, según Wayne Tomasi "...se puede decir que las comunicaciones electrónicas son la transmisión, recepción y procesamiento de información entre dos o más lugares, mediante circuitos electrónicos." (Tomasi, 2003) [9]

El nacimiento de las telecomunicaciones, como se dijo antes, se remonta a la década de 1830 con el nacimiento del telegrama, sistema que permitía el envío de letras, *codificadas* en grupos de 3 pulsos eléctricos, según el *protocolo* llamado clave Morse, que podían ser escuchados en el *receptor* por un intérprete (aun en ese tiempo, una persona) que conocía la codificación y que podía traducir o *decodificar* los mensajes en letras legibles. Los pulsos eléctricos tenían características físicas que los hacían fácilmente transportables por el *canal de comunicaciones*, razón por la cual se

podía lograr comunicación a largas distancias y en canales con mucho *ruido* y con la escasa tecnología del momento.

Como puede verse en la descripción de este sistema de telecomunicaciones aparecen varios conceptos que deberán ser explicados detalladamente más adelante, conceptos que aparecen en la descripción de todo sistema de esta naturaleza, ya que, incluso ahora, casi 200 años después, todos los sistemas de telecomunicaciones mantienen la misma estructura.

### **2.1.1. Estructura de sistema de telecomunicaciones:**

Todos los sistemas de comunicaciones se componen de tres elementos básicos, emisor, canal, y receptor. Cuando dos personas hablan, una de ellas es el emisor, que envía un mensaje auditivo a través del canal, que en este caso es el aire. El mensaje es recibido por la otra persona o receptor. Los elementos del sistema van alternándose en los roles cada vez que el interlocutor responde.

Este ejemplo puede complicarse mucho más hablando en detalle de cada uno de los componentes que pueden ser identificados. El esquema más usado agrega dos elementos, “...*la fuente de información* antes del transmisor y el *destino de la información*.” (Tomasi, 2003). [9]

Como se dijo antes, se puede hablar de telecomunicaciones cuando el elemento físico que transporta los mensajes en una señal eléctrica.

Para que esto sea posible es necesario convertir los mensajes en la señal eléctrica que se va a enviar. En la Figura 1 se presenta un diagrama de bloques simple de un sistema de telecomunicaciones.

**2.1.1.1. Transductor:** es un elemento de sistemas de telecomunicaciones que genera señales eléctricas a partir de señales de otra naturaleza, por ejemplo el micrófono, es un elemento que transforma señales auditivas (ondas mecánicas del aire) en señales eléctricas. En la Figura 1, el transductor formaría parte del Transmisor, pudiendo ser encontrado en la primera etapa del mismo (es el primer componente del Transmisor). [9]

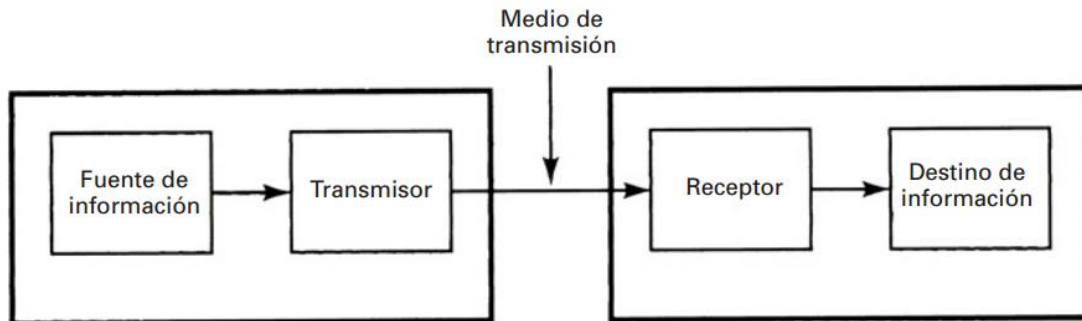


Figura 1: Diagrama de bloques simplificado de sistema de telecomunicaciones. (Tomasi, 2003)

**2.1.1.2. Canal de telecomunicaciones:** en la Figura 1 el canal de comunicaciones es representado como el Medio de transmisión, en el cual se propaga la señal portadora de información, en este caso, una señal electromagnética. Los canales o medios de transmisión pueden ser guiados, como cables o guías de ondas; o puede ser el espacio libre, como es el caso de enlaces inalámbricos. (Tomasi, 2003) [9]

**2.1.1.3. Fuente de información y destino de la información:** la Fuente es el origen de la información que se pueda representar como una señal para ser transmitida y el destino es el dispositivo o persona que desea obtener esta información, en el ejemplo de dos personas hablando por teléfono la fuente de información es la persona que habla y el destino la que escucha. Estos dos elementos son los propiciadores de la comunicación y pueden ser vistos como los usuarios del sistema de telecomunicaciones. (Tomasi, 2003)

**2.1.1.4. Transmisor:** es el elemento del sistema de telecomunicaciones que acondiciona la información proveniente de la Fuente, en una señal apta para ser transportada por el canal. “Un transmisor es un conjunto de uno o más dispositivos o circuitos electrónicos que convierte la información de la fuente original en una señal que se presta más a su transmisión a través de determinado medio de transmisión.” (Tomasi, 2003) [9]

**2.1.1.5. Receptor:** Análogamente, “Un receptor es un conjunto de dispositivos y circuitos electrónicos que acepta del medio de transmisión las señales transmitidas y las reconvierte a su forma original.” (Tomasi, 2003) [9]

## 2.2. Modulación

Es el proceso en el cual se modifican las características físicas de una señal acorde con otra señal de información que se desea transmitir. La razón de esto es que la señal portadora es más apta para viajar por el canal (normalmente esta tiene mayor frecuencia). En este proceso, la señal de información o *moduladora* se “imprime” en la señal portadora y es esta última la que será transmitida finalmente. Los canales tienen diferentes características físicas que hacen que este proceso sea necesario para una correcta difusión de la señal a través del canal. (Tomasi, 2003) [9]

Otra razón por la cual hacer modulación es para solventar el problema del canal compartido (hacer posible la *multiplexación*). Muchas veces no se dispone de un canal único y totalmente disponible para la comunicación entre dos puntos, el caso más común es que este canal sea compartido entre varios usuarios, incluso en canales guiados resulta poco rentable instalar un solo canal (cable o fibra óptica) para cada usuario, así que se debe compartir el canal con otros usuarios e incluso con otros servicios. Este hecho es inevitable cuando se habla de canales no guiados, ya que el espacio libre es totalmente compartido y donde se encuentran infinidad de servicios de difusión.

La técnica para evitar estos obstáculos es la modulación. En este proceso se emplea una señal, que es más apta para viajar en el canal (ya sea por nivel de potencia, frecuencia, etc.) para que esta sea la portadora de la información que se desea enviar. Las portadoras son generadas para que no se interfieran entre sí, de manera que, si a distintos usuarios se les asignan distintas portadoras será posible el uso compartido del canal.

Según el tipo de señal manejada, se puede hablar de modulación analógica o digital y según la forma en que la información es “impresa” en la señal portadora y se distinguen distintos tipos de modulación, por ejemplo: FM (la información está contenida en la frecuencia de la señal portadora), AM (en la amplitud), PM (en la fase), PSK (versión digital de modulación por fase), etc.

## 2.3. Codificación

El nacimiento de la era digital trajo consigo el cambio en el tratamiento de las señales eléctricas usadas en comunicaciones. Se demostró que es suficiente usar menos información que la contenida en una señal analógica (completa) manteniendo la calidad de servicio. Por esto, se

popularizó el uso de las comunicaciones digitales y se desarrollaron procesos de digitalización de las señales portadoras para su transmisión.

Anteriormente las señales portadoras se enviaban de forma analógica, de hecho algunas de estas técnicas aún son usadas para algunos sistemas, por ejemplo, la transmisión de radio difusión en las bandas de FM y AM es usada en forma analógica a pesar de las nuevas tecnologías, esto es debido a razones económicas y a la alta penetración en el mercado que lograron estas técnicas. Pero la tendencia apunta hacia una digitalización total de las comunicaciones dados los beneficios que esto conlleva.

Son muchos los beneficios de la digitalización, uno de ellos es la posibilidad de multiplexado por división de tiempo o TDM, una solución para el problema de canal compartido. Otro beneficio es el manejo de las señales con computadores digitales, que permite uso de algoritmos y procesamiento más complejos e inteligentes. (Abramson, 1981) [10]

## **2.4. Digitalización**

En la naturaleza las señales, como cantidades físicas que son, se encuentran en forma analógica, es decir, pueden tomar cualquier valor (con infinitos decimales) y ocupar el tiempo de forma continua. La digitalización es el proceso por el cual esa señal analógica se pasa a una forma digital, lo que permite reducir la memoria necesaria para su manipulación (es imposible el procesamiento de señales analógicas por computadores digitales) y el mejor aprovechamiento del tiempo de transmisión. Si se toma en cuenta el hecho de que el cerebro humano es incapaz de detectar toda la información contenida en una señal continua, es posible esta reducción de información sin producir pérdidas de calidad de servicio, es decir, deterioro perceptible de la información recibida. La digitalización involucra principalmente tres procesos, el muestreo, la cuantificación y la codificación. (Tomasi, 2003) [9]

### **2.4.1. Muestreo:**

Es el proceso en el cual se toman muestras de una señal analógica con una frecuencia llamada frecuencia de muestreo, obteniéndose luego del mismo, una señal discreta en el tiempo. (Tomasi, 2003) [9]

#### **2.4.2. Cuantificación:**

Es el proceso en el cual se limita la cantidad de valores que puede tomar una señal, ya sea analógica o discreta. Una forma de ejemplificar esto es cuando se aplica la función “parte entera” a una curva, esto elimina todos los posibles valores decimales. La reducción de los posibles valores de una señal, que permite su manipulación de forma digital, también reduce la precisión de la misma. En este ejemplo ya no es posible tener la precisión de los decimales. En este sentido, es necesario estudiar la cuantificación más adecuada según la aplicación del sistema de telecomunicaciones. [9]

#### **2.4.3. Codificación:**

La cuantificación típicamente conlleva a un proceso de codificación, esto es, la asignación de un código a cada uno de los posibles valores definidos anteriormente. En la mayoría de las telecomunicaciones digitales se utiliza un código binario, a cada valor de la señal se le asigna un número binario que representa a cada una de las muestras de la señal. Finalmente estos son los códigos que van a ser transmitidos por el sistema de comunicaciones. En este punto vale la pena decir que existen varios tipos de codificación. (Abramson, 1981) [10]

**2.4.3.1. Codificación de línea:** la codificación de línea es también llamada modulación en banda base. Se le llama así al código que se aplica a una cadena de datos o bits que van a ser transmitidos en una línea de telecomunicaciones. El objetivo de este código es eliminar algunos de los inconvenientes que aparecen en los sistemas eléctricos, tales como, la corriente continua (o pedestal) en la línea, sensibilidad a la fase, fallas de sincronismo en comunicación serial, interferencia inter-simbólica (ISI), y probabilidad de error. Haciendo una selección adecuada del código de línea es posible mitigar todos estos problemas que dificultan la comunicación en los sistemas. (Abramson, 1981) [10]

Algunos ejemplos de códigos de línea son: NRZ, RZ-polar, RZ-bipolar, código Mánchester, etc. Que pueden verse en la Figura 2.

**2.4.3.2. Codificación de fuente:** la codificación de fuente está más relacionada con los datos que genera la fuente de información, el “vocabulario” con el que se trabaja, es decir, el conjunto completo de símbolos que pueden ser generados por la fuente y las probabilidades de aparición de cada uno de estos símbolos. (Tomasi, 2003)



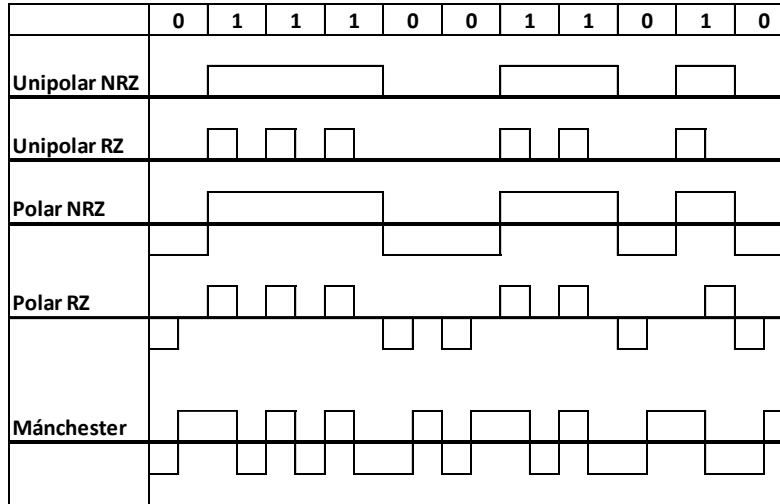


Figura 2: Ejemplos de códigos de línea.

Este concepto será de gran importancia en este trabajo ya que uno de los objetivos es generar un código de fuente adecuado para el conjunto completo de fonemas del Español, así como la obtención de las probabilidades de ocurrencia de los mismos en el habla cotidiana. En este caso, la fuente de información, que sería una persona hablando, y el sistema puede conocer todo el “abecedario de símbolos” (fonemas) que pueden ser producidos. Y es parte de este trabajo diseñar un código que garantice el mínimo uso del canal (mayor eficiencia). [9]

**2.4.3.2.1. Código Huffman:** es un tipo de codificación de fuente en la cual se conoce todo el compendio de símbolos que puede producir la fuente, y además las probabilidades de aparición de cada uno de ellos. Para llevar a cabo la generación del código, se hace uso de un algoritmo bien detallado, desarrollado por David A. Huffman, que produce un código de longitud variable.

El código Huffman es un ejemplo de código prefijo, es decir, es un código de longitud variable que cumple con la siguiente propiedad: el código de ningún carácter es prefijo de ningún otro código, también llamada “propiedad de prefijo”. Esto permite que el receptor pueda determinar cuál es el carácter recibido a pesar de tener longitud pseudo-aleatoria y sin necesitar un carácter de parada.

En este algoritmo se deben ordenar los caracteres disponiéndolos de mayor a menos probabilidad de aparición, se suman las probabilidades de los dos últimos caracteres, se asigna a cada uno de estos un bit diferente (al primero un cero y al otro un uno), seguidamente se asumen que estos dos caracteres son uno solo con una probabilidad de aparición que es igual a la suma calculada, y se repite el procedimiento, acumulando los bits que se asignan a cada carácter. El algoritmo se detiene cuando todas las probabilidades han sido sumadas. (Abramson, 1981) [10]

#### **2.4.4. Multiplexado:**

Se refiere al uso de un mismo medio o canal de comunicación de forma compartida por múltiples señales, sin que estas se mezclen (o siendo posible la posterior recuperación de cada una). Existen muchas formas de multiplexado, por ejemplo, FDM o multiplexado por división de frecuencia (usada en señales analógicas como Radio) y TDM o multiplexado por división de tiempo (solo para señales digitales), esta es una de las mayores ventajas del uso de señales discretas en tiempo ya que, aprovechando la reducción de información que se necesita transmitir, se permite el uso del mismo canal por más de una señal, a diferentes intervalos de tiempo. Esta ventaja será de importancia para este trabajo más adelante. (Tomasi, 2003) [9]

## **2.5. Estándar o protocolo de comunicaciones**

Un protocolo de comunicaciones no es más que las reglas que tanto el receptor como el transmisor deben conocer y seguir para que la comunicación se lleve a cabo con éxito. En las telecomunicaciones son de mucha importancia ya que la creación de nuevos protocolos puede aumentar ampliamente la eficiencia en las comunicaciones. Este es el objetivo final del trabajo actual.

### **2.5.1. Protocolos digitales:**

**2.5.1.1. PCM:** El estándar PCM, Modulación por Impulsos Codificados, se rige bajo la recomendación G.711 de la ITU-T, que especifica una forma de codificación ampliamente usada en telecomunicaciones, específicamente para señales de voz.

Este estándar se basa en el procesamiento del audio de forma digital lo que involucra un proceso de digitalización y el envío de las muestras digitales en lugar de manipular el audio completo.

Para digitalizar la señal de audio se pasa por las etapas de muestreo, cuantificación y codificación, típica de toda digitalización. En este caso, el estándar especifica las características de cada una de estas etapas de la siguiente forma:

- Se establece la velocidad de muestreo en ocho mil muestras por segundo (8.000 muestras por segundo, con una tolerancia de  $\pm 50$  ppm) lo que es equivalente a una muestra cada 125  $\mu$ s, cumpliendo con el teorema de Nyquist de muestreo y tomando como ancho de banda máximo de la voz humana en 4 kHz (lo que involucra una frecuencia de muestreo mínima de 8 kHz).
- Para la cuantificación, el uso de 8 bits de resolución para cada muestra de audio, esto es, el rango de amplitud posible se divide en segmentos, y se aplica una reducción de los posibles valores de la señal analógica aplicando una función piso por ejemplo. Para esta división, se aplica la ley A o la ley  $\mu$  a convenir, según la región. Estas leyes son descritas detalladamente en la recomendación G.711 de la UIT-T. (1988).

La aplicación de estas leyes modifica el detalle con el cual se codifican las diferentes amplitudes de la señal, dándole más importancia a los segmentos donde es más probable obtener la señal de voz. La recomendación especifica también que el primer bit enviado siempre es el bit más significativo y el menos significativo de último lugar.

PCM se usa para manejo de audio sin importar la naturaleza de los mensajes, de manera que el tipo de información enviada es transparente al procedimiento de codificación. [11]

**2.5.1.2. MIDI:** el protocolo MIDI (*musical instrument digital interface*) fue desarrollado en los años 80 para la comunicación entre instrumentos musicales y está documentado en los estándares AES10-1991 y ANSI s4.43-1991. Con éste, es posible el envío de paquetes de datos a diferentes instrumentos en plena ejecución musical y así lograr el control de múltiples dispositivos desde un computador o sintetizador.

En cuanto al hardware que da sustento al protocolo MIDI cabe destacar que es simple y de bajo costo. La trama utiliza un bit de comienzo (*start*), 8 bits de datos, y dos bits de parada (*stop*), funcionando a una velocidad de 31,25 kbaudios. MIDI utiliza un

bucle de corriente de 1,5 mA. Con conectores tipo DIN de 5 patas, hembra en los equipos y macho en los cables. En la Figura 3 puede verse el diagrama de conexiones de entradas y salidas. También se puede ver que existe una salida llamada “Salida MIDI thru” esta no es más que la réplica de la “Entrada MIDI” lo que permite conectar más de dos equipos.

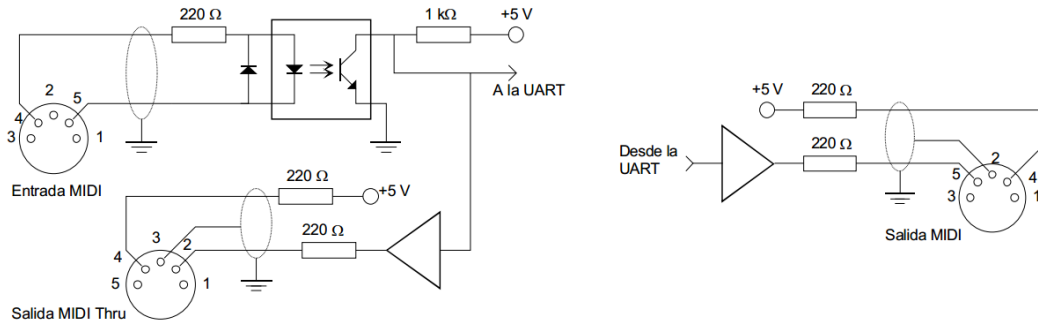


Figura 3: Diagrama de puertos de entrada y salida MIDI. [12]

El cable usado debe ser de máximo 15 m, de par trenzado y blindado. Un ejemplo de conexión en protocolo MIDI se ve en la Figura 4, donde se aprecia que no todos los equipos tienen la salida “Salida thru”, por lo que se recomienda el uso de un equipo “Expansor MIDI”, sobre todo cuando se cuenta con sistemas de más de tres conexiones en la cuales suele haber problemas de comunicación.

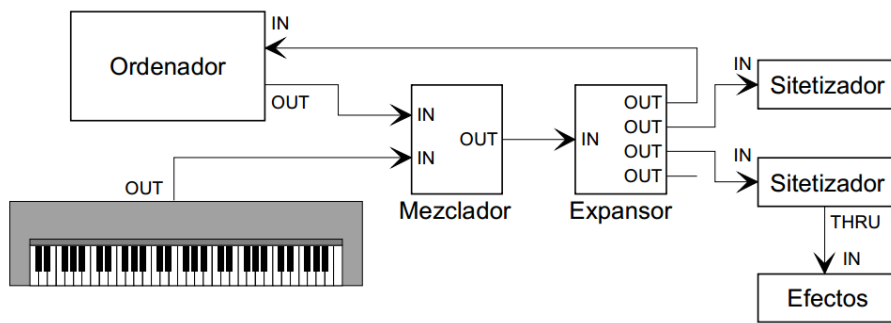


Figura 4: Diagrama de interconexiones MIDI. [12]

La estructura de la trama de mensajes del protocolo MIDI puede verse en la Figura 5 donde se ve que está compuesta de 3 palabras de 8 bits cada una. La primera palabra o

byte de estado siempre comienza con un 1 (MSB) seguido de 3 bits que determinan el tipo de mensaje que se envía, es decir se cuenta con 8 tipos diferentes de mensajes posibles, luego 4 bits para indicar el canal al que se envía el mensaje. En MIDI cada dispositivo se puede asignar a cada canal o incluso varios al mismo canal, se admiten hasta 16 canales diferentes en una conexión. La siguiente palabra o el primer byte de datos siempre comienza con un 0 (MBS) seguido de 7 bits que indican los datos correspondientes según el tipo de mensaje y acción que se va a ejecutar, por ejemplo, la nota a tocar, o el efecto a aplicar, etc.

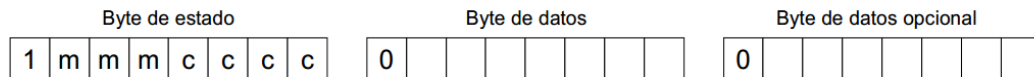


Figura 5: Esquema de la trama MIDI. [12]

Como se dijo antes, existe la posibilidad de enviar 8 tipos diferentes de mensajes, la lista completa se presenta en la Figura 6, donde puede verse que el segundo byte de datos es opcional ya que algunos mensajes no lo utiliza, por ejemplo, la acción de “cambio de programa”. También se observa que el último tipo de mensaje es opcional, configurable para las necesidades de cada sistema.

Byte de estado	Nº bytes de datos	Descripción	Byte de datos 1	Byte de datos 2
1000nnnn	2	Desactivación de nota	Altura	Velocidad
1001nnnn	2	Activación de nota	Altura	Velocidad
1010nnnn	2	Expresión de nota Post-pulsación	Altura	Presión
1011nnnn	2	Cambio de control	Tipo de control	Intensidad
1100nnnn	1	Cambio de programa	Programa	
1101nnnn	1	Expresión de canal	Presión	
1110nnnn	2	Cambio de tono 'Pitch Bend'	MSB	LSB
1111xxxx	variable	Mensajes de sistema		

Figura 6: Lista de mensajes MIDI. [12]

Es posible enviar mensajes sin determinar el canal, para estos casos el protocolo entiende que el mensaje debe ser entregado a todos los canales, por lo que aplica para todos los dispositivos. Los mensajes más usados son los de “activación de nota” y “desactivación de nota”, en los cuales, como se puede ver en la Figura 6, se requiere la especificación de los dos bytes de datos, el primero para determinar la nota a ser activada o desactivada, contando con 7 bits, que permite escoger entre 128 notas posibles. El segundo byte de datos determina la velocidad con la que se debe activar la nota, o lo que es lo mismo, la dureza con la que se tocaría la tecla del teclado. Esta palabra debe ser enviada a pesar de que el teclado no cuente con una implementación de sensibilidad. De forma similar trabajan los mensajes de tipo “desactivación de nota”.

Una característica importante de MIDI es que si el byte de estatus es idéntico al del mensaje anterior, el protocolo permite su omisión, lo que se refleja en un aumento de eficiencia considerable.

**2.5.1.3. OSC:** este protocolo (*Open Sound Control*) aparece como una mejora y reemplazo de MIDI, siendo superior en capacidades y velocidad, pero bastante más complejo. Su principal característica es que usa las técnicas modernas de redes de datos, como empaquetado UDP y TCP, para hacer posible la conexión entre computadores o sintetizadores modernos y algunos instrumentos que soportan el estándar. Es posible usarlo para conectar en un mismo sistema controladores de luces, teclados, y sintetizadores, además tiene capacidad para un número ilimitado de canales. Incluye marca de tiempo en los mensajes, útil para guardar registros así como un identificador único para cada mensaje. [13]

Dada su alta complejidad y limitación en cuanto a los equipos de soporte, este protocolo no será explicado ampliamente ya que no será de mucha utilidad.

**2.5.1.4 MADI:** son las siglas de *Multi channel Audio Digital Interface* (interfaz multicanal de audio digital), es un estándar profesional de transmisión de datos para señales de audio digital por un solo cable, documentado en los estándares AES 10-1991 y ANSI s4.43-1991. Aunque el estándar se planteó desde mediados de los 80s, fue redactado finalmente hasta el año 1992 y con una última revisión en el 2003 en el estándar AES10-

2003. El permite la transmisión hasta de hasta 64 canales de audio con 24 bit de resolución y 96 kHz de muestreo, transmitidos en serie en un único cable coaxial, permitiendo una transmisión a distancias mayores de 50 m. También permite la transmisión sobre fibra óptica lo que aumenta la distancia permitida hasta 2 km, con una resolución de 32 bits por cada muestra de audio contando los bits de control correspondientes.

Como en protocolos anteriores, MADI permite la transmisión de varios canales por un solo cable, todo estos multiplexados en tiempo. En la Figura 7 puede verse el esquema del sistema MADI sobre cable coaxial.

La trama especificada consta de 56 canales numerados del 0 al 55, cada uno transporta una muestra de audio y los bit de control relacionados con esta. En cada ciclo de muestreo el primer canal enviado es el canal 0, intercalados con los símbolos de sincronización que dependiendo de la configuración pueden intercalarse cada uno o cada dos canales. Cada canal consiste en 32 bits numerados desde el 0 al 31, de los cuales 24 son los bits de datos, ya sean de audio o no (dependiendo del bit de audio/no-audio) estos son los bits que van desde el 4 hasta el 27 inclusive. Los bits anteriores (del 0 al 3) son bits de control que identifican cada bloque como bloque de datos o de sincronización, habilitan el canal e identifican las sub tramas A o B. Los bits finales, se identifican como V, U, C, P y determinan, la validez de la trama, si está en uso el contenedor de datos, el estatus, y la paridad respectivamente. Los bits de datos son transmitidos con el bit menos significativo primero (bit 4) y el más significativo al final (bit 27). En la Figura 8 puede ver la trama de cada canal. [14]

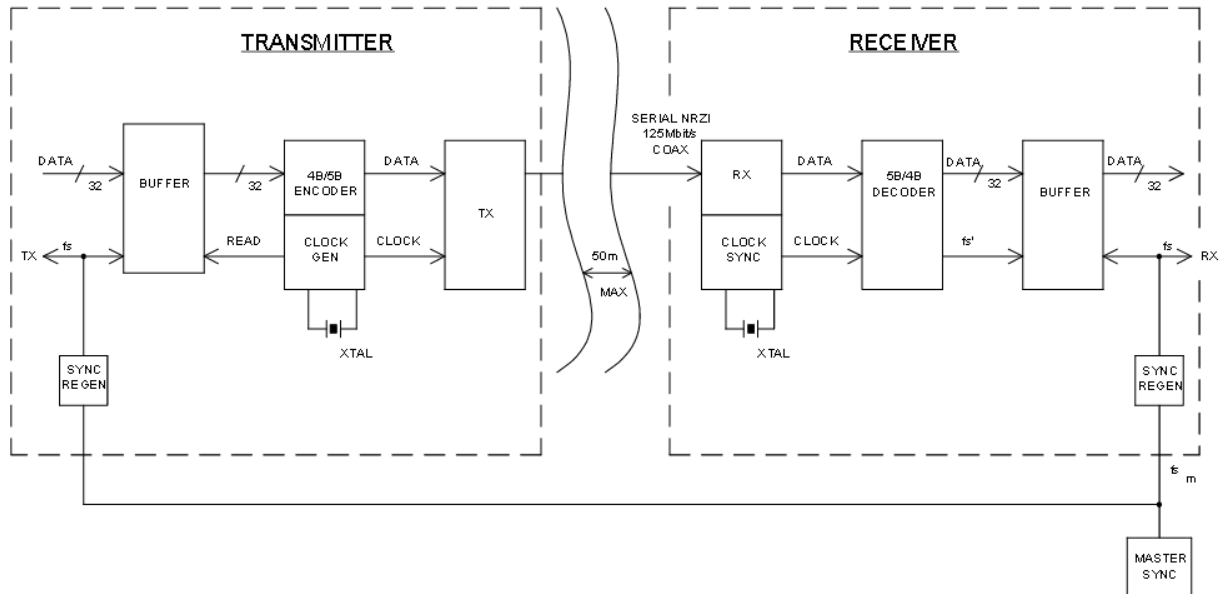


Figura 7: Diagrama de bloques de comunicación MAD. [14]

Se debe tomar en cuenta que algunos de estos bits (el bit 2 por ejemplo) son usados para lograr compatibilidad con el estándar anterior (AES3-1985).

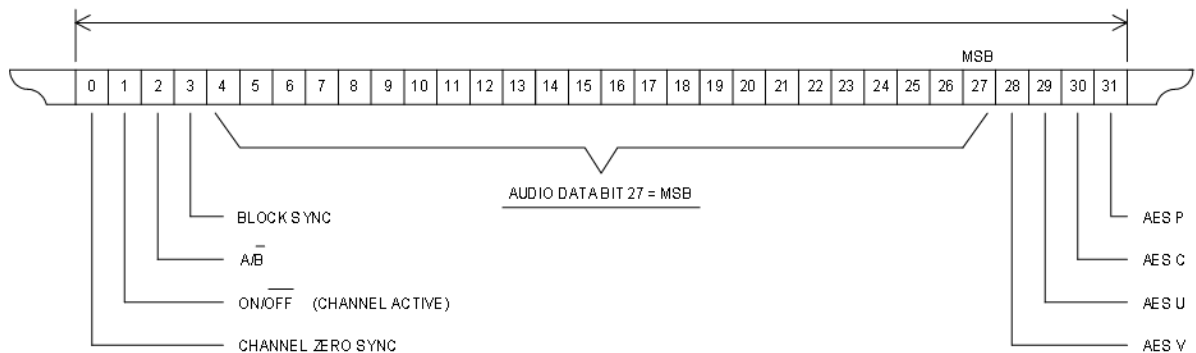


Figura 8: Trama MADi por canal. [14]

La transmisión de los canales se hace de forma serial usando codificación NRZI sin polarización (no retorno a cero invertido) con el esquema de codificación siguiente: cada trama de canal se separa en 8 palabras de 4 bits cada una, a cada una de estas palabras se le asigna un código de 5 bits según la Figura 9, con lo que se obtiene un código más amplio que permite implementar símbolos extras de control como por ejemplo el de sincronización. Finalmente cada palabras es transmitida en el orden desde el bit 0 hasta el bit 39.



El símbolo de sincronización está determinado de forma fija como 11000 10001, dos palabras de 5 bits cada una que se deben insertar al menos una vez por periodo de la trama con fines de sincronismo. El símbolo de sincronización es enviado desde la izquierda (como se lee) y debe ser repetido al final de la trama antes de que el último canal sea transmitido.

4-Bit data	5-Bit encoded data
0000	11110
0001	01001
0010	10100
0011	10101
0100	01010
0101	01011
0110	01110
0111	01111
1000	10010
1001	10011
1010	10110
1011	10111
1100	11010
1101	11011
1110	11100
1111	11101

**Figura 9: Codificación de expansión MADI. [14]**

La frecuencia de muestreo se encuentra en un rango entre 32 kHz y 48 kHz y la tasa de transmisión debe estar en los 125 Mbps, no obstante dada la constitución de la trama, la tasa de transmisión de datos es de 100 Mbps. En una configuración máxima en velocidad usando los 56 canales y una velocidad de muestreo de 48 kHz se obtienen 96,768 Mbps.

La interfaz física consta de un cable coaxial de 75  $\Omega$ . En la Figura 10 se observa el esquema de conexión del buffer transmisor y en la Figura 11, el buffer del receptor, ambos con cable coaxial. [14]

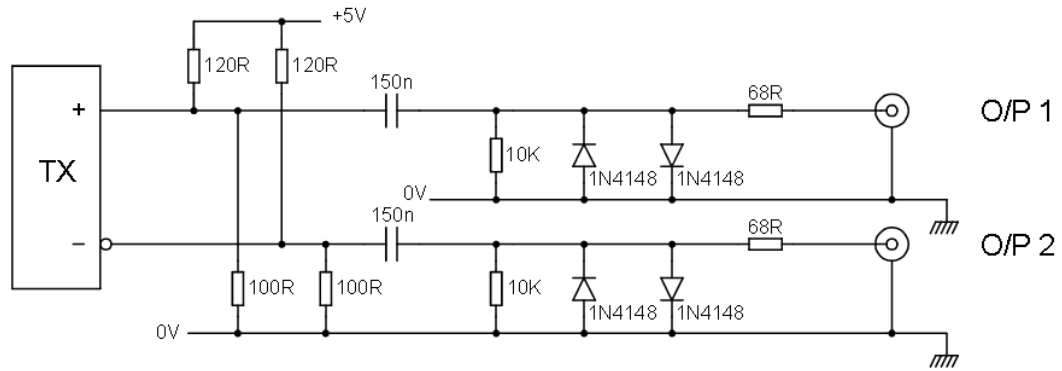


Figura 10: Esquema de puerto transmisor MADI. [14]

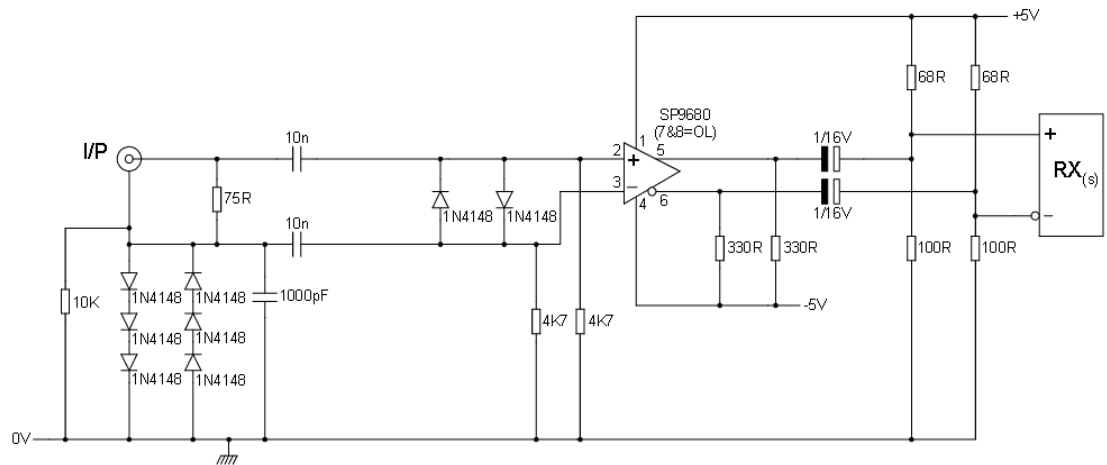


Figura 11: Esquema de puerto receptor MADI. [14]

**2.5.1.5 MPEG:** (*Moving Picture Experts Group*) es un grupo de trabajo especializado de la organización internacional de la normalización o ISO y la comisión electrónica internacional o IEC, creado en el año 1988 con el objetivo del desarrollo de la familia de estándares que llevan su nombre y que regularizan el manejo de datos audiovisuales. Este protocolo presenta una mejora en cuando al paradigma de codificación ya que se basa en el diseño de un codificador complejo que se encarga de la mayor parte de la inteligencia del proceso, y un decodificador mucho más simple que solo repite tareas simples para interpretar la trama de datos que recibe. El “cómo” interpretar la data, le es también transmitida en forma de metadata, por lo que es posible cambiar o actualizar las tramas (cambiar el codificador), manteniendo la compatibilidad con los decodificadores. Esta idea

es muy útil para sistemas de *broadcasting* por ejemplo, donde se tienen múltiples receptores.

Este estándar también toma en cuenta las reglas sobre multiplexado de audio con video dentro de un canal, así como el multiplexado de diferentes canales.

El grupo de trabajo MPEG está conformado por varios sub-grupos que se dedican a partes específicas del protocolo, en particular es importante el grupo encargado del audio. El estándar MPEG-1, el primero publicado por el grupo, que fue diseñado para implementación en audio y video en discos compactos o CD, incorporó el diseño de la codificación de audio MP3 o MPEG-1 *Audio Layer*. Este estándar fue muy popular en audio y aun hoy es muy usado para el almacenamiento de audio.

Las distintas actualizaciones y estándares propuestos por el grupo MPEG presentan mejoras que van acorde con la demanda del mercado y las necesidades de los usuarios. En espacial es de interés el estándar MPEG-7 para este trabajo. [15]

**2.5.1.6 MPEG-7:** su título completo, Interfaz de Descripción de Contenido Multimedia, surge en el año 2002 como parte de los estándares producidos por el grupo MPEG, específicamente para la regularización y manejo de metadatos en las tramas de datos multimedia. Una vez conformados los estándares anteriores quedando establecidos los métodos para el manejo de datos audiovisuales, y dado el aumento considerable de la información disponible y que debe ser manejada (en internet o sistemas muy grandes) comienza a ser más compleja la indexación de contenidos para poder catalogarlos, filtrarlos, indexarlos, buscarlos, ordenarlos, etc. Surge entonces la necesidad de generar descripciones de estos datos de forma de poder manejarlos fácilmente, una suerte de etiqueta donde puede ser obtenida la información sobre el contenido de un paquete sin necesidad de abrirlo.

El estándar MPEG-7 incorpora una serie de descriptores para ayudar a caracterizar el contenido de los datos en cuanto a su creación, formato usado, derechos de autor, y sobre el contenido en sí, estos últimos puede variar en nivel de abstracción según las necesidades del sistema de indexado.

MPEG-7 consta de 8 partes que lo componen, estas incluyen descriptores del sistema, visuales, de esquemas de lenguaje, de audio, etc. En particular es de interés el manejo de las componentes de audio. Como se ha dicho, este esquema, específicamente en la parte de audio, brinda una forma de analizar y caracterizar un contenido de audio de

modo que sea posible reconocerlo de entre otras señales de audio, lo que va a resultar muy útil ya que responde a la interrogante ¿qué características de una señal de audio es necesario enviar para permitir el reconocimiento de una voz de otra? Para esto se usan los siguientes descriptores: *Basic Temporal* describe la forma temporal de la señal (con los descriptores *Audio Wave Form* y *Audio Power*), *Basic Spectral* describe la forma espectral de la señal (sus descriptores, *Audio Spectrum Envelope*, *Audio Spectrum Centroid*, *Audio Spectrum Spread*, *Audio Spectrum Flatness*), *Signal parameters* usado para señales periódicas (*Audio Fundamental Frequency* y *Audio Harmocity*), *Timbral Temporal* incorpora descriptores que, extraídos a partir de la forma temporal de la señal, ayudan a identificar el timbre musical del sonido (*Log Attack Time* y *Temporal Centroid*), *Timbral Spectral* también ayudan a caracterizar el timbre pero a partir de la forma del espectro de la señal (*Spectral Centroid* y *Audio Spectrum Centroid*), estos últimos son de especial interés en este trabajo. Los siguientes descriptores determinan la forma y características estadísticas del espectro de la señal (*Harmonic Spectral Centroid*, *Harmonic Spectral Deviation*, *Harmonic Spectral Spread*, *Harmonic spectral Variation*), *Spectral Basis* se usan para la indexación eficiente del audio (*Audio Spectrum basis*, *Audio Spectrum Projection*) y finalmente el descriptor de los silencios *Silence Segment* el cual puede ser usado para la toma de decisión de procesar o no un segmento. [15]

## 2.6. Estructura del lenguaje hablado

### 2.6.1. Fonema:

Es una construcción abstracta de la idea que genera un sonido en la mente del escucha. El sonido debe estar correspondido con la comunicación hablada y por lo tanto con un idioma específico. El fonema es el elemento más pequeño en el que se pueden dividir los componentes que conforman el habla, logrando distinción de significados. Comúnmente los fonemas son tratados en pares, o bi-fonemas ya que por sí solo son difíciles de identificar.

“Fonema: La unidad lingüística más pequeña con capacidad para diferenciar significados... Unidad lingüística abstracta que tiene en cada caso una realización concreta determinada por los factores de variación en la producción del habla.

Codificamos los mensajes mediante fonemas, pero los producimos y los percibimos mediante los sonidos.” (Gil, 2007, pág. 540). [21]

### **2.6.2. Grafemas:**

Es la representación gráfica de un fonema dentro del lenguaje escrito de cada idioma. En otras palabras son las letras que, en el lenguaje escrito, representan a los fonemas que se intercambian entre extremos de la comunicación. [17]

### **2.6.3. Alófono:**

Es la división mínima del sonido que produce un hablante cuando se comunica de forma hablada, un alófono se asocia directamente con un fonema una vez es escuchado e interpretado por el oyente. [17]

### **2.6.4. Representación fonética:**

Es la forma escrita de los sonidos del habla. En un idioma en concreto, representa los sonidos que deben ser pronunciados para que la comunicación en dicho idioma se lleve a cabo. Existen varios alfabetos fonéticos con diferentes símbolos y reglas para tal fin. Por ejemplo el Alfabeto fonético internacional o IPA (siglas en inglés), el alfabeto fonético americanista o AFA, El alfabeto fonético de la Revista de Filología Española (RFE) o alfabeto hispánico, etc. Este trabajo se limita al uso del IPA ya que es el alfabeto más universal y que abarca los sonidos posibles de todas las lenguas e idiomas. Además el compendio de reglas brinda mayores detalles en la representación del sonido, como se verá más adelante.

## **6.5. Alfabeto Fonético Internacional:**

De sus siglas en inglés, el IPA es un compendio de símbolos y reglas necesarias para representar de forma regularizada, precisa y única, todos los fonemas de los idiomas conocidos hablados, de manera tal que éstos puedan ser comparados entre sí (en cuanto a pronunciación) sin que la lectura se vea afectada por la lengua nativa del lector ni por la ortografía del lenguaje. [16]

Es comúnmente usado por lingüistas y estudiosos del lenguaje, así como por estudiantes de lenguajes extranjeros ya que este representa las pronunciaciones de las palabras independientemente del idioma del lector.

El IPA fue publicado por primera vez en el año 1888 por un grupo de lingüistas que desde dos años antes conformaban la Asociación Fonética Internacional, su líder era el lingüista francés Paul Passy. Desde su creación el IPA ha sufrido varias modificaciones, la mayoría de estas constan de simples actualizaciones de símbolos y caracteres o también la actualización de los ejemplos de fuentes. Su última modificación fue en el año 2005 y es la que se usa actualmente. La versión completa puede verse en los anexos de este trabajo (Anexo A). En la Figura 12 se pueden ver algunos de los símbolos que conforman este alfabeto.

Alfabeto Fonético Internacional - Consonantes											
	Bilabial	Labio dental	Dental	Alveolar	Post Alveolar	Retrof.	Palatal	Velar	Uvular	Faring.	Glotal
Oclusiva	p b		t d			ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ	n			ɳ	ɲ	ŋ	ɴ		
Vibrante múltiple	ʙ		r						ʀ		
Vibrante simple			r			ɽ					
Fricativa	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Fricativa lateral			ɬ ɮ								
Aproximante		ʋ	ɹ			ɻ	j	ɰ			
Aproximante lateral			l			ɭ	ʎ	ʟ			
Oclusiva eyectiva	pʰ		tʰ			tʰ	cʰ	kʰ	qʰ		
Implosiva	ɓ ɗ		ɗ ɗ				ɟ ɟ	ɠ ɠ	ɢ ɢ		

Figura 12: Fonemas consonánticos pulmonícos del Alfabeto Fonético Internacional. [16]

Algo importante de destacar es que el IPA incluye símbolos para la duración de los fonemas y para la acentuación o énfasis con que se pronuncian, detalles que más adelante son tomados en cuenta en este trabajo.

**2.6.5.1. Caracteres suprasegmentales del IPA:** este es un grupo de caracteres del IPA que se anexan al mismo para caracterizar los elementos del habla que se refieren a la duración, acento, ritmo, entonación y tonos. En general las características suprasegmentales son las que afectan a estructuras mayores que el fonema (sílabas o palabras), pero algunas de estas pueden tener importancia en la pronunciación dentro de un mismo fonema, por ejemplo, un aumento del tono o el volumen dentro del mismo fonema. En la Figura 13 se presenta la tabla completa de los mismos extraída del IPA. [16]

En esta figura, se ve que el alfabeto contiene caracteres para el acento, primario y secundario, caracteres de duración, y reducción en los silencios entre palabras, de tonos dentro de fonemas e incluso dentro de palabras, etc.

Al f a b e t o F o n é t i c o I n t e r n a c i o n a l		
Suprasegmentales	Tonos y acentos	
' Acento primario	NIVEL	CONTORNO
ˈ Acento secundario		
ː Larga	ẽ ˘ Extra-alto	ẽ ˆ Subiendo
ˑ Semi-larga	é ˘ Alto	ẽ ˆ Cayendo
˘ Extra-corta	ē ˘ Medio	ẽ ˆ Alto subiendo
· División de sílaba	è ˘ Bajo	ẽ ˘ Bajo subiendo
Grupo menor	ë ˘ Extra-bajo	ẽ ˘ Subiendo-Bajando
Grupo mayor (entonacional)	↓ "Downstep"	↗ Subida global
˘ Ligadura (ausencia de división)	↑ "Upstep"	↘ Bajada global

Figura 13: Símbolos de características suprasegmentales del IPA. [16]

## 2.7. Español de Venezuela

Este trabajo se centró en el español de Venezuela, lo que permite simplificar la lista de fonemas del español genérico. La razón de esta simplificación es el uso de muletillas o vicios del lenguaje muy arraigadas en el habla común venezolana, tales como el seseo y el yeísmo.

### **2.7.1. Seseo:**

El seseo una deformación lingüística del español, en la cual el sonido de los fonemas /θ/ y /s/ (IPA), no se distinguen. Esta forma de hablar es común en algunas regiones del sur de España y casi todos los países latinoamericanos, incluido claramente Venezuela.

### **2.7.2. Yeísmo:**

De forma similar al seseo, el yeísmo representa la deformación del fonema /ʒ/ (fricativo post-alveolar) pronunciándolo de forma idéntica al fonema /ʎ/ (lateral proximante), en este caso ambos fonemas se pronuncian de igual forma, siendo la última las más popular en la región. Este fenómeno ocurre en muchos países latinoamericanos tales como Venezuela y México. En algunos países, como Argentina ambos fonemas se deforman hasta aproximarse más al fonema /ʃ/, que se corresponde con los grafemas [ch] del Español, o más aun, al [sh] del Inglés.

### **2.7.3. Betacismo venezolano:**

El Español venezolano también se caracteriza por la no distinción del fonema /v/, siendo sustituido, en casi todos los casos, por el oclusivo bilabial /b/. A pesar de que esta no es una regla formalmente aceptada, la mezcla de estos fonemas no conlleva a confusión lógica o de significado y en todos los casos, el significado puede determinarse inequívocamente a partir del contexto. Esta es la razón por la que en este trabajo se permite la omisión del fonema /v/, en favor del /b/ que aparece con mucha mayor frecuencia en el habla cotidiana.

En la Tabla 1 se pueden observar los fonemas consonánticos que fueron seleccionados, en su representación IPA, acompañados del grafema o letra correspondiente al Español escrito venezolano, que como puede verse, en algunos casos el sonido depende de la posición del grafema en la palabra los cuales se identifican con un asterisco (\*). Se incluye también el modo y punto de articulación de fonema.



Tabla 1: Fonemas consonánticos seleccionados.

Fonema IPA	Grafema español	Tipo de fonema	Modo de articulación	Punto de articulación
/p/	p	consonántico	oclusiva	bilabial
/b/	b	consonántico	oclusiva	bilabial
/t/	t	consonántico	oclusiva	labidental
/d/	d	consonántico	oclusiva	labidental
/k/	c*, k, q*	consonántico	oclusiva	velar
/g/	g*, j*	consonántico	oclusiva	velar
/x/	j	consonántico	fricativa	velar
/ʃ/	ch	consonántico	fricativa	post-alveolar
/l/	l	consonántico	proximante lateral	alveolar
/ʎ/	ll, y*	consonántico	proximante lateral	palatal
/r/	rr	consonántico	vibrante múltiple	alveolar
/r/	r	consonántico	vibrante simple	alveolar
/m/	m	consonántico	nasal	bilabial
/n/	n	consonántico	nasal	alveolar
/ɲ/	ñ	consonántico	nasal	palatal
/f/	f	consonántico	fricativa	labidental
/s/	s, z, c*	consonántico	fricativa	alveolar

En la Tabla 2 se presentan los fonemas vocálicos y su clasificación según la apertura bucal y proximidad de articulación.

Tabla 2: Fonemas vocálicos seleccionados.

Fonema IPA	Grafema español	Tipo de fonema	Anterioridad	Abertura
/a/	a	vocálico	central	abierta
/e/	e	vocálico	anterior	semi-cerrada
/i/	i	vocálico	anterior	carrada
/o/	o	vocálico	posterior	semi-cerrada
/u/	u	vocálico	posterior	cerrada

## **2.8. Características del sonido hablado**

El sonido es una onda transversal mecánica, conformada por perturbaciones de las partículas de aire que se desplaza alejándose de la fuente de vibraciones que la origina. Juana Gil en su libro “Los sonidos del lenguaje” (1988), define dos fases del movimiento ondulatorio, “Compresión: Fase del ciclo vibratorio de la onda sonora durante la que se produce el aumento de la presión aérea... ..Rarefacción: Fase del ciclo vibratorio durante la cual se produce un descenso en la presión aérea...” (p. 148). Posteriormente, la autora define una onda sonora como “... propagación de una perturbación a través de un medio material como el aire, en forma de una serie de compresiones y rarefacciones alternas que afectan a cada una de las partículas componentes de este medio” (p. 150). [17]

Dicho esto, una señal de audio es el resultado de una perturbación de las partículas de un medio, que se propaga por el mismo. En el caso de ser una onda sonora simple o tono puro, la perturbación se propaga con frecuencia y amplitud única y constante (sin tomar en cuenta pérdidas y atenuaciones).

Físicamente los elementos que caracterizan una señal de audio son tiempo (durante el cual se percibe la perturbación), frecuencia (número de perturbaciones percibidas por unidad de tiempo) y amplitud (distancia entre punto de reposo y punto máximo de desplazamiento de las partículas conformantes).

Las señales de voz son señales de audio compuestas, productos de la suma de un número determinado de señales de audio simples, que tienen diferentes características físicas. [17]

### **2.8.1. Frecuencia fundamental:**

Es la frecuencia del tono de menor frecuencia que compone la señal de audio compleja. “Onda sonora simple de frecuencia más baja entre las que forman la onda sonora compleja” (Gil, 1988, pág. 85). La frecuencia fundamental concuerda con la frecuencia de apertura y cierre de las cuerdas vocales, por lo tanto esta puede ser modificada con el fonema que se pronuncie. [17]

### **2.8.2. Armónicos:**

Son los tonos cuya frecuencia es un múltiplo entero de la frecuencia fundamental. Estos acompañan al tono fundamental y aportan complejidad al sonido. [19]

### **2.8.3. Formantes:**

“Zona de la escala de frecuencias en la que un sonido presenta una mayor concentración de energía. También puede definirse como cada una de las resonancias del conducto vocal” (Gil, 1988, pág. 148). Los formantes son el producto del resonar de la onda de sonido en los conductos bucales y de la glotis, reforzando las amplitudes de grupos armónicos en frecuencias específicas según la forma física y tamaño de dichos conductos. Es decir, los formantes son propios del hablante. [17]

### **2.8.4. Timbre:**

Según algunos autores, es un atributo del sonido que permite identificar dos de estos que tengan el mismo tono, tiempo y volumen. En realidad es una característica compleja, que involucra múltiples medidas físicas o atributos distintivos y de naturaleza evaluativa del sonido, así por ejemplo, una misma nota musical tocada con dos guitarras con las mismas características (fuerza en la cuerda, tiempo, velocidad, músico, etc.) pueden distinguirse en el timbre, aportando mejor calidad de sonido a una guitarra sobre la otra. El timbre es un identificador de la fuente que produce el sonido. Este es producto de múltiples características del sonido pero se pueden destacar algunos factores principales que lo definen. [19]

**2.8.4.1. Espectro:** es la distribución energética de las señales simples que componen la señal compleja de audio. El espectro, de forma práctica, suele determinarse con la Transformada de Fourier lo que arroja el resultado de cuanta energía conforma la señal y en qué frecuencias se encuentran. [19]

**2.8.4.2. Envoltente temporal:** que se puede interpretar como la línea imaginaria que une las amplitudes máximas de la señal vista en el tiempo. Ésta en la práctica puede calcularse con la ayuda de cuatro medidas de tiempo: ataque (tiempo que tarda la señal en alcanzar el máxima amplitud), decaimiento (tiempo que tarda en bajar hasta el punto de reposo), sostenimiento (tiempo que se mantiene), liberación (tiempo que tarda en extinguirse el sonido). Estas cuatro medidas de tiempo pueden caracterizar en forma mínima pero suficiente la envoltente temporal de la señal. [20]

**2.8.4.3. Los transitorios:** son señales o tonos de muy corta vida que aparecen mayormente durante el ataque y decaimiento de la señal de audio pero estos suelen atenuarse rápidamente, dejando los sonidos fundamentales y otros componentes. [20]

### 2.8.5. Velocidad de elocución:

“...es el número de elementos fónicos (sonidos y pausas) que se pronuncian en una unidad de tiempo determinada” (p. 548). Gil, J. (2007). Como la intuición lo indica, es la velocidad a las que se producen fonemas al hablar. [21]

No existe un número único para esta característica del habla, ya que esta varía de acuerdo con el idioma, la región donde se hable el mismo idioma o el hablante, incluso varía según el motivo por el que se habla (por ejemplo publicidades de radio en las que el tiempo al aire es costoso). Algunos autores establecen rangos de velocidades para delimitar el “habla normal”. Según Emma Rodero (2012) este rango debe estar, para el español, entre las 160 y 200 palabras por minuto o wpm por sus siglas en Inglés. Este rango varía ligeramente entre diferentes autores. Ann Utterback (2000) establece que el rango es de 145 hasta 180 wpm. John Laver (1994) propone 160 a 240 wpm. Ronald Carver (1992) establece, según experimentación, que el límite de comprensión humana es de 45 fonemas por segundo (fps), a pesar de demostrar que no es posible hablar a esta velocidad. Por otro lado el tamaño promedio de una palabra en español es de aproximadamente cuatro fonemas (Quilis & Fernández, 1969), lo que permite hacer un equivalente de esta velocidad en fonemas por minuto. [22]. [23].

En la Tabla 3 se presenta la aproximación obtenida al cruzar los rangos comentados. Se pueden ver los tiempos necesarios para representar un fonema a tres velocidades diferentes de elocución, tomando en cuenta los extremos descritos, y para la velocidad máxima teórica de comprensión.

Tabla 3: Rangos de velocidades de elocución.

	Rango normal de habla			Máxima
Velocidad (wpm)	145	180	240	675
Fonemas promedio por palabra	4	4	4	4
Fonemas por minuto	580	720	960	2700
Fonemas por segundo	9,67	12	16	45
Tiempo por fonema (ms)	103,5	83,3	62,5	22,2

### 2.8.6. Intensidad del sonido:

Existen varios conceptos relacionados con la intensidad del sonido que deben ser explorados y distinguidos. Formalmente la intensidad del sonido es la cantidad física que mide la energía transportada por una onda de sonido en cada segundo (potencia), esta se mide en unidades

de vatio entre metros cuadrados ( $W/m^2$ ) y se calcula con la Ecuación (1), donde P es potencia emitida por la fuente y A es área de propagación normal. [18]

$$I = \frac{P}{A} \quad (1)$$

Para el caso de una onda esférica, la expresión puede ser escrita de la forma que muestra la Ecuación (2).

$$I = \frac{P}{4\pi r^2} \quad (2)$$

**2.8.6.1. Potencia acústica:** es la potencia emitida por una determinada fuente acústica independientemente del entorno en el que se encuentra. Se mide en términos de presión, típicamente la unidad usada es el pascal (Pa). Esta cantidad está relacionada directamente con la amplitud de la onda sonora, y la percepción de la misma por parte del oyente es lo que se conoce como volumen. [18]

**2.8.6.2. Volumen:** es la percepción humana que se tiene de la potencia de un sonido. El oído humano es sensible no en una escala lineal sino logarítmica, es decir, los cambios en el nivel de potencia que pueden ser percibidos por el oído corresponden a una escala logarítmica. Por lo tanto tiene sentido que la unidad para medir el volumen sea el decibel y se calcula con la Ecuación (3), usando como potencia de referencia  $w_0$  la potencia umbral de audición y donde  $w$  es la potencia del sonido y  $V$  el volumen percibido. [18]

$$V = 10 * \text{Log} \frac{w}{w_0} \quad (3)$$

**2.8.6.3. Potencia umbral de audición (PUA):** es la potencia mínima que debe emitir una fuente de sonido para activar una sensación en el oído humano o para ser detectada por el mismo. El PUA depende de la edad del oyente, como parece indicarlo la lógica, pero también depende de la frecuencia del sonido. Sonidos de frecuencias más bajas tienden a tener PUA más bajos y de forma contraria, sonidos más agudos o de frecuencias más altas tienen PUA mayores. Este se mide en unidades logarítmicas audiométricas que sitúan la referencia de la escala (los 0 dB) en los  $10^{-12} \frac{W}{m^2}$  hablando en términos de intensidad de sonido (sonoridad) o en los 20  $\mu\text{Pa}$  en términos de potencia acústica. En la Figura 14 puede verse la dependencia de la frecuencia y el PUA. [24]

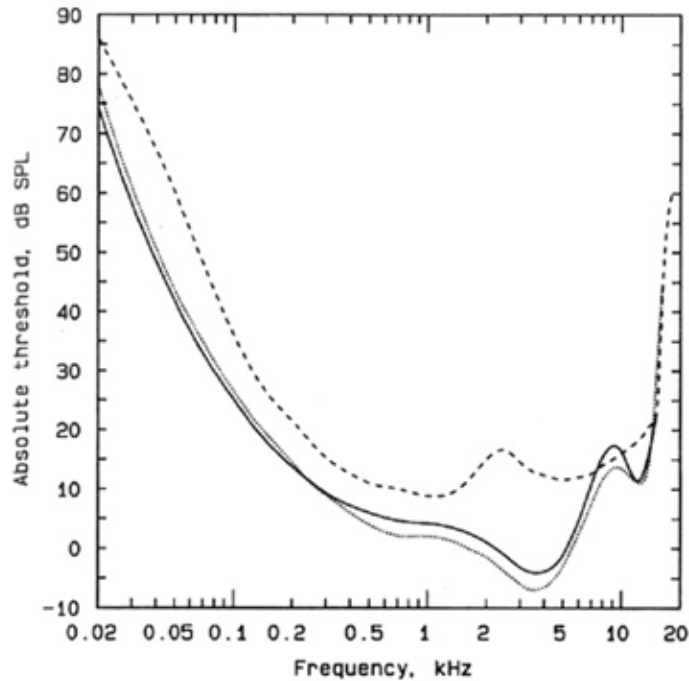


Figura 14: Curva PUA vs Frecuencia del sonido (Moore). [24]

Esta curva fue extraída del libro “An Introduction to the Psychology of Hearing” (Moore, 1977), desarrolladas por medio de experimentación práctica. La curva punteada fue obtenida con la medición directa sobre sujetos, es decir, de la percepción de una persona, y las continuas se obtuvieron en ausencia del sujeto, con micrófonos ubicados en el lugar en el que se encontraría la cabeza del sujeto. [24]

La Figura 15 se puede ver la curva obtenida por Hugo Fastl y Eberhard Zwicker (2007), en las que se incorporan los límites superiores para las intensidades. Como se ve, se representa en línea punteada un límite de riesgo y por encima una línea entrecortada del límite de dolor, también calculada en ausencia del sujeto. En estos estudios se planteas los límites del habla dentro del rango de 100Hz y 5kHz y entre los 20dB y 80dB de intensidad. [25]

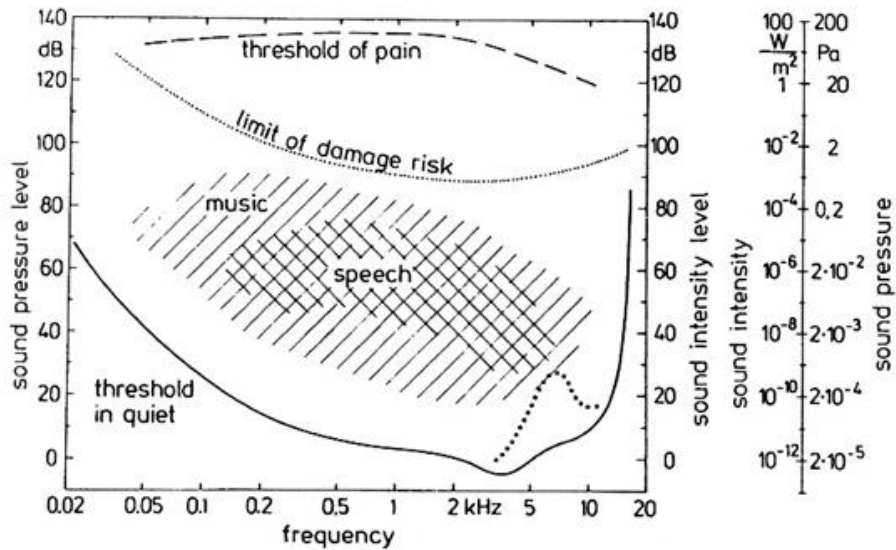


Figura 15: Curva PUA contra frecuencia del sonido (Fastl y Zwicker). [25]

### 2.8.7. Umbral de frecuencia auditiva:

Es el rango de frecuencias que es capaz de activar la sensibilidad el oído humano. Este rango puede variar según el autor. En la investigación realizada el rango más grande encontrado fue de entre 20Hz y 20kHz, que se observa en la Figura 15. Por otra parte, vale la pena tomar en cuenta que el rango de frecuencias capaz de ser producido por el sistema bucal humano, es mucho menor, y ya que se pretende el diseño de un sistema especializado para el transporte del habla, no vale la pena exigir al sistema que soporte frecuencias fuera de este rango. Esta consideración dará con resultado un sistema más eficiente en cuanto al tamaño de la trama. El rango de habla encontrado por los autores Fastl y Zwicker está entre los 100Hz y 5kHz, como se dijo anteriormente. [25]

**2.8.7.1. Mínima variación perceptible de frecuencia (MVPF):** Es el cambio de frecuencia mínimo que puede ser detectado por el oído humano, según H. Fastl y E. Zwicker (2007) esta MVPF es de 3.6Hz mínimo hasta los 500Hz como se puede ver en la Figura 16. [25]

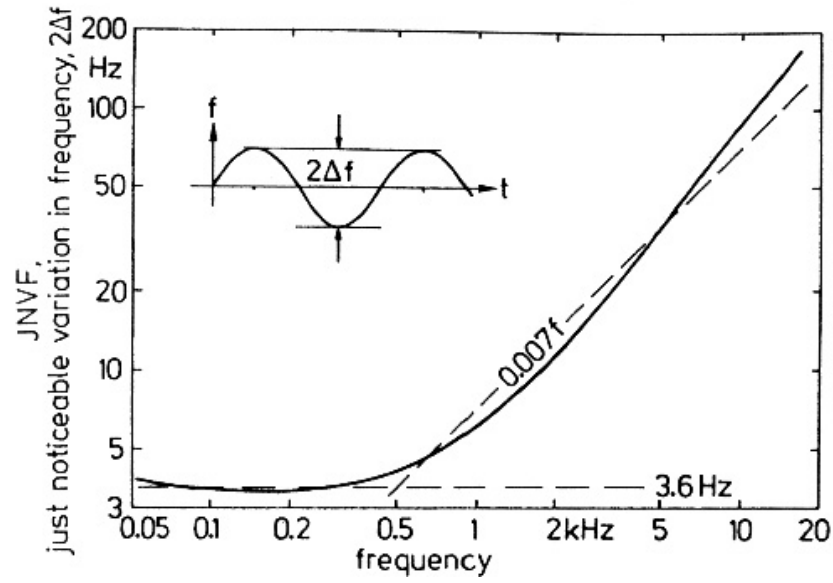


Figura 16: Mínima variación perceptible de frecuencia. [25]

Para valores superiores de frecuencia el MVPF es aún mayor, es decir, para el oído humano es más difícil distinguir cambios cuando mayor es la frecuencia fundamental.



## CAPITULO III

### MARCO METODOLÓGICO

Según Fidiás Arias, en su libro “El proyecto de investigación” (2006), este trabajo de investigación clasifica como una investigación documental de tipo correlacional, ya que en esta se realiza una comparación entre variables independientes. En este caso, las variables, se refiere a las características comparadas de ambos protocolos, Codificación Por Fonema (CPF) y Modulación por Codificación de Pulso (Pulse Code Modulation, PCM). Sin embargo, los datos comparados provienen de fuentes tanto primarias, las características obtenidas en el diseño de la trama CPF; como secundarias, las obtenidas de la recomendación G.711 que define el estándar PCM. [26]

En este capítulo se describe la metodología seguida a lo largo de este trabajo, recorriendo la realización de cada uno de los objetivos de la investigación.

#### **3.1. Realización de la librería fonética**

Producto de la investigación previa, se determinó que la cantidad mínima de fonemas necesarios para representar el español venezolano es de 17 fonemas consonánticos y 5 vocálicos (Cava, 2016) como pudo verse en las tablas anteriores (Tabla 1 y Tabla 2 del capítulo II). [27]

Una vez determinada esta lista, se seleccionó un programa de radio público y de divulgación libre que tuviera las características deseadas, es decir, que contara con diálogo entre dos o más personas de forma improvisada, además de lenguaje popular pero lo suficientemente formal para una pronunciación comprensible y lo más genérico posible. El programa seleccionado fue “La entrevista”, programa de opinión, de aproximadamente una hora de duración (1 hora, 06 min, 47 seg), de la emisora Radio Caracas Radio, difundido en vivo el día 24 de octubre del 2016.

El programa de radio fue transcrito en español (ver Anexo B), haciendo énfasis en la forma de pronunciación de las palabras sobre la ortografía formal.

**Tabla 4: Frecuencia de aparición fonética en el programa de radio**

<b>Fonema IPA</b>	<b>Cantidad</b>	<b>Porcentaje</b>
/a/	5133	12,55
/e/	5806	14,19
/i/	3004	7,34
/o/	4099	10,02
/u/	1872	4,58
/p/	996	2,43
/b/	832	2,03
/t/	2146	5,25
/d/	2072	5,06
/k/	1406	3,44
/g/	293	0,72
/x/	227	0,55
/ʃ/	158	0,39
/l/	1927	4,71
/λ/	218	0,53
/r/	293	0,72
/r/	2210	5,4
/m/	1081	2,64
/n/	3030	7,41
/ɲ/	69	0,17
/f/	224	0,55
/s/	3816	9,33
<b>Total</b>	<b>40912</b>	<b>100,01</b>

De la transcripción del programa se realizó la Tabla 4, en la cual se puede ver el número de apariciones de cada uno de los fonemas, así como la frecuencia de cada uno con respecto al total en el programa. De la lectura y análisis de esta tabla, se obtiene la Tabla 5, donde se puede ver, que los fonemas más comunes son los vocálicos, estos constituyen un 48.68%, casi la mitad del total, a pesar de ser solo 5 fonemas. También se observa que los fonemas menos comunes son el /ʃ/ (grafema [ch] del español), /ɲ/ (grafema [ñ] del español) y /r/ (grafemas [rr] o [r] al comienzo de la palabra).

Del análisis de las tablas anteriores (Tabla 4 y Tabla 5) se puede extraer un dato que será relevante más adelante. En la Tabla 6 se observa un estimado de la cantidad de fonemas por minuto que se emitieron en el programa. En un programa de radio en el cual desea pasar poco tiempo en

silencio y que se mantiene una velocidad de habla cuidadosamente uniforme y comprensible, es interesante obtener un aproximado de la velocidad fonética promedio, o dicho de otra manera, la cantidad de fonemas que se emiten por unidad de tiempo. El resultado es de aproximadamente 682 fonemas por minuto.

**Tabla 5 Frecuencia de aparición fonética por tipo de fonema.**

<b>Duración aprox. 60 min</b>	<b>Cantidad</b>	<b>Porcentaje</b>
<b>F. vocálicos</b>	19914	48,68
<b>F. consonánticos</b>	20998	51,32
<b>Total Fonemas</b>	40912	100

**Tabla 6: Fonemas emitidos por minuto en el programa de radio.**

<b>Duración aprox. (min)</b>	60
<b>total fonemas</b>	40912
<b>Fonemas por minuto (aprox.)</b>	681,87

En abril de 2003, Hernán Pérez publicó su trabajo llamado “Frecuencia de fonemas”, para la Universidad de Concepción en Chile. A pesar de que este trabajo es extranjero, e involucra un español diferente al del presente estudio, se presentará una comparación con los resultados estadísticos obtenidos en dicho trabajo. En su estudio, Pérez usa noticias de televisión chilena formando un corpus de más de 75000 fonemas, y los compara con cuatro trabajos previos de lingüistas del español de otros países que también estudiaban frecuencia fonética del español. Además, la tabla presentada por Pérez se corresponde con los fonemas seleccionados en este trabajo, es decir, hizo uso de los mismos criterios para la selección de los fonemas usados en Chile, lo que la hace comparable directamente. La Tabla 7 contiene la comparación de sus resultados con respecto a los obtenidos en este trabajo. [28]

Tabla 7: Tabla comparativa de resultados de frecuencia fonética.

Fonema IPA	Ruiz	Pérez	Error porcentual
/a/	12,55	12,31	1,91
/e/	14,19	14,13	0,42
/i/	7,34	7,46	1,61
/o/	10,02	9,28	7,39
/u/	4,58	3,05	33,41
/p/	2,43	2,58	5,81
/b/	2,03	1,92	5,42
/t/	5,25	4,92	6,28
/d/	5,06	4,84	4,35
/k/	3,44	3,98	13,57
/g/	0,72	0,94	23,4
/x/	0,55	0,74	25,68
/ʃ/	0,39	0,32	19,95
/l/	4,71	5,05	6,73
/ʎ/	0,53	0,69	23,19
/r/	0,72	0,64	11,11
/r/	5,4	6,19	12,76
/m/	2,64	2,62	0,76
/n/	7,41	7,78	4,75
/ɲ/	0,17	0,24	29,17
/f/	0,55	0,75	26,67
/s/	9,33	9,61	2,91
<b>Total</b>	<b>100,01</b>	<b>100,04</b>	

Como puede verse en la Tabla 7, salvo algunos casos particulares, los errores se mantienen por debajo del 10%, y más aún en los fonemas más importantes, los de mayor frecuencia, el error es menos al 5% (como el caso del /e/, /a/, etc.). Los casos de errores de más de 20% se presentan en casos de fonemas con muy baja frecuencia (menores de 1%), como el caso de /ɲ/ y /f/, este error puede deberse a una diferencia de pocas apariciones de más en un trabajo respecto al otro. El único caso que sobresale, es el de /u/ que siendo un fonema vocálico y de alta frecuencia presenta un error de más de 30%, esto puede deberse a la diferencia del idioma inter regional, o a un caso particular del presente trabajo, ya que en comparación con los demás resultados mostrados por Pérez, tales como: (Navarro, 1946) y (Quilis & Fernández, 1969), se obtiene un error parecido.

No obstante, dada la muestra reducida de este trabajo, y dado que la comparación se realizó contra resultados de trabajos extranjeros (español de diferentes regiones), los resultados obtenidos son muy satisfactorios, ya que muestran una estructura estadística de los fonemas del español bastante uniforme y acorde con trabajos previos. Además de la confirmación de los criterios de selección de la lista de fonemas.

### **3.2. Características fonéticas a transmitir**

Dado que los fonemas son el producto de la percepción de un sonido, es de esperarse que la transmisión de los mismos deba contener toda la información necesaria para identificar el sonido que los produjo (o bien permitir su reproducción).

En el diseño de este sistema se pretende pre-interpretar el sonido antes de ser transmitido, es decir, los fonemas van a ser generados antes de la transmisión, en el dispositivo de transmisión y no en el cerebro del oyente, como normalmente ocurre. Por lo tanto no es necesario el envío del audio en sí, he ahí la innovación del sistema. Aun así, resultaría poco natural la recepción de la información totalmente desligada del sonido que la generó y ya que se diseña un sistema de telecomunicaciones de habla humana y para humanos, se decidió la incorporación de datos extras que complementen la información, para una mejor reconstrucción de audio en el receptor que aporten familiaridad humana en los mensajes transmitidos.

Como se dijo anteriormente, para hacer posible el reconocimiento de un sonido son necesarias 4 características: duración (tiempo), tono (frecuencia), volumen o intensidad (amplitud) y timbre. Este último resulta ser una característica compleja que se dejará en una modalidad opcional del sistema, como se verá más adelante.

#### **3.2.1. Tiempo**

Como se observó en la Tabla 3, la velocidad máxima de producción de fonemas no alcanza los 45 fonemas por segundo, tomando este como límite superior de velocidad, se puede establecer la cota de tiempo mínimo a ser expresado. El resultado es 22 ms por fonema a una velocidad de habla, que en teoría, no podría ser alcanzada.

En este punto vale la pena recordar algunos datos que quedaron en espera anteriormente. Como se vio en la Tabla 5, en la transcripción realizada para este trabajo, se

obtuvo un total de 40.912 fonemas, si se toma la velocidad media obtenida en la Tabla 3 de 180 wpm o su equivalente en fonemas por minuto (fpm) se obtiene el resultado de la Tabla 8.

**Tabla 8: Densidad fonética del programa de radio utilizado**

Fonemas	Duración (min)	fpm	wpm
40912	60	681,87	170,5

La velocidad fonética promedio aproximada del programa transcrito es de 170,5 wpm (bajo la aproximación de cuatro fonemas por palabra), lo que concuerda con los rangos de tiempo y velocidades usadas.

La cota superior de tiempo se establecerá tomando en cuenta que se puede tratar de un fonema vocálico, por ejemplo el /a/, y ya que este puede ser pronunciado durante un tiempo relativamente arbitrario, el límite se seleccionó según conviene en la trama, por lo menos de un segundo. Con 6 bits es posible representar máximo 1,408 s. Ver Ecuación 4.

$$2^6 \times 22 \text{ ms} = 1,408 \text{ s} \quad (4)$$

Lo que permite la transmisión de fonemas de entre 22 ms y 1,4 s de duración. Este rango concuerda con lo discutido y permite el manejo de velocidades tan bajas que no deberían ser alcanzadas en una conversación normal.

### 3.2.2. Intensidad:

Esta medida física se refiere a la amplitud de la onda de sonido (tamaño de la perturbación), musicalmente se relaciona con el volumen del sonido o cuan fuerte es percibido por el escucha.

Como se vio anteriormente (en la Figura 15), H. Fastl y E. Zwicker (2007), establecen los límites de intensidad para el habla entre los 20 y 80dB [19], si se toma en cuenta el 0dB en los  $10^{-12} \frac{W}{m^2}$ , lo que plantea un rango disponible de 60dB. Para obtener este rango en la trama se necesitan 6 bits. Por la Ecuación 5 se obtiene que es posible transmitir 64dB en el rango esperado. Dado que el rango es asimétrico, el receptor debe recordar sumar 20dB al número recibido para obtener el valor de intensidad que se debe reproducir.

$$2^6 = 64 \text{ dB} \quad (5)$$

Así, el rango manejado por el sistema es de 20dB a 84dB.

### 3.2.3. Tono o frecuencia:

El trabajo experimental en acústica humana de H. Fastl y E. Zwicker (2007) determinan el rango de audición entre 20 Hz y 20 kHz. Pero el rango de habla se establece entre los 100 Hz y 5 kHz, lo que es de mayor peso para los objetivos de este trabajo. Aun así, este rango resulta grande en comparación con otros autores.

El estándar PCM establece como límite superior de frecuencia del habla en 4 kHz, más específicamente de la recomendación G.711 de la ITU-T (1988), sabemos que la frecuencia de muestreo es de 8000 muestras por segundo que se deben al rango máximo de 4 kHz en la frecuencia del habla. Ya que la comparación se realiza con este estándar se decidió reducir el rango hasta 4 kHz.

Este rango puede ser simplificado aún más tomando en cuenta la mínima variación perceptible de frecuencia, lo que permite transmitir los cambios de frecuencia con apreciación máxima de 4 Hz. Ya que el oído no es perceptible a cambios más pequeños.

La Ecuación 6 describe el cálculo de la cantidad mínima de niveles que deben ser transmitidos manteniendo el rango y asumiendo una apreciación de 4 Hz.

$$\frac{4000\text{Hz} - 100\text{Hz}}{4\text{Hz}} = 975 \text{ niveles} \quad (6)$$

Para ello son necesarios mínimo 10 bits, con los que se pueden determinar 1024 niveles de frecuencia. Si se establece el mínimo en 100Hz y que cada nivel representa 4 Hz, se obtiene el valor máximo transmitible de frecuencia. En la Ecuación 7 puede verse en detalle.

$$2^{10} * 4\text{Hz} + 100\text{Hz} = 4196 \text{ Hz} \quad (7)$$

En cuanto al límite superior, se tiene un rango de frecuencia incluso mayor que la que se puede transmitir en PCM, aunque no se tiene precisión menor de 4 Hz.

### 3.2.4. Elementos suprasegmentales:

Como se habló en el capítulo II, las características suprasegmentales se enfocan en identificar el tono, tiempo, pausas, acentos, ritmo e intensidades con las que se pronuncian los fonemas. Algunas de estas características pueden obtenerse indirectamente de las otras, por ejemplo, el acento es una característica propia de la palabra y se puede determinar del tono y énfasis con que se pronuncian los fonemas que la conforman. Pero ¿qué pasa cuando el fonema cambia de tono antes de terminar de ser pronunciado? Un fonema que se esté pronunciando lentamente (tiempo prolongado) puede ser susceptible de presentar cambios en el tono o la intensidad, (típicamente combinaciones de ellas) para este caso se decidió implementar algunos de los caracteres suprasegmentales del IPA, los que actúan dentro de la jerarquía del fonema (y no de palabras, como el acento) y que no pueda ser extraído de los otros campos que ya se estén enviando.

En la Figura 17 vemos los caracteres seleccionados. Estos son los casos en los que el fonema cambia de tono mientras es pronunciado, aumentando, decayendo (*rising* y *falling*). Y en los casos de *peakin* y *dipping* en los cuales el tono aumenta y disminuye; o disminuye y aumenta respectivamente.

<i>Contour-tone examples:</i>		
ě	↑	Rising
ê	↓	Falling
ẽ	↗	Peaking
ẽ	↘	Dipping

Figura 17: Caracteres suprasegmentales IPA implementados. [16]



Como se puede observar, son cuatro los caracteres que deben ser agregados a la información, más uno que represente el caso neutro, en el que no hay variaciones en la pronunciación del fonema, es decir cinco casos. Para ello se reservan 3 bits más en la trama.

### 3.2.5. Código identificador del fonema:

En cuanto a la selección del código implementado para identificar el tipo de fonema, se utilizó un código Huffman con los 22 fonemas español de Venezuela y la probabilidad de ocurrencia obtenida para cada uno.

En la Tabla 9 se pueden ver los códigos obtenidos luego de aplicar el algoritmo de Huffman a la lista de fonema.

Tabla 9: Codificación Huffman para la lista de fonemas

Número	Fonema API	Probabilidad	Código
1	/e/	14,19	010
2	/a/	12,55	011
3	/o/	10,02	110
4	/s/	9,33	111
5	/n/	7,41	0010
6	/i/	7,34	0011
7	/r/	5,40	1001
8	/t/	5,25	1010
9	/d/	5,06	1011
10	/l/	4,71	00000
11	/u/	4,58	00010
12	/k/	3,44	10000
13	/m/	2,64	10001
14	/p/	2,43	000010
15	/b/	2,03	000110
16	/g/	0,72	0001110
17	/r/	0,72	0001111
18	/x/	0,55	00001101
19	/f/	0,55	00001110
20	/λ/	0,53	00001111
21	/ʃ/	0,39	000011000
22	/ɲ/	0,17	000011001

El código cumple con las propiedades de Huffman y arroja una longitud de código promedio de 3,9147 bits, calculada con la ecuación 8, donde  $L_p$  es la longitud promedio,  $P_n$  es la probabilidad de aparición del n-esimo carácter,  $l_n$  es la longitud de ese carácter, y  $N$  es la cantidad de caracteres del diccionario.

$$L_p = \frac{\sum_1^n P_n * l_n}{N} \quad (8)$$

Esta longitud resulta más pequeña que la necesaria si no se usa el código Huffman, es decir, implementando un código de longitud fija, en ese caso se necesitarían 5 bits para la identificación de cada fonema.

### 3.2.6. Construcción de la trama:

La trama consta de grupos de secuencias concatenadas que se generan para cada fonema que se va a transmitir. Cada grupo está conformado por 5 campos: Primero un bit de control que siempre se encuentra en uno, que encabeza cada grupo (cada fonema). El identificador de fonema, que como vimos es una cantidad variable que va entre 3 y 9 bits, pero con longitud promedio de 4 bits aproximadamente, debido a las probabilidades de ocurrencia. El campo de tiempo, que dice durante cuánto tiempo se pronunció el fonema, este es un campo de 6 bits. El campo intensidad, también de 6 bits. Campo de tono abarca 10 bits. Y finalmente el campo para suprasegmentales que abarca 3 bits. La Figura 18 muestra la trama para un grupo, es decir, para un fonema  $n$ .

c	Fonema	Tiempo	Intensidad	Tono	S-S
1	n				
1	3<n<9	6	6	10	3

Figura 18: Trama detallada para un fonema a transmitir

Ya que el campo de fonema es variable, el tamaño para cada fonema puede estar en el rango de 29 y 35 bits por cada fonema. Pero con longitud promedio de 30 bits.

### 3.3. Comparación con el estándar PCM

#### 3.3.1. Programa generador de fonemas:

La realización de las pruebas de comparación se llevó a cabo con ayuda de un paquete de cálculo computacional. El software de cálculo utilizado fue MATLAB, del cual la Escuela de Ingeniería Eléctrica cuenta con licencia para la versión R2008. Vale la pena destacar que el código implementado es ejecutable desde otros paquetes de cálculo, incluso otros lenguajes de programación ya que no se usó ninguna de las funciones propias de MATLAB.

Se diseñó un código, que trabaja en dos modalidades. La primera es para la generación pseudo-aleatoria de fonemas (según las probabilidades obtenidas anteriormente) en una cantidad especificada por el usuario. Cabe destacar que esta generación de la lista de fonemas es acorde a unas reglas básicas de estructuración del idioma que se tomaron en cuenta, por ejemplo, las probabilidades cambian dependiendo del carácter que se acaba de generar (carácter anterior). De esta forma no se generan listas muy largas de fonemas consonánticos que tengan alta probabilidad de ocurrencia ya que una vez generado uno de estos, las probabilidades se ajustan para el próximo fonema.

Los fonemas fueron separados en cuatro conjuntos:

El primer conjunto está conformado por los fonemas que, por sí solos, pueden formar una palabra, o que típicamente son fonemas finalizadores de palabras, estos no afectan a las probabilidades del próximo fonema ya que en principio, el comienzo de la próxima palabra podría ser cualquier fonema. Los fonemas de este grupo son los vocálicos, /a/, /e/, /i/, /o/, /u/ y los consonánticos /s/, /n/, /l/, /r/.

El segundo conjunto, está conformado de fonemas consonánticos que solo pueden aparecer seguidos de una vocal, esto afecta la probabilidad de ocurrencia ya que el 100% de dicha probabilidad debe repartirse entre las 5 vocales. Estos Fonemas son los /m/, /r/, /x/, /ɲ/, /λ/ y /ʃ/.

El tercer conjunto lo conforman los fonemas /t/ y /d/, que pueden ir precediendo a los cinco fonemas vocálicos pero también al fonema /r/, en palabras como “**tragar**”, “**dragón**”, etc.

Finalmente el conjunto número cuatro se compone de los fonemas restantes /p/, /b/, /k/, /g/, /f/, que pueden venir acompañados de los cinco vocálicos además de los fonemas /r/ y /l/, en palabras como “**plato**”, “**grama**”, “**iglesia**”, etc.

En la Tabla 10 se puede ver la conformación de estos grupos acompañados de las nuevas probabilidades de ocurrencia que conlleva cada uno.

**Tabla 10: Grupos fonéticos del programa generador.**

Fonema actual	Conjunto	Prox. fonemas posibles	Probabilidad
/a/, /e/, /i/, /o/, /u/, /s/, /n/, /l/, /r/	1	Todos	No afecta
/m/, /r/, /x/, /ʎ/, /ʃ/, /ɲ/	2	/a/	25,79
		/e/	29,15
		/i/	15,08
		/o/	20,58
		/u/	9,41
/t/, /d/	3	/a/	23,21
		/e/	26,24
		/i/	13,57
		/o/	18,53
		/u/	8,47
/p/, /b/, /k/, /g/, /f/	4	/r/	9,99
		/a/	21,35
		/e/	24,14
		/i/	12,44
		/o/	17,04
		/u/	7,79
		/l/	8,01

Posteriormente el programa genera la codificación de los fonemas según lo visto anteriormente, asignando un código Hoffman correspondiente al fonema y reservando los campos fijos de información así como el bit de control inicial de cada grupo.

La segunda modalidad de trabajo del programa es por introducción directa de los datos (fonemas), para esto se solicita al usuario una frase o palabras de forma escrita, y el programa extrae de esta frase los fonemas que se van a transmitir. Se toma en cuenta que el carácter espacio representa silencios o pausas en el habla y por lo tanto no deben ser transmitidos por el sistema (así como los casos espaciales como las letras “h”).

En la Figura 19 se presenta el diagrama de flujo del programa en la modalidad de lista pseudo-aleatoria, y en la Figura 20 se ve el diagrama de flujo del mismo en la modalidad de entrada directa de datos.

Una segunda parte del programa genera la trama de bits a transmitir concatenando todos los grupos (uno para cada fonema) y hace el conteo de la cantidad de bits utilizados.

Además se hace una aproximación de cómo sería la transmisión de estos mismos fonemas en el estándar PCM, en cuanto a la cantidad de bits (ya que para PCM el tipo de fonema es transparente). Para ello se asume que el hablante produce estos fonemas a una velocidad de elocución fija de 720 fpm, que como se dijo anteriormente, es la velocidad promedio de habla.

El programa presenta la comparación entre las dos tramas, en cuanto a la cantidad de bits utilizados y la eficiencia de la trama diseñada respecto a la trama PCM.

El código completo descrito anteriormente se presenta en los anexos de este trabajo (Anexo C).

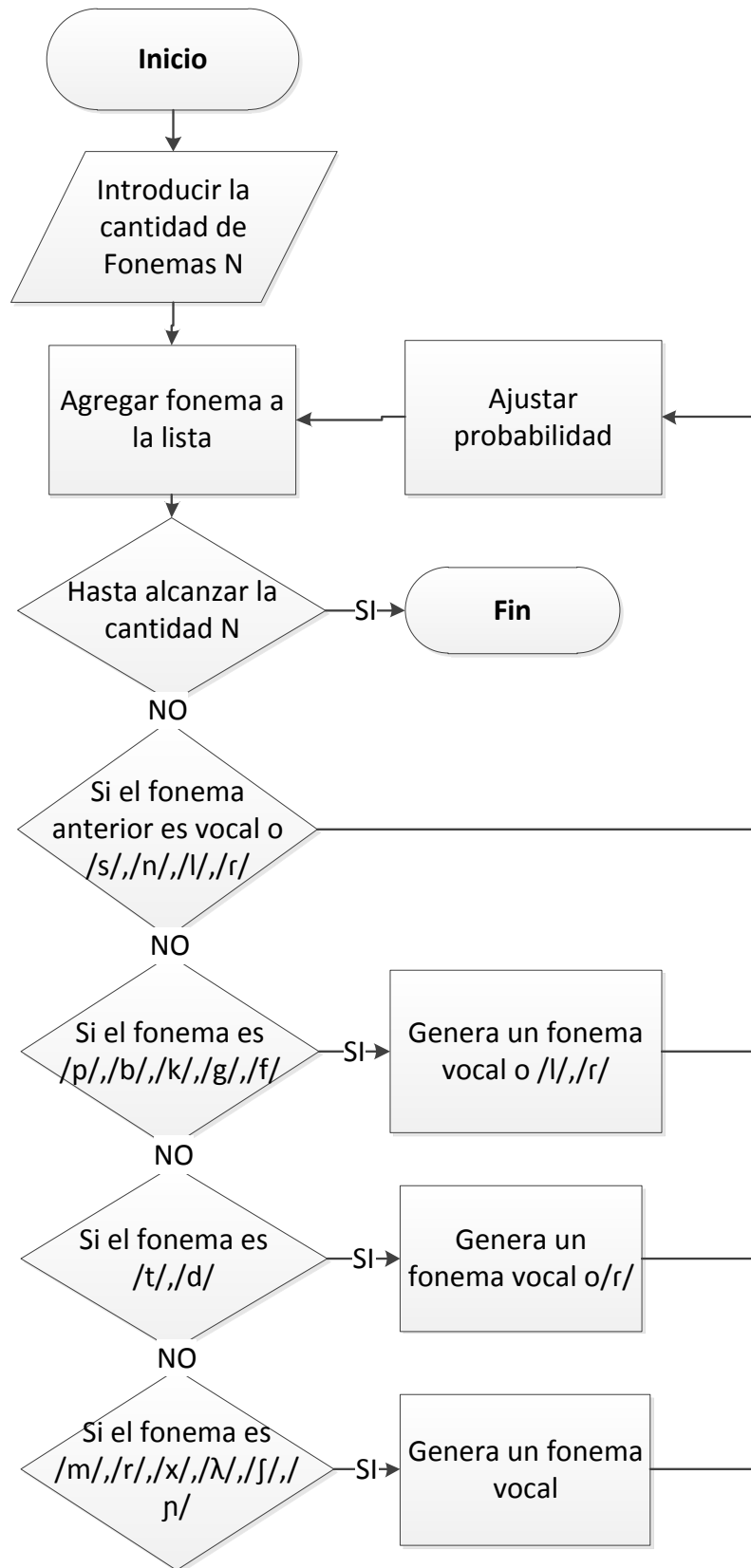


Figura 19: Diagrama de flujo de programa generador en modo aleatorio.

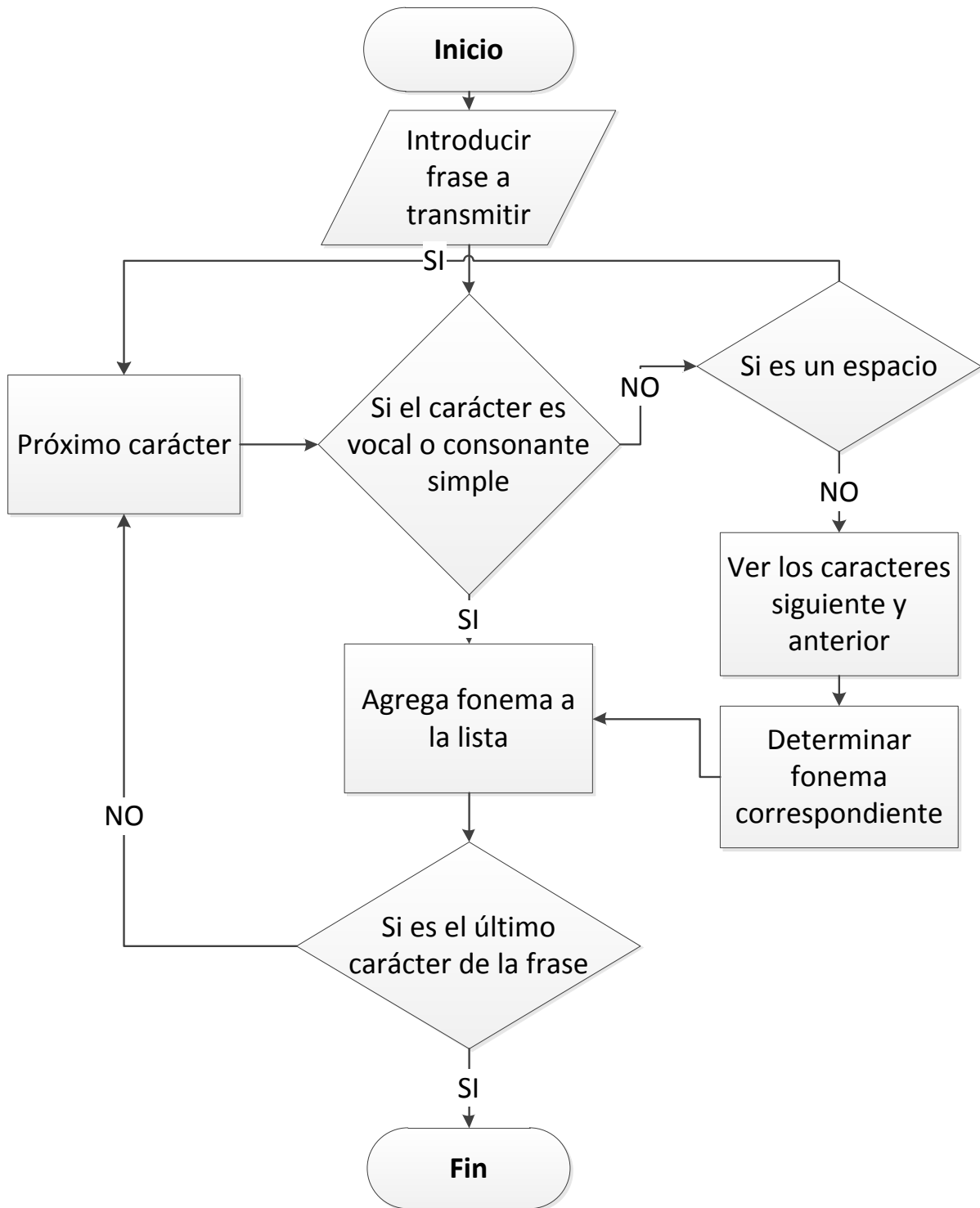


Figura 20: Diagrama de flujo de programa generador en modo directo.

### 3.3.2. Pruebas de comparación:

**3.3.2.1. Verificación del programa:** la primera prueba se realizó con la palabra “casa” introducida al programa en modalidad directa para realizar la comparación básica y para verificar el correcto funcionamiento del programa.

#### 3.3.2.2. Comparación con PCM:

**3.3.2.2.1. Tamaño de la trama:** para esta prueba se realizaron 10 ejecuciones del programa en modalidad pseudo-aleatoria, para listas creciente desde 100 hasta 1000 fonemas, con aumentos de 100 fonemas cada ejecución. Con sus respectivas comparaciones respecto al tamaño de trama correspondiente PCM, tomando la velocidad de elocución de 180 wpm.

**3.3.2.2.2. Eficiencia:** para las ejecuciones anteriores y con ayuda de la ecuación 9 se calculó la eficiencia de la trama de Codificación Por Fonemas (CPF) en comparación con la trama PCM.

$$E = \left(1 - \frac{\text{Trama CPF}}{\text{Trama PCM}}\right) * 100\% \quad (9)$$

**3.3.2.2.3. Velocidad de transmisión:** la velocidad de transmisión definida formalmente depende del sistema físico en el cual se implementa. Para realizar una comparación entre los protocolos es necesario establecer que ambos se implementan en el mismo sistema físico, es decir, a la misma velocidad de transmisión binaria (bits por segundo), lo que además determina la duración del bit. De esta manera es posible comparar la capacidad de compresión de la trama CPF con respecto a la PCM.

**3.3.2.2.4. Velocidad de información:** debido a la diferencia con la que se produce y se transmite la información útil en ambos protocolos es conveniente definir un concepto que ayude a realizar la comparación. Si se fija la cantidad de información que produce el usuario y ésta se traduce en cantidad de bits que abarca por unidad de tiempo (un segundo), la velocidad de información es la



cantidad de bits transmitidos en un segundo que se requiere para transmitir esa información.

En PCM esta velocidad de información es igual a la velocidad de transmisión, ya que esta última es siempre la misma independientemente de la información que contiene.

En el caso de CPF el cálculo de la velocidad de información no es tan directo ya que se trata de una codificación que depende de la velocidad con la que se producen los fonemas (velocidad de habla del usuario) y por lo tanto esta velocidad no es constante, como en PCM. Es necesario usar los valores promedio encontrados en resultados anteriores para hacer un estimado de este valor. Usando promedio de velocidad de habla (en fonemas por segundo) y el tamaño promedio de la trama (en bits por fonema) se puede calcular este valor de velocidad de información promedio de la trama CPF.

**3.3.2.2.5. Capacidad de multiplexado de usuarios:** dada la gran diferencia entre la velocidad de transmisión de cada bit y las de pronunciación de los fonemas, la característica determinante en la capacidad del estándar para el multiplexado de usuarios es la disminución del tamaño de tramas. Ya que el sistema empleará más tiempo “escuchando” y determinando los fonemas que transmitiéndolos, la condición más probable es que sea el sistema el que espere por el usuario y no al contrario, dado que los tiempos dependen de las velocidades de habla.

Una forma de hacer esta estimación es observando en la Tabla 3 los valores de velocidad de habla. Tomando el caso más conservador, es decir, calculando la capacidad para el caso más desfavorable, que es a velocidad máxima de habla, se puede calcular cada cuanto tiempo debe tener acceso al canal un mismo usuario. Además, en un canal PCM con velocidad de transmisión de 64 kbps, se transmite un bit cada 15,6  $\mu$ s. Si se asume esta misma velocidad física de transmisión para el esquema CPF, y usando también la longitud promedio de trama por fonema (usando la longitud promedio del código Huffman), se puede calcular durante cuanto tiempo ocupa el canal cada uno de los usuarios. El intervalo de tiempo restante es el que se tiene disponible para mutiplexar otros usuarios en el mismo canal.

La anterior es la forma más conservadora de calcular este rango, ya que se supone a los usuarios hablando a una velocidad máxima en todo momento y en un sistema realista esto no debería pasar. La forma más eficiente es diseñar un sistema que se adapte a la tasa de transmisión y pueda adaptar la cantidad de usuarios que soporta según la misma.

Otra manera de estimar la capacidad del canal es comparando los resultados que serán obtenidos luego de las pruebas de comparación de tamaños de tramas, ya que sabiendo la relación de tamaños la comparación debe ser directa. Este punto se abordará más adelante.

### **3.3.2.3. Cambios en el esquema:**

**3.3.2.3.1. Cambio en la velocidad de habla:** ya que este es un parámetro tomado como estándar, se consideró importante observar el comportamiento del sistema y de los resultados anteriores cambiando la velocidad de elocución. Lo esperado es que esto produzca alteraciones en la eficiencia de la trama CPF respecto a la PCM. Los límites seleccionados fueron los descritos en el capítulo II como: máxima velocidad inteligible, 22 ms por fonema que se traduce en 681 wpm, y la mínima velocidad permitida por el esquema en el campo de tiempo, es decir, 1.4 segundos por fonema, lo que se traduce en 10 wpm aproximadamente. Ambas velocidades están fuera del rango de cualquier conversación normal, pero aun así es importante observar el comportamiento del esquema en estos valores extremos.

## **3.4. Retardo en transmisión de fonemas:**

Una característica inherente del sistema diseñado, es que debe haber un retraso considerable en la transmisión de los fonemas con respecto a la generación de los mismos. Este retraso está asociado al tiempo que debe tardar el equipo transmisor en reconocer el fonema. A pesar de que éste sea muy eficiente y trabaje a una velocidad muy alta, es necesario que el fonema sea pronunciado totalmente antes ser identificado y luego transmitido. Se deduce que fonemas más largos, o pronunciados a velocidades menores acarrearán mayores retrasos. En la Figura 21 se ilustra esta característica, como puede verse, la trama de cada fonema no es transmitido sino hasta el final de la pronunciación del mismo.

Para estimar este retardo se puede tomar la velocidad de habla mínima según la Tabla 3, de manera de obtener un resultado desfavorable pero dentro de los rangos de habla. Observando la velocidad mínima de producción de fonemas se obtiene el tiempo máximo de pronunciación del mismo y por lo tanto el tiempo máximo del retardo en su transmisión, dentro de los rangos normales de habla. Por otro lado, un resultado más extremo (desfavorable) es posible si se toma en cuenta el tiempo máximo de duración de fonema que admite la trama, este valor es de 1,4 s (campo de tiempo a capacidad máxima).

## CAPITULO IV

### RESULTADOS

En este capítulo se presentan uno por uno los resultados obtenidos a lo largo de las pruebas de comparación entre las tramas CPF y PCM descritas en el capítulo III.

#### 4.1.Verificación del programa generador de fonemas

La primera prueba se realizó introduciendo la palabra “casa” en modo directo al programa generador, en la Figura 21 se puede ver el resultado de esta prueba, además de la comparación entre la trama CPF y la PCM. Esta figura es de carácter ilustrativo y se presenta la señal analógica correspondiente al audio de la palabra “casa”, pueden verse los dos aumentos de amplitud en la señal que se corresponden con la pronunciación de los dos bi-fonemas que conforman la palabra: /k//a/ y /s//a/, rodeados por segmentos de ruido ambiental en los cuales hay silencio del hablante.

En el esquema CPF la transmisión de información se produce solo en los segmentos en los que se genera un fonema, y no hay transmisión durante los silencios. No así en el caso del esquema PCM, en el cual se transmiten muestras de audio sin importar si la información es de ruido ambiental o voz del hablante, esta diferencia se ve en la Figura 21 (a pesar de que la figura no está en escala) donde vemos la comparación entre las tramas de PCM y CPF para el segmento de audio mostrado. Para transmitir este audio se necesitan 38.400 bits en PCM y solo 118 bits en CPF (60 para los fonemas /k/ y /a/, y 58 para los fonemas /s/ y /a/). En la Figura 22 se hace un acercamiento a la trama de los fonemas /k/ y /a/ donde se ve detalladamente a qué corresponde los 60 bits usados.

Debe tomarse en cuenta que la no transmisión de información durante los silencios permite el uso del canal para otros fines, como bits extras de redundancia o también la atención a otro usuario.

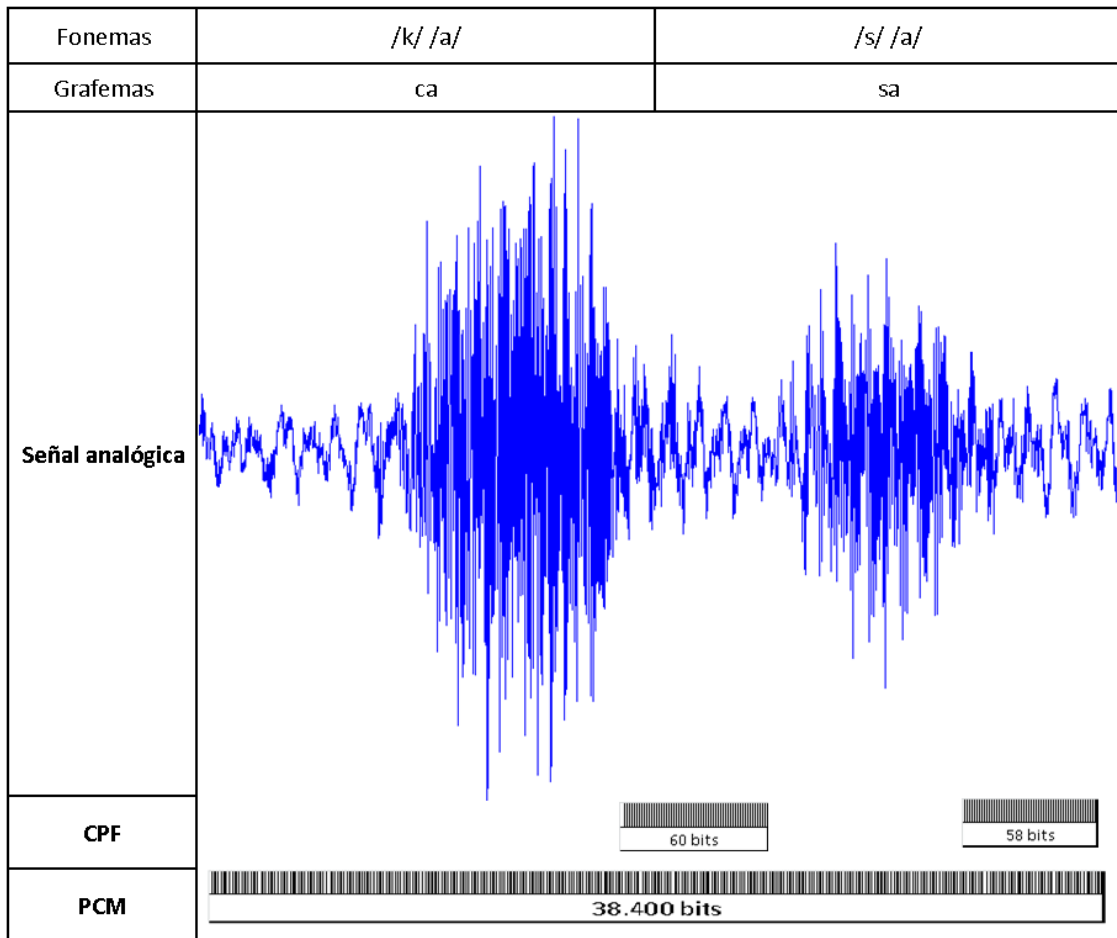


Figura 21: Comparación entre señal analógica, trama PCM y trama CPF.

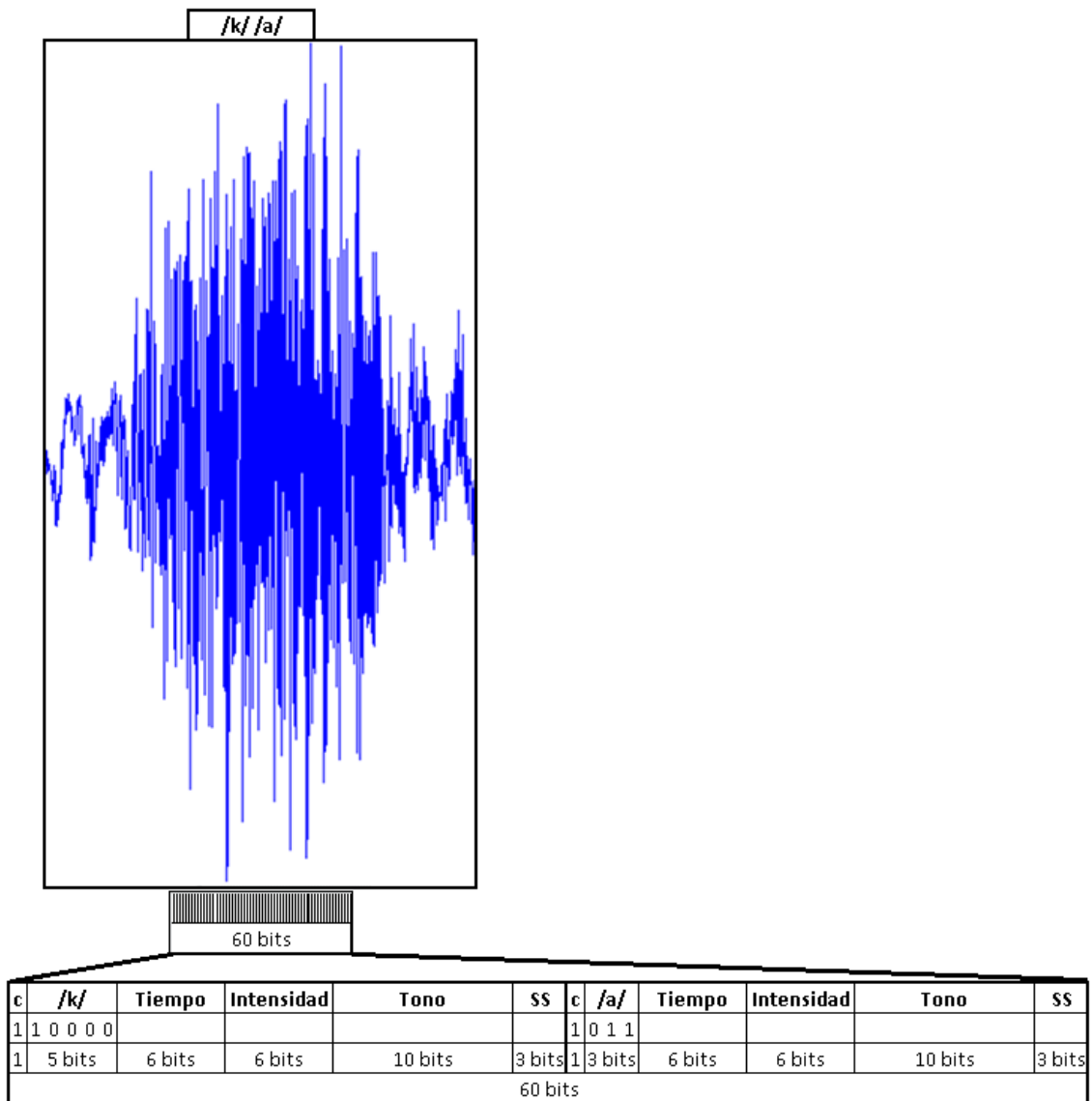


Figura 22: Trama CPF detallada para fonemas /k/ y /a/.

## 4.2.Comparación con PCM.

### 4.2.1. Tamaño de la trama:

Como se dijo en el capítulo III, se realizaron 10 ejecuciones del programa generador de fonemas, los resultados se presentan en la Tabla 11. En esta se ve la cantidad absoluta de bits que miden las tramas en cada estándar, así como la relación de tamaño entre las mismas calculada con la Ecuación 10. El valor promedio obtenido es de 179,19 con una desviación de 0,18. De esta manera es posible afirmar que la trama diseñada es aproximadamente 179 veces más pequeña que la trama PCM.

$$Relación = \frac{\# \text{ bit totales transmitidos Trama PCM}}{\# \text{ bit totales transmitidos Trama CPF}} \quad (10)$$

Tabla 11: Comparación de tramas PCM y CPF.

Tamaño de la trama (bits)			Comparación	
n° Fonemas	PCM	CPF	Relación	Eficiencia
100	533333	2980	178,97	99,441
200	1066667	5954	179,15	99,442
300	1605333	8952	179,33	99,442
400	2133333	11892	179,39	99,443
500	2666667	14895	179,03	99,441
600	3200000	17868	179,09	99,442
700	3733333	20829	179,24	99,442
800	4266667	23764	179,54	99,443
900	4800000	26774	179,28	99,442
1000	5333333	29802	178,96	99,441

En la Figura 23 se ve la representación gráfica de la Tabla 11 para las primeras tres entradas (casos de 100, 200 y 300 fonemas).

Como puede verse, la trama CPF se mantiene en un orden aproximado de 179 veces más corta que la trama PCM. La pequeña variación de estas medidas se debe a que la trama CPF no es de longitud fija por lo que se ve una desviación en los resultados con la generación de cada nuevo fonema. En la Figura 24 se observa esta variación con mejor detalle.

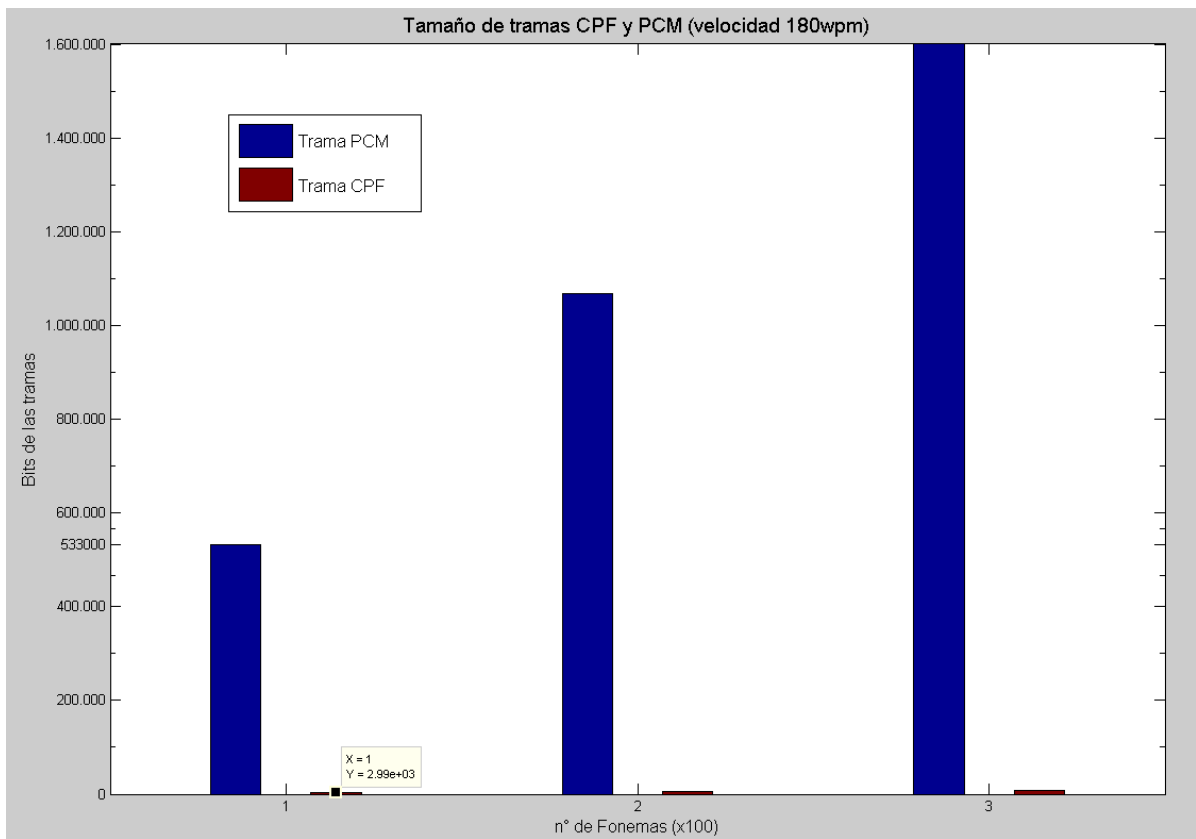
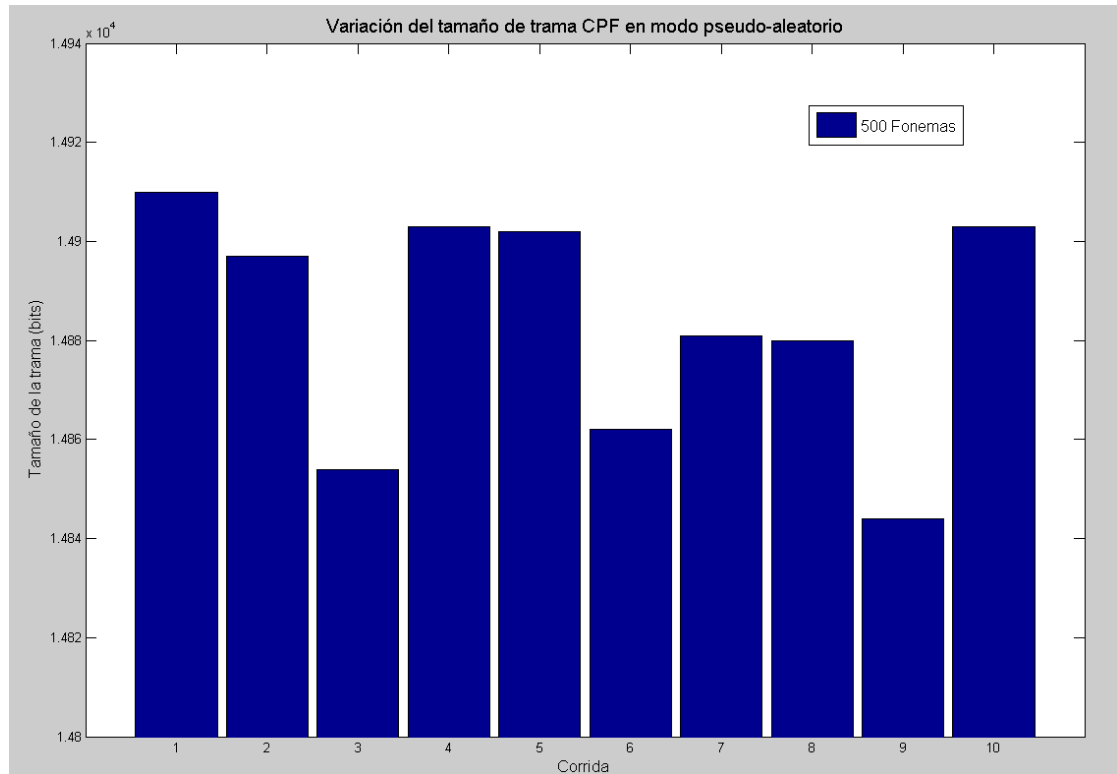


Figura 23: Comparación de tamaños de trama PCM y CPF





**Figura 24: Tamaños de trama CPF en generación aleatoria.**

#### **4.2.2. Velocidad de transmisión:**

La velocidad de transmisión se asume igual en ambas tramas para realizar una comparación adecuada. Esta se tomó asumiendo la velocidad binaria a la que opera PCM, de 64 kbps. Lo que equivale a una duración de bit de  $15,6 \mu\text{s}$ .

#### **4.2.3. Velocidad de información:**

En el caso de la trama PCM, como se comentó en el capítulo III, la velocidad de información es de 64 kbps, igual a su velocidad de transmisión.

Para determinar la velocidad de información en la trama CPF se usó la información de la Tabla 3, donde se plantea una velocidad promedio de producción de fonemas de 12 fpm. Por otro lado, la longitud promedio del código Huffman es aproximadamente 4 bits, calculada con la Ecuación 8, lo que resulta es que el tamaño de la trama CPF promedio por fonema es de 30

bits. Con la Ecuación 11, presentada a continuación, se puede calcular la velocidad de información promedio de la trama CPF.

$$12 \frac{\text{fonemas}}{\text{segundo}} * 30 \frac{\text{bits}}{\text{fonema}} = 360 \frac{\text{bits}}{\text{segundo}} \quad (11)$$

Es importante destacar que, dado que la cantidad de información que se transmite en ese segundo es igual en ambas tramas, es más eficiente la trama cuya velocidad de información sea menor, pues necesita menos bits para transmitir la misma información por unidad de tiempo.

#### **4.2.4. Eficiencia:**

En la Tabla 11 también se puede observar el cálculo de la eficiencia para cada una de las ejecuciones. Este valor fue calculado con la Ecuación 9 (capítulo III) para cada entrada de la tabla. El valor promedio obtenido es de 99,44% con una desviación de 0,0006%. Como se ve en la Figura 25, la eficiencia es casi constante para las diferentes corridas y prácticamente independiente del número de fonemas.

Se puede afirmar entonces que la eficiencia de la trama CPF diseñada es de 99,44% sobre la trama PCM.

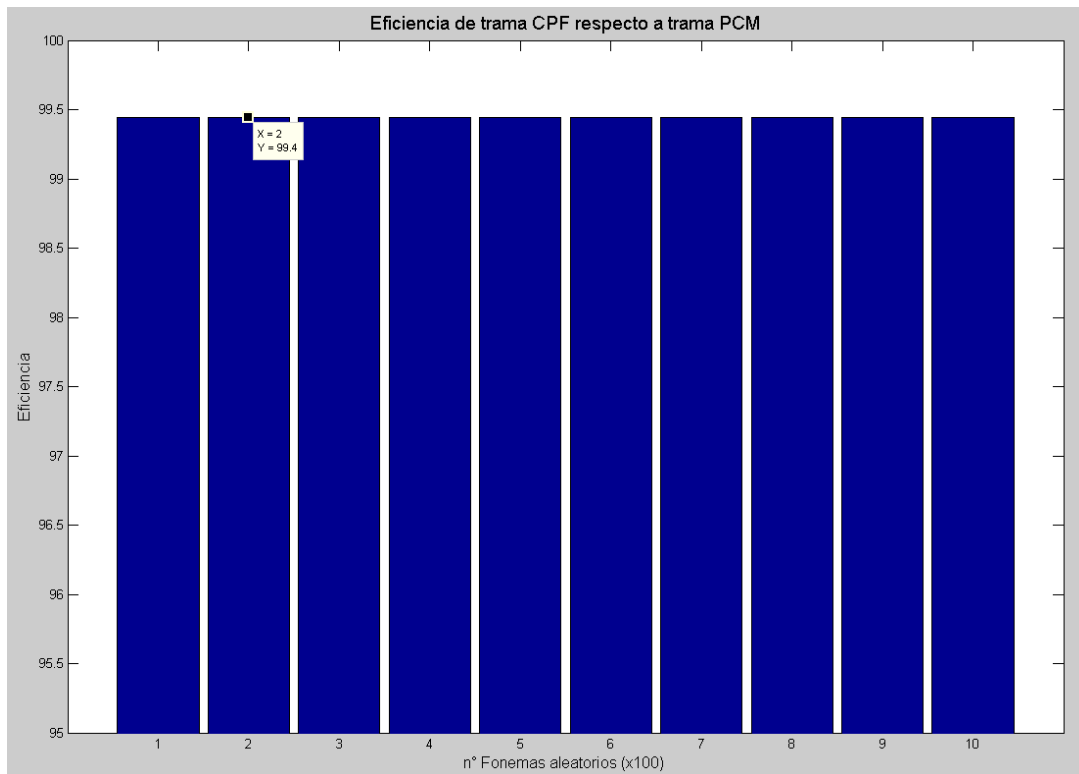


Figura 25: Eficiencia de trama CPF respecto a la trama PCM.

#### 4.2.5. Multiplexado de usuarios:

Si se toma en cuenta un canal de transmisión PCM que tiene velocidad fija de 64 kbps, se deduce que éste trasmite un bit cada 15,6  $\mu$ s, este es el tiempo máximo de duración de bit, ya que se debe tomar esta misma velocidad física de transmisión para la comparación. Y recordando que la longitud promedio de trama CPF (usando la longitud promedio del código Huffman) es de 30 bits por fonema, se obtiene el tiempo durante el cual cada usuario ocupa el canal (Ecuación 12).

$$30 \text{ bit} * 15,6 \mu\text{s} = 469 \mu\text{s} \quad (12)$$

Por otro lado, se sabe (Tabla 3) que la velocidad de habla máxima es de 45 fps. Esto dice que cada usuario debe tener acceso al canal 45 veces cada segundo, durante 469  $\mu$ s. El tiempo entre cada transmisión del mismo usuario es de 22,2 ms (el inverso de 45 fps). En la Figura 26 se puede ver el diagrama que ayuda a entender esta operación, donde se presenta la

ventana de tiempo entre dos transmisiones de un usuario A. El resultado obtenido con la Ecuación 13 es que, es posible introducir 47 usuarios usando una trama CPF.

$$\frac{22,2 \text{ ms}}{469 \mu\text{s}} = 47,3 \quad (13)$$



Figura 26 : Ventana de tiempo de transmisión en la trama CPF.

Si toma en cuenta un esquema de transmisión E1 de PCM, en el cual se agrupan 30 canales de voz (30 usuarios), más 2 canales de señalización, conformando una trama de 32 canales. Cada canal PCM transmite a 64 kbps, la operación anterior es válida y directamente escalable. El resultado es que se pueden introducir 1410 usuarios en un esquema E1 usando la trama CPF a una velocidad de habla máxima, como se muestra en la Ecuación 14.

$$30_{\text{canales}} * 47_{\text{usuarios}} = 1410 \quad (14)$$

### 4.3. Cambios en los parámetros del esquema

#### 4.3.1. Velocidad del habla:

**4.3.1.1. Velocidad de 680 wpm:** esta es la velocidad máxima soportada por el esquema, la modificación de este parámetro afecta directamente en la comparación ya que el usuario habla (genera fonemas) más rápido, y por lo tanto el esquema CPF debe enviar bits más rápido para mantener el ritmo de la conversación. A pesar de eso el esquema PCM no cambia su tasa de transmisión de bits ya que, como se sabe, la información enviada es transparente al estándar.

Se realizó nuevamente la ejecución del programa generador de fonemas para una lista de 100 fonemas y aumentando la lista en 100 fonemas más en cada ejecución, como en el caso anterior, esta vez con una velocidad de habla de 680 wpm. Los resultados obtenidos pueden verse en la Tabla 12.

**Tabla 12: Resultados obtenidos a velocidad máxima.**

<b>Velocidad de 680wpm</b>				
<b>Tamaño de la trama (bits)</b>			<b>Comparación</b>	
<b>n° Fonemas</b>	<b>PCM</b>	<b>CPF</b>	<b>Relación</b>	<b>Eficiencia</b>
100	142222	2969	47,90	97,912
200	285867	5983	47,78	97,907
300	426667	8950	47,67	97,902
400	568889	11869	47,93	97,914
500	711111	14898	47,73	97,905
600	853333	17843	47,82	97,909
700	995556	20841	47,77	97,907
800	1139200	23820	47,83	97,909
900	1280000	26807	47,75	97,906
1000	1423644	29806	47,76	97,906

Los valores de la columna de la trama CPF son aproximadamente iguales a los de la Tabla 11, a velocidad de habla promedio de 180 wpm, ya que el esquema CPF debe transmitir toda la información, independientemente de la velocidad a la que se pronuncia. La columna de trama PCM presenta valores más pequeños a los obtenidos en la Tabla 11, como es de esperar, ya que si el usuario genera los fonemas más rápidamente hace uso del sistema por menos tiempo así que la trama PCM es más pequeña.

La relación entre ambas tramas se ve reducida a un promedio de 47,79 veces más corta la trama CPF comparada con la trama PCM con una desviación de 0,07. Lo que involucra una disminución de la eficiencia bajando hasta un promedio de 97,91% con desviación estándar de 0,003%.

El esquema propuesto sigue siendo considerablemente más eficiente a pesar de la disminución de eficiencia a una velocidad máxima que no debería ser alcanzada en la práctica, y que en teoría, es la mínima eficiencia que se obtendría.

**4.3.1.2. Velocidad de 10 wpm:** esta es la mínima velocidad de habla soportada por el sistema. En caso de un usuario que genera fonemas más largos que 1,4 s el sistema determinará que se trata de dos fonemas diferentes y deberá generar dos tramas distintas con el mismo código de fonema. En este caso limite la eficiencia debe aumentar (comparada con PCM) dada la misma lógica vista anteriormente a pesar de que no se trate de información útil (la pronunciación de un fonema durante más de un segundo y medio es poco práctica).

Los resultados obtenidos se presentan en la Tabla 13. En cuanto a las columnas del tamaño de trama se obtuvo el resultado contrario al caso anterior, la trama PCM es más larga, conteniendo la misma información (trama CPF igual) lo que aumenta la relación calculada entre tramas en un factor de 17 veces mayor que en la Tabla 11 (caso promedio). La eficiencia obtenida en este caso es de 99,97% con desviación estándar de  $5,82 \times 10^{-5}\%$ .

**Tabla 13: Resultados obtenidos a velocidad mínima.**

<b>Velocidad de 10wpm</b>				
<b>Tamaño de la trama (bits)</b>			<b>Comparación</b>	
<b>n° Fonemas</b>	<b>PCM</b>	<b>CPF</b>	<b>Relación</b>	<b>Eficiencia</b>
100	9142857	2978	3070,13	99,967
200	18285714	5960	3068,07	99,967
300	27428571	8943	3067,04	99,967
400	36571429	11870	3081,00	99,968
500	45714286	14917	3064,58	99,967
600	54948571	17911	3067,87	99,967
700	64091429	20924	3063,06	99,967
800	73142857	23834	3068,85	99,967
900	82285714	26791	3071,39	99,967
1000	91428571	29711	3077,26	99,968

#### **4.4. Retardo en transmisión de fonemas:**

Como se explicó el final del capítulo III, el retardo en transmisión de fonemas está relacionado con el tiempo que dura su pronunciación. De la Tabla 3 se obtiene la velocidad de habla más baja permitida dentro de una conversación normal, aproximadamente 9 fps. Con la Ecuación 15 se calculó el tiempo máximo que debe durar este retraso para cada fonema.

$$9^{-1} \frac{\text{fonema}}{s} = 0.111 s \quad (15)$$

El resultado es aproximadamente 111 ms de atraso para cada fonema, a la velocidad de habla más baja.

También es importante hacer este cálculo para el caso más desfavorable soportado por el sistema, correspondiente con un tiempo de pronunciación de un fonema de 1,4 s. En este caso el tiempo de atraso es directamente 1,4 s.

Vale la pena destacar que este retraso no afecta a la comunicación ya que se compensa con el tiempo que el receptor debe tardar reproduciendo el fonema.

## CAPITULO V

### CONCLUSIONES Y RECOMENDACIONES

#### 5.1. Conclusiones

La aplicación de sistemas de telecomunicaciones diseñados especialmente para el transporte de un idioma específico produce una mejor eficiencia en el uso del ancho de banda.

Con el uso del protocolo diseñado, es posible implementar tramas 178 veces más pequeñas que con un esquema PCM, logrando mantener la comunicación, con una eficiencia superior.

La no transmisión de datos durante los momentos de silencio, representa gran parte de la mejora en eficiencia. Haciendo uso del canal solo cuando es necesario, y liberándolo para otros fines cuando el interlocutor no pronuncia sonidos importantes.

El sistema es altamente dependiente de la velocidad de habla de los interlocutores, un hablante más rápido ocupa el canal más veces por unidad de tiempo, lo que afecta la comparación con el estándar PCM. A pesar de ello, a velocidad máxima de habla, la eficiencia se mantiene sobre el 97%.

La gran eficiencia lograda es posible gracias al recorte de información no vital para la comunicación, pero esta información extra puede ser importante para algunos tipos de comunicaciones humanas más familiares. El sistema es idóneo para situaciones donde la comunicación es de mayor importancia que estos aspectos de familiaridad en la voz.

Para lograr comunicaciones en ambientes con niveles extremos de ruido o distancias extremadamente largas (atenuación), es posible hacer uso del ancho de banda ahorrado para introducir métodos de redundancia muy robustos que en otros sistemas podría resultar demasiado ineficientes.



## 5.2. Recomendaciones

Es recomendable la investigación sobre datos menos básicos de la comunicación, que ayuden a identificar al hablante entre otras personas, lo que permitiría la implementación de un sistema más “cálido” en cuanto a comunicación personal, manteniendo un alto nivel de eficiencia. Es posible adaptar la trama para que la información extra sea opcional y ajustable a las necesidades.

Se recomienda evaluar la implementación de campos con información biométrica a la trama y utilizar estándares como MPEG-7, para identificación de los interlocutores. La implementación de estas técnicas, a pesar de alargar la trama, puede representar grandes ventajas, no solo en reproducción de una voz más parecida a la del hablante, sino también para permitir el manejo de conversaciones seguras, decodificada solo para las personas autorizadas.

Se deben tomar en cuenta posibles mejoras en eficiencia, aplicando la trama a estructuras fonéticas más complejas y no a fonemas individuales. De esta forma se podría reducir la cantidad de paquetes enviados. Este caso debe ser estudiado, ya que los nuevos paquetes podrían ser más complejos y compensar la ganancia lograda con el alargamiento de los campos de la trama.

Se recomienda realizar un estudio estadístico formal de frecuencia fonética del español de Venezuela y basar en éste el sistema de comunicación final. Este aporte podría significar el uso de una codificación más precisa y mejorar el nivel de eficiencia.

# MATERIALES DE REFERENCIA CONSULTADOS

## 1. Referencias bibliográficas

- [1] Martínez Celdrán, E. (1996). *El sonido en la comunicación humana*. Barcelona: Octaedro.
- [2] Stallng, W. (2004). *Comunicaciones y redes de computadoras*. (6. Edición, Ed.) Pearson Educación.
- [3] López Martín, A. (2000). *Ingeniería de Ondas Formatos de Audio Digital*. Valladolid: Universidad de Valladolid.[http://www.indiana.edu/~emusic/etext/MIDI/chapter3\\_MIDI3.shtml](http://www.indiana.edu/~emusic/etext/MIDI/chapter3_MIDI3.shtml)
- [4] (2011). Recuperado el 02 de 2017, de Open Sound Control: <http://opensoundcontrol.org/introduction-osc>
- [5] Espinosa, L. (2003). *Compresión de voz para transmisión en redes de datos (Tesis de maestría)*. Universidad Autonoma de Nuevo Leon.
- [6] Monroy, M. (2001). *Codificación de voz en subbandas (Tesis de maestría)*. Universidad Autonoma de Nuevo Leon.
- [7] Gonzales, R., & Mejorada, A. (2008). *Técnicas de grabación y reproducción de sonido (Tesis de grado)*. México D.F.: Instituto Politécnico Nacional.
- [8] Puertas, J. (2000). *Robustez en reconocimiento fonético de voz para aplicaciones telefónicas (Tesis doctoral)*. Madrid: Universidad Politécnica de Madrid.
- [9] Tomasi, W. (2003). *Sistemas de Comunicaciones Electrónicas*. (4ta, Ed.) Phoenix, Arizona: Pearson Educación.
- [10] Abramson, N. (1981). *Teoría de la información y la codificación*. (5ta, Ed.) Madrid: Paraninfo.
- [11] UTI-T. (1988). Recomendacion G.711. En *Libro Azul* (pág. Fascículo III.4).
- [12] Huber, D. M. (2007). *The MIDI Manual: A Practical Guide to MIDI in the Project Studio*. New York: Focal Press.
- [13](2011). Recuperado el 02 de 2017, de Open Sound Control: <http://opensoundcontrol.org/introduction-osc>
- [14] Ferrer, F. (s.f.). Recuperado el 03 de 2017, de Protocol: <https://protocools.wordpress.com/>

- [15] The moving picture expert group. (s.f.). *MPEG*. Recuperado el 02 de 2017, de <http://mpeg.chiariglione.org/>
- [16] International Phonetic Association. (2005). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- [17] Gil, J. (1988). *Los sonidos del lenguaje*. Madrid: Síntesis.
- [18] Schiffman, H. (2001). *La Percepción Sensorial*. (5ta, Ed.) Limusa Wiley.
- [19] Pedro, D. d. (1990). *Teoría completa de la música*. Madrid: Real Musical.
- [20] Grabner, H. (2001). *Teoría general de la música*. Madrid: Ediciones AKAL.
- [21] Gil, J. (2007). *Fonética para profesores de español: De la teoría a la práctica*. Madrid: Arco-Libros.
- [22] Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.
- [23] Carver, R. (1992). Reading rate: Theory, research and practical implications. *Journal of Reading* 36, 84–95.
- [24] Moore, B. (1977). *An Introduction to the Psychology of Hearing*. (6ta, Ed.) Cambridge: Emerald.
- [25] Fastl, H., & Zwicker, E. (2007). *Psychoacoustics, Facts and Models*. Múnich: Springer.
- [26] Arias, F. (2006). *El proyecto de investigación, introducción a la metodología científica*. (5ta ed.). Caracas, Venezuela: Episteme, C.A.
- [27] Cava, V. (julio de 2016). Recuperado el Febrero de 2017, de Contraclave:  
<http://www.contraclave.es/lengua/espanol%20venezuela.pdf>
- [28] Pérez, H. E. (2003). *Frecuencia de fonemas*. Santiago de Chile: Universidad de Concepción.

## 2. Bibliografía

- (2011). Recuperado el 02 de 2017, de Open Sound Control:  
<http://opensoundcontrol.org/introduction-osc>
- Abramson, N. (1981). *Teoría de la información y la codificación*. (5ta, Ed.) Madrid: Paraninfo.
- Alarcos Llorach, E. (1965). *Fonología española*. Madrid: Gredos.
- Arias, F. (2006). *El proyecto de investigación, introducción a la metodología científica*. (5ta ed.) Caracas, Venezuela: Episteme, C.A.
- Canellada, M. J., & Kuhlmann Madsen, J. (1987). *Pronunciación del español: lengua hablada y literaria*. Castilla.
- Carver, R. (1992). Reading rate: Theory, research and practical implications. *Journal of Reading* 36, 84–95.
- Cava, V. (julio de 2016). Recuperado el Febrero de 2017, de Contraclave:  
<http://www.contraclave.es/lengua/espanol%20venezuela.pdf>
- Espinosa, L. (2003). *Compresión de voz para transmisión en redes de datos (Tesis de maestría)*. Universidad Autonoma de Nuevo Leon.
- Fastl, H., & Zwicker, E. (2007). *Psychoacoustics, Facts and Models*. Múnich: Springer.
- Ferrer, F. (s.f.). Recuperado el 03 de 2017, de Protocol: <https://protocools.wordpress.com/>
- Gil, J. (1988). *Los sonidos del lenguaje*. Madrid: Síntesis.
- Gil, J. (2007). *Fonética para profesores de español: De la teoría a la práctica*. Madrid: Arco-Libros.
- Gonzales, R., & Mejorada, A. (2008). *Técnicas de grabación y reproducción de sonido (Tesis de grado)*. México D.F.: Instituto Politécnico Nacional.
- Grabner, H. (2001). *Teoría general de la música*. Madrid: Ediciones AKAL.
- Huber, D. M. (2007). *The MIDI Manual: A Practical Guide to MIDI in the Project Studio*. New York: Focal Press.
- International Phonetic Association. (2005). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.
- López Martín, A. (2000). *Ingeniería de Ondas Formatos de Audio Digital*. Valladolid: Universidad de Valladolid.

- Martínez Celadrán, E. (1996). *El sonido en la comunicación humana*. Barcelona: Octaedro.
- Monroy, M. (2001). *Codificación de voz en subbandas (Tesis de maestría)*. Universidad Autónoma de Nuevo León.
- Moore, B. (1977). *An Introduction to the Psychology of Hearing*. (6ta, Ed.) Cambridge: Emerald.
- Navarro, T. (1946). *Escala de frecuencia de los fonemas españoles*. Syracuse: Estudio de fonología española.
- Navarro, T. (2004). *Manual de pronunciación española*. Madrid: Servicios Editoriales, S.A.
- Pedro, D. d. (1990). *Teoría completa de la música*. Madrid: Real Musical.
- Pérez, H. E. (2003). *Frecuencia de fonemas*. Santiago de Chile: Universidad de Concepción.
- Promotora Española de Lingüística*. (2013). Recuperado el 05 de 2017, de <http://www.proel.org/index.php?pagina=mundo/fonetico>
- Puertas, J. (2000). *Robustez en reconocimiento fonético de voz para aplicaciones telefónicas (Tesis doctoral)*. Madrid: Universidad Politécnica de Madrid.
- Quilis, A., & Fernández, J. (1969). *Curso de fonética y fonología españolas para estudiantes angloamericanos*. Madrid: CSIC.
- Real Academia Española. (2005). *Diccionario panhispánico de dudas*. Madrid.
- Rodero, E. (2012). A comparative analysis of speech rate and perception in radio bulletins. (S. Sarangi, Ed.) *Text & Talk* 32(3), 391-411.
- Schiffman, H. (2001). *La Percepción Sensorial*. (5ta, Ed.) Limusa Wiley.
- Stalling, W. (2004). *Comunicaciones y redes de computadoras*. (6. Edición, Ed.) Pearson Educación.
- The moving picture expert group. (s.f.). *MPEG*. Recuperado el 02 de 2017, de <http://mpeg.chiariglione.org/>
- Tomasi, W. (2003). *Sistemas de Comunicaciones Electrónicas*. (4ta, Ed.) Phoenix, Arizona: Pearson Educación.
- UTI-T. (1988). Recomendación G.711. En *Libro Azul* (pág. Fascículo III.4).
- Utterback, A. (2000). *Broadcast voice handbook*. Chicago: Bonus Book.

### **3. Glosario:**

ISO: Organización internacional de estandarización.

PCM: modulación por pulsos codificados.

AES3: Audio Engineering Society.

MATLAB: software de cálculo matricial.

G.711: recomendación ITU para el estándar PCM.

ITU: Unión internacional de telecomunicaciones.

ITU-T: sector de Normalización de las Telecomunicaciones de la UIT.

CPF: codificación por fonemas.

fpm: fonemas por minuto.

fps: fonemas por segundo.

fph: fonemas por hora.

PUA: potencial umbral de audición.

wpm: palabras por minuto.

wph: palabras por hora.

wps: palabras por segundo.

IPA: Alfabeto fonético internacional.

AFA: Alfabeto fonético americano.

RFE: Revista fonética española.

UDP: protocolo de datagrama de usuario.

TCP: protocolo de control de transmisión.

MSB: bit más significativo.

LSB: bit menos significativo.

FM: frecuencia modulada.