



FACULTAD DE INGENIERÍA Y CIENCIAS AGROPECUARIAS / INGENIERÍA
EN SONIDO Y ACÚSTICA

DISEÑO DE UN SOFTWARE QUE ANALICE EL COMPORTAMIENTO DE
FORMANTES EN UN GRUPO DE SÍLABAS DEL CASTELLANO PARA
REALIZAR UNA REPRESENTACIÓN VISUAL DEL MOVIMIENTO DE LOS
LABIOS, UTILIZANDO REDES NEURONALES ARTIFICIALES.

Trabajo de Titulación presentado en conformidad con los requisitos
establecidos para optar por el título de Ingeniero en Sonido y Acústica

Profesor Guía

Ing. José Francisco Lucio Naranjo, M.Sc.

Autores

Paúl Andrés Ruiz Jara

Diana Vergel Peña

Año

2012

DECLARACIÓN DEL PROFESOR GUÍA

“Declaro haber dirigido este trabajo a través de reuniones periódicas con los estudiantes, orientando sus conocimientos y competencias para un eficiente desarrollo del tema escogido, y dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación.”

.....

José Francisco, Lucio Naranjo
Ingeniero en Sistemas, M. Sc.

170721174-2

DECLARACIÓN DE AUTORÍA DEL ESTUDIANTE

“Declaramos que este trabajo es original, de nuestra autoría, que se han citado las fuentes correspondientes y que en su ejecución se respetaron las disposiciones legales que protegen los derechos de autor vigentes.”

.....

Paúl Andrés, Ruiz Jara
070416002-7

.....

Diana, Vergel Peña
172016056-1

AGRADECIMIENTOS

Dios es el que me ha permitido llegar hasta este punto de mi vida, y le agradezco por haberme dado la oportunidad de estudiar, y por ayudarme a culminar este trabajo.

Diana Vergel Peña

AGRADECIMIENTOS

Agradezco primero a Dios por permitirme estudiar, ya que aún en los momentos más difíciles de mi vida pude salir adelante con su bendición. Agradezco también a mis padres por su incondicional apoyo y por su inmenso amor. Agradezco a mi novia por su incomparable ayuda, ya que sin ella este trabajo no hubiera podido ser una realidad.

Paúl Andrés Ruiz Jara

DEDICATORIA

Dedico esta tesis primero a Dios y a mis padres Nelson Ruiz Córdoba y Mariana Jara Moreira. También la dedico a todas las personas que me apoyaron desde el primer momento y permitieron que este sueño se haga realidad.

Paúl Andrés Ruiz Jara

RESUMEN

El presente trabajo consiste en un estudio de las características de los formantes del habla humana, abarcando los procesos inmersos en dicho fenómeno desde su generación, hasta su grabación y posterior análisis, considerando varias cualidades propias del habla, y generalidades descriptivas de su interpretación. Es por esto que se realizó el diseño de un software que posee la capacidad de analizar las principales características las sílabas del habla hispana latinoamericana, enfocando dicho análisis únicamente en las compuestas por una consonante y una vocal. Lo anterior permite realizar una identificación de la sílaba hablada, presentando como resultado una representación visual del movimiento de los labios, por medio de una animación simple.

Como característica principal del diseño propuesto, se debe mencionar la utilización de una Red Neuronal Artificial, la cual se encargó de asociar las características y patrones de comportamiento espectral de los formantes de cada fonema hablado, constituyéndose como la herramienta de identificación utilizada. Para todo esto, se empleó la herramienta matemática MATLAB, en donde se diseñó el software por medio de la aplicación de funciones del “Artificial Neural Network Toolbox”, y una interfaz gráfica en donde se pueden realizar tareas básicas como cargar archivos, visualizar las animaciones, etc.

ABSTRACT

The present project involves a formants characteristics study of the human speech, covering the processes involved since the beginning, the recording and its further analysis; considering several speech qualities and descriptive characteristics of its interpretation. According to this, this work shows a software design that is capable to analyze the Latin-American-speaking syllables main characteristics, focused only in those who have a consonant-vowel combination. This let to identify the spoken syllable, showing a simple animation where is represented the lips movement.

The main characteristic in the proposed design is the use of an Artificial Neural Network, which was responsible of associate the formants spectral patterns of each spoken phoneme, constituting the identification tool used. Because of this, it was necessary to use the mathematic tool MATLAB, where the software was designed by the "Artificial Neural Network Toolbox" functions application, and a graphic interface where simple tasks can be done.

ÍNDICE

Introducción.....	1
1. Marco teórico.....	6
1.1 La voz humana	6
1.1.1 El aparato fonatorio.....	6
1.1.1.1 Anatomía y funcionamiento del aparato fonatorio.....	7
1.1.2 Características generales de la voz	10
1.2 El lenguaje hablado y sus características	11
1.2.1 Principales conceptos de lenguaje.....	12
1.2.1.1 Fonología y Fonética.	12
1.2.1.2 La entonación	13
1.2.2 Características del lenguaje hablado	14
1.2.2.1 Fonemas y alófonos.....	15
1.2.2.2 Vocales y consonantes	17
1.3 Formantes y sus características en los fonemas.....	24
1.3.1 El tracto vocal	24
1.3.2 Formantes.....	25
1.3.2.1 Características de los formantes.....	25
1.3.2.2 Formantes en las vocales	28
1.4 Conceptos generales de Redes Neuronales Artificiales	30
1.4.1 Elementos de una ANN.	30
1.4.1.1 Unidades de procesamiento.	30
1.4.1.2 Conexiones	31
1.4.1.3 Entradas y salidas.....	32
1.4.1.4 Capas	33

1.4.2. Cálculo computacional.....	35
1.4.3 Propiedades de las ANN.....	37
1.4.4 Entrenamiento.....	37
1.4.4.1 Entrenamiento supervisado.	38
1.4.4.2 Entrenamiento no supervisado.	39
1.5 Software usado en el cálculo de formantes.....	39
1.5.1 Software PRAAT.....	39
1.5.1.2 Cálculo de formantes por medio del algoritmo de burgh.....	40
1.5.2 Grabación de audio en PRAAT.....	41
1.5.3 Formas de visualización de información en PRAAT.....	42
2. Desarrollo experimental.	47
2.1 Estudio de los fonemas.....	47
2.1.1 Definición de los fonemas a analizar.	48
2.1.2 Procedimientos usados para la obtención de información.....	53
2.1.2.1 Interpretación de la información obtenida.	54
2.1.3 Cálculo de diferencias entre formantes.....	65
2.1.4 Análisis de los formantes de las consonantes.	68
2.1.4.1 Oclusivas.....	72
2.1.4.2 Fricativas.....	82
2.1.4.3 Laterales.....	87
2.1.4.4 Vibrantes.....	89
2.1.4.5 Nasales.....	91
2.2. Implementación de las Redes Neuronales Artificiales.....	95
2.2.1 Descripción de las etapas realizadas.....	95
2.2.1.1 Especificación de nomenclatura.....	96

2.2.1.2 Identificación de la información.....	97
2.2.1.3 Arreglo para el reconocimiento de nomenclatura.....	101
2.2.1.5 Topología usada en las RNA.....	107
2.2.1.6 Reducción de datos erróneos de formantes.....	108
2.2.1.7 Entrenamiento de las RNA.....	114
2.2.2 Resultados del entrenamiento de las RNA.	120
2.2.2.1 Resultados de las consonantes vibrantes sonoras [r/, /rr/]	120
2.2.2.2 Resultados de las consonantes nasales sonoras [m/, /n/, /ñ/] .	121
2.2.2.3 Resultados de las consonantes fricativas sordas [s/, /f/, /j/].....	122
2.2.2.4 Resultados de las consonantes oclusivas sordas [p/, /t/, /k/]...	123
2.2.2.5 Resultados de las consonantes africada sorda y lateral sonora [ch/, /l/, /ll/].....	124
2.2.2.6 Resultados de las consonantes oclusivas sonoras [b/, /g/, /d/]...	125
2.2.3 Resultados generales de identificación.....	126
2.2.4 Reconocimiento de las seis Redes Neuronales Artificiales en conjunto	133
2.2.5 Elaboración de la animación de las diferentes aberturas de la boca	140
2.2.5.1 Resultados de la animación de la boca	146
2.2.6 Interfaz gráfica del software diseñado	147
3. Análisis económico.....	151
3.1 Costo de elementos empleados.....	151
3.2 Relación costo-beneficio	152
4. Conclusiones y Recomendaciones	153
Referencias	163

Anexos 165

Introducción

El habla humana es un proceso complejo en donde intervienen distintos componentes del cuerpo humano, cada uno de los cuales cumple una función específica para la caracterización final del sonido propio de cada persona, denominado como voz. Una tarea anhelada por la humanidad en base al desarrollo tecnológico que se ha obtenido en las últimas décadas, es poder relacionar de manera directa al habla humana con máquinas o computadoras. Este objetivo es mucho más complejo de lo que parece, ya que existen muchas variables que son poco previsibles en el proceso del habla y dificultan la obtención de resultados satisfactorios.

El análisis de voz es un método muy usado en diversos ámbitos profesionales y experimentales en la actualidad. En ese sentido existe una gran variedad de aplicaciones que pueden, entre otras cosas, identificar a una persona por su voz, reconocer comandos simples, etc. Sin embargo, es evidente que es difícil encontrar aplicaciones que tengan como finalidad analizar aspectos característicos y determinantes de los fonemas del habla hispana, de manera que sea posible identificar patrones de reconocimiento que se puedan relacionar con animaciones, las cuales tengan concordancia con señales de audio que contienen voz humana.

Actualmente existen varios métodos de síntesis sonora que abarcan modelamiento de formantes del habla humana. Varios de estos han tenido resultados muy realistas en base a la sonoridad obtenida, sin embargo aún existen deficiencias en los procesos, propias de un fenómeno extremadamente complejo. Estableciendo un orden contrario al de la síntesis sonora, en el presente trabajo se pretende realizar un estudio de las características de los formantes presentes en la voz humana, para poder reconocer la sílaba hablada y emitir una animación simple que represente de mejor manera el movimiento de los labios al momento de hablar.

Antecedentes

La representación del habla de una imagen humana a partir de la información de un archivo de audio, es una acción que no se ha desarrollado extensivamente dentro de los diferentes ámbitos audiovisuales de hoy en día. Tal vez la mayor implementación de este tipo en el campo audiovisual, es que una imagen humana, ya sea de animación o caricatura, realice un movimiento simple comprendido por dos acciones básicas como abrir y cerrar la boca. Es por esto, que desde un punto de vista no solo innovador sino también funcional, sería muy favorable para el desarrollo tecnológico, que los movimientos de la boca de una imagen digital humana sean representados realísticamente por el hecho de asociar una señal de audio de voz hablada con dicha imagen. Esta representación se puede generar por medio de una serie de procesos entrelazados, como el análisis espectral de la voz, enfocado primordialmente en el estudio de los formantes producidos en cada sílaba pronunciada, para producir así una secuencia de imágenes que representen el movimiento que se genera en la boca al momento en que un humano habla.

Objetivos

Objetivos generales

- Realizar un estudio de las características espectrales que poseen los fonemas del castellano, enfocados en el habla de Ecuador. Dicho estudio comprenderá el análisis de aspectos de generación del fonema por medio del aparato fonatorio, y principalmente se enfocará en los formantes originados de cada fonema hablado, abarcando su distribución en el tiempo así como su variación al momento en que se articulan dos fonemas, formando una sílaba bisílaba.
- Diseñar un software que esté en capacidad de comparar una información de entrada compuesta por formantes, relacionando sus características con un resultado que permita identificar la sílaba que generó tales formantes. Para esto se hará uso de una Red Neuronal Artificial (RNA) que, en base a un entrenamiento previo, al ser excitada por los patrones presentes en cada fonema del castellano, arroje una

animación simple del movimiento de los labios simulando el proceso del habla en concordancia con la sílaba identificada.

Objetivos específicos

- Hacer un estudio específico de los fonemas del castellano en base a su clasificación, y lograr determinar la influencia de los componentes analizados en los procesos necesarios para su identificación.
- Obtener los valores de formantes de cada fonema considerado y realizar un estudio de su caracterización, logrando identificar los patrones de comportamiento de cada uno.
- Realizar una matriz con las combinaciones reales posibles entre consonante y vocal (CV) pertenecientes al idioma castellano. Para esto se descartarán aquellas letras que comprendan un mismo fonema, se eliminarán los elementos que no sean utilizados en el habla ecuatoriana y se retirarán aquellas sílabas que no sean aceptadas por el idioma castellano debido a su estructuración.
- Realizar un análisis visual detallado del movimiento de la boca al pronunciar cada una de las sílabas pertenecientes a la matriz mencionada anteriormente, de manera de poder usar dicha información para identificar la similitud visual en el movimiento de los labios, y simplificar así la cantidad de animaciones.

Hipótesis

La hipótesis que se consideró en el presente proyecto se basa en poder obtener una relación entre una señal de audio generada por una voz humana, y una animación que represente el movimiento de los labios acorde a la sílaba hablada.

Esta hipótesis plantea lo siguiente: los formantes son suficientes para obtener un reconocimiento tal de los fonemas del castellano, que sirva para representar una simulación de la abertura de la boca al hablar.

Esto se basa principalmente en la relación existente entre los fonemas y los formantes producidos en ellos, ya que la voz humana posee una caracterización acústica distinta en función de algunos factores; de entre los cuales, se centrará únicamente en los formantes.

Alcance

En el actual proyecto se presenta una parte del diseño de un software que sea capaz de reconocer patrones en el comportamiento de formantes de sílabas compuestas únicamente por una consonante y una vocal del castellano; y que esté en la capacidad de emitir como resultado una animación de la labialización correspondiente a la pronunciación de la sílaba hablada. Cabe recalcar que este diseño comprende la etapa inicial del desarrollo de una posible aplicación informática de reconocimiento del habla, ya que se omitieron casos de habla continua y palabras interconectadas, debido a complejidades que se presentan por distintas diferencias tonales en pronunciación del hablante, así como la variación de la velocidad del habla, coarticulación entre sílabas y palabras, etc.

El diseño propuesto abarca únicamente la descripción de las etapas necesarias para lograr los objetivos planteados, así como los procesos realizados en cada una de estas.

En el presente trabajo se usaron a seis personas modelos (comprendidas por tres hombres y tres mujeres) para la grabación del conjunto de sílabas. Dichas grabaciones fueron almacenadas y analizadas, obteniendo los formantes de cada sílaba y su comportamiento.

Debido a que uno de los objetivos principales del presente estudio es desarrollar una aplicación que muestre una animación, la cual reproduzca el movimiento de labios correspondiente a una sílaba almacenada en un archivo de audio; se debe mencionar que, en base a la existencia de varios fonemas del castellano que poseen movimientos de labios similares, la identificación puede confundir fonemas. Sin embargo, ya que únicamente se persiguen fines netamente visuales, se minimizaron detalles complejos de reconocimiento que no conllevaban a diferencias en la representación visual.

Justificación

Este proyecto constituye un trabajo investigativo con fines experimentales. Sin embargo, los resultados obtenidos podrían servir como punto de partida para desarrollo de aplicaciones que persigan fines similares. Actualmente, se han hecho distintos estudios en varias universidades, los cuales han conllevado al desarrollo de aplicaciones capaces de determinar parámetros importantes de la voz humana, y hasta ahora se siguen investigando otros aspectos relacionados. Por lo tanto, el software propuesto y el estudio realizado de los fonemas del habla y sus formantes, pretende contribuir con algunos de los parámetros que faciliten la investigación sobre el comportamiento del aparato fonatorio relacionado al sonido fonético que producen los seres humanos.

Se espera que con este proyecto se pueda marcar tendencias innovadoras en cuanto a eventos sonoros relacionados con movimientos de distintos tipos, de modo que en un futuro se pueda desarrollar una herramienta informática que permita integrar de manera más real imágenes o animaciones con señales de audio, en los ámbitos de entretenimiento y comunicación audiovisual a distancia, principalmente. Se debe considerar la necesidad de una futura etapa de desarrollo, para que el proyecto se pueda consolidar.

La principal aplicación del software es que una representación gráfica humana hable en concordancia con una señal de audio que se esté emitiendo. En primera instancia el diseño del software abarca la implementación de animaciones simples; sin embargo, para etapas posteriores se podrían aprovechar tecnologías de animación avanzadas para usar imágenes de distintas personas y emular de manera más real los movimientos bucales.

Capítulo I.- Marco teórico

1.1.- La voz humana

La voz humana es el conjunto de sonidos que se obtiene al realizar el proceso del habla, en el cual intervienen varios elementos del cuerpo humano. A partir del estudio de la voz, se han generado varios inventos que han logrado un gran impacto en el desarrollo de la humanidad, tales como: la telefonía, en el ámbito de telecomunicaciones; softwares de reconocimiento de voz, en el ámbito de seguridad, plugins de procesamiento de voz, en el ámbito de producción musical, etc.

En Rosero (2009, p.102) se define a la voz humana como “un sonido complejo formado por una frecuencia fundamental y un gran número de armónicos y sobretonos”; además, se menciona que la frecuencia fundamental está directamente relacionada con las características de las cuerdas vocales de cada persona.

El conjunto de sonidos que conforman la voz provienen del denominado aparato fonatorio. Cada individuo posee características totalmente diferentes en su cuerpo con relación a otros seres humanos; y la voz no es la excepción.

Toda persona posee una voz específica, determinada en términos sonoros por su altura (relacionada con la frecuencia fundamental), por su timbre (relacionado con los armónicos y con los sobretonos) y por su intensidad (Rosero, 2009). Sin embargo, los factores físicos que influyen en la característica tonal y frecuencial de la voz son muchos, generados por un sistema mecánico-acústico muy complejo.

1.1.1 El aparato fonatorio

En términos generales, según las leyes de la Acústica, todo sonido es generado mediante la participación de tres elementos, los cuales son: un cuerpo que entre en vibración, un medio de propagación elástico, y una caja de resonancia (Rosero, 2009). En el caso de la voz, estos tres elementos se encuentran presentes en el cuerpo humano en el *aparato fonatorio*. El aparato fonatorio es un conjunto de elementos (órganos) que cumplen una función

específica en la generación del habla, los cuales son: las cuerdas vocales, en las cuales se produce la vibración; el flujo de aire emitido desde los pulmones, que es el medio de propagación del sonido; y la cavidad torácica, que es la caja de resonancia.

Además, también se encuentran otros elementos que son muy importantes en este proceso, como son: la faringe; la glotis; las cavidades oral y nasal; y los elementos articulatorios, tales como: los labios, los dientes, el alvéolo, el paladar, el velo del paladar, y la lengua (Miyara, 2004).

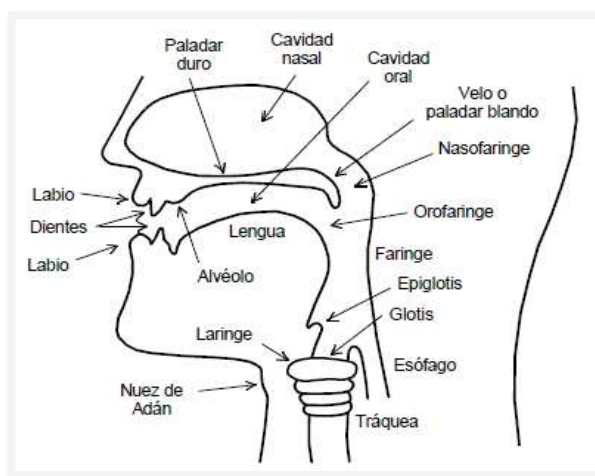


Figura 1.

Corte esquemático del aparato fonatorio. Tomado de Miyara, 2004, p. 4.

1.1.1.1 Anatomía y funcionamiento del aparato fonatorio

El primer elemento involucrado en la generación del habla, son los pulmones, ya que es en ellos en donde se genera la energía necesaria para que el proceso acústico empiece, debido a que sin un elemento elástico (en este caso, el aire) no se puede generar un sonido.

El segundo elemento del proceso, pero no menos importante, son las cuerdas vocales; las cuales son membranas que se encuentran dentro de la laringe. (Miyara, 2004). Existen cuatro cuerdas vocales, las cuales se dividen en dos superiores, que no participan en la articulación de la voz; y dos inferiores, que

son las responsables de la generación de la voz. Las cuerdas vocales están unidas al cartílago tiroides por delante, y a los cartílagos aritenoides por detrás. La abertura presente entre ambas cuerdas, se denomina glotis.

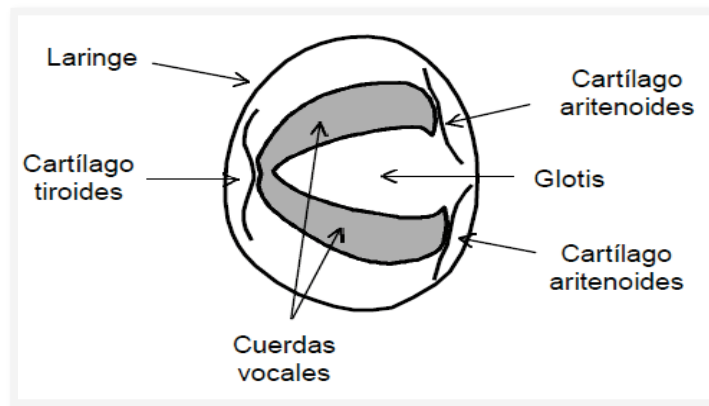


Figura 2.

Corte esquemático de la laringe, según un plano horizontal.

Tomado de Miyara, 2004, p. 5.

Según (Miyara, 2004), y como se puede observar en la figura anterior, cuando las cuerdas vocales se encuentran separadas, la glotis tiene una forma triangular. En esta condición, el aire pasa libremente y no se produce sonido alguno. Cuando la glotis comienza a cerrarse, el aire que proviene de los pulmones comienza a experimentar una ligera obstrucción, emitiéndose un ruido conocido como aspiración. Esto sucede en los sonidos denominados “aspirados” (como la /j/ del castellano). A medida que la glotis se sigue cerrando, las cuerdas vocales adoptan una vibración a modo de lengüetas, produciéndose un sonido tonal o periódico. La frecuencia de este sonido depende de algunos factores, como: la masa y el tamaño de las cuerdas vocales, la tensión que estas posean al momento de la vibración, y la velocidad del flujo del aire que proviene de los pulmones.

Las cuerdas vocales pueden tensarse para producir sonidos agudos, o distenderse para producir sonidos graves. Es decir, cuando la tensión en las cuerdas vocales aumenta, la frecuencia también aumenta. Además de todo esto, si la glotis es de mayor tamaño, la vibración es de menor frecuencia. Esta

última condición, entre otras, caracteriza el tono de voz que poseen hombres y mujeres, ya que en los hombres el tamaño promedio de la glotis es mayor, y por lo tanto su voz es más grave. Además de todo lo anterior, cuando la glotis es obturada completamente, no se produce sonido alguno.

Encima de la glotis se encuentra un cartílago en la faringe que se denomina epiglotis, el cual permite tapar la glotis cuando se ingiere alimento, para evitar que este se introduzca en el tracto respiratorio. Durante la respiración y la fonación (emisión de sonido), la epiglotis se encuentra separada de la glotis, permitiendo así la circulación normal del flujo de aire.

La parte del aparato fonatorio que incluye las cavidades faríngea, oral y nasal, junto con los elementos articulatorios, se denomina cavidad supraglótica. En cambio, la tráquea, los bronquios y los pulmones, se denominan cavidades infraglóticas, ya que se encuentran por debajo de la laringe.

La mayoría de elementos de la cavidad supraglótica, se pueden controlar a voluntad, de manera que se modifiquen los sonidos producidos por las cuerdas vocales; esto es lo que normalmente realiza un ser humano al momento de hablar.

Enfocando al proceso del habla de una manera técnica, se puede mencionar que está compuesto por varios procesos (Miyara, 2004), como son:

- La *emisión*. Esta etapa abarca todos los procesos necesarios para que el sonido sea generado.
- El *moldeamiento* o *filtrado*. Es el proceso en el que se modifica el espectro del sonido. Esta etapa tiene lugar en las cuatro cavidades supraglóticas principales: la faringe, la cavidad nasal, la cavidad oral y la cavidad labial. Además, es propia de cada persona pero maneja tendencias generalizadas entre hombres, y entre mujeres. Las cavidades supraglóticas cumplen la función de ser resonadores acústicos, los cuales enfatizan determinadas bandas de frecuencia del espectro generado por las cuerdas vocales.

- La *articulación*. Es un proceso en donde se modifican los sonidos a nivel temporal. Dicha modificación se encuentra relacionada con la emisión de los mismos y está caracterizada por el lugar en el tracto vocal en donde se produce, los elementos que intervienen, y el modo en que se produce.

En síntesis, se puede concluir que el sonido producido en las cuerdas vocales necesita ser amplificado para poder ser audible, y este proceso lo realizan los resonadores. Luego de esto, el sonido es moldeado por los elementos de las cavidades supraglóticas, en donde destacan: el paladar, la lengua, los dientes, los labios, entre otros; y finalmente es articulado por medio de alguna obstrucción hacia el flujo de aire.

1.1.2 Características generales de la voz

La voz es una de las expresiones humanas en donde más se pone de manifiesto las características propias de cada individuo, las cuales abarcan tanto a las anatómicas, como a las anímicas.

Las características sonoras de la voz están directamente relacionadas con distintas propiedades acústicas, como son: el tono, el timbre, y la intensidad. Estas propiedades sirven para establecer diferencias sonora entre cada voz. A continuación se presenta una pequeña descripción:

- La *altura* de una voz está relacionada con la percepción del tono, por lo que depende de la frecuencia.
- La *intensidad* depende de la amplitud que alcancen las vibraciones, esto depende netamente de lo que permita el aparato fonatorio de cada persona.
- El *timbre* concreto de cada voz está relacionado con los armónicos que acompañan a su frecuencia fundamental.

La voz humana tiene un ancho de banda, que generalmente oscila entre los 80 y 1000 Hz, aunque su eficiencia es mayor entre los 200 y 700 Hz (Rosero, 2009). Cabe recalcar que estos valores corresponden únicamente a las frecuencias fundamentales, y no consideran armónicos ni sobretonos.

1.2.- El lenguaje hablado y sus características

El habla humana ha sido desde la antigüedad el medio de comunicación entre seres humanos usado por excelencia. El habla se puede definir como el acto de seleccionar signos lingüísticos propios del lenguaje, y organizarlos a través de ciertas reglas (Miyara, 2004).

Debido a la necesidad de comunicación entre seres humanos existente desde la antigüedad, fue necesario desarrollar un método por el cual los sonidos producidos puedan transmitir un mensaje. De aquí es donde nació el concepto de *lenguaje*, el cuál es definido según la Real Academia Española (RAE) como “conjunto de sonidos articulados con que el hombre manifiesta lo que piensa o siente”.

Por consiguiente, se puede definir al lenguaje como el medio o método que utiliza el ser humano para comunicarse entre sí; además, se debe mencionar que está compuesto por dos partes generales (Frías, 2001): una parte de emisión y articulación (denominada ‘Fonética’); y otra de percepción o interpretación (denominada ‘Fonología’).

Debido a que el lenguaje está compuesto por sonidos, como ya se mencionó, es necesario poder clasificar a dichos sonidos según sus características, de aquí es donde se manifiesta el concepto de vocales y consonantes; sin embargo, es necesario primero definir a la unidad fónica ideal mínima del lenguaje, la cual se denomina *fonema* (Miyara, 2004). El fonema es el descriptor de la percepción de los sonidos. La realización de un fonema es un sonido, el cual es representado entre barras; por ejemplo /t/ representa al sonido realizado al pronunciar la letra ‘t’.

Los fonemas en el castellano son dos: las vocales y las consonantes. Usando estos dos fonemas y combinaciones entre ellos se generan sílabas, y a su vez las combinaciones de una o más de estas, generan palabras; las cuales indican un mensaje que transmite información entre seres humanos. La sílaba es una emisión de voz, y según la psicolingüística, es la unidad menor que percibe el

oído humano; de esta forma se establece que lo que percibe el ser humano son sílabas, no fonemas.

La necesidad de la humanidad de poder transmitir sus ideas, la llevó a generar maneras de describir o especificar el lenguaje de una forma 'tangible'. Es por esto que el lenguaje, como medio de comunicación, también posee una parte escrita. Normalmente todo humano en su niñez aprende primero a hablar que a escribir, sin embargo, hoy en día en muchas sociedades es considerado primero el lenguaje escrito; y al lenguaje hablado se lo suele considerar como una forma de representar al escrito, es decir, en un segundo plano de importancia.

Esto se debe, muy probablemente, a que en la actualidad el lenguaje hablado posee muchas diferencias entre personas de distintos continentes, regiones, países, y hasta ciudades de un mismo país; por lo que se suele llegar a pensar que es un método con muchas deficiencias (Holmes, 2001).

Enfocando el proceso del habla desde un punto de vista comunicativo, se puede establecer a dos sujetos inmersos en él, los cuales son: el *emisor* y el *receptor* (Miyara, 2004).

El emisor es el que "piensa" el mensaje y lo traduce en emisión acústica por medio del aparato fonatorio. El receptor es el que "recibe" el mensaje, y por medio del aparato auditivo capta las ondas sonoras y las transforma en impulsos nerviosos, los cuales son después interpretados por el cerebro; cumpliéndose así el ciclo básico que debe tener un mensaje para poder transmitir alguna información.

1.2.1 Principales conceptos de lenguaje

1.2.1.1 Fonología y Fonética

La Fonología y Fonética son dos ramas de la lingüística que se encargan de estudiar a los sonidos presentes en el lenguaje hablado. Ambas abarcan aspectos diferentes del habla, y son complementarias entre sí (Frías, 2001).

La Fonología abarca el estudio de todos los sonidos del lenguaje en cuanto a su carácter simbólico o de representación mental. Procede detectando regularidades o recurrencias en dichos sonidos y sus combinaciones, y haciendo abstracción de las pequeñas diferencias debidas a la individualidad de cada hablante y de características suprasegmentales, como lo son: la entonación, el acento (que puede ser tónico, por aumento de la intensidad; y agógico, por aumento de la duración), entre otros. Cada uno de los sonidos abstractos así identificados, corresponde a un fonema. Uno de los objetivos de la fonología, es acotar al máximo la cantidad de fonemas requeridos para representar cada idioma de una manera suficientemente precisa.

La Fonética, en tanto, se refiere a los sonidos en el habla, incluyendo su producción acústica y los procesos físicos y fisiológicos de emisión y articulación involucrados (Miyara, 2004).

Esta ciencia se encarga de estudiar experimentalmente los mecanismos de producción y percepción de los sonidos utilizados en el habla, a través del análisis acústico, articulatorio y perceptivo. Se ocupa, por consiguiente, de las realizaciones de los fonemas. Dicho de otra manera, la fonética se encarga del estudio de la sonoridad que se emite por la pronunciación de palabras, y dicha emisión sonora es conocida como *fonos*, los cuales son más numerosos que los fonemas (Frías, 2001).

1.2.1.2 La entonación

En el habla cotidiana de los seres humanos, es muy común encontrar distintas variaciones en la pronunciación que realizan las personas al momento de comunicar alguna idea. La *entonación* se puede definir como el conjunto de variaciones en la pronunciación de una o varias palabras (Frías, 2001).

Al hablar, el tono de voz generalmente no es constante, ya que posee aumentos o reducciones, para facilitar la transmisión de información. Se puede decir que existen tres distintas clasificaciones que involucran diferentes entonaciones: el enunciado, la pregunta, y la exclamación.

En el lenguaje escrito, estas clasificaciones se expresan por medio de signos gráficos, sin embargo en el lenguaje oral se expresan con inflexiones de voz o cambios de tono.

El tonema

El *tonema* es la unidad de medida de la entonación. Cada subida, bajada o mantenimiento del tono, es un tonema. Aunque algunos autores difieren en su clasificación, generalmente se pueden especificar a tres (Frías, 2001):

- Ascendente (↑), la variación del tono se hace en base al aumento de frecuencia.
- Horizontal (→), no existe variación de tono; se tiene la percepción de frecuencia constante.
- Descendente (↓), se manifiesta por inflexiones de voz; la frecuencia baja.

Grupo fónico

En términos generales, el habla cotidiana siempre se realiza con pausas. Dichas pausas se pueden clasificar en dos tipos: cortas y largas (Frías, 2001).

Las pausas cortas se usan para separar cláusulas o sintagmas del resto de un enunciado. Las pausas largas se usan para respirar y separar oraciones. Cada una de estas, sea corta o larga, se usan para separar *grupos fónicos*. Un grupo fónico se puede definir como el conjunto de palabras dentro de una frase u oración, que poseen una relación directa entre ellas, para la obtención de la idea final que se desea transmitir; por ejemplo, se tiene la siguiente oración: “(La estación del tren) (se encuentra cerrada) (desde ayer)”. En este ejemplo, cada grupo fónico se encuentra separado por paréntesis, y como se mencionó anteriormente, todos ellos conforman la idea general de la oración.

1.2.2 Características del lenguaje hablado

A medida que el ser humano iba comprendiendo la complejidad de representar sus ideas por medio del habla, y posteriormente de la escritura; se fueron diseñando modelos descriptivos que facilitarían la comprensión y utilización del lenguaje. Es por esto que nacieron los conceptos gramaticales de letras,

sílabas, palabras, frases, oraciones, abecedarios, entre otros; sin embargo, debido a la diferencia de culturas y regiones, se fueron desarrollando distintos lenguajes, lo que derivó en la inclusión del término *idioma*, el cual se refiere a la forma que posee cierto sector geográfico para expresar un determinado lenguaje. Un mismo lenguaje también posee distintas modalidades, dependientes de las diferentes regiones en donde se desarrolle el habla de dicho lenguaje, las cuales son llamadas *dialectos*.

1.2.2.1 Fonemas y alófonos

Como se mencionó anteriormente, los fonemas son la unidad mínima del lenguaje. A diferencia de las letras del abecedario, las cuales son únicas, ya que se diferencian una de otra por su símbolo; existen fonemas que pueden representar a varias letras, o dicho de otras palabras, un mismo sonido (fonema) puede ser usado al pronunciar letras que se escriben diferente. Este es el caso de la letra “k” y la letra “q”, ya que ambas corresponden al mismo fonema /k/.

En el abecedario del castellano existen 29 letras, y aunque el número de fonemas también es de 29; no se cumple el posible pensamiento de que a cada letra le corresponde un fonema, o viceversa. Es aquí donde se incluye el concepto de *alófono*. Un alófono es una variación de un fonema específico (Miyara, 2004), la cual por lo general depende de la ubicación del fonema dentro de la sílaba o palabra pronunciada. Por ejemplo, la letra “n” (representada por el fonema /n/) posee un alófono, el cual se produce cuando dicho fonema se encuentra al inicio, o al final de una sílaba. Por ejemplo, la palabra “nota” posee un fonema /n/ inicial; en cambio la palabra “planta” posee un fonema /ŋ/ que se encuentra luego de una vocal, por lo que se denomina “postvocálico”, y su sonido es diferente al de la palabra anterior.

Existen algunas variaciones en las propiedades acústicas de ciertos alófonos, representando algunos fonemas específicos. Estas diferencias pueden deberse a influencias del sector geográfico en donde la persona halla crecido o se encuentre viviendo, como es el caso del fonema /s/ en el Ecuador, referente a personas de la región litoral en contraste con personas de la región interandina.

Este alófono se presenta principalmente en el caso en que la letra “s” se encuentre antes de otra consonante. Por ejemplo, dos personas de las regiones mencionadas no pronuncian de igual manera la palabra “costa”.

Variaciones como la citada anteriormente se producen con mucha frecuencia en distintos países y sectores a nivel mundial. En el caso del idioma inglés, por citar un ejemplo, existen muchas diferencias en la pronunciación de una misma vocal (fenómeno que no se produce en el castellano), dependiendo de la palabra que se quiera decir. Un ejemplo de esto es el caso de las palabras “dole” y “doll”, las cuales se diferencian únicamente por la sonoridad de la vocal intermedia; ejemplos como el anterior existen muchos a nivel mundial, y en distintos idiomas. Con respecto al castellano, en las vocales solo existen dos alófonos correspondientes a las denominadas vocales cerradas, es decir la vocal “i” y la vocal “u”. Estos alófonos, si bien no son realizados con mucha frecuencia por la población latinoamericana, se producen debido a la unión de dos vocales seguidas, la cual es conocida como *diptongo*. Un ejemplo de esto es la pronunciación de la palabra “hueso” (en donde la vocal “u” corresponde al fonema /w/).

A diferencia del castellano propio de España, en Ecuador (y en la mayoría de países de Latinoamérica) no se pronuncia el fonema correspondiente a la letra “z”; además, en la mayoría de la población ecuatoriana no se diferencia el fonema /λ/ correspondiente a la letra “ll” (por ejemplo de la palabra ‘lluvia’), del fonema /dʒ/ correspondiente a la letra “y” (por ejemplo de la palabra ‘yate’). Caso contrario sucede en el caso de las letras “b” y “v”, ya que a diferencia de lo que se pudiera llegar a creer, ambas se pronuncian de igual manera (Miyara, 2004); sin embargo, cada una de ellas poseen dos fonemas diferentes dependiendo de su ubicación dentro de una palabra. Un ejemplo de esto corresponde a las palabras “base” y “labor” (en donde la consonante “b” posee una pronunciación ligeramente distinta dependiendo de su ubicación).

A continuación, en la tabla 1 se muestran todos los fonemas presentes en el castellano; añadiendo además algunos ejemplos por cada caso especificado. Se debe recalcar que todos los símbolos fonéticos presentados en la siguiente

tabla, pertenecen al denominado *Alfabeto Fonético Internacional* (IPA, por sus siglas en inglés), los cuales permiten representar de una manera inequívoca a los fonemas, de manera independiente al idioma que correspondan.

En este caso se presentan los que corresponden únicamente al castellano:

Tabla 1.

Fonemas utilizados en el castellano.

Fonemas castellanos					
Sonido	Ejemplo	Sonido	Ejemplo	Sonido	Ejemplo
[p]	p aso	[θ]	zor z al, láp iz	[ɲ]	ma ñ ana, ñ o ñ o
[b]	b ase, v ena	[s]	s olo, c osa	[dʒ]	y o, Yape y ú
[β]	lab or , lav ar	[x]	g iro, jar ab e	[j]	bien, bi ó logo
[t]	t res, cant o	[tʃ]	he ch o, Ch ubut	[w]	h u eso, bu it re
[d]	d ama, and ar	[r]	ard er , jar ab e		
[ð]	ced r o, verd ad	[rr]	per r o, ro j o	[a]	ca m a
[k]	c aso, disc o	[l]	lo ab le, fiel	[e]	es per a, ver
[g]	g ula, g oma	[λ]	ll an to, call e	[i]	vine, ir is
[ɣ]	ag ua , neg ro	[m]	ma m á, ámb ar	[o]	lor o , pos
[f]	fin o , tif ón	[n]	n en e, jov en	[u]	burl a , hurac án

Tomado de Miyara, 2004, p.9.

1.2.2.2 Vocales y consonantes

El concepto de letras proviene de la inclusión de los denominados *signos lingüísticos*. Los signos lingüísticos forman parte de lenguaje y poseen dos clasificaciones: *significado* y *significante* (Miyara, 2004). El significado es el concepto que se desea comunicar, mientras que el significante es la imagen que se da al significado, la cual puede ser: gráfica (i.e. letras), o acústica (i.e. fonemas).

Desde un punto de vista mecánico-acústico, (Miyara, 2004) establece que las vocales son aquellos sonidos generados por la vibración de las cuerdas vocales, cuyo flujo de aire no posee ningún obstáculo u obstrucción a lo largo de todo el aparato fonatorio; en cambio, las consonantes son sonidos que pueden ser o no generados por la vibración de las cuerdas vocales, pero que sí poseen obstrucción en su flujo de aire.

En el castellano, a diferencia de otros idiomas, las vocales sí pueden constituir palabras completas; caso contrario al de las consonantes, las cuales siempre deben ir acompañadas de una vocal para poder formar una sílaba y consecuentemente una palabra. La estructura silábica en los distintos idiomas a nivel mundial, permite que una sílaba esté compuesta por al menos una vocal y una consonante (o viceversa); sin embargo, existen idiomas que permiten la presencia de dos o más consonantes juntas, tanto al inicio como al final de una palabra (Holmes, 2001). Cabe recalcar que una sílaba no puede poseer dos fonemas de vocales juntos (a excepción del caso específico del diptongo), pero existen idiomas en que las sílabas no poseen ningún fonema de vocal; como por ejemplo el idioma inglés, en palabras como: "button" y "little"; en donde personas de algunas regiones de Inglaterra (por citar un ejemplo), no pronuncian un sonido de vocal antes de la consonante final de la segunda sílaba.

Las vocales en castellano son cinco, y se clasifican en: abiertas y cerradas (aunque algunos autores suelen especificar también otra clasificación denominada "medias"). Las vocales abiertas son la /a/, /e/, y /o/; mientras que las vocales cerradas son la /i/ y /u/. Además, a las vocales abiertas también se las suele denominar como "fuertes", y a las cerradas como "débiles".

Para la generación de vocales, los elementos articulatorios que intervienen son: la lengua, los labios, el paladar blando, y la mandíbula inferior. Además de la clasificación expuesta anteriormente, también se puede realizar otra clasificación en base a las posiciones de la lengua, tanto vertical (elevación) como horizontal (avance):

Tabla 2.

Clasificación de las vocales del castellano según la posición de la lengua.

Posición vertical	Tipo de vocal	Posición horizontal (avance)		
		Anterior	Central	Posterior
Alta	Cerrada	i		u
Media	Media	e		o
Baja	Abierta		a	

Tomado de Miyara, 2004, p. 8.

Por otro lado, las consonantes en castellano son 24 (existen algunas que no se pronuncian, como la 'h'), y se las clasifica según el fonema que las representa, de dos maneras distintas (Miyara, 2004):

- Según el lugar o punto de articulación, que se refiere a los elementos articulatorios que intervienen para generar un fonema.
- Según el modo de articulación, que se refiere a los procesos internos que se realizan para producir un fonema.

Entonces, según lo expuesto anteriormente, se pueden clasificar a las consonantes debido al punto de articulación, como:

- *Bilabiales*: existe oposición de ambos labios.
- *Labiodentales*: hay una oposición de los dientes superiores con el labio inferior.
- *Linguodentales*: hay una oposición de la punta de la lengua con los dientes superiores.
- *Alveolares*: existe oposición de la punta de la lengua con la región alveolar.
- *Palatales*: se produce una oposición de la lengua con el paladar duro.
- *Velares*: hay una oposición de la parte posterior de la lengua con el paladar blando.
- *Glotaes*: hay una articulación en la propia glotis.

Además, según el modo de articulación, se las puede clasificar como:

- *Oclusivas* o *explosivas*: la salida del aire se cierra momentáneamente por completo.
- *Fricativas*: el aire sale atravesando un espacio estrecho.
- *Africadas*: oclusión seguida por fricación.
- *Laterales* o *líquidas*: la lengua obstruye el centro de la boca, y el aire sale por los lados.
- *Vibrantes*: la lengua vibra cerrando el paso del aire intermitentemente.
- *Aproximantes*: La obstrucción es muy estrecha de modo que no llega a producir turbulencia.

Además de todo esto, existen dos clasificaciones generales de las consonantes, por su sonoridad. Dichas clasificaciones se deben principalmente al lugar por donde pasa el aire al momento de hablar (i.e. el tipo de resonador que se encuentra por encima de la laringe). Estas son:

- *Nasales*: El aire pasa por la cavidad nasal.
- *Orales*: El aire pasa por la cavidad oral, o boca.

En la tabla 3 se muestran todas las clasificaciones mencionadas anteriormente:

Tabla 3.

Clasificación de las consonantes del castellano según el lugar y el modo de articulación, y la sonoridad.

Lugar de articulación	Modo de articulación								
	Oral								Nasal
	Oclusiva		Fricativa		Africada	Lateral	Vibrante	Aproximante	Sonora
	Sorda	Sonora	Sorda	Sonora	Sorda	Sonora	Sonora	Sonora	
Bilabial	p	b, v		b, v				w	m
Labiodental			f						
Linguodental			z	d					
Alveolar	t	d	s	y	ch	l	r, rr		n
Palatal				(y)	(ch)	ll		i	ñ
Velar	k	g	j	g					
Glotal			h						

Tomado de Miyara, 2004, p. 7.

Cabe recalcar que los fonemas en donde participa la vibración de las cuerdas vocales, se denominan sonoros o tonales; mientras que aquellos en donde no existe dicha participación, se denominan sordos (Miyara, 2004).

Como se puede notar en la tabla 3, existen algunas consonantes que son descritas dos veces, lo que significa que poseen dos modos de articulación distintos, o en otras palabras, corresponden a dos fonemas diferentes. Esto se debe a la posición en la sílaba en que se encuentren, por ejemplo:

- La “b” o “v”, inicial y postnasal; corresponde al fonema /b/, y es oclusiva;
- La “b” o “v”, postvocálica, postlateral, y postvibrante; corresponde al fonema /β/, y es fricativa.
- La “g” inicial, postnasal, y postlateral; corresponde al fonema /g/, y es oclusiva.

- La “g” postvocálica y postvibrante; corresponde al fonema /ɣ/, y es fricativa.
- La “d” inicial o postnasal; corresponde al fonema /d/, y es oclusiva.
- La “d” postvocálica y postvibrante; corresponde al fonema /ð/, y es fricativa.

1.2.2.3 Diptongo

El diptongo es la unión de una vocal débil con una fuerte, dentro de una misma sílaba (Frías, 2001). En términos generales se considera al diptongo como una sola vocal.

Existen en el castellano dos tipos de diptongos: creciente y decreciente. Cuando la vocal débil va delante, el diptongo es creciente; mientras que si la vocal va detrás, es decreciente. Cuando esto ocurre, se considera a la vocal débil como “semiconsonante”, y su fonema cambia, representando a la “i” como /j/, y a la “u” como /w/. La clasificación de diptongos se puede hacer de la siguiente forma:

Tabla 4.

Clasificación de los diptongos del castellano.

	<i>laʎ</i>	<i>leʎ</i>	<i>loʎ</i>	<i>liʎ</i>	<i>luʎ</i>
creciente	"ia" → /ja/	"ie" → /je/	"io" → /jo/		
	"ua" → /wa/	"ue" → /we/	"uo" → /wo/		
decreciente	"ai" → /ai/	"ei" → /ei/	"oi" → /oi/		
	"au" → /au/	"eu" → /eu/	"ou" → /ou/		
homogéneo				"ui" → /wi/	"iu" → /ju/

Tomado de Frías, 2001, p. 5.

Dentro de los diptongos enunciados en la tabla 4, se debe recalcar que algunos de ellos son muy raros en el castellano latinoamericano, como son: /ou/, /oi/, /ei/, entre otros. Se debe mencionar que cuando el acento recae en la vocal débil, entonces ya no se considera diptongo, sino hiato.

1.2.2.4 La sílaba

Como se mencionó anteriormente, la sílaba es una emisión de voz, y está compuesta por vocal(es) y consonante(s). En el castellano, la sílaba debe tener una vocal obligatoriamente, aunque puede o no contener consonante(s). Además, también puede contener diptongo, el cual se considera como una vocal.

Clases de sílabas

Las sílabas se pueden clasificar según su terminación, ya sea esta, en vocal o en consonante. Existen dos tipos de sílabas según este criterio, y son:

- *Abiertas o libres*: son aquellas que terminan en vocal.
- *Cerradas o trabadas*: son aquellas que terminan en consonante.

Las dos clases de sílabas expuestas anteriormente, pueden ser usadas en cualquier tipo de combinación para formar palabras; sin embargo, las sílabas del castellano no pueden estar compuestas por todas las combinaciones vocal-consonante (VC) existentes.

Existen nueve posibles combinaciones, las cuales se muestran en la tabla 5 (resaltadas en negro):

Tabla 5.

Combinaciones vocal-consonante posibles en el castellano.

(V)	<i>a-ho-ra, dí-a</i>
(CV)	<i>pa-sa, tra-je</i>
(VC)	<i>ab-side, al-to</i>
(CVC)	<i>dos, sín-te-sis</i>
(CCV)	<i>pre-sa, gra-sas</i>
(VCC)	<i>ins-truc-ción, abs-ten-ción</i>
(CVCC)	<i>cons-ti-tu-ción, bi-ceps</i>
(CCVC)	<i>pren-sa, gran-de</i>
(CCVCC)	<i>trans-por-te</i>

Tomado de Frías, 2001, p. 15-16.

1.3.- Formantes y sus características en los fonemas

1.3.1 El tracto vocal

El tracto vocal es un segmento del aparato fonatorio que está constituido por la cavidad oral, cavidad nasal, faringe y laringe. Dentro de estas cavidades están los órganos de articulación, los cuales pueden ser divididos en *activos* y *pasivos*. Los órganos articulatorios activos son: la lengua, la mandíbula, el velo del paladar y los labios; mientras que los pasivos son: los dientes, el paladar duro y el maxilar superior. A través de la modificación y diferentes posiciones que adoptan los órganos articulatorios, el tracto vocal posee distintas formas o configuraciones que actúan como filtros acústicos para el sonido producido. Cada configuración diferente del tracto vocal constituye un filtro diferente y por consiguiente el sonido escuchado es distinto (Guzmán, 2010).

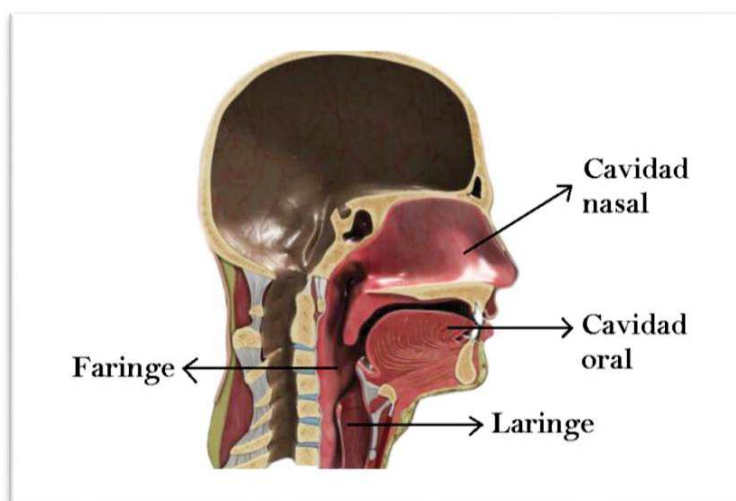


Figura 3.

Estructura del tracto vocal. Adaptado de Guzmán, 2010, p. 1.

Los cambios anatómicos del tracto vocal están basados principalmente en dos elementos: su longitud, y los distintos diámetros transversales a lo largo de él. Dependiendo del largo y de los diámetros transversales, el tracto vocal actuará como un molde o filtro acústico. Este filtro está determinado acústicamente por la función de transferencia, la que a su vez está determinada por los valores de los formantes del tracto vocal (Guzmán, 2010).

1.3.2 Formantes

Para realizar un estudio acerca de la voz humana, es necesario mencionar que los sonidos generados en el aparato fonatorio son modificados de acuerdo a diversos factores, entre los cuales se puede mencionar: distintas configuraciones del tracto vocal, obstrucción de los elementos articulatorios, entre otros. Todo esto desemboca en la obtención de una característica muy importante en los sonidos hablados por los seres humanos, denominada *formantes*.

“Los formantes son las resonancias propias de cualquier elemento que tenga la capacidad de resonar (vibrar). Un formante es un pico de intensidad en el espectro de un sonido” (Guzmán, 2010, p.1). Se los podría definir también como la concentración de energía que se da en una determinada frecuencia o banda de frecuencia.

Técnicamente, los formantes son bandas de frecuencia donde se concentra la mayor parte de la energía de un sonido. Realizando un enfoque específico en el tracto vocal, se debe mencionar que los formantes son producidos por el proceso de filtrado que realizan los resonadores que lo conforman.

1.3.2.1 Características de los formantes

Los formantes pueden ser descritos según tres parámetros:

- El centro de frecuencia.
- El ancho de banda.
- La energía.

Al modificar la forma del tracto vocal, los tres elementos mencionados anteriormente son modificados en diferente medida, y por consiguiente la función de filtro y el sonido final, sufren variaciones. Los armónicos provenientes del sonido laríngeo son reforzados o atenuados por los formantes. De esta forma se establece que los armónicos que poseen valores de frecuencia más distantes a los formantes, no son tan amplificados como aquellos que poseen frecuencias cercanas a las de los formantes.

Los formantes son mencionados por nomenclatura como: F1, F2, F3, F4, etc. La numeración establecida, corresponde al valor mas bajo de frecuencia para el primer formante, y prosigue en sentido ascendente para los demás. En el estudio del habla, generalmente poseen mayor importancia los formantes F1 y F2, aunque pueden existir algunos casos en que F3 tenga protagonismo. Los formantes superiores a estos no brindan información importante en la mayoría de fonemas.

Existen dos formas básicas para modificar los valores de frecuencia de los formantes del tracto vocal según (Guzmán, 2010):

- La primera es mediante un alargamiento, o una reducción del tracto vocal. El alargamiento puede ser realizado por medio de un descenso laríngeo, una protrusión labial, o ambas juntas. En este caso, el valor de los dos primeros formantes es menor. En cambio, la reducción se puede realizar por medio de una retrusión labial o una elevación de la laringe. Para este caso, el valor de los dos primeros formantes es mayor.
- La segunda es a través de un estrechamiento, o un distanciamiento de los labios. Cuando los labios se encuentran cerca entre ellos, el valor de los formantes es menor. Por el contrario, a medida que los labios se distancien entre ellos, el valor de los formantes será mayor.

Además de la longitud del tracto vocal, las frecuencias de los formantes pueden ser modificadas de acuerdo a los diferentes diámetros transversales del tracto vocal; es decir, la contracción o la dilatación del conducto vocal, incide en la frecuencia de todos los formantes de manera distinta. Estas constricciones y dilataciones, ese deben principalmente de la posición mandibular y lingual, siguiendo las siguientes normas generales:

La apertura mandibular influye directamente en la frecuencia del primer formante, el cual aumenta cuando existe una mayor apertura.

- La frecuencia del segundo formante depende principalmente de la forma del cuerpo de la lengua.
- La frecuencia del tercer formante varía con la posición del ápice lingual.

- Para los formantes superiores no existe una teoría específica acerca de sus variaciones.

Es de suma importancia mencionar que los formantes no son picos de resonancia estáticos. La información temporal es indispensable para poder tener una idea clara de las características de cada fonema. Los formantes poseen un comportamiento distinto a lo largo del tiempo de generación del fonema hablado, lo que permite considerar mucha mayor información para un caso de reconocimiento del habla. Basándose en esto, cabe mencionar que una característica fundamental para el caso de unión entre dos fonemas, es la variación de formantes que existe en dicha unión. De aquí es donde se introduce el término “transición” (Massone, 1988), el cual se refiere a la curva de formantes (ya sea ascendente o descendente) que existe en la unión entre dos fonemas. Esta transición es de suma importancia para caracterizar un fonema, ya que en muchos casos, los formantes de una misma consonante varían dependiendo de la vocal siguiente, y viceversa.

A continuación se muestra un gráfico en donde se puede apreciar el comportamiento temporal de los formantes de la sílaba /ga/, pronunciada por un hombre:

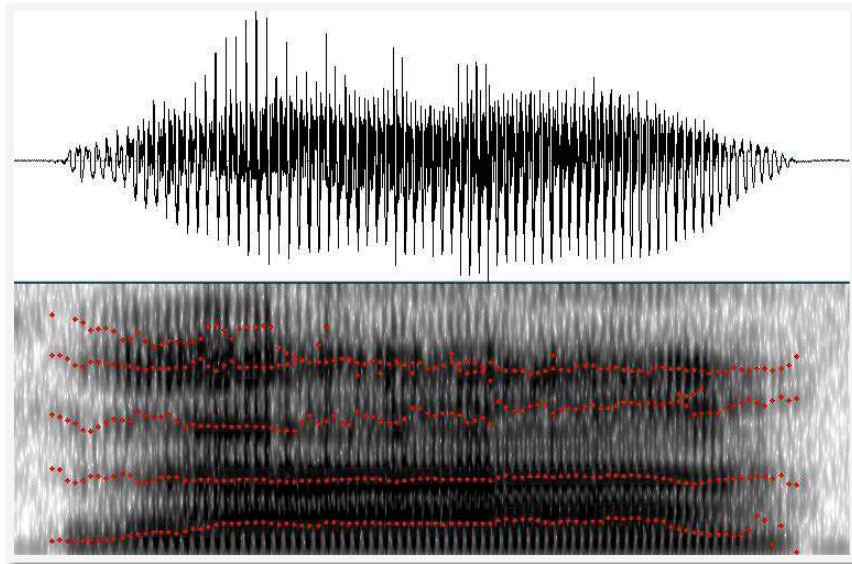


Figura 4.

Gráfico de formantes de la sílaba /ga/.

Como se puede observar en el gráfico anterior, existen dos secciones distintas. La sección superior corresponde a la forma de onda de la sílaba mencionada, en el dominio amplitud vs. tiempo. En cambio, en la sección inferior, se muestra un espectrograma en donde se resaltan los formantes como puntos rojos que forman distintas curvas.

1.3.2.2 Formantes en las vocales

Como ya se mencionó, los formantes poseen un protagonismo estelar al momento de analizar los sonidos del habla. A continuación, se detalla una explicación más descriptiva sobre las variaciones que se pueden producir en los dos primeros formantes de los fonemas de las vocales (Guzmán, 2010):

- Formante uno:
 - El primer formante (F1) varía directamente en relación a la apertura mandibular, es decir, mientras más abierta esté la mandíbula, más alto será el valor de F1, y por consiguiente, mientras más cerrada esté la cavidad oral, el valor de F1 será menor.
 - F1 también varía inversamente proporcional a la altura de la lengua. A medida que la lengua sube, F1 disminuye su valor; por el contrario, cuando la lengua se hace más plana y desciende, el valor de F1 aumenta. Un ejemplo de los aspectos anteriores es el caso de la vocal /a/, debido a que esta posee el valor de F1 más alto de todas las vocales del castellano. Para producir la vocal /a/ la mandíbula debe descender y la lengua debe estar plana en el piso de la boca, de esta forma se logra que F1 tenga un valor elevado de frecuencia, característico de la vocal /a/. Por el contrario, las vocales *altas* como la /i/ y la /u/ tienen baja frecuencia de F1 debido a que la mandíbula se encuentra más cerrada y la lengua en ubicación ascendente.
- Formante dos:
 - La segunda formante varía con la dimensión *antero-posterior* de la lengua, es decir, en qué posición dentro de la cavidad oral se encuentra la lengua.

A medida que la lengua se acerca a la boca (posición anterior), el valor de F2 asciende. Por el contrario, para obtener una disminución del valor de F2, la lengua debe dirigirse hacia la zona posterior de la cavidad oral. Un ejemplo del caso mencionado es la vocal /i/. Para que sea posible la producción de dicha vocal, la lengua debe ir hacia la zona anterior (ya que el punto de mayor constricción para esta vocal es justamente dicha zona) y por lo tanto el valor de F2 aumenta. Contrario a la vocal /i/, se tiene el caso de la vocal /u/. Para que se produzca el sonido de esta vocal, la lengua debe tener su punto de mayor constricción en la zona posterior de la cavidad oral, y de esta forma el valor de F2 es menor.

A continuación, en la figura 5 se muestra el comportamiento normal que posee la lengua al generar los fonemas /i/, /a/, y /u/:

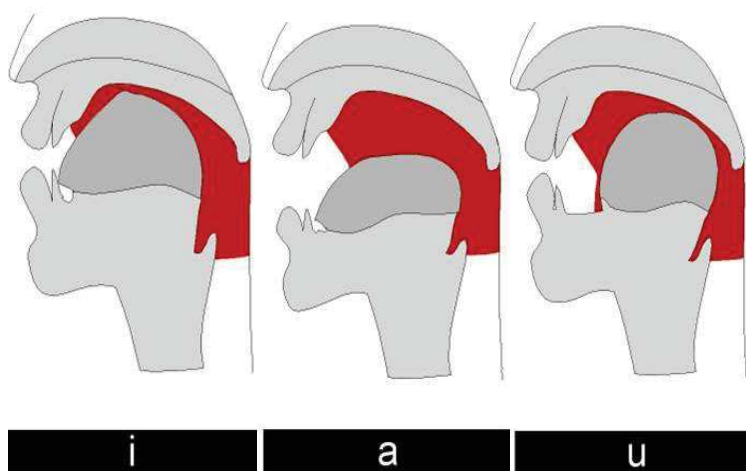


Figura 5.

Posición de la lengua en los fonemas /i/, /a/, y /u/ del castellano.

Tomado de Guzmán, 2010, p. 1.

A lo largo de la historia, han sido desarrollados a nivel mundial varios estudios acerca de los formantes de las vocales, mostrando resultados promedio muy específicos.

Como era de suponer, debido a que cada vocal implica un proceso de generación interno distinto, los valores de los formantes que caracterizan a dichos fonemas, también son diferentes.

1.4.- Conceptos generales de Redes Neuronales Artificiales

Una Red Neuronal Artificial (ANN, por sus siglas en inglés) es un modelo matemático computacional que se encuentra compuesto por un conjunto de neuronas artificiales interconectadas entre sí. Una ANN está basada en la estructura de una red neuronal biológica. Una neurona biológica se puede definir como una unidad procesadora que recibe y envía información.

Las Redes Neuronales Artificiales se caracterizan principalmente por asemejarse a la capacidad de interpretación del cerebro humano, por lo que su utilización en reconocimiento de patrones es muy empleada en la actualidad.

Aunque existen varios tipos de redes neuronales, todas ellas poseen cuatro atributos básicos (Tebelskis, 1995), los cuales son:

- Conjuntos de elementos o unidades de procesamiento.
- Conjuntos de conexiones.
- Un procedimiento de cálculo computacional.
- Un procedimiento de entrenamiento.

1.4.1 Elementos de una ANN

1.4.1.1 Unidades de procesamiento

Cada neurona artificial presente en una ANN, es denominada como *unidad o elemento procesador* (PE, por sus siglas en inglés); por lo que cada PE es análogo a una neurona biológica del cerebro humano (Basogain, 2000).

Todos y cada uno de los PE operan de manera simultánea y paralela, y son la base del modelo computacional. A cada instante de tiempo, cada PE calcula una función escalar de sus entradas y transmite ese resultado hacia sus vecinos.

Los PE pueden ser de tres tipos:

- *De entrada.*- aquellos que reciben los datos del medio.
- *Ocultos.*- son aquellos que transforman internamente la representación de los datos.
- *De salida.*- son aquellos que representan las decisiones tomadas.

En la representación de redes neuronales, y como se puede observar en la figura anterior, los PE (o neuronas) son definidos por medio de círculos.

1.4.1.2 Conexiones

Todos los PE de una red neuronal artificial son organizados por una topología específica en un conjunto de conexiones, las cuales se dibujan como líneas. Cada conexión posee un valor o peso, el cual describe la influencia que posee cierto PE sobre sus vecinos, de manera que un valor positivo hace que un PE excite a otro, mientras que un valor negativo hace que lo inhiba.

Las conexiones son generalmente unidireccionales, especificándose desde las entradas hacia las salidas, aunque existen casos en donde pueden ser bidireccionales. Una red puede poseer conexiones de cualquier tipo de topología, las cuales ofrecen mejores resultados para aplicaciones específicas.

Aunque las topologías pueden ser de diversos tipos, existen cuatro principales (Tebelskis, 1995), como son: no estructurada, por capas, recurrente, y modular.

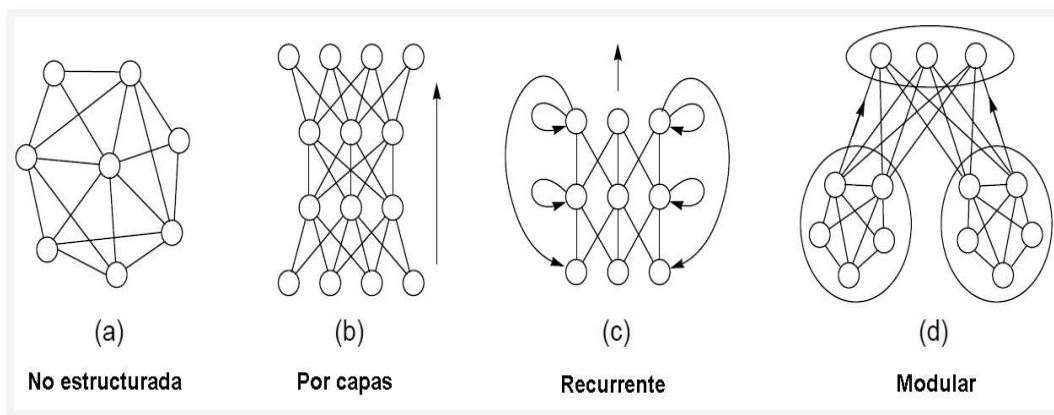


Figura 6.

Principales topologías de ANN. Adaptado de Tebelskis, 1995, p. 29.

1.4.1.3 Entradas y salidas

Las neuronas artificiales o PE, simulan el comportamiento de una neurona biológica. Una neurona biológica está compuesta por: un conjunto de entradas, denominadas *dendritas*; y una salida, denominada *axón*. El axón de cada neurona se conecta a las dendritas de otras neuronas, por medio de uniones denominadas *sinapsis* (Basogain, 2000). Haciendo una comparación entre una neurona biológica y una artificial, se puede mencionar que:

- Las dendritas son análogas a las entradas.
- El axón es análogo a las salidas.
- La sinapsis corresponde a los pesos sinápticos (o conexiones).

La figura 7 muestra una comparación gráfica entre la estructura de una neurona biológica y una artificial:

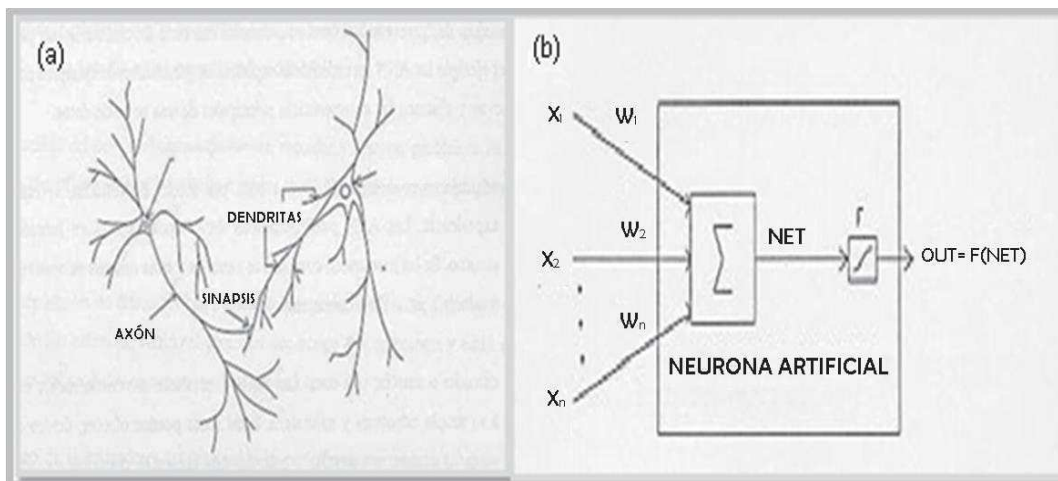


Figura 7.

Comparación entre la estructura de una neurona biológica (a) y artificial (b).

Adaptado de Basogain, 2000, pp. 3,15.

En términos generales, las redes neuronales son *no lineales*, es decir, los vectores de salida no pueden ser mapeados directamente por los vectores de entrada. Cada una de las entradas de una neurona artificial es multiplicada por su peso o ponderación correspondiente. Todas las entradas de este tipo de neuronas, una vez que se ponderan, se suman y se determina así el nivel de

activación que esta posee. Una representación vectorial del funcionamiento básico de una neurona artificial (Basogain, 2000) se puede indicar como:

$$NET = X*W \quad (\text{Ecuación 1})$$

En donde NET corresponde a la salida, X al vector de entrada, y W al vector de pesos o ponderaciones. La señal de salida NET es procesada por una función de activación F para producir la señal de salida de la neurona, denominada como OUT . La función F puede ser una función lineal, no lineal, o umbral. La función no lineal es la que simula con mayor exactitud las características de transferencia no lineales existentes en las neuronas biológicas.

1.4.1.4 Capas

En una red neuronal artificial, los PE son generalmente organizados en distintos niveles o capas, las cuales consisten básicamente en un grupo de neuronas. Cada capa posee un nombre distinto dependiendo de las unidades que en ella se encuentren. Generalmente existen dos capas que se conectan con el mundo exterior (Basogain, 2000): la capa de entrada (que es donde se presentan los datos a la red), y la capa de salida (que ofrece la respuesta de la red con respecto a las entradas). El resto de capas se las denomina como “ocultas”.

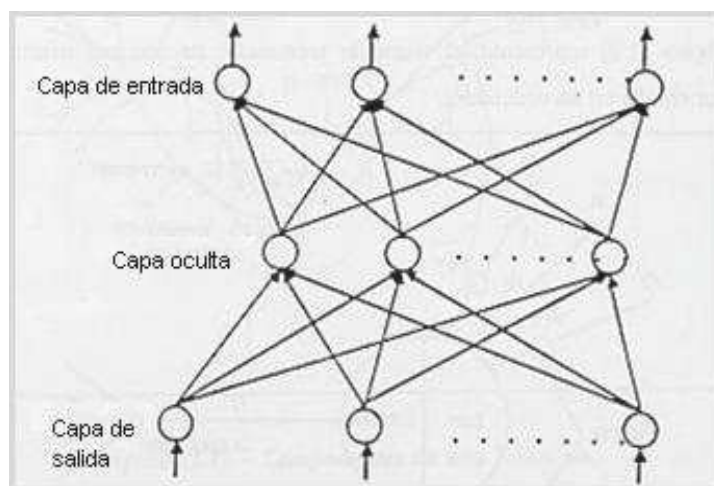


Figura 8.

Arquitectura de una red neuronal simple. Tomado de Basogain, 2000, p. 4.

La capacidad de cálculo de una ANN proviene de las distintas conexiones de las neuronas artificiales. La red más simple consiste en un grupo de neuronas ordenadas en una sola capa. En la figura 9 se muestra un diagrama en donde se especifica la arquitectura de una red neuronal simple; se debe recalcar que los nodos circulares no se consideran constituyentes de una capa, ya que sólo son distribuidores de las entradas:

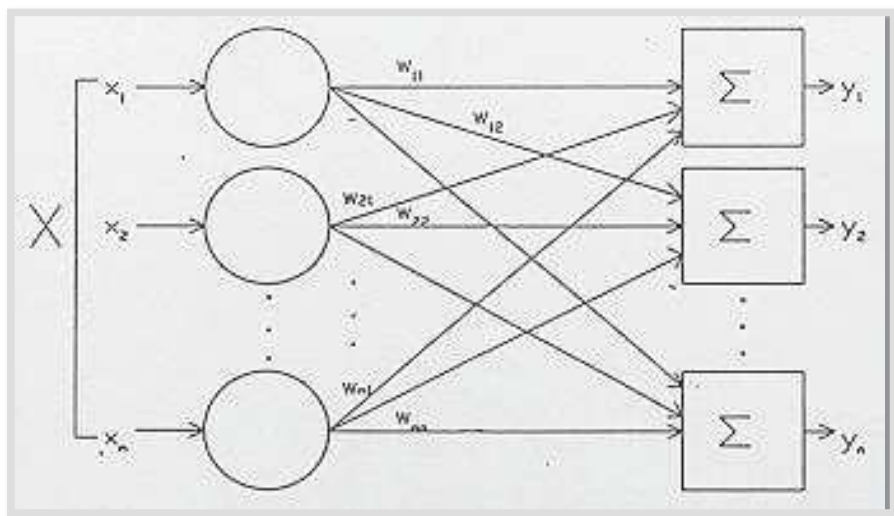


Figura 9.

Diagrama de una red neuronal de una sola capa. Tomado de Basogain, 2000, p. 16.

Como se puede observar en el gráfico anterior, cada una de las entradas está conectada a cada neurona artificial a través de su peso correspondiente. Normalmente las redes más complejas y más grandes ofrecen mejores prestaciones en el cálculo computacional que las redes simples. Las configuraciones de las redes construidas presentan aspectos muy diferentes pero el ordenamiento de las neuronas en capas o niveles siempre será el aspecto en común.

En el caso de redes multicapa, la configuración es tal que la salida de una capa es la entrada de la siguiente, y así se forma un grupo de capas simples en cascada. En términos generales se conoce que las redes multicapa presentan mejores resultados que las redes de una sola capa. Esto se debe a que las

ANN no tienen la información almacenada en un único lugar, sino que está distribuida a lo largo de toda la red (Basogain, 2000).

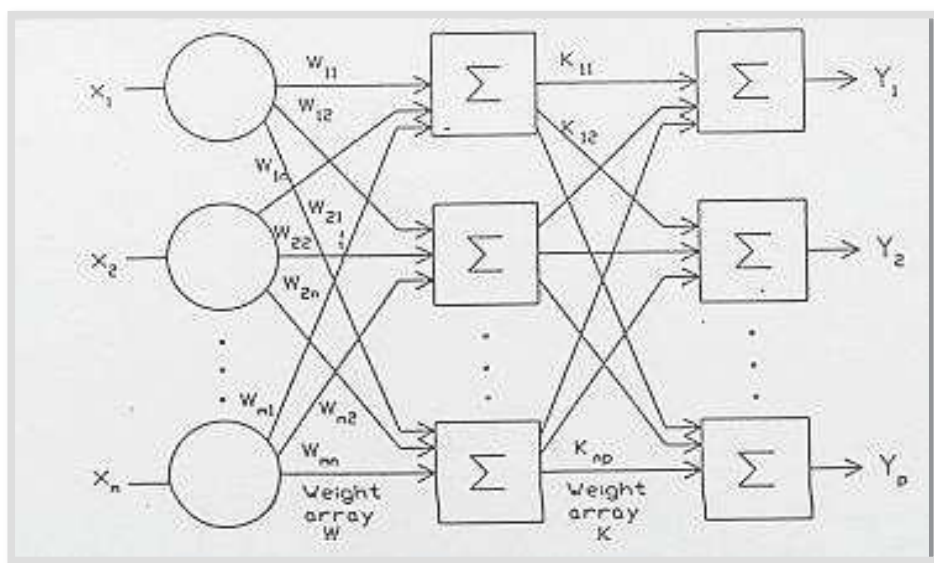


Figura 10.

Diagrama de una red neuronal multicapa. Tomado de Basogain, 2000, p. 16.

1.4.2. Cálculo computacional

Un computador tradicional posee un CPU (*Control Process Unit*), el cual se encarga de realizar todos los cálculos por medio de la ejecución de las instrucciones de secuencia programadas en un algoritmo; en donde se pueden realizar un gran número de comandos básicos ejecutados secuencialmente en sincronismo con el reloj del sistema. En contraste con esto, en la computación neuronal cada unidad PE solo puede realizar un cálculo, o en algún caso específico, unos pocos. Es por esto que la potencia del procesado de una ANN depende del número de conexiones actualizadas por segundo durante el proceso de aprendizaje de la red (Basogain, 2000).

El cálculo computacional de una ANN es un proceso complejo compuesto de varias etapas. Este cálculo comienza cuando se le presenta a la red un patrón de entrada. Luego de esto los PE son activados y se computa la información de dos maneras distintas, según sea el caso: de manera *sincronizada*, es decir

todos a la vez en un sistema en paralelo; o de manera *no sincronizada*, es decir cada unidad a la vez (Tebelskis, 1995).

Un PE dado, es usualmente segmentado en dos estados. El primero consiste en el cómputo de la información como una unidad de entrada de la red; mientras que el segundo consiste en el cómputo de la activación de su salida como una función de la entrada de la red.

En un caso general, la entrada de la red x_j para la unidad j , corresponde a la suma de sus entradas, según la siguiente ecuación:

$$x_j = \sum_i y_i w_{ji} \quad (\text{Ecuación 2})$$

Donde y_i corresponde a la activación de la salida de una unidad de entrada; mientras que w_{ji} es el peso de una unidad i hasta una j .

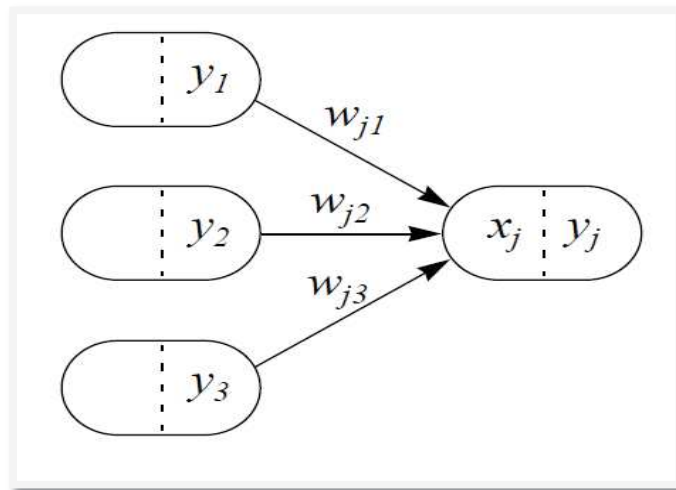


Figura 11.

Esquema simple del cómputo de unidades de activación en una ANN, donde x = entrada de la red, y = activación. Tomado de Tebelskis, 1995, p. 31.

En la computación neuronal las ANN generan sus propias reglas; las cuales se derivan del aprendizaje resultante de los ejemplos que se les muestran durante la etapa de entrenamiento.

1.4.3 Propiedades de las ANN

Las Redes Neuronales Artificiales poseen tres propiedades principales, como son:

- Aprendizaje.
- Generalización.
- Abstracción.

Las ANN pueden aprender de los patrones de entrada que se les presente, de manera de poder asociarlos con los resultados requeridos. Una característica importante para esto es la forma en que las ANN almacenan la información, debido a que el conocimiento de estas redes está distribuido a lo largo de todas las conexiones ponderadas que poseen, o pesos sinápticos.

Además de lo anterior, las ANN poseen la capacidad de generalizar la información de entrada, ya que debido a la naturaleza de su memoria, estas redes pueden responder adecuadamente cuando se tiene una entrada incompleta o con ruido.

Otra característica de las ANN, es la tolerancia a la falta, la cual se refiere al hecho de que el comportamiento de una red puede ser mínimamente modificado en caso de que varios elementos procesadores resultaran destruidos, o si se alteraran las conexiones; es por esto que aunque el comportamiento varíe, el sistema en general no deja de funcionar (Basogain, 2000).

1.4.4 Entrenamiento

El aprendizaje de una ANN se consigue a través de un entrenamiento dedicado, el cual consiste en adaptar o cambiar los pesos de las conexiones, dependiendo de la respuesta que se obtenga de los ejemplos de entrada, y eventualmente también en respuesta a las salidas deseadas, en caso de requerirlo.

El entrenamiento de una red neuronal permite obtener el comportamiento computacional deseado. Este proceso no involucra la modificación de la topología que esta posea (i.e. añadiendo o reduciendo conexiones), sin

embargo, algunas veces es mejor cambiar la topología cuando el entrenamiento no resulta adecuado, ya que esto conlleva a un mejoramiento tanto de la velocidad de aprendizaje de la red, como de la capacidad de generalizar que esta posee.

Además de esto, todos los pesos de sus conexiones solo pueden ser asignados por un proceso interactivo, el cual requiere múltiples pasos de entrenamiento. Cada paso es denominado como *interacción* o *época*.

Por otro lado, debido a que el conocimiento que va adquiriendo la red es acumulado y distribuido alrededor de sus conexiones, según (Basogain, 2000) los pesos de estas deben ser modificados de manera de que el aprendizaje obtenido no se destruya; para esto se usa una constante (ε) llamada *tasa de aprendizaje*, la cual controla la magnitud de modificación de los pesos. Se debe recalcar que es indispensable encontrar un valor correcto de ε , ya que si el valor es muy bajo, el aprendizaje de la red será demasiado lento, en cambio un valor muy alto hará que el aprendizaje anterior se pierda. A pesar de todo esto, la única manera de encontrar este valor es de forma empírica, por el método de prueba y error.

Los algoritmos de entrenamiento o ajuste de los valores de los pesos de conexiones de una ANN, se pueden clasificar en dos tipos: *supervisado* y *no supervisado* (Basogain, 2000).

1.4.4.1 Entrenamiento supervisado

En este tipo de entrenamiento se le especifica a la ANN cuáles son los objetivos de salida para los patrones de entrada que se le indicó.

El entrenamiento de la red consiste básicamente en presentarle un vector de entrada, obtener la salida que esta produzca, y comparar dicha salida con la deseada. Una vez hecho esto, se obtiene el error que existe entre la salida real de la red y la deseada, y por medio de dicho error se realimenta a la red y se cambian los pesos de las conexiones, para minimizar así el error obtenido.

Las parejas de vectores del conjunto de entrenamiento se aplican de manera cíclica y secuencialmente. Una vez que el error es calculado, se ajustan los pesos de cada pareja hasta que el error para todo el conjunto de entrenamiento sea mínimo.

1.4.4.2 Entrenamiento no supervisado

El entrenamiento no supervisado se caracteriza por no especificar a la red los vectores de salida deseados, ya que en este tipo de entrenamiento la ANN debe detectar las regularidades en los datos de entrada por sí sola. En este caso, el conjunto de vectores de entrenamiento consisten únicamente en los vectores de entrada.

En el proceso de entrenamiento de este tipo, las propiedades estadísticas de los vectores de entrenamiento son extraídas por la red, de tal manera que los vectores similares sean agrupados entre sí.

1.5.- Software usado en el cálculo de formantes

1.5.1 Software PRAAT

PRAAT es un software libre encargado del estudio fonético del habla humana. Este software fue diseñado por Paul Boersma y David Weenink, miembros del Instituto de Ciencias Fonéticas de la Universidad de Ámsterdam. La principal función de este software es analizar, sintetizar y manipular el habla humana por medio de distintos algoritmos de análisis y decodificación de señales, enfocadas principalmente en audio.

El software PRAAT permite visualizar algunas características importantes de las señales de audio, tales como: su forma de onda, espectrograma, sonograma, gráfica de frecuencia fundamental en función del tiempo, gráfica de intensidad del habla, y gráfica de los formantes en función del tiempo.

PRAAT posee varias opciones de manipulación del audio, ya sea grabado o cargado por el software. Todas estas opciones se encuentran en la parte derecha de la ventana principal del software y varían dependiendo del tipo de "objeto" que se cargue. Un objeto de PRAAT es un archivo que contiene

cualquier tipo de información compatible con el software; este archivo no se almacena en el disco duro del computador hasta que el usuario así lo realice en el menú *Save*.

Actualmente terceras personas y usuarios en general utilizan el software para realizar aplicaciones y estudios de la voz humana, ya sea para creaciones de software de reconocimiento de voz, como también para el estudio del comportamiento espectral de la voz.

Dentro de todas las opciones que PRAAT brinda al usuario, la que se usó principalmente en el desarrollo de esta tesis, es el cálculo de formantes. PRAAT permite elegir distintas formas de calcular formantes, pero para el presente proyecto se utilizó el método por defecto que el software emplea. La señal de audio de la cual se desea analizar las formantes puede proceder de un archivo previamente grabado y cargado por medio del software; o puede ser grabado directamente desde PRAAT en distintas frecuencias de muestreo.

1.5.1.2 Cálculo de formantes por medio del algoritmo de burgh

El cálculo de formantes que PRAAT realiza por defecto es por medio del algoritmo de *burgh*, descrito por (Childers, 1978).

El sonido cargado en la ventana de objetos de PRAAT es resampleado en el doble de frecuencia del máximo valor de formantes (en Hz) que el usuario elija en las configuraciones de este algoritmo (5000 Hz en el caso de un hombre promedio, y 5500 Hz para una mujer). Luego de esto se aplica un algoritmo de *pre-énfasis* de +6 dB por octava. El objetivo de este paso es que se logre un espectro lineal, debido a que el espectro de las vocales tiende a caer en lo inverso a esa proporción.

Una vez cumplido el proceso anterior, se aplica el algoritmo de burgh como tal, en donde se obtienen los coeficientes LPC (Linear Prediction Code), y luego de esto se consiguen los valores de formantes.

1.5.2 Grabación de audio en PRAAT

Para la grabación de los fonemas se hizo uso de las siguientes herramientas:

- Un micrófono dinámico Shure SM57.
- Una interfaz de audio AVID Mbox3.
- Software PRAAT versión 5.3.04 instalado en un computador con Windows7.

Las grabaciones de audio fueron realizadas directamente desde el software PRAAT para poder tener una mayor facilidad de acceso a la información almacenada y poder analizarla de manera más rápida.

La figura 12 presenta una vista de la interfaz gráfica del software PRAAT en la etapa de grabación de audio:

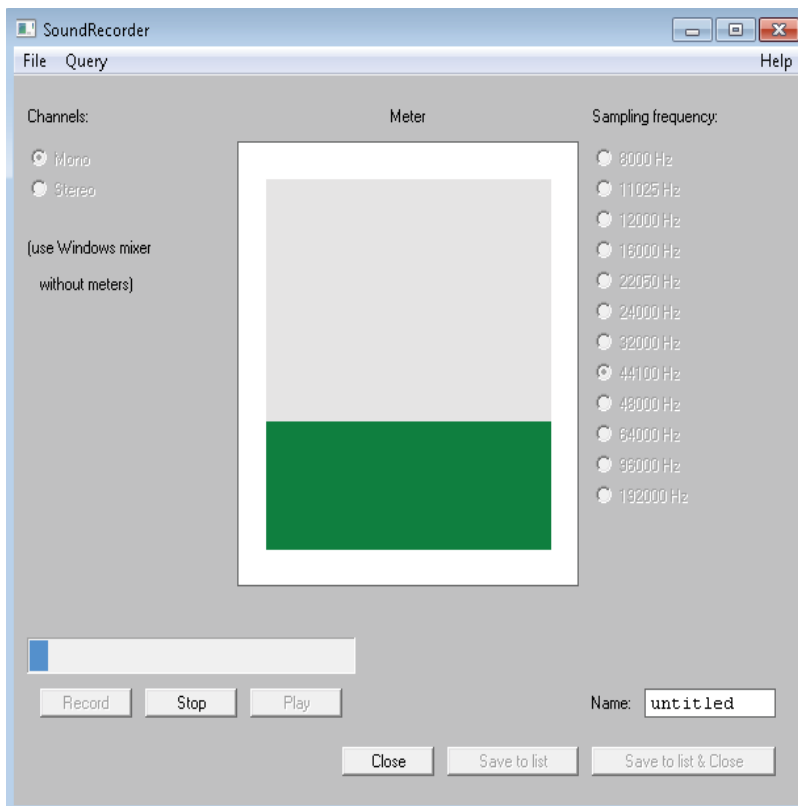


Figura 12.

Ventana de grabación en PRAAT.

Las grabaciones fueron realizadas en formato mono y con una frecuencia de muestreo de 44100 Hz. Estas grabaciones se realizaron por medio del software PRAAT, debido a una mayor facilidad al momento de manejar el audio directamente desde dicho software, tanto en la etapa de grabación como en la de su posterior análisis.

En la figura 12 se puede observar la ventana de grabación de PRAAT en el instante en que se está realizando una grabación. Esta ventana es muy simple y posee opciones sencillas, como:

- Elegir si el tipo de grabación será mono o estéreo.
- Elegir frecuencia de muestreo.
- Poner nombre a la grabación.
- Empezar o pausar la grabación.
- Guardar la grabación en la ventana de objetos.
- Visualizar el nivel de grabación por medio de un medidor visual (“*meter*”).

Las grabaciones realizadas fueron almacenadas en la ventana de objetos de PRAAT la cual es su ventana principal, en donde se pueden realizar las operaciones de análisis y visualización que el usuario requiera.

Una vez realizada la grabación, se le da un nombre al archivo de audio grabado y se lo transfiere a la ventana de objetos por medio del botón “Save to list”.

1.5.3 Formas de visualización de información en PRAAT

PRAAT permite realizar varios tipos de análisis de los objetos cargados en su lista de objetos (ver figura 13), dependiendo del tipo que este posea.

Como se mencionó con anterioridad, un objeto de PRAAT es un archivo almacenado en la RAM que posee un tipo de información compatible con el software, la cual puede ser información de audio, gráficos, texto, entre otras.

A continuación se presenta una gráfica en donde se puede apreciar un listado de objetos cargados en PRAAT, los cuales corresponden a los sonidos de todas las sílabas del castellano analizadas en el presente proyecto:

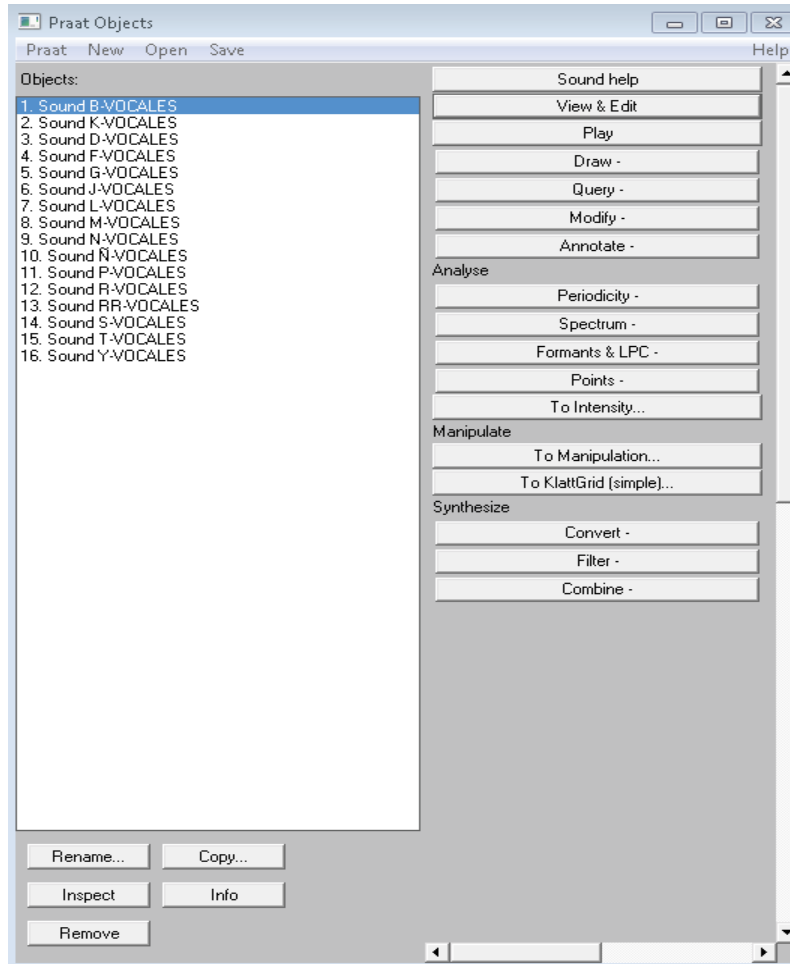


Figura 13.

Lista de objetos de la ventana principal de PRAAT.

PRAAT se caracteriza por permitir la visualización de distintos parámetros y características de los archivos de audio a analizar por parte del usuario.

Dentro de estos se puede mencionar a:

- Frecuencia fundamental en relación al tiempo (denominada como “pitch”).
- Intensidad en función del tiempo.

- Valor de frecuencia de los formantes a lo largo del tiempo.
- Pulsos calculados por el software en el dominio amplitud vs tiempo.

Para poder visualizar la forma de onda de un sonido específico, así como un espectrograma, la gráfica de formantes, y todos los parámetros mencionados anteriormente, se debe seleccionar el objeto y presionar la opción “View & Edit” situada en la parte lateral de la ventana principal del software.

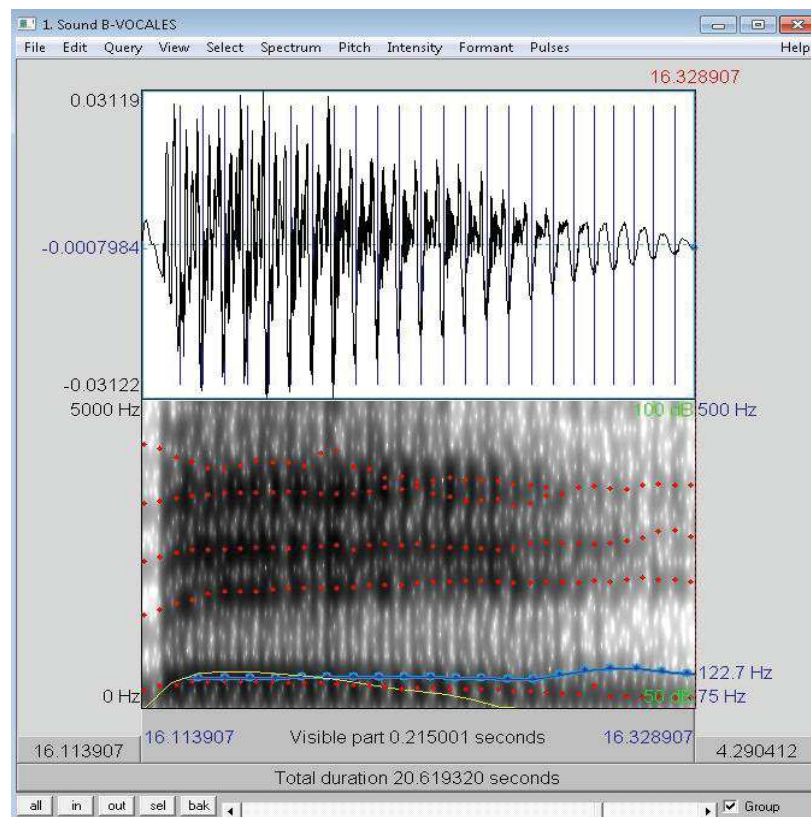


Figura 14.

Ventana de edición de un archivo de audio de PRAAT.

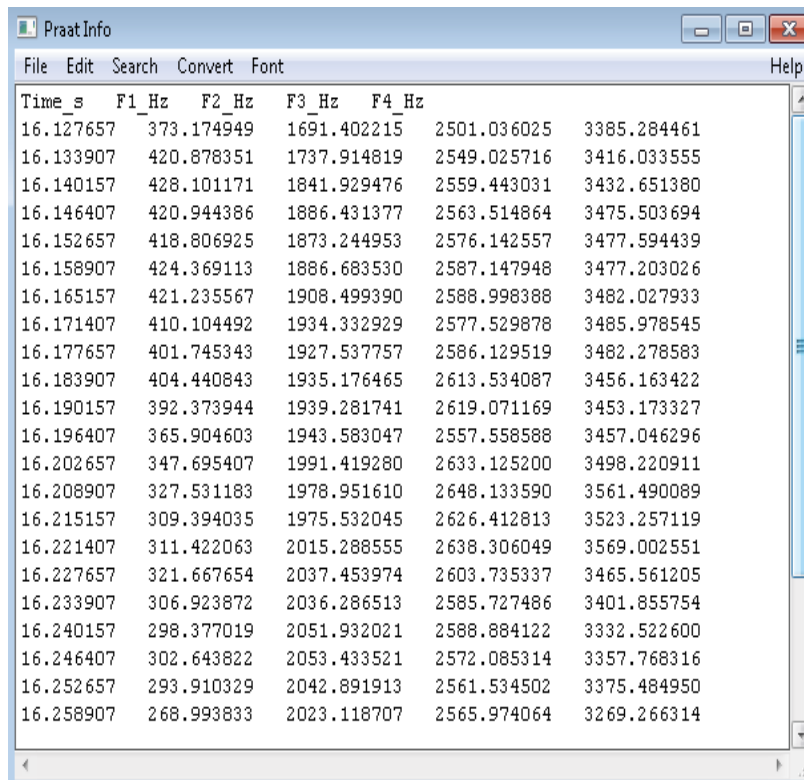
En la figura 14 se puede observar la ventana de edición de PRAAT. En la sección superior la forma de onda del archivo de audio grabado, así como los pulsos de dicha señal dibujados como líneas verticales de color azul. Mientras que en la sección inferior se puede observar el espectrograma y las curvas de formantes, frecuencia fundamental e intensidad. Las cinco curvas graficadas con puntos de color rojo a lo largo del eje temporal corresponden a las curvas de formantes, siendo la que está situada en la parte inferior la que corresponde

al primer formante, la siguiente en sentido ascendente corresponde al segundo formante, y así sucesivamente hasta la curva correspondiente al máximo número de formantes que el usuario haya especificado en configuraciones del menú Formant (por defecto este valor viene establecido en cinco).

Además de las curvas de formantes, también se encuentra la gráfica de frecuencia fundamental dibujada de color azul a lo largo del eje temporal, y la curva de intensidad con que se pronunció el fonema, la cual al igual que las demás curvas mencionadas, también varía en función del tiempo. Esta curva está dibujada de color amarillo en la sección inferior.

PRAAT permite visualizar los valores requeridos tanto de formantes como de frecuencia fundamental directamente desde la gráfica. Para esto se puede ubicar al cursor del mouse en la parte deseada de la gráfica y en la parte inferior derecha de la ventana se mostrará el valor de frecuencia fundamental en el instante de tiempo seleccionado, mientras que en la parte inferior izquierda el correspondiente a la curva de formantes que se halla seleccionado.

Para un análisis más preciso PRAAT también permite al usuario hacer una selección de cualquier porción de la gráfica de la sección inferior de su ventana de edición, y seleccionar la opción correspondiente en la barra de menú al parámetro que desea analizar a fondo, ya sea pitch, intensidad, formantes, o pulsos. Una vez elegido el menú correspondiente se debe seleccionar la opción denominada como “formant listing” (en este caso se usó al menú “Formant” ya que es el utilizado mayoritariamente en este proyecto). Una vez hecho esto se despliega una ventana en donde se muestra la información de formantes por intervalo de tiempo a lo largo de toda la selección establecida.



The screenshot shows the PraatInfo window with a menu bar (File, Edit, Search, Convert, Font, Help) and a table of formant data. The table has five columns: Time_s, F1_Hz, F2_Hz, F3_Hz, and F4_Hz. The data is as follows:

Time_s	F1_Hz	F2_Hz	F3_Hz	F4_Hz
16.127657	373.174949	1691.402215	2501.036025	3385.284461
16.133907	420.878351	1737.914819	2549.025716	3416.033555
16.140157	428.101171	1841.929476	2559.443031	3432.651380
16.146407	420.944386	1886.431377	2563.514864	3475.503694
16.152657	418.806925	1873.244953	2576.142557	3477.594439
16.158907	424.369113	1886.683530	2587.147948	3477.203026
16.165157	421.235567	1908.499390	2588.998388	3482.027933
16.171407	410.104492	1934.332929	2577.529878	3485.978545
16.177657	401.745343	1927.537757	2586.129519	3482.278583
16.183907	404.440843	1935.176465	2613.534087	3456.163422
16.190157	392.373944	1939.281741	2619.071169	3453.173327
16.196407	365.904603	1943.583047	2557.558588	3457.046296
16.202657	347.695407	1991.419280	2633.125200	3498.220911
16.208907	327.531183	1978.951610	2648.133590	3561.490089
16.215157	309.394035	1975.532045	2626.412813	3523.257119
16.221407	311.422063	2015.288555	2638.306049	3569.002551
16.227657	321.667654	2037.453974	2603.735337	3465.561205
16.233907	306.923872	2036.286513	2585.727486	3401.855754
16.240157	298.377019	2051.932021	2588.884122	3332.522600
16.246407	302.643822	2053.433521	2572.085314	3357.768316
16.252657	293.910329	2042.891913	2561.534502	3375.484950
16.258907	268.993833	2023.118707	2565.974064	3269.266314

Figura 15.

Listado de formantes de una selección en software PRAAT.

Como se puede observar en la figura 15, el intervalo de tiempo en el que PRAAT realiza la obtención de formantes corresponde aproximadamente a 6 mseg.

Capítulo II.- Desarrollo experimental

2.1 Estudio de los fonemas

El habla humana es un proceso que involucra varias connotaciones de distintas índoles. En términos generales posee características de generación comunes en todos los seres humanos, sin embargo su caracterización final muchas veces depende de ciertos factores específicos de cada persona. Si bien el principal objetivo de un ser humano al momento de hablar es transmitir información, generalmente una misma información no se transmite de igual manera por todas las personas.

En base a todo lo anterior, la principal consideración que se realizó al momento de empezar a desarrollar el presente proyecto, fue establecer hasta qué punto se centraría el análisis del habla. Considerar el 100% de análisis de este fenómeno hubiera sido inviable de realizar en base a las condiciones presentadas. Es por esto que se decidió únicamente escoger los casos de sílabas compuestas por una consonante y una vocal (CV), en dicho orden específico.

Al centrar el estudio de los fonemas en sílabas CV se descartaron varios casos de mayor complejidad de análisis como son: habla continua, oraciones, frases, palabras, y sílabas compuestas por más de dos letras. En condiciones generales los seres humanos no expresan sus ideas hablando sílabas aisladas, sino por medio de “habla continua”, la cual consiste en la pronunciación de palabras y/o frases interconectadas entre sí, a distintas velocidades y pausas, dependiendo de cada persona y considerando principalmente algunos factores como influencia geográfica y edad.

Centrándose en el caso de estudio escogido para este trabajo, vale la pena mencionar que si bien la unidad fónica ideal mínima del lenguaje es el fonema, la sílaba es la unidad menor que percibe un ser humano (Miyara, 2004). Esta es la principal razón que respalda el enfoque del presente proyecto. Sin embargo, al comienzo del desarrollo de este trabajo fue necesario definir el tamaño mínimo que se consideraría en una sílaba.

Al hablar del tamaño de una sílaba se debe interpretar como la cantidad de fonemas presentes en ella, es decir, cuántas letras posee. En base a la estructura del castellano, se establece que una sílaba puede poseer hasta cinco letras, y que su principal característica es la presencia de por lo menos una vocal.

Los casos de sílabas compuestas por cinco y cuatro letras no son muy usados en el castellano de Ecuador, por lo que para simplificar el estudio se decidió descartarlos. De la misma manera tampoco se consideraron las sílabas compuestas por tres letras, ya que aunque dichos casos son muy comunes en el castellano, involucran mucha mayor complejidad de análisis. Esto se basó en el hecho de que la articulación entre consonantes juntas no siempre se realiza de la misma forma en distintas personas. Esto se produce debido a ciertos factores de orden geográfico, fisiológico, entre otros, que influyen en la forma en que una persona pronuncia ciertos fonemas cuando se encuentran enlazados con otros.

Una vez planteados y definidos los objetivos y límites del proyecto, se procedió a establecer una metodología para poder desarrollarlo. El primer paso que se llevó a cabo para comenzar el desarrollo del proyecto, consistió en definir qué fonemas del castellano serán usados para el análisis de sus características acústicas. Para esto se procedió a realizar un estudio a fondo de algunos factores predominantes en ellos.

2.1.1 Definición de los fonemas a analizar

Cada fonema del castellano constituye un sonido específico y diferente, debido a que el proceso de generación interno del cuerpo humano en cada uno de ellos es distinto. Adicionalmente, existen algunos factores que influyeron en la delimitación de los fonemas que conformarían la base de análisis de este proyecto, los cuales se describirán con detalle más adelante.

Centrándose en la labialización de cada fonema del castellano, existen algunos que poseen un movimiento de labios similar, por lo que se decidió usar únicamente a uno de los posibles fonemas que posean esta característica.

Además, también se procedió a descartar aquellos fonemas que no son usados en el castellano latinoamericano, enfocado específicamente en Ecuador.

Las sílabas que se usaron en el análisis fueron resultado de la elaboración de una matriz con combinaciones de sílabas CV, como se muestra en la tabla 6:

Tabla 6.

Matriz final de sílabas CV.

	A	E	I	O	U
B	BA	BE	BI	BO	BU
CH	CHA	CHE	CHI	CHO	CHU
D	DA	DE	DI	DO	DU
F	FA	FE	FI	FO	FU
G	GA	GE	GI	GO	GU
J	JA	JE	JI	JO	JU
K	KA	KE	KI	KO	KU
L	LA	LE	LI	LO	LU
M	MA	ME	MI	MO	MU
N	NA	NE	NI	NO	NU
Ñ	ÑA	ÑE	ÑI	ÑO	ÑU
P	PA	PE	PI	PO	PU
R	RA	RE	RI	RO	RU
RR	RRA	RRE	RRI	RRO	RRU
S	SA	SE	SI	SO	SU
T	TA	TE	TI	TO	TU
Y	YA	YE	YI	YO	YU

Cabe recalcar que para poder obtener la matriz presentada anteriormente fue necesario realizar distintos pasos, considerando el análisis de varios factores determinantes que se describirán a continuación:

El primer paso consistió en la elaboración de una matriz “global” de sílabas, que contenga todas las posibles combinaciones de letras que conformen una

sílaba en el castellano, considerando a una consonante siempre en posición inicial. Sin embargo, esto conllevó a la obtención de una matriz de cuarenta filas y setenta columnas, con alrededor de 2800 posibilidades de sílaba, en donde se incluían sílabas que no presentaban coherencia en el idioma castellano.

Posteriormente se optó por eliminar los casos de sílabas de cinco y cuatro letras, para dejar únicamente las combinaciones de dos y tres letras. En esta ocasión se obtuvieron 190 posibilidades, reduciendo en gran manera la cantidad de posibilidades expuesta anteriormente. Para este caso se usaron todas las consonantes de manera independiente unidas con cada una de las cinco vocales del castellano. Además de esto, se consideraron uniones de consonantes fricativas y oclusivas, con laterales y vibrantes, por constituir posibilidades reales del castellano (a excepción de la unión “TL”, la cual casi no es usada en palabras de este lenguaje).

Tabla 7.

Matriz de sílabas formadas por dos y tres letras.

	A	E	I	O	U
B	BA	BE	BI	BO	BU
C	CA	CE	CI	CO	CU
D	DA	DE	DI	DO	DU
F	FA	FE	FI	FO	FU
G	GA	GE	GI	GO	GU
J	JA	JE	JI	JO	JU
K	KA	KE	KI	KO	KU
L	LA	LE	LI	LO	LU
LL	LLA	LLE	LLI	LLO	LLU
M	MA	ME	MI	MO	MU
N	NA	NE	NI	NO	NU
Ñ	ÑA	ÑE	ÑI	ÑO	ÑU
P	PA	PE	PI	PO	PU

Q	QA	QE	QI	QO	QU
R	RA	RE	RI	RO	RU
RR	RRA	RRE	RRI	RRO	RRU
S	SA	SE	SI	SO	SU
T	TA	TE	TI	TO	TU
V	VA	VE	VI	VO	VU
W	WA	WE	WI	WO	WU
X	XA	XE	XI	XO	XU
Y	YA	YE	YI	YO	YU
Z	ZA	ZE	ZI	ZO	ZU
BL	BLA	BLE	BLI	BLO	BLU
BR	BRA	BRE	BRI	BRO	BRU
CH	CHA	CHE	CHI	CHO	CHU
CL	CLA	CLE	CLI	CLO	CLU
CR	CRA	CRE	CRI	CRO	CRU
DR	DRA	DRE	DRI	DRO	DRU
FL	FLA	FLE	FLI	FLO	FLU
FR	FRA	FRE	FRI	FRO	FRU
GL	GLA	GLE	GLI	GLO	GLU
GR	GRA	GRE	GRI	GRO	GRU
PL	PLA	PLE	PLI	PLO	PLU
PR	PRA	PRE	PRI	PRO	PRU
TL	TLA	TLE	TLI	TLO	TLU
TR	TRA	TRE	TRI	TRO	TRU

Para llegar a la matriz final presentada en la tabla 7 se descartaron las sílabas compuestas por tres letras debido a las razones expuestas anteriormente, y además se realizó un análisis de todos los fonemas del alfabeto castellano, buscando descartar aquellos que no constituyen una opción real o común en el habla ecuatoriana. Las letras descartadas son las siguientes:

- Sílabas compuestas por “c” en unión con vocales. Ya que esta consonante constituye en realidad dos fonemas diferentes: /k/ y /s/, se la

descartó debido a que en su lugar se decidió usar las consonantes “k” y “s”, las cuales corresponden a los fonemas especificados anteriormente. De igual manera se descartó la letra “q”, debido a que también corresponde al fonema /k/.

- Sílabas compuestas por “ll” en unión con vocales. En este caso se decidió descartar a esta consonante debido a que en el habla común de un gran sector de la población ecuatoriana, es muy raro diferenciar la pronunciación entre las letras “ll” y “y”, aunque estas dos constituyen dos fonemas distintos, como son: /λ/ y /dʒ/ respectivamente. Adicionalmente, también se descartó el caso de pronunciación “sh”, ya que no constituye una condición real en Ecuador.
- Sílabas compuestas por la letra “v” en unión con vocales. Ya que tanto la consonante “b” como la “v” corresponden a un mismo fonema del castellano (a diferencia de lo que normalmente se piensa), se decidió únicamente dejar a las sílabas compuestas por “b” ya que aunque gramaticalmente son dos letras distintas, en el aspecto fonético no existen diferencias, por lo que un análisis de ambas consonantes sería redundante.
- Sílabas compuestas por la letra “w” en unión con vocales. Se decidió descartar este fonema debido a su mínima utilización en el castellano ecuatoriano, ya que esta se produce principalmente en palabras procedentes del idioma inglés (como: “Walter”, “Washington”, etc.). Además, en un sector de la población ecuatoriana se suele considerar a este fonema como la unión de dos fonemas: /g/, y /u/.
- Sílabas compuestas por la letra “x” en unión con vocales. Debido a su poca utilización se decidió descartar a esta consonante.
- Sílabas compuestas por “z” en unión con vocales, ya que la pronunciación del fonema correspondiente a esta letra no se utiliza en el continente latinoamericano, debido a que en su lugar se emplea al fonema /s/.

Si bien el castellano es un lenguaje que posee una riqueza enorme en la estructuración de sus sílabas y palabras, no se debe pensar que el estudio

fonético y acústico de sílabas compuestas únicamente por dos fonemas constituye un trabajo trivial. Para los fines planteados en este proyecto, se considera que es mejor enfocarse en este tipo de estructura de sílabas, debido a que en estas combinaciones se minimizan en grandes proporciones las diferencias de pronunciación, entonación, y velocidad que cada persona puede generar; por lo que se considera una opción muy acertada ya que de esta manera se podría interpretar que se está realizando un estudio del lenguaje en su manera más “limpia” o general, sin variaciones en pronunciación de personas, lugares geográficos, etc. Aunque es también muy válida la acotación de que para estudiar el lenguaje castellano en su manera más natural, sería necesario realizar dicho estudio en su país origen.

Se debe recalcar que además de todas las sílabas descritas anteriormente, también se procedió a realizar un estudio independiente de cada una de las vocales del castellano, debido a su importancia en la conformación de las sílabas; y poder tener así mayor información acerca de su comportamiento individual y al momento de juntarse con las consonantes en cada caso correspondiente.

2.1.2 Procedimientos usados para la obtención de información

En función de lograr analizar a una sílaba compuesta por estos dos fonemas distintos (consonante-vocal), fue necesario primero realizar un análisis específico de cada fonema por separado, y luego esto centrarse en la manera en que varía el comportamiento independiente de cada fonema al unirse con el fonema aledaño.

Lo primero que se hizo fue grabar a seis personas distintas, compuestas por tres hombres y tres mujeres de diferentes edades, diciendo cada una de las sílabas de la matriz reducida presentada en la tabla 7.

Este procedimiento se llevó a cabo con la finalidad de poder evaluar el comportamiento de formantes de cada fonema, pero teniendo en consideración los valores que aparecen en todos los casos para poder establecer comparaciones.

2.1.2.1 Interpretación de la información obtenida

Una vez realizadas las grabaciones de las vocales por separado y de las sílabas CV, se procedió a visualizar las curvas de formantes y extraer la información de los valores correspondientes a cada formante.

Como ya se mencionó con anterioridad, un formante es un pico de resonancia en cierta frecuencia o banda de frecuencia que se produce debido a distintos factores fisiológicos de los órganos presentes en cada aparato fonatorio, como: tamaño, anchura, etc.; y también debido a los procesos internos que realiza el cuerpo humano para producir cierto fonema, ya que por medio de dichos procesos también se varían los tamaños, anchuras, distancias, etc., de los órganos del aparato fonatorio.

Lo anterior conlleva al pensamiento de que cada persona posee información de formantes distinta, debido a que cada aparato fonatorio posee dimensiones específicas para cada persona. Si bien esto es cierto, también es verdad que cada fonema producido posee un comportamiento de formantes muy similar entre las personas. En el caso de las mujeres debido al menor tamaño promedio de su laringe, tráquea, entre otros elementos en comparación con los hombres, es de esperarse que los valores de formantes obtenidos sean altos. En la tabla 8 se muestra una comparación entre los valores promedio de los dos primeros formantes de la vocal “a” en hombres y mujeres.

Tabla 8.

Valores promedio obtenidos de F1 y F2 de /a/ en hombres y mujeres.

Vocal	Valor promedio de F1 (Hz)		Valor promedio de F2 (Hz)	
	Hombre	Mujer	Hombre	Mujer
a	700	820	1.300	1.550

Análisis de los dos primeros formantes de las vocales

Las vocales en el castellano son cinco. Las cuales no poseen variaciones significativas en su pronunciación en diferentes combinaciones de palabras a

diferencia de otros idiomas (como en el inglés). La única variación en la pronunciación de vocales del castellano se produce en el caso de diptongo. Cuando se produce la unión de una vocal cerrada con una abierta, la vocal cerrada corresponde a un fonema diferente. En el caso de la “i”, el fonema correspondiente es /j/; mientras que en el caso de la “u” es /w/.

Las vocales poseen características propias de elaboración en el aparato fonatorio que las diferencian de las consonantes. Desde un punto de vista mecánico-acústico, se establece que las vocales son aquellos sonidos generados por la vibración de las cuerdas vocales, cuyo flujo de aire no posee ningún obstáculo u obstrucción a lo largo de todo el aparato fonatorio (Miyara, 2004). Es por esta característica que las vocales son consideradas como fonemas sonoros, ya que siempre involucran la vibración de las cuerdas vocales.

Como ya se mencionó anteriormente, en la generación de vocales intervienen algunos elementos articulatorios dentro del aparato fonatorio, los cuales son: la lengua, los labios, el paladar blando y la mandíbula inferior. Dentro de los elementos involucrados en la elaboración de estos fonemas, se realizará un principal enfoque en la posición de la lengua y la postura de los labios dentro del análisis elaborado. En términos generales, se puede decir que F1 (primer formante) da la característica sonora principal a la mayoría de fonemas, mientras que F2 (segundo formante) corresponde a una cualidad de “moldeamiento” del sonido (Guzmán, 2010).

Comparación entre vocales

Las vocales generalmente suelen ser clasificadas según dos características: la postura de los labios y la posición de la lengua. Según (Frías, 2001) se pueden establecer las siguientes consideraciones:

- La vocal “a” es considerada como una vocal abierta de acuerdo a la postura de los labios. En cambio, tomando en cuenta la posición de la lengua también se la podría considerar como central, ya que la lengua

se ubica en la parte céntrica de la cavidad oral. Por estas dos razones expuestas se conoce a esta vocal como una vocal central abierta.

- La vocal “e” es considerada como una vocal abierta por la postura de los labios. Además, también es considerada como una vocal inicial-central por la posición que adopta la lengua. Este fonema se caracteriza porque la lengua adopta una posición anterior-central debido a que tiende a acercarse más a la sección anterior o delantera de la cavidad oral.
- La vocal “i” es considerada como una vocal cerrada por la postura de los labios. Además, también es considerada como una vocal inicial debido a la posición que adopta la lengua. Este fonema se caracteriza porque la lengua adopta una posición anterior o delantera debido a que se encuentra ubicada en la parte delantera de la cavidad oral. Es la vocal que posee la característica única de que la lengua tome una posición completamente delantera. Por estas características mencionadas, se considera al fonema /i/ como una vocal inicial cerrada.
- La vocal /o/ es considerada como una vocal abierta por la postura de los labios. Además, también es considerada como una vocal central-final debido a la posición que adopta la lengua, ya que en la producción de este fonema esta adopta una posición central anterior, tendiendo a irse hacia la parte posterior de la cavidad oral.
- La vocal /u/ es considerada como una vocal cerrada por la postura de los labios, y como vocal inicial debido a la posición que adopta la lengua. Este fonema se caracteriza porque la lengua adopta una posición final o posterior, debido a que tiende a ubicarse en la parte posterior de la cavidad oral. Por estas características mencionadas, se considera al fonema /u/ como una vocal final cerrada.

Con respecto al análisis de los dos primeros formantes realizado en (Guzmán, 2010) y el desarrollado en el presente trabajo, se pueden establecer las siguientes observaciones con respecto a cada vocal:

- El comportamiento de formantes de la vocal “a” presenta una particularidad, ya que posee el valor de frecuencia de F1 más alto de

todas las vocales. Esto se debe principalmente a la posición descendente de la lengua y la mandíbula, como se explicó anteriormente. A medida que la lengua desciende en la cavidad oral, mayor es el valor de F1. Todos los valores (en Hz) de F1 poseen un rango específico, esto se debe a que el comportamiento de los formantes y sus valores varían en función del tiempo, y también por las características fisiológicas de cada aparato fonatorio. En el caso de F2, esta vocal presenta valores intermedios en comparación a las demás vocales.

En términos generales los valores de F1 siempre se encuentran en rangos más bajos que los de F2.

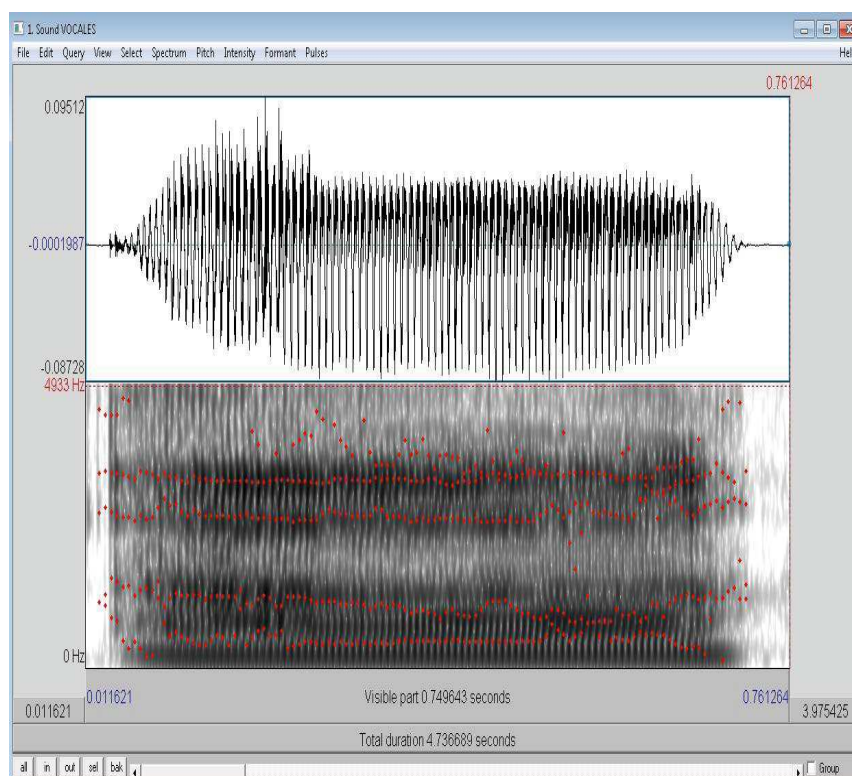


Figura 16.

Formantes del fonema /a/ hablado por un hombre.

- En el análisis de formantes de la vocal “e” se presentan algunas observaciones. La primera consiste en la curva de F1, ya que este fonema posee un valor de frecuencia de F1 más bajo que el fonema /a/

pero más alto que las demás vocales, a excepción del fonema /o/. Esto se debe principalmente a la posición vertical que adopta la lengua en la cavidad oral, ya que asciende en comparación al fonema /a/ por lo que su valor de F1 se reduce.

En el caso de F2, esta vocal presenta valores muy altos en comparación a las demás vocales, a excepción de la /i/. La principal razón de esto es la posición anterior de la lengua, ya que al producirse este fonema se tiende a acercar la lengua hacia la parte delantera de la cavidad oral.

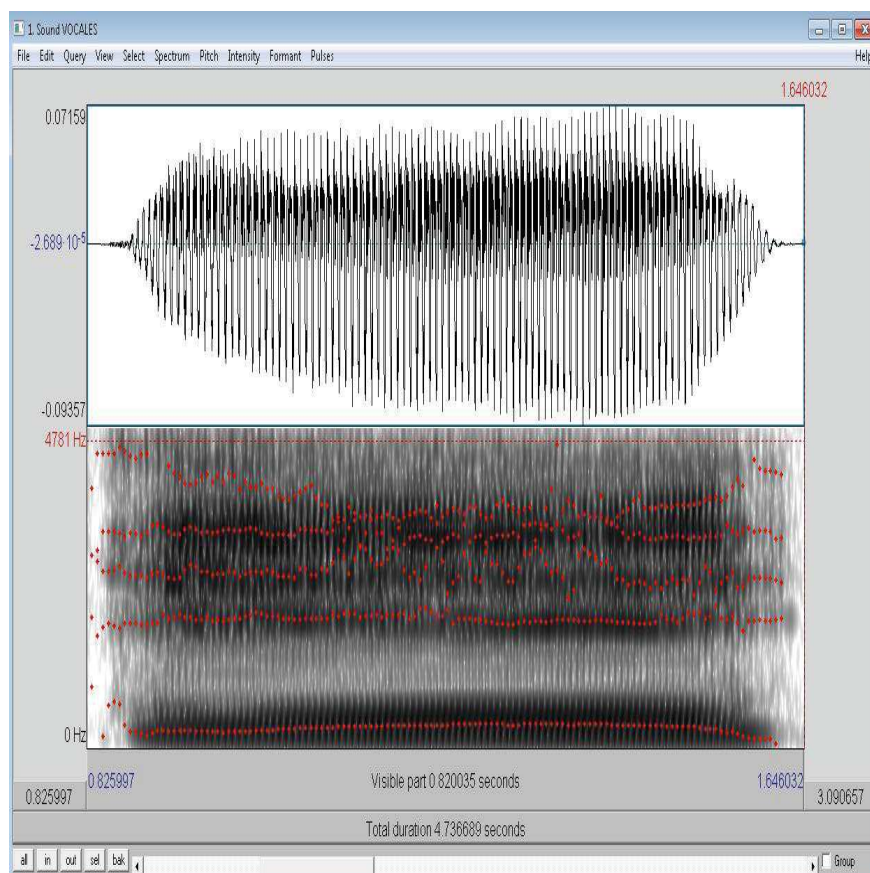


Figura 17.

Formantes del fonema /e/ hablado por un hombre.

- La vocal “i” presenta un comportamiento de formantes con las siguientes observaciones: La curva de F1 posee los valores más bajos de todas las vocales, junto con la vocal “u”. Esto se debe a la posición vertical de la

lengua, ya que la elevación que esta posee al pronunciar este fonema es mucho mayor que en el caso del fonema /a/, por ejemplo.

En el caso de F2, esta vocal presenta la curva con valores más altos en comparación a las demás vocales. La principal razón de esto es la posición anterior de la lengua en la cavidad oral.

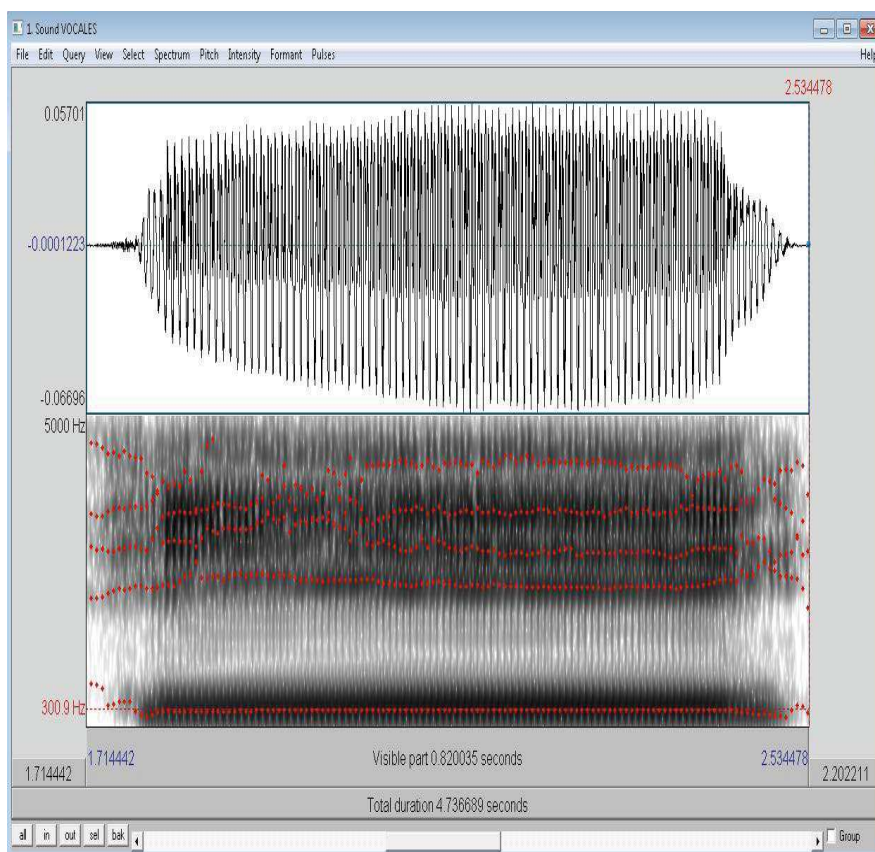


Figura 18.

Formantes del fonema /i/ hablado por un hombre.

- En el análisis de formantes de la vocal “o” se pueden mencionar las siguientes observaciones: La curva de F1 posee valores un poco más altos que los dos casos anteriores, únicamente superados por el caso del fonema /a/, como ya se mencionó anteriormente. Esto se debe principalmente a la posición vertical de la lengua ya que la elevación que esta posee al pronunciar este fonema es casi tan reducida como en el caso de la /a/.

En el caso de F2, esta vocal presenta la curva con valores más bajos junto con la vocal “u” en comparación a las demás vocales. La principal razón de esto es la posición central anterior de la lengua en la cavidad oral.

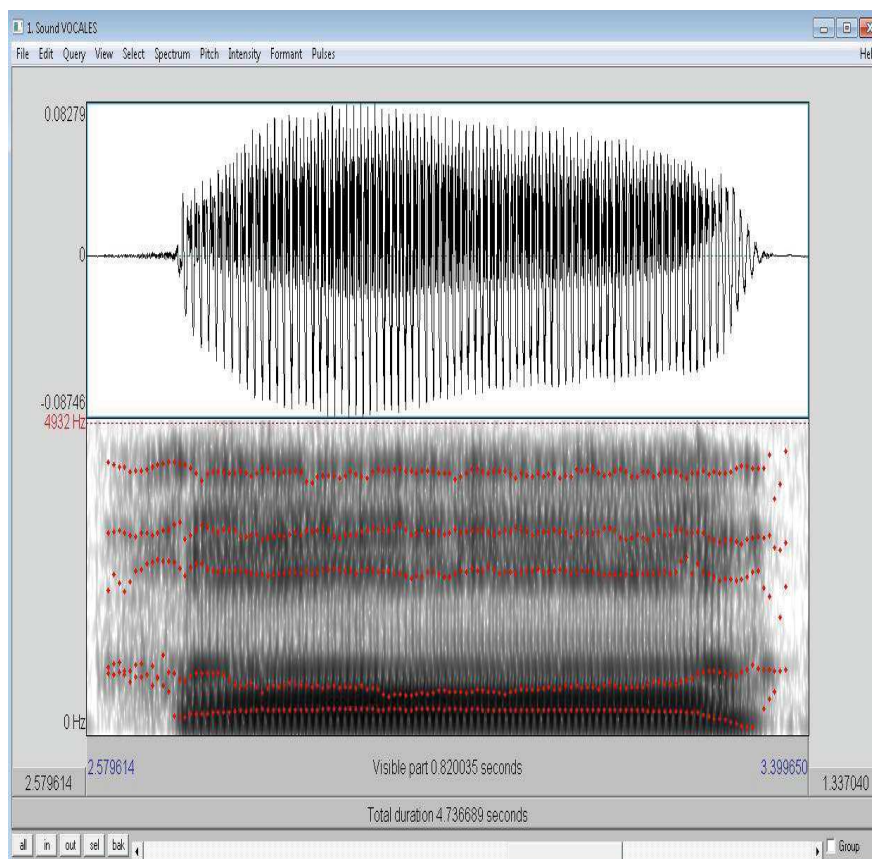


Figura 19.

Formantes del fonema /o/ hablado por un hombre.

- En el caso de la vocal “u” se pueden mencionar las siguientes observaciones en el comportamiento de sus formantes: La curva de F1 posee los valores de frecuencia más bajos de todas las vocales junto con la vocal “i”, como ya se mencionó anteriormente. Esto se produce debido a la posición vertical de la lengua, o sea la elevación que esta posee al pronunciar este fonema.

En el caso de F2, esta vocal presenta la curva con valores más bajos en comparación a las demás vocales junto con la vocal “o”. La principal

razón de esto es la posición posterior que adopta la lengua en la cavidad oral.

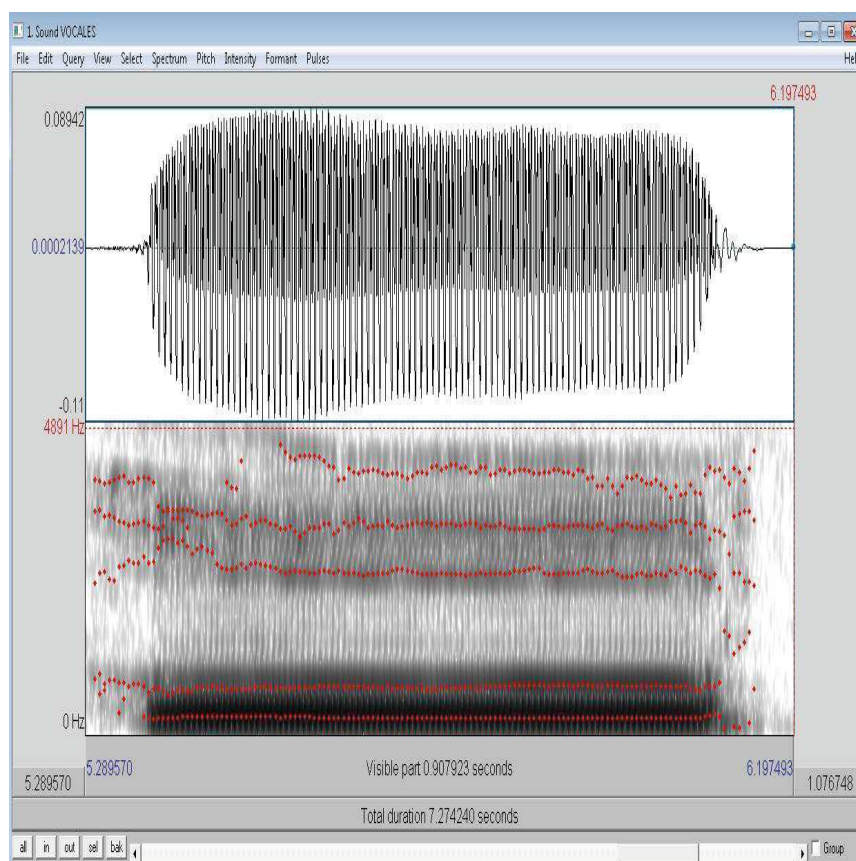


Figura 20.

Formantes del fonema /u/ hablado por un hombre.

Es necesario dejar constancia de que los valores de formantes (en Hz) no poseen una relación directa con la frecuencia fundamental con que se pronuncie un fonema. Debido a esto no debe asociarse una sonoridad aguda de cierta vocal, con sus valores de frecuencia de formantes. Un ejemplo de esto corresponde al caso del fonema /i/, el cual tiende a ser más agudo que el fonema /o/, sin embargo su valor de F1 es menor que este último.

También es importante aclarar que en las figuras anteriores, existen ciertos formantes que no se encuentran dentro de una curva específica y pueden visualizarse como “saltos” en las tendencias. Estas variaciones no constituyen una representación real de un formante sino que deben ser consideradas como

errores de cálculo del LPC por parte del software PRAAT, el cual advierte acerca de ciertos casos en los que se podría producir este tipo de saltos, y menciona al usuario que no los considere para su análisis.

La separación entre las curvas de F1 y F2 también constituye un factor de diferenciación entre las vocales. Como observaciones generales se pueden mencionar las siguientes:

- En el fonema /a/ las curvas de F1 y F2 siempre se encuentran relativamente juntas, con una separación aproximada entre 300 y 400 Hz.
- En el caso del fonema /e/ la principal observación es la gran separación entre la curva de F1 y de F2. Los valores correspondientes a dicha separación oscilan entre 1400 y 1600 Hz.
- La diferencia entre las curvas de F1 y F2 en el fonema /i/ es la mayor de todas las vocales. El intervalo de diferencia presente entre estas dos curvas se encuentra entre 2000 y 2200 Hz aproximadamente.
- En el caso del fonema /o/ se puede observar que la curva de F1 y F2 se encuentran muy cercanas, debido a toda la explicación dada anteriormente sobre las posiciones de la lengua. Los valores de este intervalo oscilan entre 250 y 350 Hz aproximadamente.
- Finalmente en el caso de la /u/, se puede observar que las curvas de F1 y F2 se encuentran muy próximas entre sí debido a las posiciones que adopta la lengua, como ya se mencionó.

Tener una idea clara del comportamiento temporal de los formantes de cada fonema, sirve para poder clarificar el comportamiento específico de cada uno de ellos. Sin embargo, también es de mucha ayuda obtener un valor que describa una información general para poder hacer comparaciones más simples. Considerar valores promedio de formantes es una buena opción para poder visualizar mejor los resultados y poder clarificar tendencias generales.

La tabla 9 muestra una comparación entre los valores mínimos y máximos de F1 y F2 en las vocales.

Tabla 9.

Valores límite de F1 y F2 en las vocales del castellano.

Vocal	F1 (Hz)		F2 (Hz)	
	Valor mínimo	Valor máximo	Valor mínimo	Valor máximo
a	660	860	900	2300
e	450	650	1600	2500
i	230	430	1600	3000
o	430	680	600	1600
u	200	420	500	1500

2.1.2.2 Observaciones generales del análisis de vocales

Como se pudo observar en el apartado anterior, cada vocal posee características específicas del comportamiento de sus formantes. En base a esto se puede afirmar que sus cualidades son únicas y además se podría generalizar las tendencias de comportamiento con valores promedio para poder establecer de una mejor manera las diferencias entre ellas.

En la figura 21 se presenta una gráfica en donde se generaliza el comportamiento de F1 y F2 en las vocales de una manera comparativa entre ambos formantes. El eje horizontal corresponde a los valores de frecuencia (en Hz) de F1, mientras que el eje vertical a los de F2. Además, se debe recalcar que existen cinco óvalos que corresponden a todas las vocales, y abarcan los rangos en que los valores de cada formante varían en relación a cada persona. El círculo interno de cada óvalo corresponde al valor promedio.

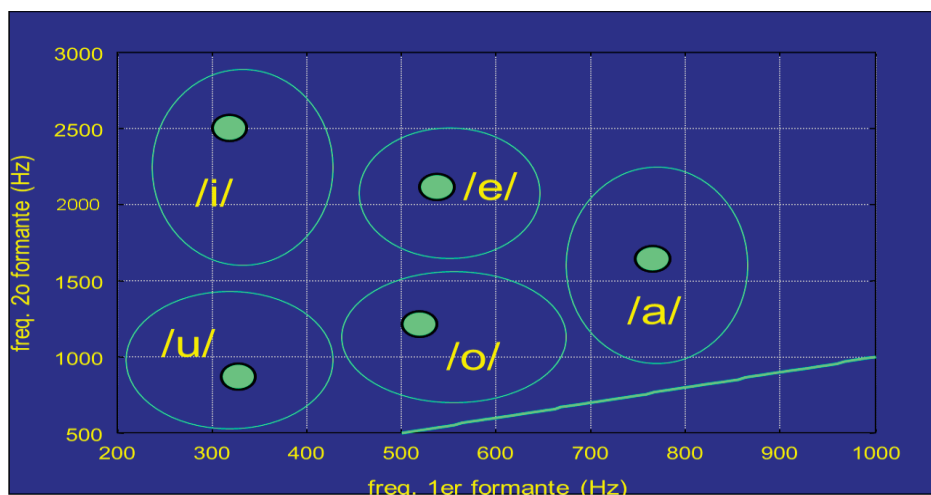


Figura 21.

Gráfica de valores promedio y límite de F1 y F2 en las vocales del castellano. Tomada de Procesamiento y comprensión de señales de audio, 2002, p. 29.

Adicionalmente a esto, la tabla 10 contiene valores promedio de F1 y F2 de las cinco vocales del castellano, habladas por tres hombres y tres mujeres evaluados en este proyecto:

Tabla 10.

Valores promedio de F1 y F2 en hombres y mujeres, obtenidos en el presente proyecto.

Vocal	F1 (Hz)		F2 (Hz)	
	Hombre	Mujer	Hombre	Mujer
a	700	820	1.300	1.550
e	410	600	2.000	2.100
i	270	340	2.300	2.350
o	380	550	850	1.030
u	290	350	760	1.020

Como se puede observar en la tabla 10, los valores de F1 y F2 siempre son más altos en las mujeres que en los hombres, lo cual corrobora la teoría expuesta anteriormente sobre las dimensiones del aparato fonatorio promedio en hombres y en mujeres.

2.1.3 Cálculo de diferencias entre formantes

Al contrario de lo que se podría pensar, cuando una vocal es dicha en conjunto con una consonante, los valores de formantes varían dependiendo del fonema que la acompañe. Por lo general la tendencia de F1 y F2 en una sílaba CV, es que los formantes de la vocal se reduzcan en comparación a su comportamiento cuando se pronuncia al fonema solo. Las variaciones dependen de algunos factores y son distintas dependiendo de la consonante que se encuentre antes. La principal conclusión realizada acerca de este comportamiento, es que debido a que la consonante es pronunciada antes que la vocal, los elementos del aparato fonatorio involucrados en la generación del fonema se comportan de una manera particular, y al momento de pronunciar la vocal, el comportamiento de los elementos en común entre ambos fonemas se ve afectado por el mínimo tiempo de duración de todo este proceso.

En base a todo el estudio realizado en el apartado anterior, se decidió buscar la existencia de alguna relación entre los formantes de una vocal específica, a pesar de ser pronunciada en combinación con distintas consonantes. Luego de analizar algunas posibilidades, se planteó la opción de identificar la diferencia aritmética entre los valores promedio de formantes adyacentes. Esta diferencia debería realizarse entre el valor promedio de cada curva de formante con su formante anterior, es decir, $F2-F1$, $F3-F2$, y $F4-F3$.

Una vez realizado este proceso, se obtuvieron resultados no tan satisfactorios en referencia a la siguiente hipótesis planteada: “a pesar de que los formantes de la vocal se vean afectados por la consonante, la diferencia entre ellos se debe mantener igual”. Esto se basó en la suposición de que la consonante afectaba a todos los formantes de una vocal en igual proporción. A continuación se muestran algunos casos de vocales dichas en conjunto en una misma sílaba CV con las consonantes: b, c, d, f, g, en donde se puede apreciar

lo mencionado. La tabla 11 corresponde a una mujer, mientras que la tabla 12 corresponde a los valores hablados por un hombre.

Tabla 11.

Diferencia de formantes de las vocales habladas por una mujer.

Vocal	Consonante	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)	F2-F1 (Hz)	F3-F2 (Hz)	F4-F3 (Hz)
a	b	897	1568	2987	4009	671	1419	1022
	c	902	1731	2953	3958	828	1223	1005
	d	810	1479	2678	3579	669	1199	901
	f	868	1569	2935	3959	701	1366	1024
	g	851	1593	2966	4002	742	1373	1037
e	b	463	2474	3011	3978	2010	538	967
	c	472	2420	2972	4003	1947	552	1031
	d	465	2418	2985	4027	1954	566	1042
	f	527	2266	2979	4150	1739	713	1172
	g	478	2435	2909	3953	1957	474	1045
i	b	313	2735	3415	4067	2422	679	652
	c	349	2702	3527	4174	2353	825	647
	d	289	2670	3321	4274	2381	651	952
	f	399	2728	3269	4239	2328	541	970
	g	299	2699	3081	3985	2400	382	904
o	b	492	945	2945	3839	454	2000	893
	c	479	1006	2997	3988	526	1991	992
	d	455	956	2863	3792	501	1906	929
	f	525	965	2927	3782	440	1962	855
	g	508	960	2899	3697	452	1938	798
u	b	384	829	2777	3922	446	1947	1146
	c	400	890	2739	4295	489	1849	1556
	d	332	842	2826	3889	510	1984	1063
	f	408	814	2609	4081	406	1795	1472
	g	358	770	2836	3947	412	2066	1111

En el caso de los hombres los valores por lo general son más bajos, pero también mantienen relativa igualdad en la diferencia de formantes.

Tabla 12.

Diferencia de formantes de las vocales habladas por un hombre.

Vocal	Consonante	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)	F2-F1 (Hz)	F3-F2 (Hz)	F4-F3 (Hz)
a	b	648	1201	2477	3464	553	1276	987
	c	576	1406	2453	3430	830	1047	977
	d	670	1386	2605	3616	715	1220	1011
	f	587	1221	2319	3242	633	1098	922
	g	625	1324	2368	3550	699	1044	1182
e	b	362	2021	2619	3599	1659	599	980
	c	391	2150	2641	3478	1759	491	837
	d	392	2049	2635	3648	1657	586	1013
	f	481	1996	2678	3461	1515	682	783
	g	383	2131	2701	3628	1748	570	927
i	b	251	2328	2977	3682	2077	650	705
	c	254	2466	3001	3561	2212	536	560
	d	255	2377	3034	3706	2122	657	672
	f	248	2256	2811	3427	2008	555	615
	g	264	2367	3161	3661	2103	794	500
o	b	393	792	2635	3423	399	1843	788
	c	369	902	2767	3654	533	1865	887
	d	430	938	2745	3505	509	1807	760
	f	423	977	2719	3577	554	1742	858
	g	409	841	2653	3467	432	1812	813
u	b	329	1012	2990	3820	683	1978	830
	c	275	754	2649	3472	479	1895	823
	d	302	913	2610	3502	611	1697	891
	f	268	944	2607	3613	676	1663	1006
	g	303	781	2756	3734	478	1975	978

Como se puede observar en las tablas 11 y 12, las diferencias de formantes presentan valores similares en la mayoría de los casos. Además, se debe mencionar que existe un margen de error aceptable, ya que nunca se podrían tener valores exactamente iguales en el 100% de los casos, debido a que cada consonante influye de manera distinta en su vocal adyacente. Sin embargo, existen algunas sílabas en donde la vocal sobrepasa los márgenes tolerables.

A pesar de lo expuesto anteriormente, se pudo confirmar que las diferencias de formantes no constituyen un patrón de reconocimiento certero para poder alimentar a la Red Neuronal Artificial que se encargará de identificar las tendencias en comportamiento de los fonemas. Las razones abarcan principalmente el hecho de que al poseer valores promediados de un comportamiento temporal variable, se está descartando una gran cantidad de información fundamental para poder plantear patrones de reconocimiento más claros y reales. Esta explicación será profundizada en capítulos posteriores en donde se describirá el entrenamiento de la ANN, así como su funcionamiento.

2.1.4 Análisis de los formantes de las consonantes

Analizar consonantes representa un análisis mucho más complejo que las vocales. Esto se debe a que es necesario realizar un estudio más profundo de los procesos inmersos en la generación de los fonemas, así como en la obtención de particularidades que se puedan presentar acerca del comportamiento de formantes, debido a que las mismas no son tan notorias como en el caso de las vocales.

Las consonantes se caracterizan por ser fonemas que presentan una obstrucción del flujo de aire en el aparato fonatorio. Además de esto, la teoría establece que una consonante puede ser sorda o sonora debido a la vibración o no de las cuerdas vocales. Esta última característica hace que el estudio sea mucho más complejo ya que existen dos clasificaciones generales de estos fonemas que se podrían considerar en el análisis.

Como ya se mencionó anteriormente, el presente proyecto se enfoca en el análisis de sílabas CV, por lo que la consonante siempre ocupará la posición

inicial de la sílaba a reconocer. Al producirse este hecho se consigue que la duración de la consonante sea más o menos constante, o en otras palabras, no sea de pronunciación prolongada como podría suceder en el caso de que se ubique al final de una sílaba. Además de todo esto, también se consigue excluir del estudio los alófonos que ciertos fonemas poseen, y así poder establecer un análisis mucho más centrado y específico de los fonemas de consonantes en posición inicial.

Otro factor importante para considerar en el análisis de sílabas CV es la coarticulación entre la consonante y la vocal adyacente. La coarticulación toma como principal referencia en la unión CV, las características que se producen al generarse dichos fonemas juntos, las cuales difieren en comparación a la pronunciación de cada fonema por separado de manera aislada.

El fenómeno de coarticulación conlleva a la generación de un factor muy importante a la hora de analizar el comportamiento de los formantes en sílabas, denominado como “transición”. La transición consiste en el cambio gradual (en la mayoría de los casos) que existe en la curva de ciertos formantes al momento en que termine la consonante y empiece la vocal, en el caso de sílabas CV. Si bien esta característica brinda una particularidad a cierta combinación de fonemas, existen ciertos casos en donde no existe una transición visible. Esto se debe principalmente a la consonante presente antes de la vocal, y aunque esto podría significar la carencia de un factor determinante para el análisis, también se puede considerar como una ayuda para la identificación de ciertas sílabas.

Las transiciones presentes en las distintas sílabas fueron objeto de análisis en el estudio de los fonemas coarticulados. Un punto fundamental para poder obtener un resultado satisfactorio en dicho análisis fue considerar de manera óptima el comienzo y finalización de las transiciones, lo cual constituyó una tarea no tan fácil de realizar en ciertos casos debido a la particularidad de comportamiento de formantes de ciertos fonemas, sobre todo sordos.

Este caso se puede especificar en la siguiente comparación entre dos fonemas distintos, como son /b/ y /s/. Para este ejemplo se presentan las figuras 22 y 23, en donde se muestran dos espectrograma en los que se puede observar el comportamiento de los formantes de ambos fonemas al conformar una sílaba con la vocal /a/, en cada uno de los casos. Entre las diferencias gráficas presentes en estas dos sílabas se puede apreciar el comportamiento distinto de formantes, así como las distintas transiciones consonante-vocal que poseen cada sílaba. Cabe recordar que son dos sílabas conformadas por la misma vocal pero en combinación de consonantes completamente distintas, ya que una es sonora (“BA”) y la otra es sorda (“SA”).

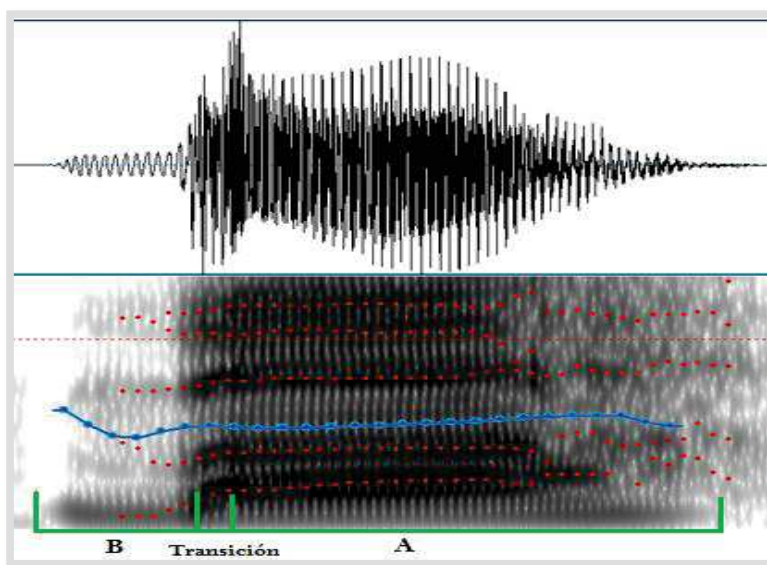


Figura 22.

Comportamiento de formantes de “ba” pronunciado por una mujer.

En la figura 22 se puede apreciar que la transición entre /b/ y /a/ es más pronunciada en el primer formante que en los demás, ya que posee una curva mucho más inclinada en comparación a formantes superiores. La tendencia de la transición en F1 es siempre ascendente, y en términos generales las transiciones de esta consonante siempre son de corta duración en comparación a las demás consonantes, como se verá más adelante.

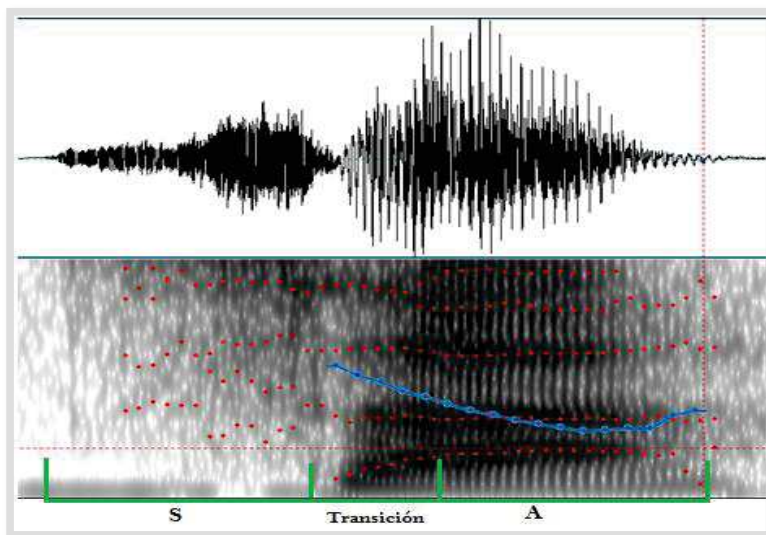


Figura 23.

Comportamiento de formantes de “sa” pronunciado por una mujer.

En la figura 23 se puede observar que la transición entre la consonante y la vocal no es tan notoria como en el caso anterior. Esto se debe principalmente a que el comportamiento de formantes de la /s/ es muy poco claro e inestable. A pesar de todo esto, en el caso de F1 no existe una transición notoria, ya que existe un “salto” brusco entre los formantes de la consonante y la vocal. Esto significa que existen consonantes (como la “s”), que no poseen una transición CV notoria, lo cual corresponde a una característica propia de identificación.

Como ya se explicó anteriormente la vocal ve afectado el comportamiento de sus formantes debido a la consonante anterior, comparado con sus características propias al pronunciarla sola. En el caso de consonantes, establecer una comparación entre el comportamiento de formantes individual de cada fonema y en conformación de sílabas, no es una tarea trivial. Esto se debe a que al ubicarse en una posición inicial, las consonantes siempre deben ir acompañadas de una vocal para poder conformar una sílaba. Además, es casi imposible analizar una consonante inicial sola debido a la corta duración que esta posee, y al intentar decir las de manera aislada muchas personas tienden a exagerar la duración y cambiar la sonoridad de los fonemas, produciéndose resultados poco reales y erróneos.

Varios autores (Fernández, 2004; Massone 1988) de distintos lugares e idiomas del mundo han realizado una variedad de estudios de las consonantes, llegando a conclusiones similares en algunos casos y diferentes en otros. A pesar de esto, el factor común en dichos estudios ha sido separar a las consonantes en distintos grupos de acuerdo a su clasificación, generalmente basándose en el modo de articulación.

A continuación se presenta un estudio de las principales características de las consonantes en base a su clasificación por el modo de articulación. En este estudio se consideran parámetros de generación de cada fonema, así como el análisis del comportamiento de los formantes en cada uno de los casos.

2.1.4.1 Oclusivas

Los fonemas oclusivos del castellano poseen una característica muy específica en su sonoridad, la cual consiste en una liberación repentina de presión justo después de una obstrucción completa del paso del aire. Por esta razón muchos autores los consideran también con el nombre de *explosivos*.

Estos fonemas se caracterizan por producirse en un tiempo muy corto, en donde los valores no sobrepasan los 120 ms de duración, y generalmente se encuentran en un rango de 80-100 ms. Todo esto influye en el análisis de sus características, ya que a diferencia de otros fonemas, al existir una duración menor la cantidad de formantes en estas consonantes también es menor.

Al igual que otras consonantes en el castellano, las oclusivas también poseen dos tipos generales de subdivisión ya que pueden ser sordas o sonoras. En términos generales se podría establecer que los formantes que influyen en la caracterización de estas consonantes son únicamente F1 y F2. Esto se debe a que dichos formantes son los que mayor información de reconocimiento presentan, ya que el comportamiento para formantes superiores es aleatorio en la mayoría de los casos.

Una consideración muy importante a tener en cuenta en el análisis de estas consonantes es el hecho de que todo fonema sonoro presenta un alineamiento mucho más estable en la curva de formantes; esto se debe entre otras cosas a

que las cuerdas vocales se encuentran participando activamente en la generación de dicho fonema. En contraste, todo fonema sordo presenta características opuestas al caso anterior.

Los fonemas oclusivos del castellano son /b/, /d/, /g/, y /p/, /t/, /k/. Los tres primeros mencionados corresponden a la subdivisión de “sonoros”, y los otros a la subdivisión de “sordos”. Para poder realizar un análisis mas a fondo de este tipo de consonantes es necesario dividir las en las dos secciones mencionadas, como se muestra a continuación.

Oclusivas sonoras

Este tipo de fonemas poseen ciertas diferencias en comparación a las otras consonantes oclusivas. La principal diferencia radica en su generación, ya que estos fonemas son sonoros; además de esto también se presentan diferencias en las características y comportamiento de los formantes. En términos generales las consonantes oclusivas sonoras poseen un comportamiento de formantes un poco más estable que las sordas.

La duración de estos fonemas depende de algunos factores como se explicará a continuación, sin embargo existen rangos que se podrían especificar para poseer una idea más clara de las características que estas consonantes poseen. Existen casos, como en el fonema /b/, en donde la cantidad de muestras de formantes no es grande debido a la corta duración de la consonante. En general el rango promedio de duración de estos fonemas es de 100-170 ms, aunque en algunos casos dichos valores podrían variar dependiendo de la pronunciación de cada persona.

En la figura 24 se presenta una gráfica de la sílaba “ba” hablada por un hombre, en donde se pueden apreciar algunas características que serán ampliadas de manera más profunda en el siguiente análisis.

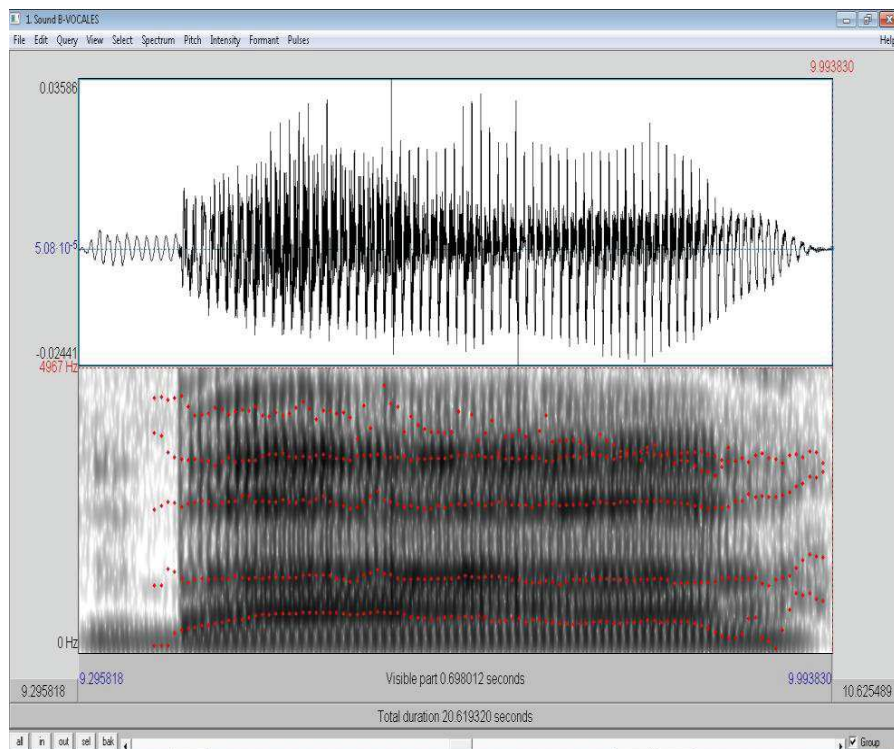


Figura 24.

Gráfica de formantes de la sílaba “ba” dicha por un hombre.

En el espectrograma de la figura 24 se puede observar que la mayor concentración de energía de formantes de la sílaba “ba” se presenta a partir de la vocal /a/, lo cual se puede apreciar en el espectrograma justo después de que se produzca una línea vertical en donde el color negro se hace más oscuro.

En el gráfico de amplitud vs tiempo ubicado en la parte superior de la ventana, se puede observar la forma de onda de la sílaba hablada. En la parte izquierda de dicho gráfico se puede ver una forma de onda de menor amplitud con ciclos un poco más periódicos que los de mayor amplitud. Dicha forma de onda corresponde a un segmento de la /b/ denominado como “murmullo nasal”. En algunas personas el murmullo nasal es mayor y en otras es menor, no existiendo una norma acerca de su duración.

Si bien esta característica es propia del fonema de modo que puede llegar a considerarse como normal (en la mayoría de los casos), su duración se relaciona directamente con los costumbrismos del hablante.

Justo después del murmullo nasal empieza la liberación de aire del fonema y es donde los formantes empiezan a alinearse con la vocal siguiente. Vale recordar que la /b/ es una consonante bilabial según el lugar de articulación ya que los labios se juntan para poder liberar el aire, lo que hace que la mayor concentración de energía se produzca a partir de ese momento y sea más notoria en el comienzo de la vocal.

En el caso de la /d/, la mayoría de observaciones realizadas para la /b/ se repiten. El murmullo nasal mencionado para el caso anterior también se produce en este fonema, siendo determinado por la característica propia del hablante. Cabe recalcar que este murmullo nasal se presenta fundamentalmente en el caso de que la consonante se encuentre en posición inicial, que es el objeto de estudio del presente proyecto.

La principal razón de esta aclaración es el hecho de que en el caso de que exista algún fonema antes de una consonante oclusiva sonora, el murmullo nasal desaparece casi en su totalidad debido a que la coarticulación presente hace que el fonema anterior prevalezca sobre él.

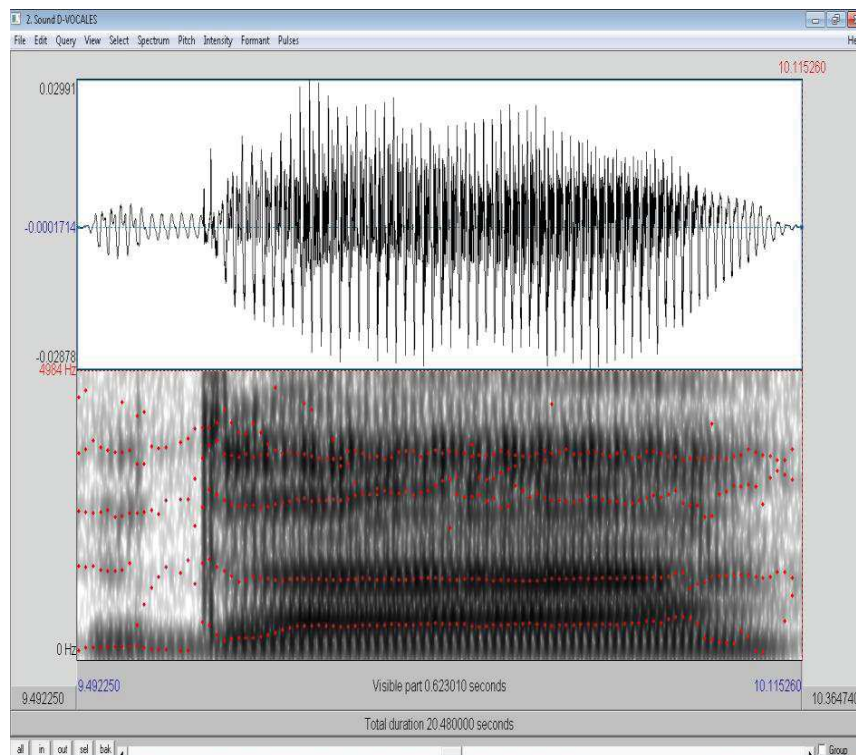


Figura 25.

Gráfica de formantes de la sílaba “da” dicha por un hombre.

Una diferencia entre la /b/ y la /d/ es que la liberación de energía no es tan fuerte en esta última como en la primera. La principal causa de esta observación se basa en que la /d/ es un fonema alveolar según el lugar de articulación, debido a que la punta de la lengua toca la región alveolar de la cavidad oral. Esto hace que no exista una obstrucción tan radical del flujo de aire como en el caso de la /b/ en donde los labios se cierran por completo. Esto origina un comportamiento un poco distinto de los formantes en la parte final del murmullo nasal, ya que la amplitud de estos es un poco mayor que en el caso anterior.

Con respecto al fonema /g/ se debe mencionar que la principal diferencia ante los fonemas anteriores es que el murmullo nasal tiende a desaparecer en la mayoría de los casos. A pesar de esto pueden existir ciertos hablantes que lo realicen con cierta frecuencia debido a los costumbrismos que puedan haber adquirido a lo largo de su niñez o por alguna influencia geográfica.

En términos generales el fonema /g/ posee mayor concentración energética de sus formantes en comparación a las otras consonantes oclusivas sonoras. Esto se puede apreciar en el espectrograma presente en la gráfica mostrada a continuación, en donde se observan zonas con mayor oscuridad en la parte izquierda de la gráfica, la cual corresponde a la consonante especificada. Además de todo lo dicho anteriormente, otra observación muy importante en esta consonante es que la transición consonante-vocal es de mayor duración que en las consonantes anteriores, principalmente en el primer formante.

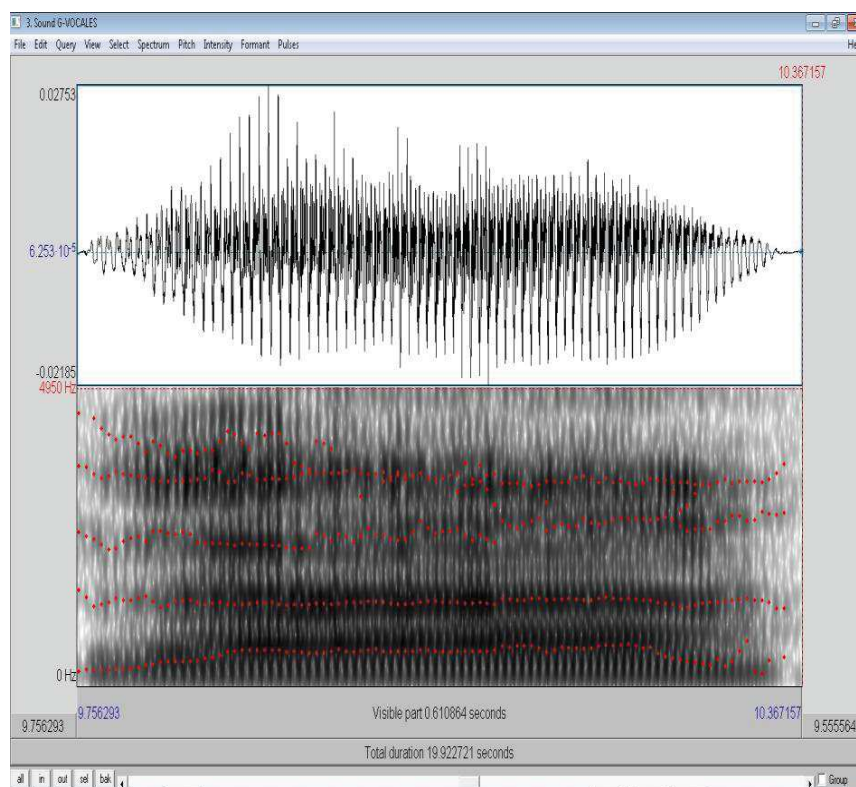


Figura 26.

Gráfica de formantes de la sílaba “ga” dicha por un hombre.

Oclusivas sordas

Los fonemas que corresponden a esta clasificación son /k/, /p/, y /t/. Al igual que en el caso anterior, existen ciertas diferencias entre cada uno de ellos como se mencionará a continuación.

El fonema /k/ corresponde, según el lugar de articulación, a la clasificación velar, debido a que la parte posterior de la lengua produce una oposición con el paladar blando. Esta característica de generación conlleva a que el comportamiento de sus formantes sea un poco distinto a los otros fonemas de la misma clasificación.

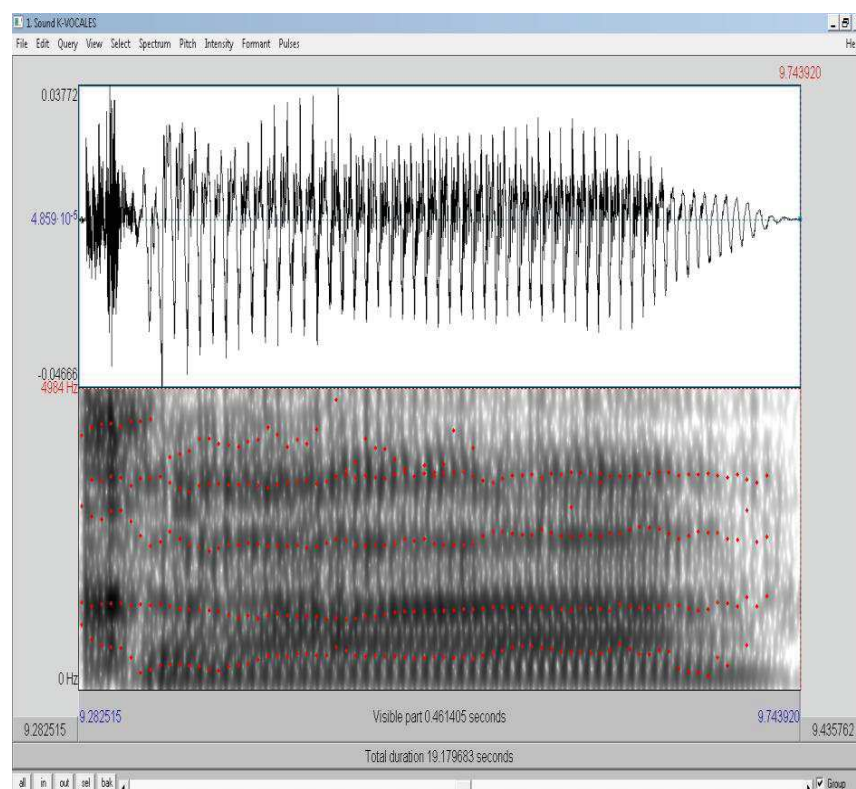


Figura 27.

Gráfica de formantes de la sílaba “ka” dicha por un hombre.

Como se puede observar en el espectrograma presente en la figura 27, la mayor concentración energética de los formantes no se produce en F1 como en casos anteriores sino que corresponde a las curvas de F2 y F5, lo cual se puede corroborar debido a la mayor obscuridad presente en dichas secciones del espectrograma.

A pesar de esto, es necesario observar que la curva de F5 de la consonante no posee una transición al momento en que aparece la vocal, esto debido a que la /a/ no posee suficiente energía en dichos formantes por lo que también se

puede notar que su curva es completamente variable y no posee estabilidad, por lo que F5 es completamente descartado para el análisis.

En base a las curvas de F1 y F2, las cuales son las más importantes para dar la caracterización del sonido, se pueden notar algunas observaciones importantes. La primera consiste en la transición consonante-vocal de F1. Como se puede observar, la frecuencia inicial de F1 siempre tiende a ser alta en comparación a las demás consonantes oclusivas. Además de esto la transición del primer formante es descendente ya que además de la característica mencionada sobre la consonante, el comportamiento de formantes de la vocal desciende en los valores de frecuencia, en este caso referente hacia el fonema /a/.

En el caso de F2 la transición suele ser casi recta, como se puede observar en la figura 27, aunque la tendencia de dicha transición puede ser descendente o ascendente en función de la vocal siguiente. Todas estas características permiten concluir que el fonema /k/ es un sonido de bastante energía en alta frecuencia.

En el caso de la consonante “p”, se debe mencionar que dicha consonante constituye tal vez el fonema más explosivo de todas las consonantes del castellano. Además de esto, su concentración energética suele considerarse de baja frecuencia, lo cual se demostrará a continuación en el análisis de la figura 28.

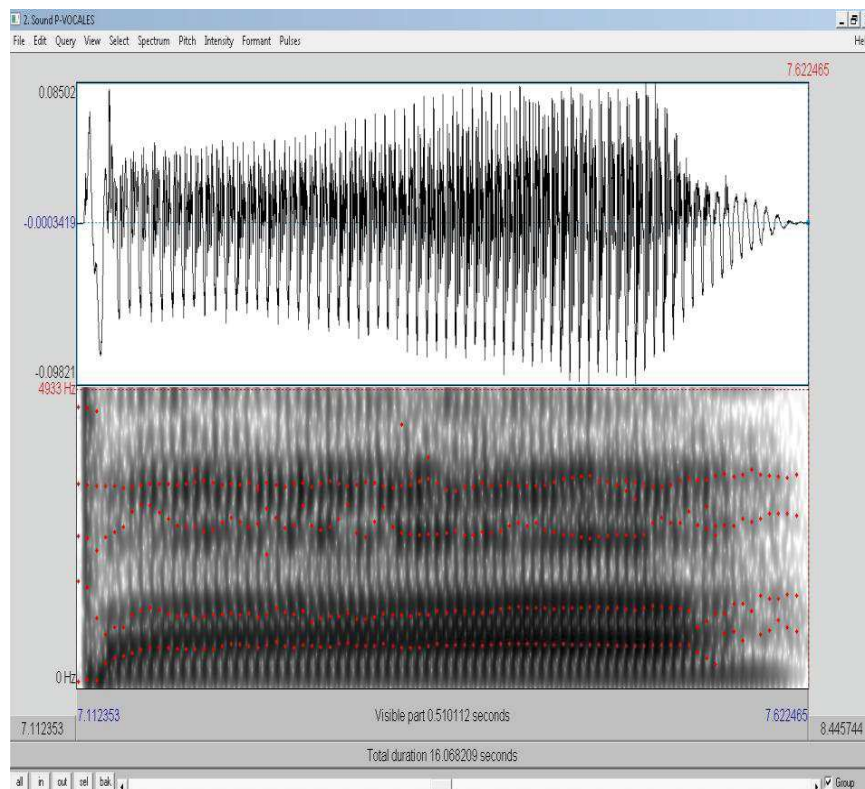


Figura 28.

Gráfica de formantes de la sílaba “pa” dicha por un hombre.

Al ser la /p/ un fonema bilabial según el lugar de articulación, es lógico pensar que al existir una oposición total al paso del flujo de aire debido a la unión de los dos labios la concentración energética será elevada.

En la gráfica de forma de onda correspondiente a la parte superior de la figura 28, se puede observar que la /p/ es un sonido sumamente explosivo con una liberación del flujo de aire repentino. La representación de la forma de onda lo demuestra claramente debido a la gran amplitud esta que posee en dicha sección.

Enfocándose en el espectrograma, se puede observar claramente que el primer formante de la consonante posee muy baja frecuencia. Además de esto también se muestra que la mayor concentración energética corresponde a F1, aunque todos los formantes poseen gran concentración energética debido a la característica explosiva de este fonema. Es necesario aclarar que no debe

relacionarse el término “explosivo” con un espectro frecuencial casi plano, sino con la liberación de energía en instantes cortos de tiempo.

Esta consonante posee una duración muy corta debido a su característica de generación, por lo que la transición consonante-vocal también es corta. La transición de F1 siempre tiende a ser ascendente, aunque en el caso de F2 la tendencia depende de la vocal siguiente.

La /t/ es una consonante alveolar según el lugar de articulación. Al no existir una oposición total del flujo de aire la concentración energética de este fonema no es tan elevada como en el caso anterior de la /p/, sin embargo su condición de consonante explosiva se mantiene. Al igual que en el caso anterior, el primer formante posee un valor de frecuencia bajo.

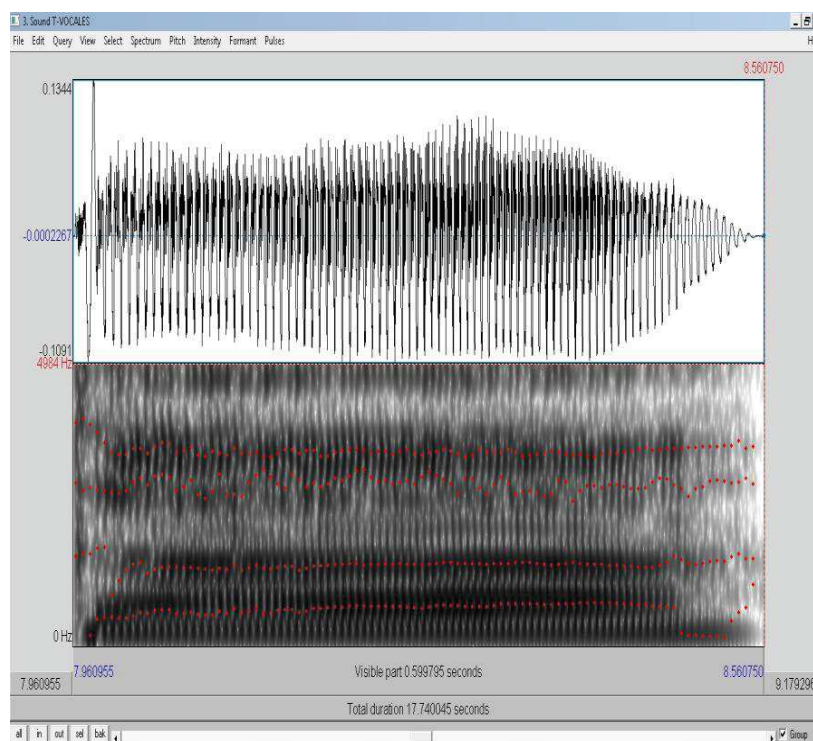


Figura 29.

Gráfica de formantes de la sílaba “ta” dicha por un hombre.

La forma de onda presente en la figura 29 permite corroborar la condición explosiva de esta consonante, ya que su amplitud es elevada. Esto se puede apreciar en la parte superior izquierda de dicha figura.

2.1.4.2 Fricativas

Las consonantes fricativas se caracterizan porque en su generación se produce una obstrucción parcial en la salida del flujo de aire. Según (Miyara, 2004) se establece que los fonemas fricativos son aquellos en donde el aire sale atravesando un espacio estrecho de la cavidad oral.

Estos fonemas poseen la misma particularidad de las consonantes oclusivas, ya que también pueden ser sonoros o sordos. Según la clasificación general de los fonemas del castellano, las consonantes fricativas son la /f/, /z/, /s/, /j/, y /b/, /d/, /y/, /g/. Siendo las cuatro primeras pertenecientes a la clasificación de “sordas” y las cuatro últimas a la clasificación de “sonoras”.

Si bien los fonemas /b/, /d/, y /g/ ya pertenecen a la clasificación de consonantes oclusivas, también se las incluye dentro de esta clasificación a dichos fonemas en posición no inicial, de entre las cuales se puede mencionar posición postvocálica, postvibrante y postlateral.

Estas consonantes no serán analizadas en esta clasificación, ya que en el presente proyecto únicamente se consideran a las consonantes en posición inicial de la sílaba, como ya se mencionó anteriormente.

Fricativas Sonoras

Dentro de esta clasificación se encuentra la /y/. Este fonema se caracteriza por que la punta de la lengua produce una oposición con la región alveolar, por lo que se la introduce en la clasificación de consonantes alveolares según el lugar de articulación (Miyara, 2004).

La generación de esta consonante involucra la vibración de las cuerdas vocales, sin embargo, su condición de alveolar produce un comportamiento un poco inestable en ciertos formantes.

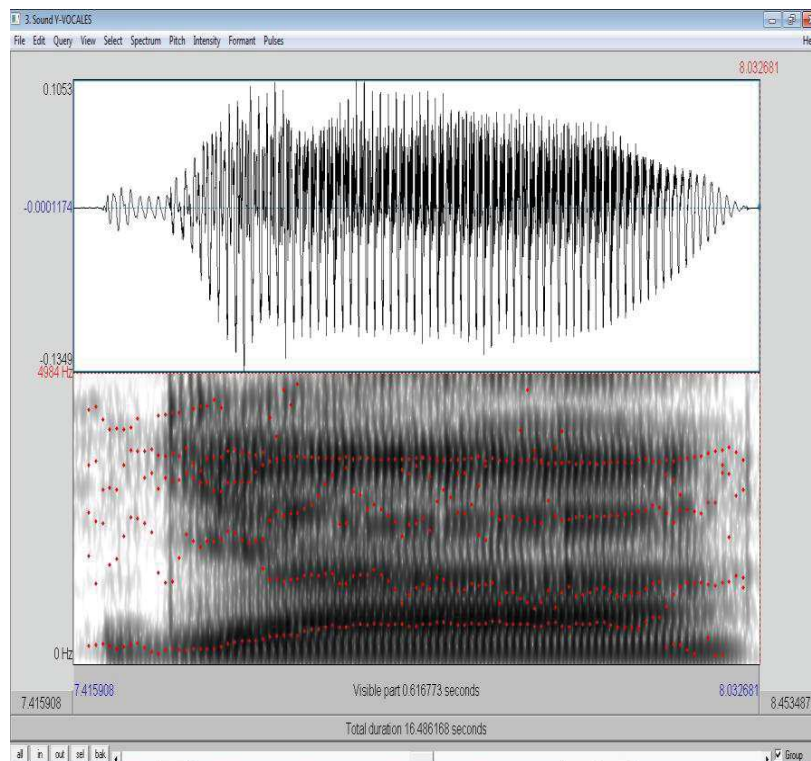


Figura 30.

Gráfica de formantes de la sílaba “ya” dicha por un hombre.

Como se puede observar en la figura 30, el primer formante es el que prevalece en la caracterización de esta consonante. Esto se puede notar rápidamente debido a dos factores principales:

El primero corresponde a la curva de formantes, ya que solo F1 posee una tendencia marcada acerca del comportamiento temporal debido a que los demás formantes poseen tendencias inestables o aleatorias.

El segundo factor corresponde a la concentración energética de los formantes. En la gráfica mostrada se puede apreciar que la curva de F1 es la única que posee una concentración energética elevada, ya que en todas los demás formantes no existe concentración energética suficiente para que los formantes puedan afectar las características del fonema.

Como se puede observar en la figura 30, la frecuencia del primer formante es baja. Con respecto a la transición consonante-vocal se puede mencionar que

su duración es un poco larga. Además, también posee una variación ascendente progresiva hasta que los formantes de la vocal se “estabilizan”. Esta característica es mayor dependiendo de la combinación con ciertas vocales, pero la tendencia es similar en todas ellas.

Fricativas Sordas

Dentro de esta clasificación se encuentran las consonantes /f/, /z/, /s/ y /j/, aunque la /z/ no será incluida en este análisis, como ya se explicó en la sección 2.1.1.

En lo referente al fonema /f/, se debe mencionar que esta consonante corresponde a la clasificación labiodental según el lugar de articulación. En su generación, se encuentran involucrados activamente los dientes superiores y el labio inferior, produciéndose una oposición entre ellos. El flujo de aire sale de manera parcial, produciendo un sonido de frecuencia relativamente alta.

Enfocándose en la parte acústica de este fonema, se puede observar que la concentración de energía de los formantes de esta consonante se presenta en alta frecuencia. El estudio realizado, además de la gráfica presentada más adelante (ver figura 31), así lo ratifica. Los formantes de este fonema poseen una tendencia poco estable, mostrando un gran número de irregularidades en las curvas de cada uno de ellos. Al ser una consonante fricativa, se puede establecer que la forma de onda de la /f/ no posee gran amplitud en comparación a otros fonemas explosivos. En términos generales la duración de esta consonante es mayor que las consonantes oclusivas. Si bien la duración de este fonema puede ser exagerada por ciertos hablantes, en condiciones normales la duración promedio no debería ser mayor que 220 ms.

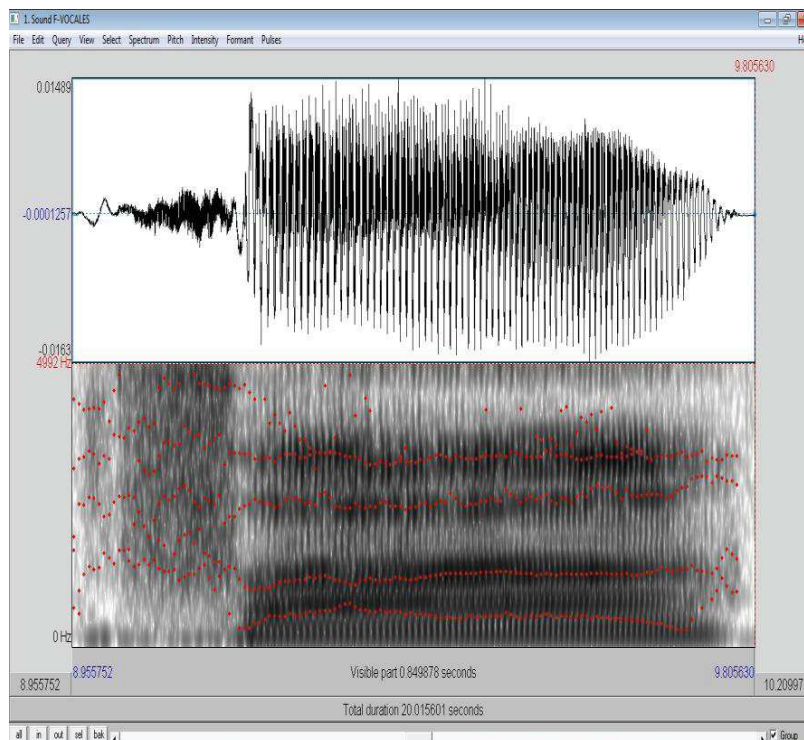


Figura 31.

Gráfica de formantes de la sílaba “fa” dicha por un hombre.

Además de todo lo dicho anteriormente, se puede observar en la figura 31 que la mayor concentración energética se produce en el quinto formante, lo cual es lógico de pensar debido a la sonoridad característica de este fonema. A pesar de esto, todos los formantes poseen una energía relativamente grande, como se puede observar en el espectrograma, debido a la coloración oscura de las curvas correspondientes a cada uno de ellos. Adicionalmente, se debe recalcar que en este tipo de consonantes es muy inusual poseer transiciones consonante-vocal notorias.

En lo referente a la consonante “s”, las consideraciones son similares a las del fonema /f/. La amplitud de esta consonante no es muy elevada, si bien la sonoridad sí es notoria (en algunos hablantes más que en otros). Además de esto, una característica similar de la /s/ con el fonema anterior, es la tendencia inestable y por momentos aleatoria que presentan las curvas de formantes a lo largo del tiempo. Otra similitud es el valor de frecuencia de sus formantes,

especialmente del primero; ya que en ambos fonemas, /s/ y /f/, la frecuencia de F1 siempre se encuentra por encima de 1000 Hz.

La duración de este fonema es similar al caso de la /f/, ya que el fonema /s/ en condiciones normales posee una duración promedio de aproximadamente 180-220 ms. Las transiciones consonante-vocal en este fonema son poco notorias y casi imperceptibles. Además, la concentración energética de los formantes del fonema /s/ también se presenta en alta frecuencia. Sin embargo, a diferencia del fonema anterior la energía presente en los primeros formantes de la /s/ es mucho menor, existiendo una mayor amplitud a partir del tercer y cuarto formante. Esto se puede apreciar en la gráfica presentada a continuación:

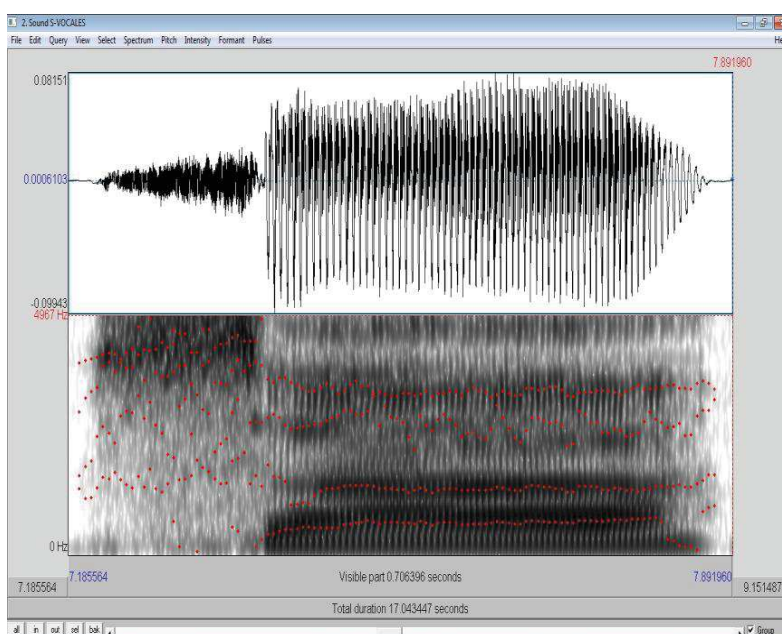


Figura 32.

Gráfica de formantes de la sílaba “sa” dicha por un hombre.

Como se puede apreciar en el espectrograma de la figura 32, existe una coloración oscura mucho más fuerte en la parte superior de la consonante. Además, aparecen ciertos formantes aislados ubicados en baja frecuencia por debajo de la curva de F1. Se debe recordar que dichos puntos no corresponden a la curva de formantes propiamente dicha, sino que deben ser considerados como errores de cálculo del software.

En el caso del fonema /j/, se deben mencionar las tendencias más estables que poseen las curvas de sus formantes. Al igual que los fonemas fricativos anteriores, la concentración energética es mayor en alta frecuencia pero en menores cantidades, como se puede observar en la figura 33:

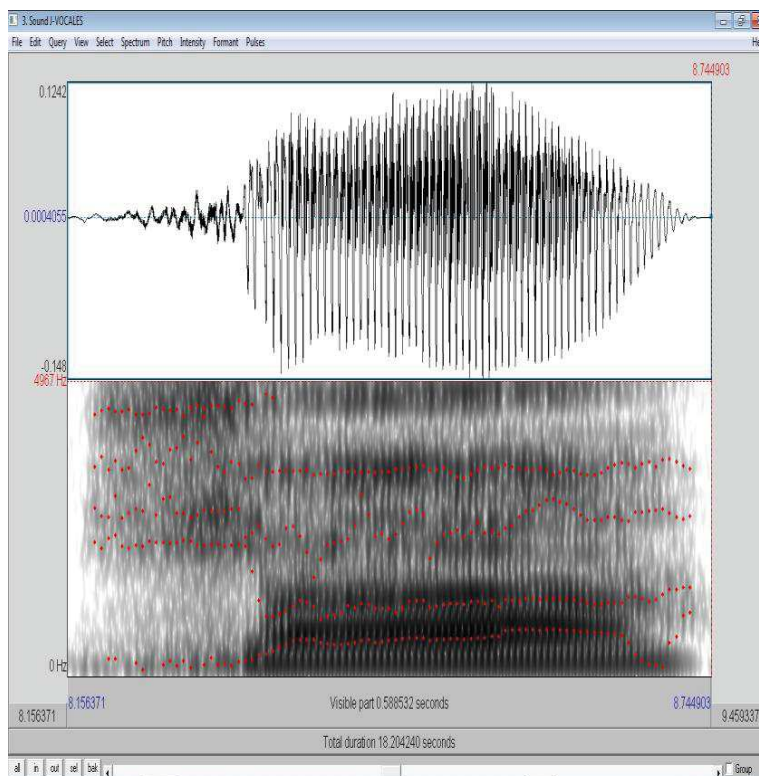


Figura 33.

Gráfica de formantes de la sílaba “ja” dicha por un hombre.

La característica principal de este fonema es que el primer formante posee una frecuencia baja y esto produce una notable separación entre las curvas de F1 y F2. Al igual que en el caso de los fonemas fricativos anteriores, esta consonante tiene la característica de poseer una relativa baja amplitud.

2.1.4.3 Laterales

Los fonemas laterales en el castellano son la /l/ y /ll/, los cuales se caracterizan principalmente por una obstrucción parcial de la lengua hacia el centro de la boca, por lo que el aire sale por los lados de esta. Para el presente análisis no se considerará a la /ll/ debido a las razones explicadas en la sección 2.1.1.

La // es un fonema que posee una característica especial en su curva de formantes, basada en la transición existente entre la consonante y la vocal. Este fonema es considerado como alveolar dentro de la clasificación según el lugar de articulación, lo cual influye en cierta manera en el comportamiento de sus formantes. La principal característica de este fonema es la poca duración que posee, ya que en términos generales ambos fonemas laterales poseen corta duración. Los valores promedio de este fonema se encuentran entre 70-90 ms.

Otra característica importante a mencionar en esta consonante es que las curvas de F1 y F2 poseen una tendencia estable. Además, es necesario recalcar que la frecuencia del primer formante siempre es baja.

Con respecto a las transiciones consonante-vocal, este fonema posee ciertas características importantes de mencionar. La transición en la curva del primer formante posee una duración no tan corta, pero la principal característica es la baja frecuencia con la que esta empieza.

Si bien la duración de una transición es un factor importante de considerar en ciertos casos, existen otros en donde la frecuencia de inicio y final de dicha transición pueden ser indispensables para obtener una caracterización del fonema. En este caso la frecuencia final de dicha transición es casi imposible de determinar debido a que el valor correspondiente depende de la vocal que se encuentre al lado de la //, por lo que se ajusta a los valores de F1 de la vocal.

Sin embargo, es necesario mencionar que es muy complicado especificar con absoluta precisión el inicio de una transición consonante-vocal, por lo tanto, también lo es identificar exactamente el valor de frecuencia en que comienza dicha transición.

A pesar de aquello, en la transición de F1 de este fonema, la frecuencia en donde esta empieza siempre es baja, y aunque no se pueda saber el valor exacto, el rango siempre se encuentra entre 200-300 Hz.

En el caso de la transición de la curva del segundo formante, la característica principal es que posee una duración un poco mayor a otras consonantes, por lo que dicha cualidad también permite diferenciar a este fonema de otros

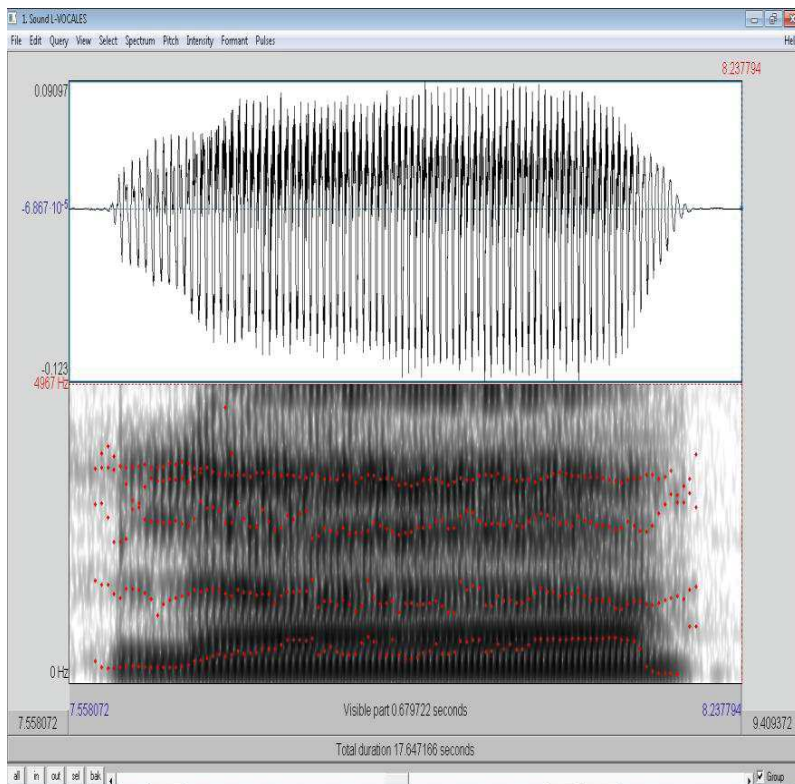


Figura 34.

Gráfica de formantes de la sílaba “la” dicha por un hombre.

2.1.4.4 Vibrantes

Los fonemas vibrantes en el castellano son dos: la /r/ y la /rr/. Estas consonantes se caracterizan por la vibración que genera la lengua al oponerse con la región alveolar de la cavidad oral. La principal diferencia en la generación de estos dos fonemas es la cantidad de vibraciones que cada uno posee. En el caso de la /r/ la lengua produce una única vibración, mientras que en el caso de la /rr/ las vibraciones son múltiples.

El comportamiento de formantes de la /r/ posee una característica distinta al de los demás fonemas del castellano. Si bien las curvas como tal no poseen alguna característica específica que produzca dicha distinción, la energía

presente en los formantes sí lo hace. Dependiendo de las vibraciones que genere la lengua, existen ciertos “vacíos” energéticos que aparecen en el espectrograma.

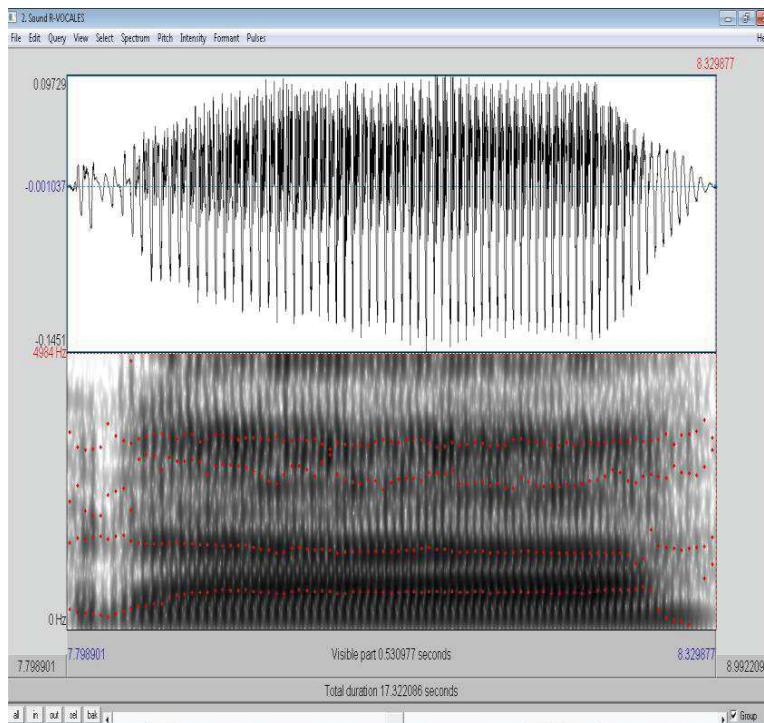


Figura 35.

Gráfica de formantes de la sílaba “ra” dicha por un hombre.

La consonante /r/ siempre es de corta duración, debido a su condición de alveolar. En lo referente a la frecuencia de las curvas de F1 y F2, la principal característica de distinción de esta consonante es la separación existente entre ellas. Dicha separación oscila entre valores de 1000 y 1300 Hz.

En las figuras 35 y 36 se pueden apreciar claramente el vacío energético mencionado anteriormente. Dicho vacío se observa en las partes que poseen ausencia de oscuridad, situadas en la parte izquierda.

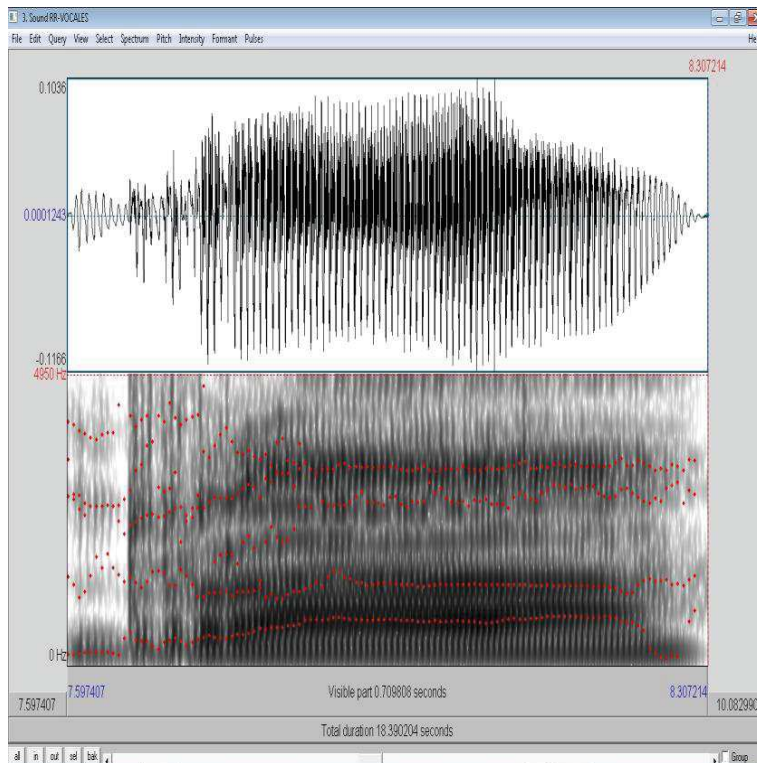


Figura 36.

Gráfica de formantes de la sílaba “rra” dicha por un hombre.

En la /rr/ se puede apreciar mayor número de “vacíos” energéticos debido a las vibraciones múltiples que produce la lengua. Estos vacíos también provocan algunas irregularidades en la curva de formantes, especialmente en F2 (ya que los formantes superiores no son considerados en el análisis).

2.1.4.5 Nasales

Las consonantes nasales del castellano son tres: la /m/, la /n/, y la /ñ/. Estos fonemas se diferencian principalmente de los demás del castellano debido al tipo de resonador principal que se involucra en su generación, en este caso la cavidad nasal en lugar de la oral. Todas las consonantes nasales del castellano son consideradas como sonoras.

Estos fonemas poseen una característica específica que los hace diferentes a la mayoría de consonantes. La transición consonante-vocal de F1 es muy corta, y en la mayoría de los casos posee una tendencia casi recta.

Enfocando el análisis en el fonema /m/, la principal diferencia con respecto a las demás consonantes nasales es el lugar de articulación que posee, ya que esta consonante es bilabial. Además de esto, otra diferencia muy importante es que la /m/ es el fonema que menos duración posee de entre todos los fonemas nasales en valores promedio, hablados en condiciones normales.

Como se mencionó anteriormente, las consonantes nasales son distintas a las demás consonantes del castellano debido a las características de comportamiento de F1 y su transición consonante-vocal casi recta y muy corta.

Sin embargo, es necesario establecer alguna diferencia notoria entre ellas mismas, ya que poseen características similares y corresponden a una clasificación única de entre todas las consonantes del castellano, debido a la condición única de que el flujo de aire pasa a través de la cavidad nasal, característica que ninguna otra consonante posee.

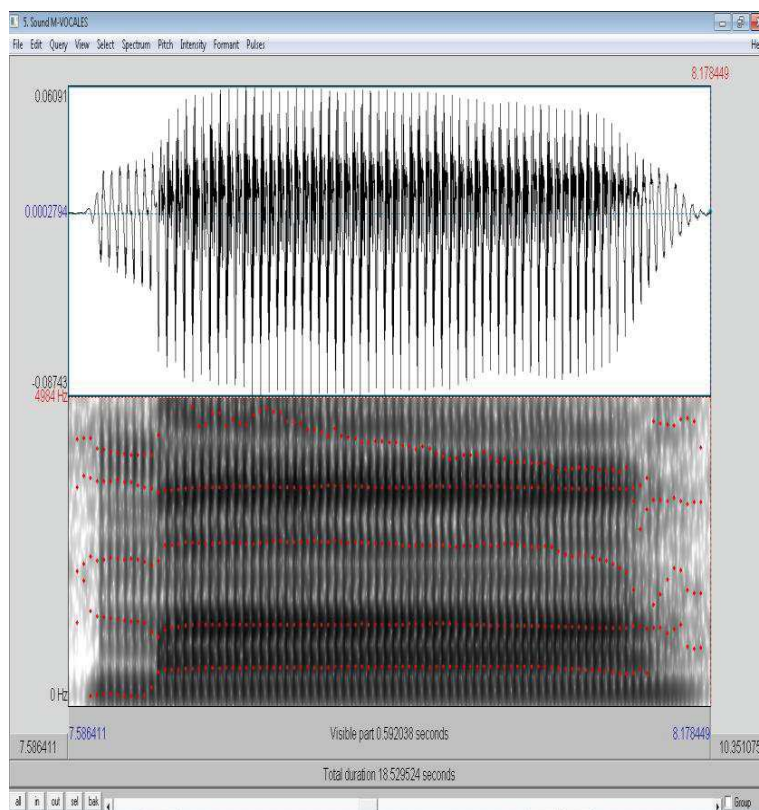


Figura 37.

Gráfica de formantes de la sílaba “ma” dicha por un hombre.

La /m/ posee una transición de F2 que la diferencia de las demás consonantes nasales. Como se puede apreciar en la figura 37, dicha transición consonante-vocal es corta.

Esta última característica es contraria al caso de la /n/, ya que en esta, la duración de dicha transición es mayor, como se observará en la figura 38.

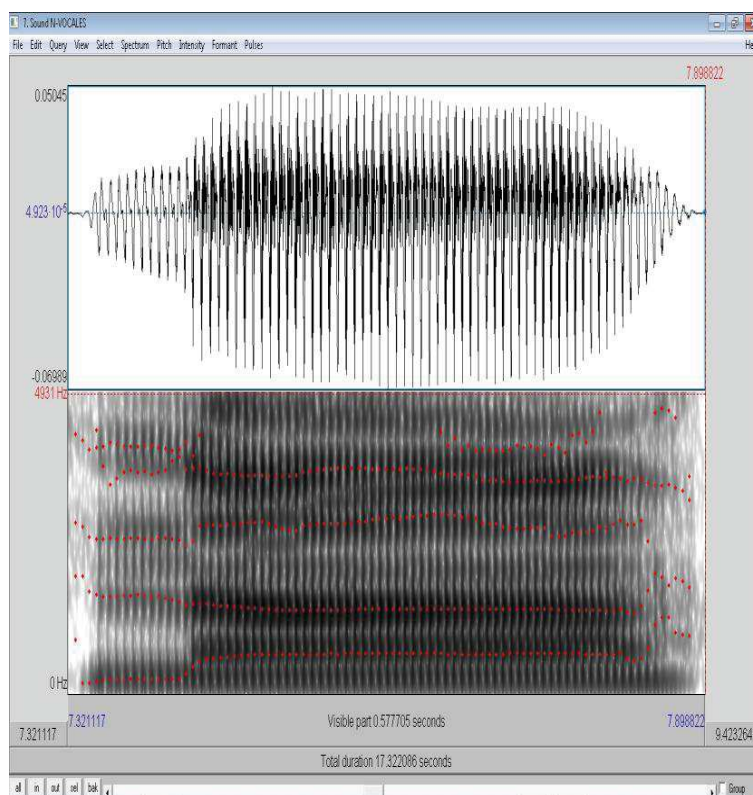


Figura 38.

Gráfica de formantes de la sílaba “na” dicha por un hombre.

La principal diferencia entre la /n/ y la /m/ es que a diferencia de esta última, la /n/ es una consonante alveolar según el lugar de articulación. Además, como se puede observar en la figura 38, la transición consonante-vocal de F2 es diferente entre ambas consonantes, ya que en la /n/ es más larga.

Otra diferencia notoria entre ambas consonantes es su duración, ya que en términos generales la /n/ es mayor a la /m/. Sin embargo, la mayor duración

entre las consonantes nasales la posee la /ñ/, como se puede observar a continuación.

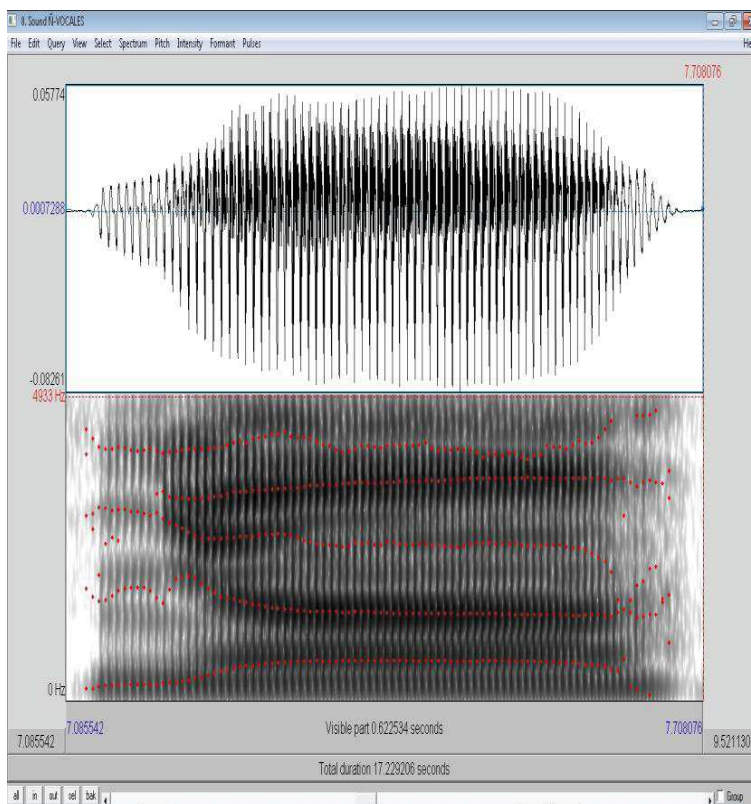


Figura 39.

Gráfica de formantes de la sílaba “ña” dicha por un hombre.

La /ñ/ es una consonante palatal según el lugar de articulación. Esta característica influye completamente en su duración así como en el comportamiento de sus formantes.

A pesar de corresponder a dos clasificaciones de índoles distintas, la condición de nasal y de palatal hacen de la /ñ/ un fonema compuesto por dos partes, si bien esto no se debe interpretar como la unión entre /n/ e /i/ que usualmente se suele enseñar a hablantes extranjeros que quieren aprender el castellano.

La primera parte corresponde a la característica específica de ser un fonema nasal, presentando características similares a las demás nasales como una tendencia recta de la curva de sus formantes, por ejemplo. La segunda parte corresponde a la condición de palatal. Esta división se puede apreciar en el

gráfico anterior como una separación existente por medio de una mayor concentración energética de los formantes, especialmente del tercero y cuarto. A partir del inicio de la sección palatal cambia el comportamiento de los formantes produciéndose transiciones más largas y graduales.

En términos generales las duraciones de cada consonante nasal comprenden los siguientes rangos promedio: /m/ entre 70-80 ms, /n/ entre 90-100 ms, y /ñ/ entre 105-115 ms.

2.2.- Implementación de las Redes Neuronales Artificiales

Para poder relacionar a los valores de formantes y sus patrones de comportamiento con los resultados visuales requeridos, fue necesario implementar un algoritmo que aplique Redes Neuronales Artificiales (RNA) del software MATLAB. La versión de este software que se usó en el presente proyecto, es la R2009a.

Este algoritmo constituye la parte fundamental del diseño de software mencionado en el presente proyecto, y se basa en todo el estudio realizado anteriormente acerca de los fonemas y sus características acústicas referentes al comportamiento de sus formantes, los cuales son el principal patrón de reconocimiento que servirá como vector de entrada para la RNA.

Se debe recalcar que en el diseño propuesto, fueron usadas seis RNA en conjunto, cada una correspondiente a una clasificación específica de las consonantes según el modo de articulación. La descripción de ellas será ampliada más adelante.

2.2.1 Descripción de las etapas realizadas

Existieron algunas secciones presentes en la parte de programación de la RNA, las cuales fueron implementadas en base a las necesidades requeridas para obtener los objetivos planteados.

A continuación se especifican cada una de ellas, así como las actividades realizadas durante el proceso.

2.2.1.1 Especificación de nomenclatura

La primera etapa realizada, luego de haber obtenido los archivos de formantes de las sílabas descritas en el capítulo anterior, fue determinar una nomenclatura específica para que la red pueda reconocer los archivos correspondientes a la información de formantes, para lo cual fue necesario realizar una programación que permita asociar dichos archivos de texto con el lenguaje propio de MATLAB.

Los datos de formantes de las sílabas deseadas obtenidas por medio del software PRAAT fueron copiados en hojas de texto formato “.txt”, debido a que el algoritmo de RNA perteneciente a MATLAB trabaja bajo dicho formato.

La nomenclatura que se utilizó para los archivos de texto con las muestras de los formantes fue “SCX[NOMBRE DE LA SÍLABA]np.txt”. La descripción de cada letra corresponde a lo siguiente:

- La letra “S” significa la palabra “sílabas”.
- La letra “C” indica el término “corregido” y se refiere a un proceso realizado en donde se corrigieron algunos errores de las muestras de formantes obtenidos del software PRAAT (este proceso será descrito más adelante).
- La “X” diferencia a estas primeras muestras de otras muestras obtenidas que se utilizaron para comprobar posteriormente la Red Neuronal Artificial.
- Luego de esto se escribe la sílaba a la que pertenece el archivo de texto en letra mayúscula.
- La letra “n” representa la velocidad de pronunciación en que se grabaron los archivos de audio. En primera instancia, estos archivos se grabaron en tres velocidades distintas, pero la que se usó finalmente para el entrenamiento de la RNA fue la considerada como “normal”.
- La letra “p” es la letra inicial del nombre de la persona que realizó las grabaciones de la pronunciación de las sílabas utilizadas para el entrenamiento de la RNA.

Para los archivos de texto con las coordenadas de los labios se utilizó la siguiente nomenclatura: CC[*CONSONANTE*] y CV[*VOCAL*], donde para el primer caso, la letra “C” inicial corresponde a la palabra “coordinada”, la siguiente “C” significa “consonante”, y luego se escribe el nombre de la consonante a la que pertenecen los datos de coordinada.

Para el segundo caso, la “V” significa vocal y seguido de esto se escribe el nombre de la vocal a la que pertenecen los datos del archivo de texto.

2.2.1.2 Identificación de la información

Una vez especificada la nomenclatura a usar para cada archivo que contiene información de formantes, se procedió a realizar una programación para que MATLAB pueda identificar el contenido de cada archivo de texto.

Las líneas de programación que se utilizaron para el reconocimiento de la nomenclatura de los archivos de texto “SCX[*NOMBRE DE LA SÍLABA*]np.txt” y el reconocimiento de las muestras de formantes que estos archivos contienen, fueron implementadas dentro de un archivo tipo “.M FILE” donde también se escribió toda la programación correspondiente al algoritmo ANN de MATLAB.

Lo primero que se procedió a realizar es la escritura de los comandos “close all”, “clear all”, “pack” y “clc” en la hoja mencionada “.M FILE”. Estos se utilizan en casi todas las hojas de programación de MATLAB para borrar los datos previos que se generan cada vez que corre la programación, liberando la memoria del computador cada vez que se ejecuta el proceso.

```

Editor - C:\Users\user\Desktop\FORMANT-VOCAL\frisorda.m
File Edit Text Go Cell Tools Debug Desktop Window Help
Stack: Base
- 1.0 + ÷ 1.1 x % % % % !
1  %RED NEURONAL DE F,S,J.
2  - close all;
3  - clear all;
4  - pack;
5  - clc;
6
7  - nop = 1; %Numero de consonantes
8  - strArray = java_array('java.lang.String', nop); %Arreglo para guardar las consonantes
9  - strArray = {'X'};
10 - opciones=cell(strArray);
11 - clear strArray; %Liberó memoria
12
13 - ncons = 3; %Numero de consonantes
14 - strArray = java_array('java.lang.String', ncons); %Arreglo para guardar las consonantes
15 - strArray = {'F','S','J'};
16 - consonantes=cell(strArray);
17 - clear strArray; %Liberó memoria
18
19 - nvoca = 5; %Numero de vocales
20 - strArray = java_array('java.lang.String', nvoca); %Arreglo para guardar las vocales
21 - strArray = {'A', 'E', 'I', 'O', 'U'};
22 - vocales=cell(strArray);
23 - clear strArray; %Liberó memoria
24
25 - npers = 1; %Numero de personas
26 - strArray = java_array('java.lang.String', npers); %Arreglo para guardar los identificadores de personas
27 - strArray = {'p'};% 'a', 'da'}; %di=diana a=argenis da=daniela
28 - personas=cell(strArray);
29 - clear strArray; %Liberó memoria
30
script Ln 1 Col 1 OVR

```

Figura 40.

Gráfica de la hoja de programación “.M FILE” para Redes Neuronales Artificiales parte (1).

Después de colocar los comandos mencionados, se procedió a indicar la cantidad de letras “X” que la nomenclatura de cada archivo de texto posee. Todos los archivos “SCX[NOMBRE DE LA SÍLABA]np.txt” están formados por solo una letra “X”, por dicha razón la variable “nop” que se encuentra en la línea siete de la figura 40 es igual a uno.

Luego de esto se realizó un arreglo de funciones de MATLAB, *strArray = java_array('java.lang.String', nop)*, para guardar a dicha letra mediante la integración de la variable “nop” dentro del arreglo. Además se especificó el o

los caracteres a MATLAB con `strArray = ('X')`, dando a entender al software que la letra "X" es la que debe reconocer en la nomenclatura.

A continuación se creó una variable con el nombre de "opciones", la cual incorpora la función "cell" y actúa sobre el "strArray", `opciones = cell(strArray)`. Este paso sirve para que se cree una matriz de celdas que junta a los caracteres. Luego de todos estos pasos se creó un comando "clear" y seguido a este se escribió el nombre de la variable "strArray" para limpiar la salida de datos que está presente y así poder utilizarla nuevamente. Este mismo proceso es repetido varias veces en la programación pero diferenciando los componentes inmersos en él, según sea el caso de consonantes, vocales, personas, y velocidades.

Se debe mencionar que para el presente trabajo se utilizó un conjunto de seis redes neuronales artificiales del tipo "Feed Forward Back Propagation", cada una con cierto número de consonantes pertenecientes a seis clasificaciones distintas, las cuales son: fricativas sordas (F, S, J), nasales sonoras (M, N, Ñ), oclusiva sorda (P, T, K), oclusiva sonora (B, D, G), lateral sonora con africada sorda (L, LL, CH), y vibrante sonora (R, RR). Por esta razón las variables "ncons" son iguales a los números tres o cuatro, dependiendo de cada ANN.

Posteriormente se repitieron los pasos anteriores para especificar el número de consonantes. Lo primero que se hizo fue modificar la variable "ncons" según el número de consonantes, luego se realizó nuevamente el arreglo de MATLAB `strArray = java.array('java.lang.'String', ncons)`, pero incorporando la variable "ncons" para que guarde las tres letras F, S, J (que pertenecen a la Red Neuronal Artificial de la figura 40).

Con "strArray" se guardan los tres caracteres para que MATLAB los especifique como letras, después de esto se creó una nueva variable llamada "consonantes" la cual incorpora la función "cell" que actúa sobre el "strArray" logrando que todos los caracteres se incorporen dentro de una matriz de celdas. Por último se creó el comando "clear" seguido del "strArray" para nuevamente liberar memoria.

Para la identificación de la vocal se creó una variable con el número de vocales llamada “nvoca” que es igual a cinco, y luego se estructuró nuevamente el arreglo de funciones `strArray = java.array('java.lang.'String', nvoca)`, pero colocando la variable “nvoca” dentro de dicho arreglo para el reconocimiento de las cinco letras. Seguidamente se especificó a MATLAB mediante el comando “strArray” las cinco letras como caracteres A, E, I, O, U. Luego de esto se creó la variable “vocales” que incorpora la función “cell” la cual especifica su función dentro del “strArray”, y finalmente se liberó memoria mediante el “clear” de “strArray”.

El último paso del proceso de identificación de la información consistió en la especificación de la cantidad de las letras “n” (de velocidad normal) y “p” (identificador de la persona que grabó las muestras) dentro del vocabulario perteneciente a los archivos de texto. Para esto se crearon las variables correspondientes para especificar el número de personas que grabaron las muestras “npers” (igual a uno), y para la velocidad de grabación que se utilizó “nvelo” (también igual a uno). Para los dos casos se realizó el arreglo de funciones de MATLAB `strArray = java.array('java.lang.'String', npers)`; y `strArray = java.array('java.lang.'String', nvelo)`, y en ambos casos se colocó a las variables “npers” y “nvelo” en el arreglo. Esto se hizo para que se pueda identificar el número de caracteres que tiene la nomenclatura con dichas letras. Seguido de esto se creó nuevamente un “strArray” especificando que los caracteres son “n” y “p”, después se escribieron dos nuevas variables con los nombres de “personas” y “velocidad” las cuales contienen a la función “cell” que actúa sobre los “strArray” para de esta forma se produzcan celdas de matrices para los caracteres.

Finalmente para los dos casos se creó un “clear” seguido del “strArray” para liberar memoria. Todas las especificaciones hechas acerca de la programación para el reconocimiento de las letras “n” y “p” se las puede observar en la figura 41.

programación que permite que todos los caracteres reconocidos y citados en el proceso anterior se junten, formando la nomenclatura completa y de esta forma se pueda reconstruir el nombre del archivo de texto que almacena la información de formantes.

Primeramente se creó, como se ve en la figura 41, una variable llamada “narchivos” (número de archivos) la cual es igual a la multiplicación entre las variables “nop”, “ncons”, “nvoca”, “nvelo” y “npers”, $narchivos = nop * ncons * nvoca * nvelo * npers$. El total de la multiplicación es igual al número de archivos de texto existentes. Seguido de esto se creó nuevamente el arreglo con “strArray” pero incorporando a la nueva variable “narchivos” la cual contiene el número total de caracteres de las variables mencionadas, $strArray = java.array('java.lang.'String', narchivos)$.

Luego de esto se realizó un algoritmo con cinco lazos “for”, donde cada uno de ellos va desde uno hasta una variable “nop”, “ncons”, “nvoca”, “nvelo” y “npers”. El comando “for” de la última variable “npers” incorpora un arreglo de funciones, las cuales dependen de un contador “cont” igual a 1. La sintaxis usada es la siguiente:

```
strArray(cont)=java.lang.String(sprintf('SC%s%s%s%s%s.txt',char(ce
ll2mat(opciones(e))),char(cell2mat(consonantes(a))),char(cell2mat(vocales(b)))
c
har(cell2mat(velocidades(c))),char(cell2mat(personas(d)))))).
```

Este arreglo también incorpora las dos letras faltantes que son “S” y “C”, y después de esto el contador aumenta su valor progresivamente hasta llegar a los máximos valores de las variables.

Por último, se elaboró un arreglo con todos los caracteres y se los ordenó hasta formar la nomenclatura propia de los archivos de texto, también el arreglo influye en la información ordenada de los formantes de las sílabas.

El “strArray” se lo colocó dentro de la función “cell” y a dicho paso se lo nombra como la variable “filename”.

```

72 - clear strArray; %Libero memoria
73 - clear cont;
74 - clear a;
75 - clear b;
76 - clear c;
77 - clear d;
78
79 - narchivoslab=ncons+nvoca;
80 - strArray = java_array('java.lang.String', narchivoslab);
81 - cont=1;
82 - for a=1:ncons
83 -     strArray(cont)=java.lang.String(sprintf('CC%s.txt',char(cell2mat(consonantes(a)))));
84 -     cont=cont+1;
85 - end
86 - for b=1:nvoca
87 -     strArray(cont)=java.lang.String(sprintf('CV%s.txt',char(cell2mat(vocales(b)))));
88 -     cont=cont+1;
89 - end
90 - filename=cell(strArray); %Arreglo con todos los nombres de archivos con la info de las coordenadas de
91 - clear strArray;
92
93 - clear a;
94 - clear b;
95 - clear cont;
96 - clear ncons;
97 - clear nvoca;
98 - clear npers;
99 - clear nvelo;
100 - clear consonantes;
101 - clear vocales;

```

Figura 42.

Gráfica de la hoja de programación “.M FILE” para Redes Neuronales Artificiales parte (3).

Para el arreglo de la nomenclatura de los archivos de texto con las distancias de los labios y su identificación en la Red Neuronal Artificial se creó una variable llamada “narchivoslab”, la cual realiza la suma entre las variables “ncons” y “nvoca”, como se ve en la figura 42. Esto debido a que a cada letra se le asignó un archivo con las distancias respectivas a su labialización. Seguido de esto se programó un arreglo de funciones para guardar los identificadores de “narchivoslab”, `strArray = java.array('java.lang.String', narchivoslab)`, luego se crearon dos lazos “for” por separado, el primero va desde uno hasta “ncons”, y el segundo va desde uno hasta “nvoca”.

Estos lazos “for” sirven para identificar los caracteres de los archivos de coordenadas de las consonantes en el caso del primer caso, y de las vocales para el segundo.

El primer lazo contiene el “strArray” con las dos letras “C” iniciales: `strArray(cont)=java.lang.String(sprintf('CC%s.txt',char(cell2mat(consonantes(a))))`), mientras que el segundo contiene la “C” inicial seguida de la “V” de vocal: `strArray(cont)=java.lang.String(sprintf('CV%s.txt',char(cell2mat(vocales(b)))))`.

El contador general “cont” va aumentando su valor cada vez que los lazos “for” cumplen con todo su proceso. Esto se debe a que dentro de cada uno de ellos se especifica `cont=cont+1`. Con todas estas líneas de programación, además de la función “cell” de estos “strArray” almacenados como una variable “filename1”, se logra el arreglo ordenado de todos los nombres de archivos con la información de distancia de abertura de la boca de cada uno de ellos.

2.2.1.4 Extracción de la información de los archivos de texto

Para lograr la obtención de la información de los formantes de los archivos de texto “SCX[NOMBRE DE LA SÍLABA]np.txt”, se programó mediante un “for” general las siguientes líneas de programación que se observan en la figura 43. Se creó un contador “a” que va desde uno hasta la variable “narchivos” (número de archivos). Dentro de este lazo se colocó una línea de programación con varias funciones de MATLAB, la cual se encarga de abrir los archivos tipo “.txt” con los caracteres de la variable “filename”; la sintaxis es: `fid=fopen(char(cell2mat(filename(a))),'rt')`.

Después de esto, mediante la función “fscanf” se especificó el número de espacios que existen en cada fila de “fid”, los cuales son 10 y terminan en un “Enter” de teclado. Dentro del lenguaje de MATLAB esto se realiza así: `%10g\n`.

```

Editor - C:\Users\user\Desktop\FORMANT-VOCAL\frisorda.m
File Edit Text Go Cell Tools Debug Desktop Window Help
Stack: Base fx
- 1.0 + ÷ 11 * % % % %
107 -
108 - for a=1:narchivos
109 -     %Crear procedimiento para extraer información del archivo llamado segun indique el arreglo autogene
110 -     fid=fopen(char(cell2mat(filename(a))), 'rt');
111 -     datemp=fscanf(fid, '%10g\n', [1,inf]);
112 -     st=fclose(fid);
113 -     tama=size(datemp,2);
114 -     mtam(a)=floor(tama(a)/2);
115 -
116 -
117 -     fid=fopen(char(cell2mat(filename1(c))), 'rt');
118 -     datemp1=fscanf(fid, '%3g\t%3g\t%3g\n', [1,inf]);
119 -     st=fclose(fid);
120 -
121 -     fid=fopen(char(cell2mat(filename1(cont2+2))), 'rt');
122 -     datemp2=fscanf(fid, '%3g\t%3g\t%3g\n', [1,inf]);
123 -     st=fclose(fid);
124 -
125 -     d=40;
126 -     for b=1:d
127 -         datemp=datemp+rand(size(datemp))*10;
128 -         P((a-1)*d+b,:)=[datemp(1:60) datemp(mtam(a)+1:mtam(a)+60)];
129 -         T((a-1)*d+b,:)=[datemp1(1,1) datemp1(1,2) datemp1(1,3) datemp2(1,1) datemp2(1,2) datemp2(1,3)];
130 -     end
131 -
132 -     cont2=cont2+1;
133 -     if cont2>5
134 -         c=c+1;
135 -         cont2=1;
136 -     end
137 - end

```

Figura 43.

Gráfica de la hoja de programación “.M FILE” para Redes Neuronales Artificiales parte (4).

La especificación del número de columnas dentro de cada archivo de texto se realiza por medio de la siguiente escritura: $[1,inf]$. Aquí se especifica que cada dato se encuentra en una fila determinada y se le dice a la red que escanee cada fila hasta que la información presente termine y siga a la siguiente.

La culminación del escaneo de los datos numéricos de formantes presentes en los archivos de texto se realizó mediante la función “fclose” de “fid”. Terminado

el escaneo general de datos, se procedió a obtener el número total de muestras dentro del “fscanf” con la utilización de la herramienta “size” la cual cumple dicha función, y luego a ese número se lo redondea hacia el entero próximo inferior mediante la herramienta “floor”. Este último paso sirve para obtener resultados mucho más precisos y no programar con decimales.

Todo el proceso descrito en el párrafo anterior se repitió para escanear los datos de las coordenadas de los labios, se reemplazó a la variable “filename” por la variable “filename1”, y se especificó que dentro de “fscanf” los datos están separados por un “Tab” de teclado y tienen tres espacios cada uno. Esto se hizo mediante la sintaxis: `'%3g\t%3g\t%3g\n'`, sin embargo en este caso se omitieron los pasos anteriores de redondeo de datos.

Se realizaron dos tipos de reconocimiento de distancias, uno para los datos de distancia de consonantes, y otra programación para la obtención de las distancias de las vocales. Para el caso de las vocales se usó un contador dos “cont2”, que por medio de un condicional “if” permite que las vocales roten y no se repitan sucesivamente.

Por último se especificó mediante la utilización de un lazo “for” que va desde uno a cuarenta, el vector de entrada P para el entrenamiento de la Red Neuronal Artificial, el cual almacena las primeras sesenta muestras de la curva de F1 y les añade un ruido matemático por medio del contador “b”; y también escoge las sesenta muestras iniciales de la curva de F2 e igualmente les añade ruido. Inmediatamente se compara dicho vector P con el vector T que representa el “target” (objetivo) al que se quiere llegar, que son los seis datos de coordenadas de los labios, tres pertenecientes a las consonantes y tres a las vocales.

Este vector T es puesto dentro del mismo “for” para la comparación de datos entre vectores, los cuales representan a las entradas y salidas deseadas de las redes, por lo que consecutivamente estos revelarán los valores deseados de coordenada por cada sílaba pronunciada.

Mediante el escaneo de todos los archivos de texto que se identificaron por su nomenclatura, se obtuvieron todos los datos de formantes y distancias de los labios presentes en ellos.

2.2.1.5 Topología usada en las RNA

La topología empleada en las RNA es la siguiente:

- Una capa de entrada, que estuvo compuesta de 120 muestras de formantes.
- Una capa oculta, que se compuso de 15 neuronas.
- Una capa de salida, que se compuso de 6 distancias.

Realizando un análisis más profundo de la topología usada y las razones por las que se decidió implementarla, se debe especificar lo siguiente:

Cada vector de entrada "P" se compuso de los valores de frecuencia pertenecientes a F1 y F2; mientras que cada vector de salida "T", contuvo las distancias de separación de los labios.

Debido a que la duración de cada fonema influye directamente en el número de formantes extraído, fue necesario establecer un número común de muestras en todas las sílabas analizadas, para poder alimentar a cada RNA. Es por esto que se seleccionaron los primeros sesenta valores de frecuencia de F1 y F2 para el vector de entrada "P". Estas sesenta muestras abarcan todos los formantes de la consonante, independientemente de su duración, sin embargo, solo corresponden a una parte de los formantes de la vocal.

La cantidad de formantes escogida, se basó en lograr una identificación precisa de las consonantes, debido a que estas poseen una mayor complejidad de reconocimiento. Para el caso de las vocales, se intentó considerar la mayor cantidad de muestras de formantes posible, sin embargo, existió la dificultad de que algunos archivos contenían una menor cantidad de muestras que otros, presentándose un error en la extracción de información por parte de MATLAB. Esto se debe a que todos los vectores de entrada de una RNA, deben tener la

misma dimensión. Es por esto que se decidió adaptarse al archivo que poseía el menor número de muestras de formantes.

Luego del vector de entrada, se encuentran las neuronas. El número de neuronas escogidas correspondió a quince. Este valor se escogió para que cada red trabaje adecuadamente en relación a los recursos del computador.

El vector de salida "T" se compuso de seis valores adimensionales, correspondientes a distancias específicas de separación de los labios, de acuerdo a la pronunciación distinta de cada sílaba. Los tres primeros valores pertenecen a la abertura de la consonante, mientras que los tres restantes corresponden a la vocal.

Todo lo anterior puede ser observado en la figura 44, en donde se presenta de manera gráfica la topología usada en cada RNA.

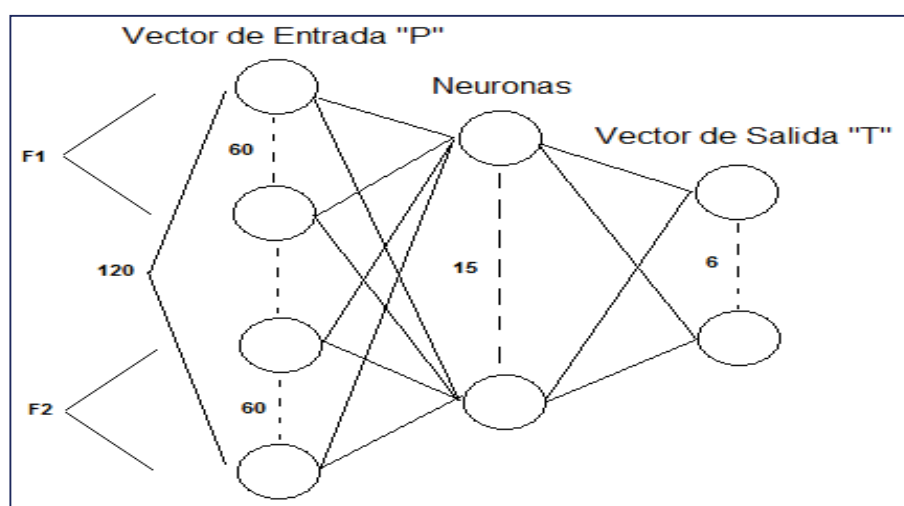


Figura 44.

Topología de las Redes Neuronales Artificiales implementadas.

2.2.1.6 Reducción de datos erróneos de formantes

Antes de proceder a la etapa de entrenamiento, fue necesario cerciorarse de que los datos de entrada para las RNA posean el menor porcentaje posible de error.

Como se mencionó en la sección 2.1.2.1, PRAAT presenta algunos datos erróneos de formantes en el cálculo del LPC, los cuales se producen por distintas causas. Aunque el software advierte al usuario de que dichos valores deben ser considerados como errores, no los elimina de la información ofrecida cuando se obtiene la lista de formantes.

Es por esto que luego de haber realizado algunas pruebas con los valores de formantes obtenidos, se decidió implementar una programación específica para que MATLAB brinde la capacidad de reducir o eliminar los errores más notorios que se presenten en los archivos de texto con la información de formantes.

Para poder implementar dicha programación fue necesario observar el comportamiento generalizado que presentan los errores, de donde se obtuvo que la principal característica es que los formantes “erróneos” se salen de la tendencia de la curva en valores muy elevados, en comparación a las diferencias de frecuencia normal que cada curva presenta.

Por dicha razón, y con el objetivo de obtener una mejor identificación de las consonantes y vocales, se elaboró un algoritmo computacional en MATLAB que se encarga de reducir los picos erróneos que presenta la curva de formantes, haciendo que la tendencia general de las curvas sea más estable.

Como se observa en la figura 45, los datos que son ingresados y definidos según la variable “d”, pasan por tres funciones de MATLAB, que son:

- “size”. Esta función se encarga de medir el tamaño de la variable “d”.
- “mean”. Se encarga de sacar el promedio de los formantes.
- “std”. Determina la desviación estándar de los datos.

```

1
2
3 function dc = desviacion(d)
4
5     n=size(d,2); %Tamaño del vector
6     m = mean(d); %Promedio
7     ds = std(d); %Desviación estandar
8
9     dife=zeros(size(d));
10    for a = 1:n-1
11        dife(a) = abs(d(a+1) - d(a));
12    end
13
14    mdif = mean(dife);
15    dsdif = std(dife);
16
17    for a = 2:n-1
18        if dife(a) >= mdif+dsdif
19            d(a)=(d(a-1)+d(a+1))/2;
20        end
21    end
22
23    if dife(1) >= mdif+dsdif
24        d(1)=(d(2)+d(3))/2;
25    end
26    if dife(n) >= mdif+dsdif
27        d(n)=(d(n-1)+d(n-2))/2;
28    end
29
30    dc=d;
31
32

```

Figura 45.

Programación de la función “desviacion”.

A partir de esta información se crea una variable llamada “dife”, la cual por medio de la función “zeros” crea una matriz de ceros a partir del tamaño que tiene “d”, la cual a su vez ingresa a un lazo “for” que va desde uno hasta el tamaño de “d” menos uno. Este primer “for” entrega la diferencia entre un valor de formante específico y su valor anterior. A esta nueva variable “dife” se la promedia con “mean” y se le determina su nueva desviación estándar con “std”.

Luego de esto se definen si las diferencias obtenidas entre formantes (dentro de la variable “dife”) son mayores o iguales al promedio más la desviación estándar. Esto se hace con un nuevo lazo “for” que va desde dos hasta nuevamente el tamaño de “d” menos uno, y con un condicional “if” dentro de él.

El principio de funcionamiento del algoritmo diseñado es el siguiente: se extrae la diferencia aritmética entre los valores de frecuencia de dos formantes adyacentes, y si dicha diferencia posee un valor mucho mayor a la tendencia normal de la curva (establecida por la varianza), se suman dichos valores y se divide el resultado para dos, de modo de poder obtener un promedio entre los formantes adyacentes.

Para poder clarificar el funcionamiento del algoritmo, se presenta la figura 46 con un ejemplo simple de reducción de errores:

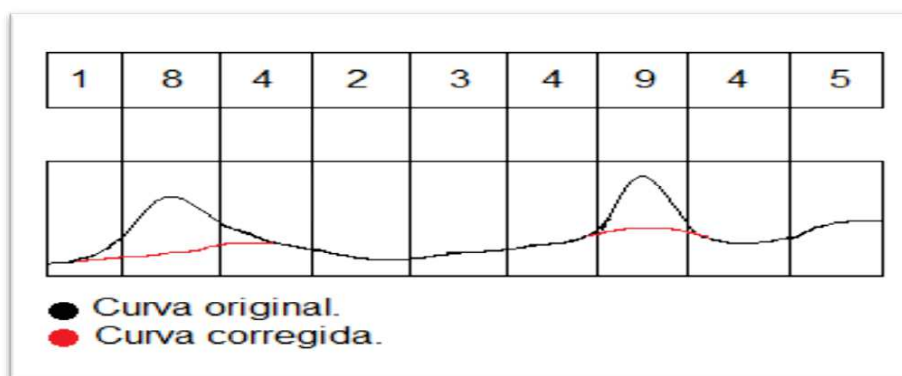


Figura 46.

Gráfico ilustrativo de corrección de errores.

Como se puede observar en la figura 46, se tiene una curva de color negro que corresponde a nueve valores distintos de amplitud a lo largo del tiempo. Dichos valores se describen en la parte superior de la figura.

La tendencia natural que se puede apreciar en la curva, es ascendente; sin embargo, existen dos valores que se salen de dicha tendencia. Estos valores pueden ser considerados como dos picos de error, y corresponden a las amplitudes ocho y nueve. Para reducir los picos y mantener la tendencia de la curva, se promedian los valores vecinos a estos, que para el primer error son

uno y cuatro, y para el segundo error son cuatro y cuatro. El promedio correspondiente arroja los valores de dos punto cinco y cuatro, respectivamente.

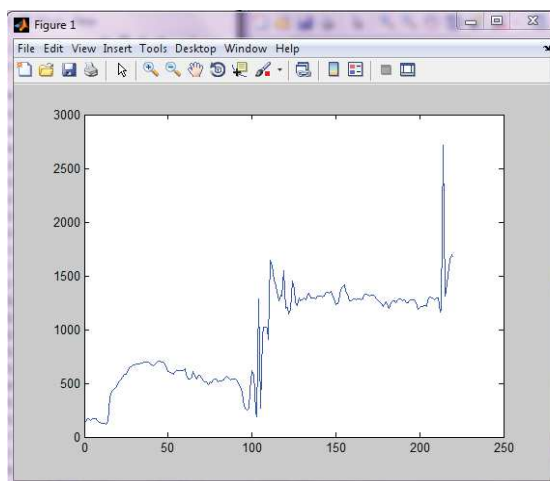
Estos valores se grafican en la curva de color rojo, la cual corresponde al resultado del algoritmo. Este proceso se repite para cada error encontrado en los valores de formantes de cada archivo de texto.

Este tipo de casos se presentan en grandes cantidades en todos los resultados ofrecidos por el software PRAAT. Es por esto que fue necesario aplicar este proceso de reducción de errores a todos los archivos de texto obtenidos, para poder así clarificar las tendencias naturales de los formantes de cada fonema.

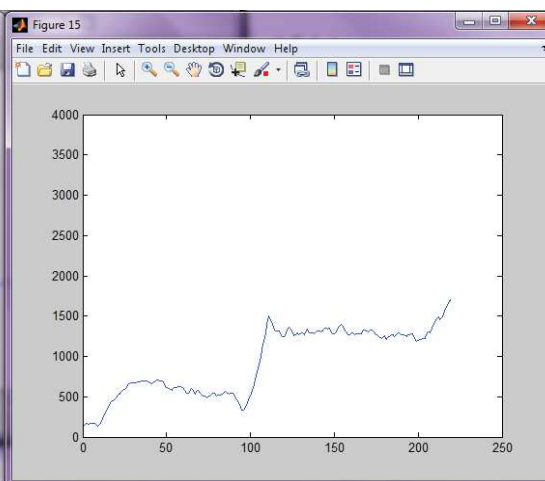
Una vez comprendido el funcionamiento básico del algoritmo, a continuación se presenta un ejemplo comparativo entre los resultados de formantes obtenidos directamente desde PRAAT, y luego de aplicar la reducción de errores. Las sílabas usadas en este ejemplo son “BA” y “SA”.

Es necesario recalcar que la parte izquierda de la figura 47 corresponde a las gráficas de formantes con errores de ambas sílabas, mientras que la parte derecha corresponde a las gráficas con reducción de errores.

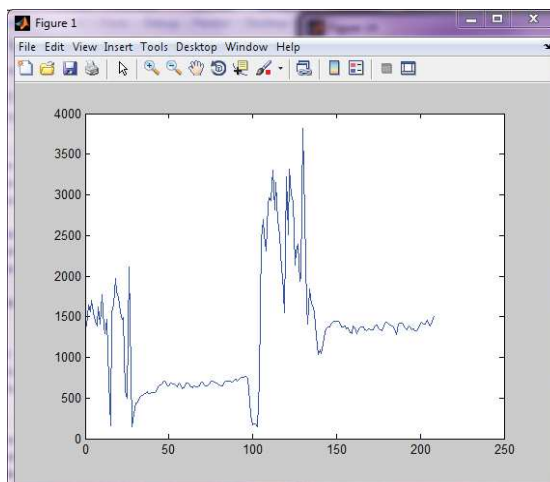
Curva de formantes con picos de error (BA).



Reducción de errores de la curva de formantes (BA).



Curva de formantes con picos de error (SA).



Reducción de errores de la curva de formantes (SA).

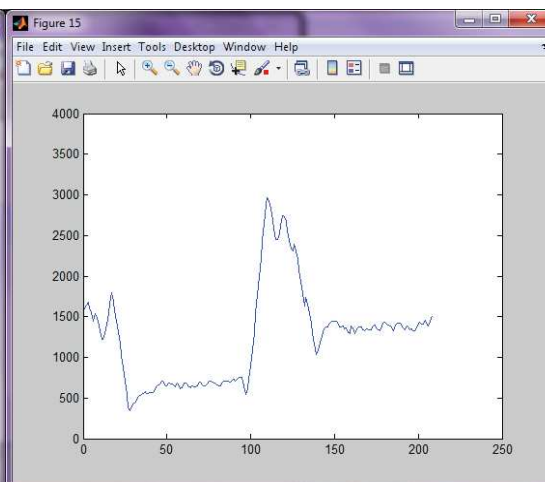


Figura 47.

Comparación entre formantes con errores (izquierda) y sin errores (derecha).

Todas las muestras de formantes presentes en los archivos "SCX[NOMBRE DE LA SÍLABA]np.txt" que se utilizaron para el entrenamiento de las seis redes, pasaron por el algoritmo de reducción de errores.

Este proceso se lo realizó previamente a la etapa de entrenamiento de las seis redes, para que los patrones de aprendizaje, como las transiciones consonante-vocal existentes y las tendencias de comportamiento temporal que

posee cada formante, sean mucho más claros y facilite el reconocimiento de las diferencias que existen entre sílabas.

La programación del algoritmo se define como una función llamada “desviacion” (sin tilde), la cual depende de dos variables: la variable “d” que representa los datos de formantes de entrada de la sílaba; y la variable “dc” que corresponde a los datos de formantes resultantes o “corregidos”. Esta programación se la realizó en una hoja “.m” de MATLAB, la cual tiene el mismo nombre que la función.

2.2.1.7 Entrenamiento de las RNA

El entrenamiento de las seis Redes Neuronales Artificiales es la parte de mayor importancia para lograr buenos resultados en cuanto al reconocimiento de las sílabas. Por dicha razón se decidió crear una variable “d” en cada una de las seis redes, que posea cuarenta vectores con información de formantes más la añadidura de ruido en cada uno de los casos. Es decir, cada sílaba comprendería cuarenta archivos con formantes, diferenciados por el ruido añadido.

De entre estos cuarenta se decidió dividir el número de archivos que cada RNA escogería para cumplir con todos los procesos necesarios en la etapa de entrenamiento, en distintos porcentajes. Los porcentajes elegidos se describen a continuación:

- 70% como el valor de entrenamiento de la red. Corresponde a 28 archivos.
- 15% para verificación. Corresponde a 6 archivos.
- 15% para prueba. Corresponde a 6 archivos.

El entrenamiento de la red corresponde a un proceso de aprendizaje supervisado, el comprende la mayor parte de los archivos totales. Este proceso implica la modificación de los pesos sinápticos, que están ubicados entre los puntos de conexión entre neuronas, en respuesta a la información de entrada para obtener el resultado deseado. Los pesos sinápticos son valores numéricos sencillos (números enteros, fraccionarios positivos o negativos), los cuales son

modificados constantemente hasta que exista una corrección favorable en el error que se produce por el proceso de entrenamiento de la red. Al culminar dicho proceso, la red está en capacidad de asimilar las tendencias de las curvas y sus transiciones consonante-vocal, logrando identificar las diferencias existentes entre sílabas.

Cada red es sometida a dos procesos adicionales al de entrenamiento, los cuales son: validación y prueba. Estos dos procesos poseen los porcentajes restantes del total de archivos, en igual proporción del 15%. El proceso de validación determina la capacidad de la red de generalizar los resultados con otras muestras, las cuales no fueron parte del entrenamiento, y así lograr la versatilidad de la misma en el reconocimiento de la sílaba.

Por otro lado, el proceso de prueba determina la precisión de la red, logrando una evaluación de cómo esta responde frente a cualquier archivo de formantes que ingrese para su identificación.

Durante todo el proceso de entrenamiento, se deben configurar algunos parámetros como son: el número de épocas, la tasa de aprendizaje y el valor mínimo de error. Las épocas son el número total de interacciones que una red debe repetir por cada proceso de entrenamiento, en este caso con las 120 muestras de formantes. La tasa de entrenamiento es una constante que permite la modificación de los pesos sinápticos en cada interacción; a mayor tasa de aprendizaje, mayor será la modificación de dichos pesos, por lo que el aprendizaje será más rápido.

Finalmente, se debe colocar el valor mínimo de error que se espera obtener en todo el proceso de entrenamiento de la red. Este valor debe ser escogido con precaución, ya que puede llegar a ser perjudicial en algunos casos, debido a la aparición de un fenómeno de "overfitting". El "overfitting" es un efecto no deseado que se produce por el exceso de entrenamiento, haciendo que la red identifique únicamente las muestras que se utilizaron en el entrenamiento y no generalice su identificación con otras que no fueron parte de él.

Como se puede observar en la figura 48, se encuentra toda la información descrita anteriormente además de los cuatro condicionales “if” que se mencionaron y que se procederá a describir a continuación.

Se debe recalcar también que se crearon cuatro contadores distintos para que el proceso planteado en los condicionales pueda ser segmentado correctamente y repetido una vez que se cumplan los cuarenta archivos de una sílaba, para proceder a la siguiente y así sucesivamente hasta que se cumplan todas. Los contadores “ctre”, “cver” y “cpru” sirven para que cada proceso pueda ser completado en los porcentajes establecidos, mientras que el contador general “cont” sirve para que este proceso se repita para cada grupo de cuarenta archivos.

En lo referente al primer condicional, el contador de entrenamiento va desde uno hasta 28 lo que sirva para coger las primeras 28 muestras de formantes del vector P en el primer proceso de entrenamiento. El segundo condicional comienza desde que el contador es igual a 28 hasta llegar a 34, luego de esto se acaba la obtención de muestras del vector P para el proceso de prueba. El tercer condicional empieza desde 34, y culmina en 40, por lo tanto cuando el contador de verificación llega al número mayor deja de considerar las muestras del vector P para el proceso de verificación; y con la última condición, el contador general se reinicia cuando pasa de 40. Todo este proceso de programación sirve para repartir de forma coherente las distintas muestras de formantes en los tres pasos importantes: entrenamiento, verificación, y prueba.

- *net.trainParam.max_fail=100*. Indica que el número máximo de errores de validación corresponderían a 100.
- *net.trainParam.show=50*. Este valor determina el número de épocas, el cuál se decidió que sea igual a 50.
- *net.trainParam.lr=0.05*. Este valor determina que la tasa de aprendizaje de la red es igual a 0.05.
- *net.trainParam.epochs=1000*. Indica que el número máximo de épocas para la etapa de entrenamiento es de 1000.
- *net.trainParam.goal=1e-9*. Indica el rendimiento objetivo de la red. Esto significa el error aceptado por el algoritmo.

Como se puede observar en la figura 49, anterior el nombre de esta ANN corresponde a “fricativasorda”, el cual se puede configurar por medio de la sintaxis: *save('fricativasorda','net')*.

La variable que almacena los resultados finales y entregados de por la red al ser excitada con los valores de “P” se llama “Y”: $Y = \text{sim}(\text{net}, P)$. Esto es la simulación o ejecución de la red a partir de datos “P”, y se realiza únicamente con redes ya entrenadas.

Luego de todos los procesos descritos anteriormente se procedió a entrenar la Red Neuronal Artificial ejecutando a la hoja de programación “.M FILE”. Esto hace que se presente una ventana con la visualización del proceso de entrenamiento de la red.

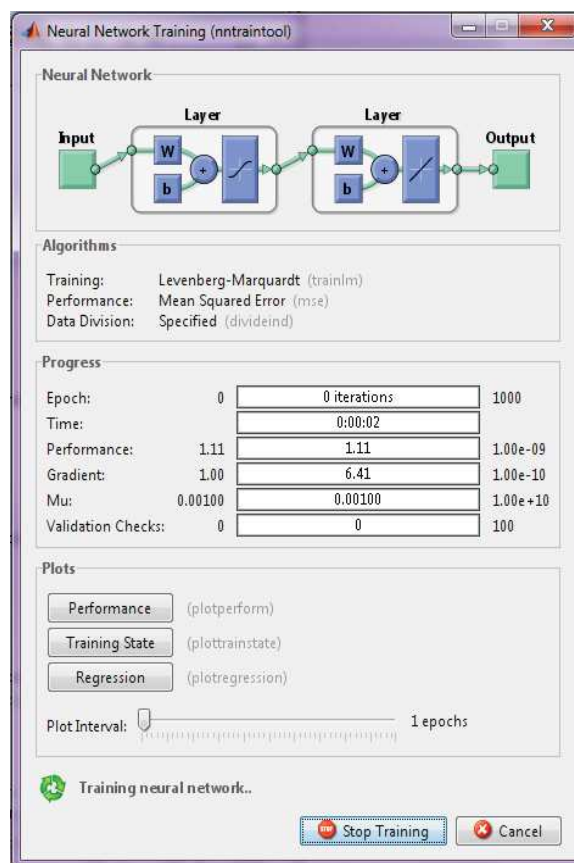


Figura 50.

Interfaz gráfica del proceso de entrenamiento de la ANN.

2.2.2 Resultados del entrenamiento de las RNA

Todos los procesos que se mencionaron en el apartado anterior, fueron implementados en las seis Redes Neuronales Artificiales y cada una presentó diferentes resultados, los cuales se explicarán a continuación.

2.2.2.1 Resultados de las consonantes vibrantes sonoras [/r/, /rr/]

La ANN que contiene las consonantes de la clasificación vibrante sonora se la nombró como “vibrante.mat”, la cual después del proceso de entrenamiento, validación y prueba, presentó el siguiente resultado:

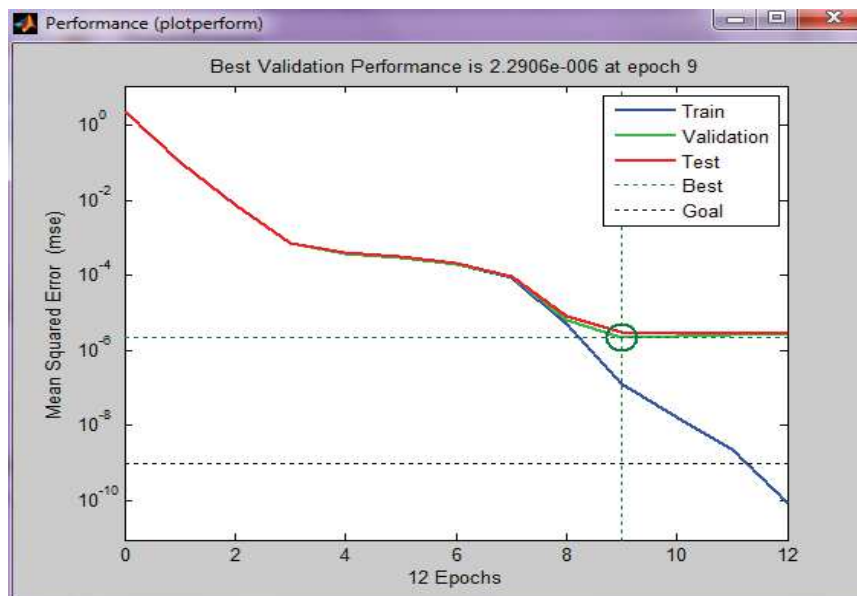


Figura 51.

Resultados de la etapa de entrenamiento de la Red “Vibrante”.

El entrenamiento de la Red Neuronal Artificial Vibrante determinó que los resultados de validación y de prueba tienen un error aproximado de 10^{-6} a partir de la época nueve, y que el resultado de entrenamiento posee un error menor a 10^{-10} . El círculo de color verde que se observa en la figura 51, determina que el proceso de entrenamiento debió culminar en la época señalada, que en este caso es la nueve. Esto radica en que en dicho punto, la respuesta de identificación de las sílabas de esa clasificación podría ser favorable, tanto para las muestras que se utilizaron al entrenar la red, como para muestras generales. Sin embargo, por las modificaciones explicadas en la sección 2.2.1.7, y al presentarse este comportamiento, la red “Vibrante” está en capacidad de reconocer únicamente las muestras con que fue entrenada.

2.2.2.2 Resultados de las consonantes nasales sonoras [m/, /n/, /ñ/]

La Red Neuronal que contiene las consonantes de la clasificación nasales sonora se la nombró como “nasales.mat”, la cual después del proceso de entrenamiento, validación y prueba, presentó el siguiente resultado:

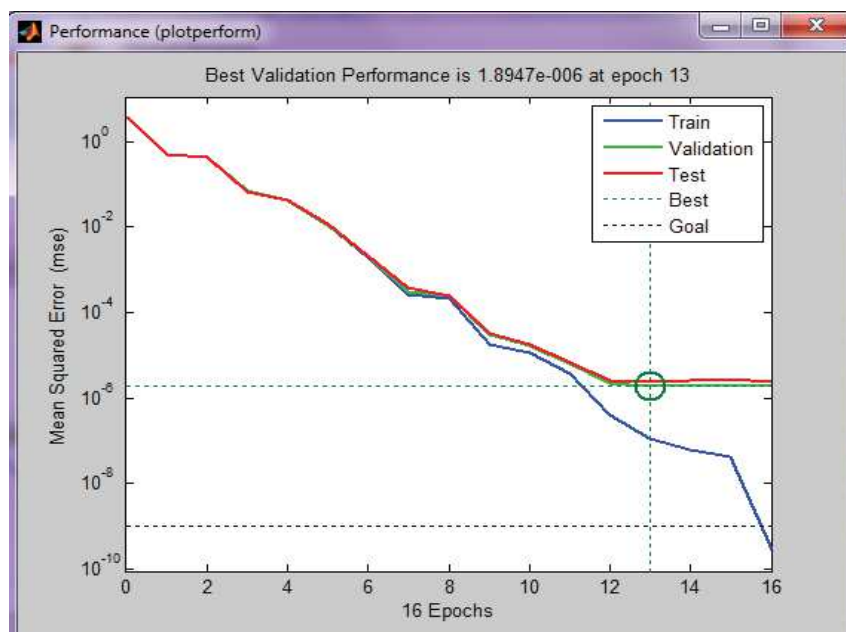


Figura 52.

Resultados de la etapa de entrenamiento de la RNA "Nasales".

Esta red presenta mejores resultados en cuanto al error producido en los procesos de validación y prueba, el cual es de 10^{-6} , y se produce a partir de la treceava época. En el proceso de entrenamiento en cambio, se obtuvo un error de 10^{-9} . La diferencia entre los dos errores resultantes que se presentan después del círculo verde, es aceptable, por lo que los resultados son buenos en cuanto a la identificación de las sílabas con consonantes nasales.

2.2.2.3 Resultados de las consonantes fricativas sordas [s/, /f/, /j/]

La Red Neuronal que contiene a las consonantes de la clasificación fricativas sordas se la nombró como "fricativasorda.mat", la cual después del proceso de entrenamiento, validación y prueba, presentó el siguiente resultado:

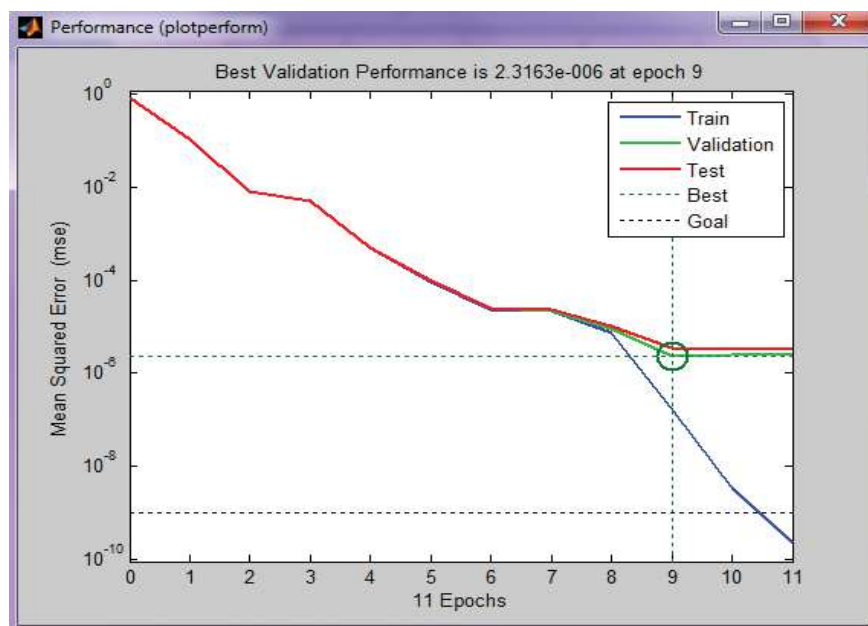


Figura 53.

Resultados de la etapa de entrenamiento de la Red "Fricativasorda".

Los procesos de validación y prueba presentan un error de 10^{-6} a partir de la novena época, mientras que los resultados en el proceso de entrenamiento determinan que el error producido es de 10^{-9} .

A pesar de que 10^{-6} puede considerarse como un error casi mínimo en la etapa de validación y prueba, también se debe evaluar la diferencia entre dicho valor con el error de entrenamiento. Además, debido a que el error se produjo en una época menor al caso anterior, también se podría presentar el efecto de "overfitting".

2.2.2.4 Resultados de las consonantes oclusivas sordas [p/, /t/, /k/]

Esta Red Neuronal se la nombró como "oclusivasorda.mat", la cual después del proceso de entrenamiento, validación y prueba, presentó el siguiente resultado:

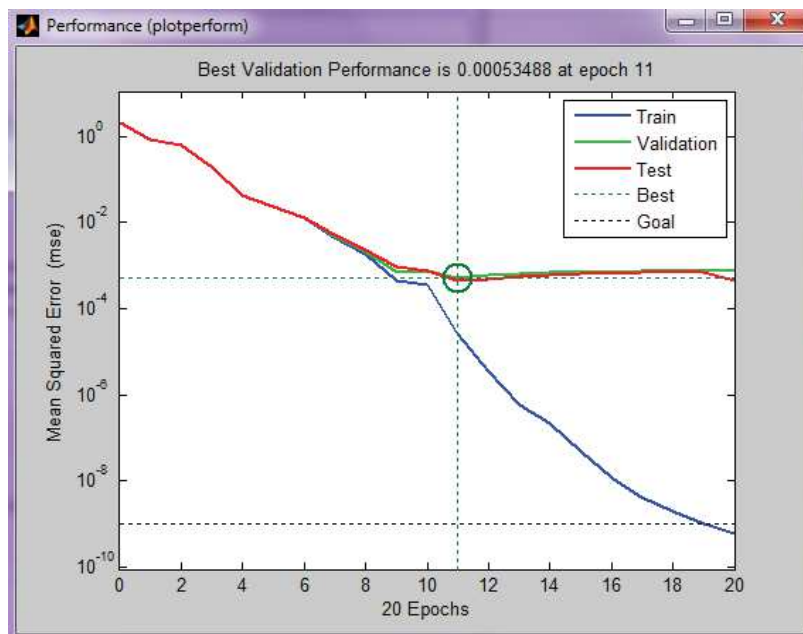


Figura 54.

Resultados de la etapa de entrenamiento de la Red “Oclusivasorda”.

Los resultados de error que presentan los procesos de validación y prueba llegan a ser de 10^{-5} , mientras que el resultado de error que presenta el proceso de entrenamiento de la red es de 10^{-10} . La diferencia entre los dos errores es muy alta, por lo que la red presenta una clara dependencia de los datos usados en el entrenamiento, dando como resultado una pobre generalización en la identificación de nuevas sílabas oclusivas sordas (debido al “overfitting”).

2.2.2.5 Resultados de las consonantes africada sorda y lateral sonora [ch/, ll, /ll/]

La Red Neuronal que contiene las consonantes africadas sordas y lateral sonora se la nombró como “latefrica.mat”, la cual después presentó el siguiente resultado:

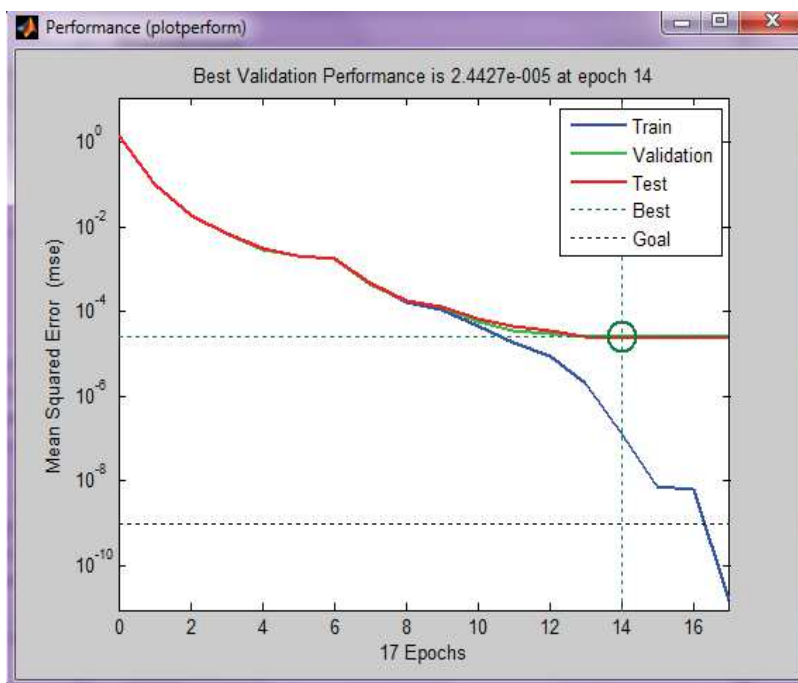


Figura 55.

Resultados de la etapa de entrenamiento de la Red “Latefrica”.

Para esta clasificación de consonantes el error obtenido en los procesos de validación y prueba fue de 10^{-5} , el cual es considerado aceptable; y el error presentado para el proceso de entrenamiento es de 10^{-11} . Al ser la diferencia entre los errores muy alta, esta red también presenta el efecto de “overfitting”.

2.2.2.6 Resultados de las consonantes oclusivas sonoras [b/, /g/, /d/]

La Red Neuronal que contiene a las consonantes de la clasificación oclusiva sonora se la nombró como “oclusivasonora.mat”, la cual después del proceso de entrenamiento, validación y prueba, presentó el siguiente resultado:

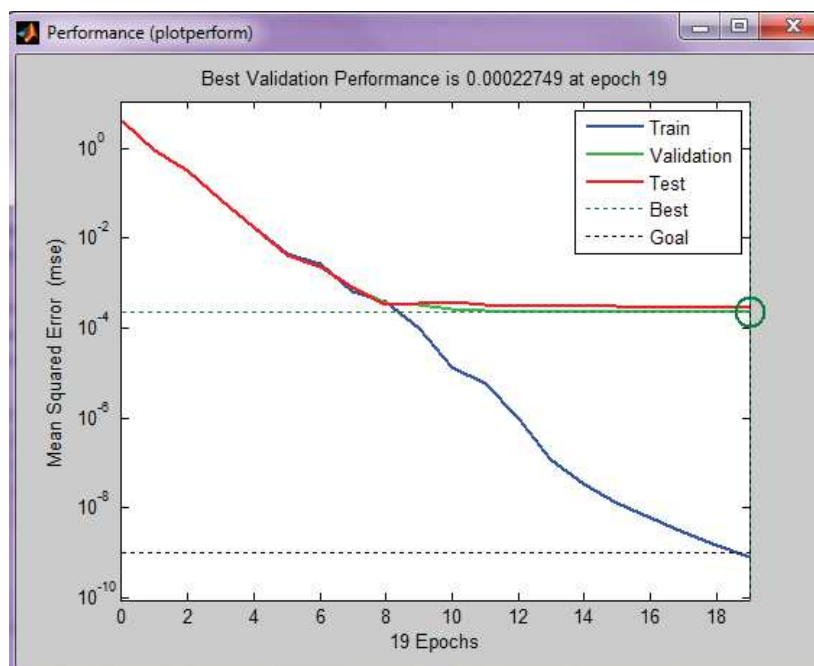


Figura 56.

Resultados de la etapa de entrenamiento de la Red “Oclusivasonora”.

Como se puede observar en la figura 56, el error para los procesos de validación y prueba es de casi 10^{-4} y se produce a partir de la octava época.

Este error es mayor a todos los errores que se presentaron en los procesos de validación y prueba de las redes anteriores. Además, al realizar la diferencia entre dicho error y el error que presenta el entrenamiento de esta red, el cual es igual a 10^{-9} ; se determina que la diferencia es muy grande, por lo tanto, la red también es dependiente de los datos que se utilizaron en el proceso de entrenamiento.

2.2.3 Resultados generales de identificación

El vector “Y” de las seis RNA presenta los resultados numéricos de reconocimiento de todas las ochenta y cinco sílabas, mediante la entrega de las seis coordenadas de abertura de la boca pertenecientes a la consonante y la vocal.

Las distancias establecidas en cada fonema deben ser consideradas como adimensionales. Estos valores fueron escogidos según una apreciación

matemática de cada abertura de la boca al pronunciar cierta sílaba, y si bien no corresponden a valores exactos, deben ser considerados como aproximaciones aceptables.

Tabla 13.

Dimensiones de la abertura de la boca de cada consonante y vocal.

LETRA	DISTANCIA 1	DISTANCIA 2	DISTANCIA 3
B	0	0	0
CH	0,6	0,5	4,5
D	0,6	0,4	4,4
F	0,4	0,3	2
G	0,8	0,6	4,5
J	0,6	0,5	4
K	0,6	0,5	4,5
L	0,9	0,8	4,5
M	0	0	0
N	0,5	0,4	3,5
ñ	0,5	0,4	3,5
P	0	0	0
R	0,9	0,6	4
RR	0,9	0,8	4,7
S	0,5	0,3	3,5
T	0,7	0,5	4
Y	0,5	0,4	3
A	0,8	0,6	4,9
E	0,8	0,5	4,6
I	0,6	0,5	3,7
O	0,5	0,3	2,9
U	0,5	0,2	2,1

Nota: Valores adimensionales.

Cada uno de estos valores fue distribuido específicamente para formar tres rectas con distinta pendiente unidas entre sí. Además, se realizó una representación opuesta de cada distancia, es decir, en el eje de coordenadas negativo, de manera de poder asemejar la forma de una boca humana, como se muestra en la figura 57.

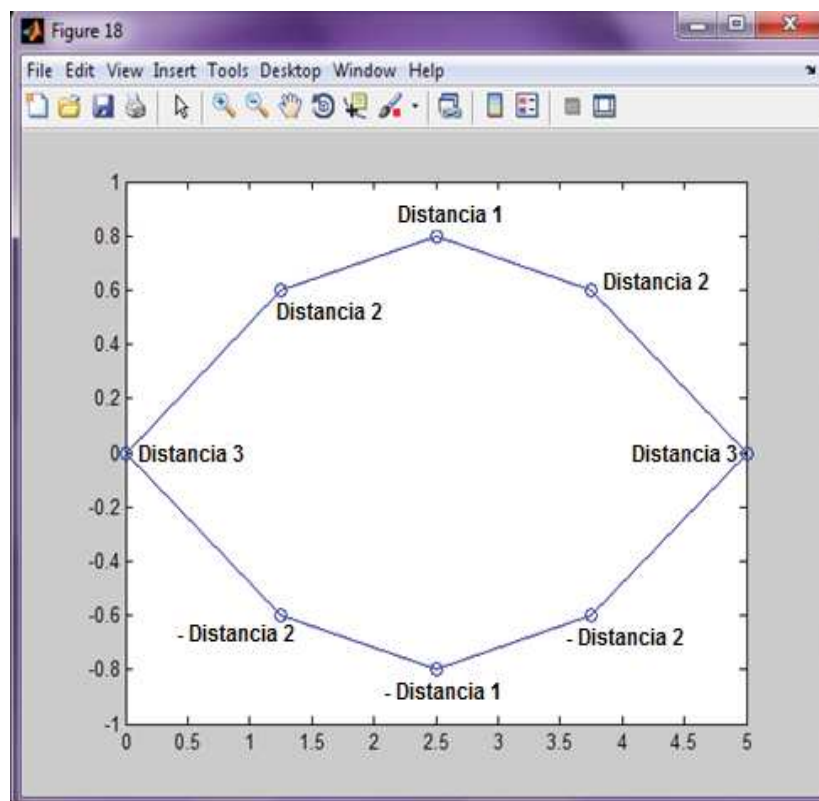


Figura 57.

Gráfica de las distancias de la boca.

En condiciones ideales estas distancias deberían ser obtenidas por el conjunto de Redes Neuronales Artificiales de manera exacta, sin embargo, eso no resulta tan sencillo en la práctica.

Los resultados que cada RNA obtuvo, fueron almacenados y distribuidos en las tablas presentadas a continuación, en donde se pueden observar los valores segmentados para cada sílaba, separando los valores para la abertura de la consonante y de la vocal.

Tabla 14.

Respuesta de la Red Neuronal Artificial "Nasales". Valores correspondientes al fonema /n/.

		NA	NE	NI	NO	NU
CONSONANTE	Distancia 1	0,5	0,5	0,5	0,5	0,31
	Distancia 2	0,4	0,4	0,4	0,4	0,25
	Distancia 3	3,5	3,5	3,5	3,5	2,19
VOCAL	Distancia 1	0,5	0,8	0,8	0,6	0,5
	Distancia 2	0,4	0,6	0,5	0,5	0,31
	Distancia 3	3,5	4,9	4,6	3,7	2,93

Nota: Valores adimensionales.

Tabla 15.

Respuesta de la Red Neuronal Artificial "Oclusivasorda". Valores correspondientes al fonema /t/.

		TA	TE	TI	TO	TU
CONSONANTE	Distancia 1	0,7	0,7	0,7	0,6	0,7
	Distancia 2	0,5	0,5	0,5	0,43	0,5
	Distancia 3	4	4,02	4,02	3,5	4
VOCAL	Distancia 1	0,6	0,8	0,8	0,6	0,5
	Distancia 2	0,5	0,6	0,5	0,5	0,3
	Distancia 3	4,5	4,89	4,6	3,67	2,9

Nota: Valores adimensionales.

Tabla 16.

Respuesta de la Red Neuronal Artificial "Oclusivasonora".
Valores correspondientes al fonema /b/.

		BA	BE	BI	BO	BU
CONSONANTE	Distancia 1	0	-0,02	0	0	0
	Distancia 2	0	-0,01	0	0	0
	Distancia 3	0	-0,12	0	0,01	0
VOCAL	Distancia 1	0,8	0,8	0,8	0,59	0,5
	Distancia 2	0,6	0,6	0,5	0,49	0,3
	Distancia 3	4,5	4,9	4,6	3,64	2,9

Nota: Valores adimensionales.

Tabla 17.

Respuesta de la Red Neuronal Artificial "Fricativasorda".
Valores correspondientes al fonema /f/.

		FA	FE	FI	FO	FU
CONSONANTE	Distancia 1	0,4	0,4	0,4	0,4	0,4
	Distancia 2	0,3	0,3	0,3	0,3	0,3
	Distancia 3	1,98	2	2,01	2	2,01
VOCAL	Distancia 1	0,6	0,8	0,8	0,6	0,5
	Distancia 2	0,5	0,6	0,5	0,5	0,3
	Distancia 3	4,02	4,9	4,59	3,69	2,89

Nota: Valores adimensionales.

Tabla 18.

Respuesta de la Red Neuronal Artificial "Latefrica".

Valores correspondientes al fonema //l/.

		LA	LE	LI	LO	LU
CONSONANTE	Distancia 1	0,9	0,9	0,9	0,89	0,9
	Distancia 2	0,8	0,8	0,8	0,79	0,8
	Distancia 3	4,55	4,5	4,5	4,48	4,5
VOCAL	Distancia 1	0,5	0,8	0,8	0,6	0,5
	Distancia 2	0,4	0,6	0,5	0,5	0,3
	Distancia 3	2,99	4,9	4,6	3,68	2,9

Nota: Valores adimensionales.

Tabla 19.

Respuesta de la Red Neuronal Artificial "Vibrante".

Valores correspondientes al fonema /r/.

		RA	RE	RI	RO	RU
CONSONANTE	Distancia 1	0,9	0,9	0,9	0,9	0,9
	Distancia 2	0,55	0,58	0,72	0,67	0,75
	Distancia 3	4,1	4,78	4,39	5,15	5,28
VOCAL	Distancia 1	0,76	1,05	1,2	1,21	1,41
	Distancia 2	0,14	-0,02	-0,21	0,56	0,45
	Distancia 3	1,01	6,17	7,17	3,34	5,72

Nota: Valores adimensionales.

Realizando una comparación entre las distancias de apertura de los labios de todas las consonantes y vocales que entregaron las seis Redes Neuronales

Artificiales, con las dimensiones de abertura medidas a cada consonante y vocal de la tabla 13, se puede determinar que el porcentaje de error obtenido total es de aproximadamente el 12%. Por consiguiente el porcentaje de identificación de las sílabas con sus respectivas coordenadas, es de 88%.

El porcentaje de reconocimiento que presentan las RNA determina que la identificación por clasificación de consonante es bastante viable y con muy buenos resultados, por lo que se puede aproximar con claridad la representación de la sílaba que se está pronunciando. A pesar de esto, se debe recalcar que este análisis comprende el resultado obtenido de la evaluación de cada red de manera independiente, y con las muestras usadas en la etapa de validación y prueba.

Es necesario dejar en constancia que para la obtención de los resultados mostrados en las tablas anteriores, se usaron las mismas muestras de sílabas empleadas para la etapa de entrenamiento de las RNA. Esto se hizo para afirmar la efectividad de las redes en cuanto al reconocimiento de las sílabas con que se entrenaron, y para corroborar lo dicho en cada análisis descrito en el apartado de resultados de entrenamiento de las RNA.

Para poder obtener resultados más claros acerca de la capacidad de reconocimiento de las redes, fue necesario realizar algunas comparaciones con otro tipo de muestras que no sean las utilizadas para el entrenamiento. Es por esto que se realizó otro análisis de reconocimiento de sílabas en las seis Redes Neuronales Artificiales ingresando datos de formantes de nuevas muestras de las mismas ochenta y cinco sílabas. Estas sílabas fueron grabadas por la misma persona y a la misma velocidad de pronunciación (considerada como "normal"). Los nuevos archivos de texto de las muestras de formantes se ingresaron y se distribuyeron en las distintas redes que se encuentran divididas por categoría de consonantes.

Los resultados indicaron que el porcentaje de error aumentó al 30%, y por consiguiente el porcentaje de reconocimiento bajó al 70%. Si bien estos resultados podrían ser considerados como aceptables en otras condiciones, al

referirse a la misma persona y en las mismas condiciones de grabación, en realidad deben ser considerados como poco favorables, ya que demuestra que las redes son muy dependientes de las muestras de entrenamiento que se le otorgaron.

2.2.4 Reconocimiento de las seis Redes Neuronales Artificiales en conjunto

Para que las seis Redes Neuronales creadas por clasificación de consonante funcionen simultáneamente, fue necesaria la creación de una función llamada "pruebafinal", la cual posee una variable de entrada de datos "n" y una variable de salida de datos "nt". Cada una de las redes posee un nombre característico, el cual es colocado en la variable de identificación de red "networkname". Luego de esto se cargan las redes con la función "load (networkname,'net')".

Cada red posee un vector de entrada "P" y un vector de salida "Y", los cuales cambian su nomenclatura dependiendo a qué red correspondan; por ejemplo, en la red "nasales" el vector de entrada es "Pn" y el vector de salida es "Yn"; mientras que en la red "vibrantes" el vector de entrada es "Pv" y el vector de salida es "Yv"; y así sucesivamente para cada red.

Cada vector de entrada de las seis Redes Neuronales Artificiales tiene involucrado la variable de entrada "n" de la función "pruebafinal", como se observa en la figura 58, por lo que si se ingresa en cada vector de entrada "P" un listado de datos de frecuencia de formantes de cualquier sílaba con el nombre de la variable de entrada "n", los valores ingresarán en las seis Redes Neuronales y estas entregarán mediante los vectores de salida "Y" los seis datos de coordenada de la sílaba que se ingresó con la variable "n".

```

1  %RESPUESTA DE LA RED NEURONAL CON LAS 6 CLASIFICACIONES
2  function nt = pruebafinal(n)
3  % close all;
4  % clear all;
5  % clc;
6
7  networkname='nasales';
8  load (networkname,'net');
9
10 % fid=fopen(n,'rt');
11 % datemp=fscanf(fid,'%10g\n',[1,inf]);
12 % st=fclose(fid);
13 tama=size(n,2);
14 mtam=floor(tama/2);
15 Pn(1,:)=[n(1:50) n(mtam+1:mtam+50)];
16 Pn=Pn';
17 Yn = sim(net,Pn);
18 %Yn1=round(Yn) %REDONDEO DE LAS MUESTRAS
19
20 networkname='vibrante';
21 load (networkname,'net');
22
23 % fid=fopen(n,'rt');
24 % datemp1=fscanf(fid,'%10g\n',[1,inf]);
25 % st=fclose(fid);
26 tamal=size(n,2);
27 mtam1=floor(tamal/2);
28 Pv(1,:)=[n(1:50) n(mtam1+1:mtam1+50)];
29 Pv=Pv';
30 Yv = sim(net,Pv);
31 %Yv1=round(Yv)

```

Figura 58.

Programación de la función “pruebafinal” parte (1).

Después de lo anterior, se realizó una programación, como se observa en la figura 59, en la que se restan los valores de coordenada de los vectores de salida de las redes “Y”, con los datos reales.

Las diferencias se realizan entre las seis coordenadas de cada una de las sílabas, por lo que las variables “Comp” y “m[sílabas]” cambian dependiendo de estas. Se debe mencionar que las seis coordenadas corresponden a las tres coordenadas descritas anteriormente con signo positivo (+) y negativo (-).

```

127 %Probar luego que funcione la red dichas condiciones.
128 %M
129 Comp=abs(Yn-[0; 0; 0; 0.8; 0.6; 4.9]);
130 mMA=mean(Comp);
131 Comp1=abs(Yn-[0; 0; 0; 0.8; 0.5; 4.6]);
132 mME=mean(Comp1);
133 Comp2=abs(Yn-[0; 0; 0; 0.6; 0.5; 3.7]);
134 mMI=mean(Comp2);
135 Comp3=abs(Yn-[0; 0; 0; 0.5; 0.3; 2.9]);
136 mMO=mean(Comp3);
137 Comp4=abs(Yn-[0; 0; 0; 0.5; 0.2; 2.1]);
138 mMU=mean(Comp4);
139 %N
140 Comp5=abs(Yn-[0.5; 0.4; 3.5; 0.8; 0.6; 4.9]);
141 mNA=mean(Comp5);
142 Comp6=abs(Yn-[0.5; 0.4; 3.5; 0.8; 0.5; 4.6]);
143 mNE=mean(Comp6);
144 Comp7=abs(Yn-[0.5; 0.4; 3.5; 0.6; 0.5; 3.7]);
145 mNI=mean(Comp7);
146 Comp8=abs(Yn-[0.5; 0.4; 3.5; 0.5; 0.3; 2.9]);
147 mNO=mean(Comp8);
148 Comp9=abs(Yn-[0.5; 0.4; 3.5; 0.5; 0.2; 2.1]);
149 mNU=mean(Comp9);
150 %ñ
151 Comp10=abs(Yn-[0.5; 0.4; 3.5; 0.8; 0.6; 4.9]);
152 mNIA=mean(Comp10);
153 Comp11=abs(Yn-[0.5; 0.4; 3.5; 0.8; 0.5; 4.6]);
154 mNIE=mean(Comp11);
155 Comp12=abs(Yn-[0.5; 0.4; 3.5; 0.6; 0.5; 3.7]);
156 mNII=mean(Comp12);
157 Comp13=abs(Yn-[0.5; 0.4; 3.5; 0.5; 0.3; 2.9]);

```

Figura 59.

Programación de la función “pruebafinal” parte (2).

Luego se realizó una comparación entre las diferencias con el condicional “if”, la cual determina que si la variable “Comp” es menor o igual al valor 0.001 entre cada una de las seis coordenadas, despliegue el nombre de la sílaba a la que le pertenecen dichas coordenadas. Esto se puede ver en la figura 60.

```

316
317 - if Comp <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
318 -     disp('MA')
319 - elseif Comp1 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
320 -     disp('ME')
321 - elseif Comp2 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
322 -     disp('MI')
323 - elseif Comp3 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
324 -     disp('MO')
325 - elseif Comp4 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
326 -     disp('MU')
327 - elseif Comp5 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
328 -     disp('NA')
329 - elseif Comp6 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
330 -     disp('NE')
331 - elseif Comp7 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
332 -     disp('NI')
333 - elseif Comp8 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
334 -     disp('NO')
335 - elseif Comp9 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
336 -     disp('NU')
337 - elseif Comp10 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
338 -     disp('ñA')
339 - elseif Comp11 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
340 -     disp('ñE')
341 - elseif Comp12 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
342 -     disp('ñI')
343 - elseif Comp13 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
344 -     disp('ñO')
345 - elseif Comp14 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
346 -     disp('ñU')

```

Figura 60.

Programación de la función “pruebafinal” parte (3).

Por último, se logró que MATLAB identificara a cada una de las sílabas por su nombre, ya que solo las lograba identificar como coordenadas o datos numéricos.

Esto se logró realizando algunos procesos. El primero consistió en definir una variable que abarque el número completo de sílabas analizadas. Esta variable se la denominó como “letras”. Cabe recordar que el número de sílabas en total es igual a ochenta y cinco.

Luego de esto se incluyó un arreglo de funciones de MATLAB: `strArray = java_array('java.lang.String', letras)`, el cual es igual Al número de todas las sílabas CV debido a la inclusión de la variable “letras”. Además de esto se creó una variable denominada como “silabas”, la cual contiene la función “cell” que actúa sobre el arreglo “strArray” definido anteriormente.

Con todas las variables “m[NOMBRE DE LA SÍLABA]” se formó una matriz con el nombre de “compfinal”. A dicha matriz se le extrajo el mínimo valor mediante la función “min”, y ese valor se lo nombró como “valorminglobal”. Esta variable es colocada dentro de un lazo “for” que va desde uno hasta el número total de sílabas, dentro del cual se coloca un condicional “if” que determina si la variable “valorminglobal” es igual a “compfinal” mediante la sintaxis: `vectfinal(a) = java.lang.String ((char(cell2mat(silabas(a))))))`.

Finalmente por medio de la función “char” se logró que se entregue el nombre de la sílaba deseada. Todo esto se puede observar en la figura 61 a partir de la línea de programación 503.

```

495 - disp ('KE')
496 - .seif Comp82 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
497 - disp ('KI')
498 - .seif Comp83 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
499 - disp ('KO')
500 - .seif Comp84 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
501 - disp ('KU')
502 - id
503 - letras=85;
504 - strArray = java_array('java.lang.String', letras);
505 - strArray = {'MA', 'ME', 'MI', 'MO', 'MU', 'NA', 'NE', 'NI', 'NO', 'NU', 'ÑA', 'ÑE', 'ÑI'}
506 - silabas = cell(strArray);
507
508 - compfinal = [mMA, mME, mMI, mMO, mMU, mNA, mNE, mNI, mNO, mNU, mÑA, mÑE, mÑI, mÑO, mÑU, m
509
510 - vectfinal = java.lang.String (cellstring algo)
511
512 - valorminglobal = min(compfinal);
513
514 - for a=1:85%Son todas las silabas, consonantes por vocales
515
516 -     if valorminglobal==compfinal(a);
517 -         vectfinal(a) = java.lang.String ((char(cell2mat(silabas(a)))));
518 -         nt=char(vectfinal);%Lared entrega la respuesta en formato string
519 -         %(vectfinal(a))
520 -     end
521
522 - end
523
524

```

Figura 61.

Programación de la función “pruebafinal” parte (4).

Los resultados de identificación de las sílabas que se obtuvieron al utilizar las seis Redes Neuronales simultáneamente son los siguientes: el porcentaje de identificación de las consonantes es del 78%, y el porcentaje de identificación de las vocales es del 13%. Esto indica que no se obtuvo un reconocimiento de la sílaba completa al utilizar las seis redes a la vez, sino que la identificación se logró en la consonante, pero no en un cien por ciento.

Una razón primordial para que este hecho se haya producido es la inclusión de únicamente sesenta muestras de formantes en el entrenamiento de la red. Esto influye principalmente en la vocal, debido a que dentro de una sílaba CV la

consonante siempre se ubica primero, por lo que las primeras muestras de formantes siempre corresponden a la consonante mientras que las restantes corresponden a la vocal. Por lo tanto, al considerar dicho número de muestras de formantes se está restando información de la vocal.

Sería lógico pensar que dado que el problema radica en la reducida cantidad de muestras que se considera, la solución óptima es aumentar el valor de muestras que la red considera para su entrenamiento. Sin embargo, esto no resulta tan fácil de realizar debido a que la cantidad de muestras de formantes depende de la duración de la sílaba y más específicamente de cada fonema, lo cual va directamente relacionado a diversos factores que influyen para que una letra sea más corta o prolongada que otra; por lo que al aumentar el número de muestras que se consideran, se presenta otro problema más complejo. Este problema consiste en que existen sílabas que poseen menos cantidad de muestras que otras (como se mencionó en la sección 2.2.1.5), por lo que al no existir el número especificado a la RNA para su entrenamiento, MATLAB arroja un error debido a falta de información.

Además de lo mencionado anteriormente, una de las principales observaciones realizadas acerca de la deficiencia en la identificación de vocales, radica en la influencia que ejercen las consonantes sobre sus formantes.

Al considerar un número reducido de formantes de una vocal, se puede llegar a “cambiar” totalmente la caracterización de una vocal por otra. La explicación de esta hipótesis es la siguiente:

Debido a que al combinar una consonante con una vocal, los formantes de la vocal tienden a reducir su frecuencia en comparación a su comportamiento normal al pronunciarla de manera aislada; pueden existir consonantes que produzcan dicha variación en una vocal en las mismas proporciones que otra consonante produce sobre otra vocal. Un ejemplo que permite clarificar esto se presenta a continuación:

La vocal “e” posee un valor promedio de F1 de 530 Hz aproximadamente, al pronunciarla sola. Sin embargo, al decirla en una sílaba /de/, su valor de F1

tiende a bajar a 400 Hz aproximadamente. Del mismo modo, la vocal “o” en una sílaba /bo/ posee el mismo valor de $F1 = 400$ Hz; por lo tanto, es muy probable que ambas vocales sean confundidas.

En conclusión, al no considerar todo el comportamiento de formantes a lo largo del tiempo, la posibilidad de confundir las vocales aumenta en gran manera, debido a que se está descartando la principal característica de diferenciación entre ellas, que es su curva de formantes a lo largo del tiempo.

2.2.5 Elaboración de la animación de las diferentes aberturas de la boca

La animación correspondiente al movimiento de los labios propios de cada fonema pronunciado, se la realizó por medio de una representación de distintos gráficos de la forma de la boca humana, cada uno de ellos con diferentes aberturas. Esto se hizo utilizando las medidas de abertura que se presentan en la tabla 13. La distancia uno corresponde a la abertura más grande que se produce al pronunciar cualquier sílaba, la dos es la abertura considerada como la más pequeña; y la tres es la medida del ancho de la boca.

Para que el gráfico obtenido de la figura 57 realice movimientos articulatorios que simulen la pronunciación de las sílabas CV, se hicieron varias líneas de programación que posicionan a las rectas en diferentes puntos y simbolizan la forma ovalada que tiene la abertura de la boca.

Lo primero fue empezar con una posición cerrada, que se representa como una recta horizontal en el eje de la abscisa, la cual va desde cero hasta el valor de la distancia tres; hasta que las rectas alcancen los valores de las distancias uno y dos, que son los límites de posición.

A los valores encontrados se los extrajo en dos variables llamadas “resp” y “resp1” como se puede ver en las líneas de programación 15 y 16 de la hoja de MATLAB presentada en la figura 62.

Estas variables se las condiciona mediante un “if”, ya que existen dos consonantes que poseen dos letras en la práctica en términos de programación, estas son la /ch/ y la /rr/. Por lo tanto en la práctica existirían dos sílabas que no poseen dos letras sino tres, por lo que el condicional “if” permite extraer la tercera letra de la sílaba.

Las letras extraídas de la sílaba se las ingresa a un nuevo condicional “if” en el que se compara la variable “cons”, la cual contiene a las letras de la sílaba, con la sílaba escrita. Cuando la condición se cumple se arrojan los datos de distancias pertenecientes a la sílaba como tal.

Después de definir las sílabas y sus distancias dentro del “if” general, se extrajeron uno por uno, los seis valores de distancia de la matriz de la sílaba resultante, de los cuales tres pertenecen a la consonante y los otros tres a la vocal, como ya se mencionó.

Los valores de las distancias uno y dos son ingresados a un lazo “for” con un contador “n” que va desde uno hasta diez. Esto indica que dichos valores de distancia inicialmente tendrán el valor de cero e irán creciendo a medida que la variable “n” del lazo “for” aumente.

y dos tomen distintos valores, y vayan en aumento hasta que alcancen los valores originales de distancia de la sílaba resultante.

Luego de esto se definió en una matriz llamada "x", con la distancia tres, la cual determina el ancho total de la boca. A dicho valor se lo distribuye en porcentajes del 25% a lo largo de todo el eje x. Esto se realizó para que las distancias uno y dos se ubiquen en los valores que pertenecen al porcentaje mencionado de la distancia tres, para que a partir de dichos puntos se pueda ejecutar las distintas posiciones de abertura de la boca.

A continuación se distribuyeron las variables "A", "B", "C" y "D", en matrices denotadas por la letra "y", las cuales se encuentran diferenciadas una de otra por un número específico.

Cada matriz "y" tiene cuatro valores, los cuales son: un cero inicial y un cero final, (los cuales definen los límites de la boca), y dos variables, las cuales se distribuyen entre "A" y "B", o "C" y "D", ya que "A" y "B" representan a las coordenadas de abertura para la vocal, y "C" y "D" representan a las distancias de abertura para la consonante. Las variables deben estar en las mismas posiciones de las distancias uno y dos, tal como en la figura 57.

variable “x”, y en el eje de las ordenadas a una matriz “y”. Ejecutando a la programación total, se logra la animación esperada del movimiento de los labios al pronunciar una sílaba.

2.2.5.1 Resultados de la animación de la boca

La animación resultante se asemeja a como un ser humano habla con normalidad, sin embargo, se tuvieron que hacer algunas adaptaciones en los cuadros de película de la programación para que el movimiento de los labios sea real. Aunque inicialmente se querían considerar 24 cuadros, se decidió quitar algunos de estos que representaban la abertura final de la consonante, porque al ejecutarse la animación, la consonante sobrepasaba en abertura a la vocal. También se quitaron algunos cuadros que representaban la abertura inicial de la vocal, porque se quería empatar los cuadros de la abertura final de la consonante con los de la abertura inicial de la vocal, y así lograr que la transición se vea como una sola abertura. Con dicha reducción, se pudo visualizar el movimiento aproximado a cómo un ser humano habla normalmente.

En la figura 65 presentada a continuación, se muestran los distintos cuadros del movimiento de abertura de la boca al pronunciar en este ejemplo la sílaba “RA”.

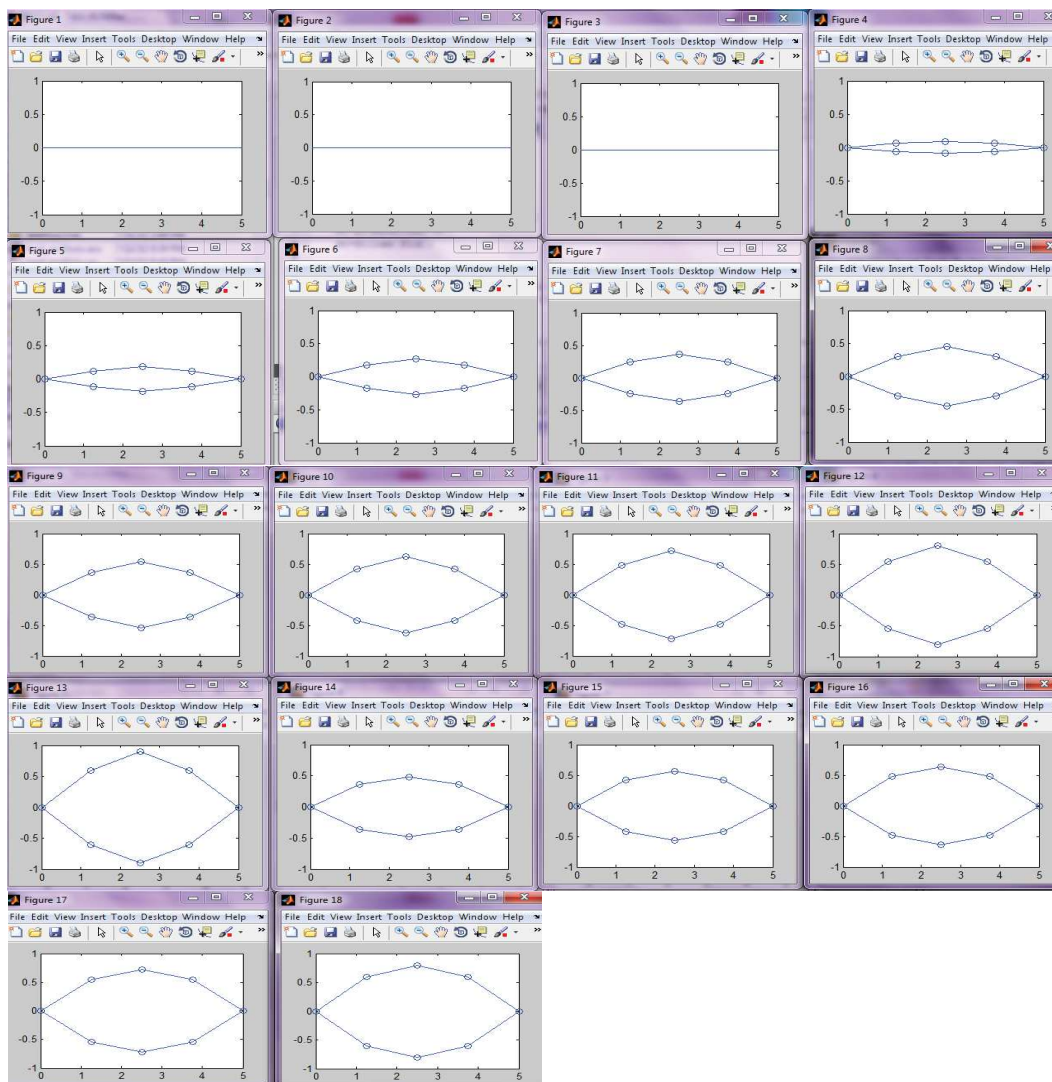


Figura 65.

Los dieciocho cuadros de abertura de la boca. Se representa la pronunciación de la sílaba “RA”.

2.2.6 Interfaz gráfica del software diseñado

La interfaz gráfica está diseñada para trabajar con archivos de texto, los cuales deben poseer la información de frecuencia de formantes de una sílaba. Dicha información puede provenir de cualquier software que entregue valores de formantes, sin embargo el software usado para el presente proyecto fue PRAAT.

Debido a lo anterior, es recomendable para obtener mejores resultados, que los archivos de texto con la información de formantes provengan de dicho software, ya que no se puede saber qué tipo de algoritmo se usó en el caso de que los datos provengan de “x” software.

PRAAT es considerado como el complemento del presente proyecto, en donde el software diseñado tiene como nombre “FORMANT-VOCAL”. PRAAT se encarga de dos procesos iniciales y que no incorpora el software “FORMANT-VOCAL”, los cuales son: grabación de archivos de audio, y determinación de los formantes de las muestras de audio pregrabados. Estos dos procesos mencionados, no se emitieron en el desarrollo del software “FORMANT-VOCAL” porque los formantes de las sílabas que se utilizaron para el estudio del reconocimiento del habla humana, se los obtuvo del análisis de cada sílaba del software PRAAT, debido a que dicha tarea no conformaba una parte de desarrollo del presente trabajo, ya que se prefirió usar una herramienta especializada para obtener dicho fin.

En la figura 66 se puede observar el diseño de la interfaz gráfica del software de reconocimiento de sílabas “FORMANT-VOCAL”. Esta interfaz posee algunas partes que serán mencionadas de manera general.

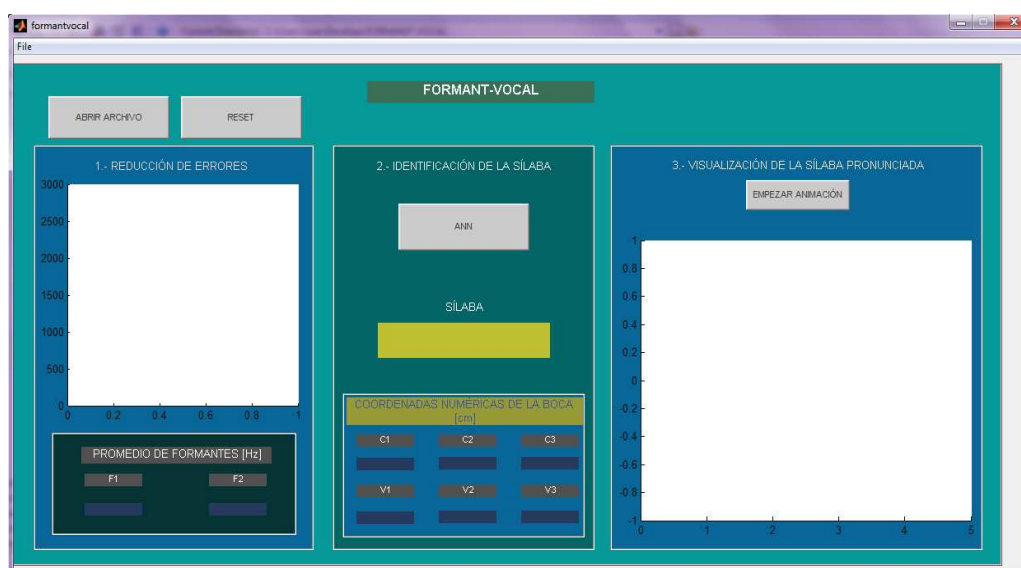


Figura 66.

Interfaz gráfica del software “FORMANT-VOCAL”.

La interfaz gráfica presenta tres secciones. La primera sección es denominada como “Reducción de errores”. Esta etapa abarca la reducción de picos de error mediante la implementación de la función “desviación”, la cual se explicó en la sección 2.2.1.6; adicional a esto se muestran los promedios de formantes F1 y F2 para las curvas de formantes ya corregidas.

La segunda sección se denomina como “Codificación de la sílaba”. En esta se identifica la sílaba de la cual se posee la información de formantes, con la implementación de las seis Redes Neuronales Artificiales en conjunto, las cuales están diseñadas para presentar la única sílaba resultante que se logró identificar en el cuadro de color amarillo ubicado en el centro de esta sección. Adicional a esto se realizó una programación en la que se muestran los datos de coordenadas uno, dos y tres, pertenecientes a la sílaba identificada.

Finalmente se tiene la tercera sección, la cual es denominada como “Visualización de la sílaba pronunciada”. En esta se presenta la animación de los labios pronunciando la sílaba reconocida.

Haciendo click en el botón “ABRIR ARCHIVO” permite abrir una ventana en la que se muestra el explorador de Windows con la ubicación por defecto del usuario. Dentro de él, el usuario debe buscar y escoger el archivo de texto (“.txt”) en la ubicación deseada, este archivo debe corresponder a la información de formantes uno y dos de la sílaba que se desea analizar. Inmediatamente después de cargar el archivo se activa la función “desviación” y se realiza el proceso de reducción de errores. Una vez terminado este proceso aparece un cuadro de diálogo que advierte acerca de que el proceso ha finalizado con éxito.

La segunda etapa se activa presionando el botón “ANN”, el cual recoge los resultados arrojados por la primera etapa de reducción de errores y mediante el uso de las Redes Neuronales, entrega la sílaba escrita a la que le pertenecen dichos resultados de formantes. Presionando el botón “Empezar Animación” se inicia la reproducción de los distintos cuadros y se puede visualizar la

animación de la sílaba resultante en dos velocidades, la primera en velocidad normal o rápida, mientras que la segunda en velocidad lenta.

Finalmente, se debe mencionar que el botón "RESET" hace que todos los resultados se borren para poder iniciar los procesos nuevamente en caso de requerirlo. Presionando este botón se borran los datos numéricos almacenados en las distintas secciones del software, así como el gráfico de reducción de errores y el último cuadro presente en la animación del movimiento de los labios de la sílaba seleccionada con anterioridad.

Capítulo III.- Análisis económico

3.1 Costo de elementos empleados

El presente trabajo de titulación corresponde un análisis investigativo que abarca aspectos de diseño de software, así como un estudio experimental de los fonemas castellanos. Por consiguiente, se emplearon en casi su totalidad a herramientas informáticas, dejando de lado otros aspectos que se pudieran presentar en trabajos de este tipo, como: salidas de campo, mediciones, etc.

El detalle de la inversión realizada en el presente proyecto se describe en la tabla 20.

Tabla 20.

Descripción de los costos correspondientes a los elementos usados.

Detalle	Unidad	Cantidad	Precio unitario (\$)	Total (\$)
Computador	u	1	1115.00	1115.00
Micrófono profesional	u	1	130.00	130.00
Interfaz de audio	u	1	200.00	200.00
Conexión a Internet	mes	3	20.00	60.00
Software Matlab R2009a	u	1	100.00	100.00
Software PRAAT	u	1	0.00	0.00
Material bibliográfico	libro	1	70.00	70.00
Asesoría	horas	5	20.00	100.00
Documentos impresos	u	5	15.00	75.00
Gastos adicionales	-	-	-	30.00
			TOTAL	1880.00

3.2 Relación costo-beneficio

El costo final del presente proyecto descrito en la tabla 20, corresponde al valor total en dólares de los elementos empleados, en caso de que no se posea de ninguno de ellos. Esto no siempre es así, ya que la mayoría de veces se tienen algunos de estos en el hogar, como un computador de mediana capacidad, por citar un ejemplo. Sin embargo, para saber el costo real de un proyecto siempre es necesario asumir que la inversión se la realizará en su totalidad para todos los elementos necesarios.

El beneficio del presente proyecto en base a la inversión realizada, se lo puede considerar como "*medio*"; ya que si bien el valor no es sumamente reducido, los beneficios obtenidos son muy considerables en base a los experimentos realizados del habla, automatización de programaciones informáticas, capacidad de análisis de formantes, etc.

Capítulo IV.- Conclusiones y Recomendaciones

El presente proyecto comprende el desarrollo de distintas etapas investigativas acerca del estudio del lenguaje castellano y el diseño de un programa que procese información acústica de los fonemas en base a la utilización de Redes Neuronales Artificiales. En lo referente a la hipótesis planteada a inicios del proyecto, se debe mencionar que después de haber analizado los fonemas de las sílabas, el método matemático para el reconocimiento de estas por medio de formantes, y habiendo obtenido los porcentajes de reconocimiento mencionados; se concluye que los formantes, por sí solos, no se los puede considerar como un factor de identificación suficiente de los fonemas del castellano. El reconocimiento obtenido en base a ellos, es mejor cuando se pronuncian los fonemas aislados, más no es suficiente en los casos de conformación de sílabas, como los planteados en este trabajo.

Además de esto, en base a todo el análisis realizado a lo largo del proyecto, se cree conveniente dividir en dos secciones las conclusiones generales obtenidas, en base a los dos objetivos generales planteados.

La primera sección corresponde al estudio de los fonemas del castellano, abarcando desde su parte fisiológica de generación, hasta su parte acústica de caracterización. En base a toda la investigación realizada acerca los procesos de generación distintos en cada fonema, así como las características de comportamiento temporal de sus formantes, se pueden establecer las siguientes conclusiones:

- Cada fonema constituye una ciencia distinta dentro del fenómeno del habla. Si bien existen clasificaciones según algunos parámetros que permiten establecer factores comunes entre diversos grupos de fonemas, cada uno de ellos está formado por un conjunto de procesos de generación internos del cuerpo humano que determinan de manera específica el comportamiento de los formantes a lo largo del tiempo. Aunque un fonema posea procesos de elaboración en el aparato fonatorio similares a otro, siempre existen diferencias que hacen de cada uno un

universo completamente distinto y complejo, por lo cual es complicado segmentar y enlistar sus características de manera exacta. A pesar de esto, a lo largo del tiempo que duró el desarrollo del presente trabajo, considerando las distintas etapas investigativas y experimentales presentes en él; se pudo realizar un estudio específico de cada uno de los fonemas del castellano en base a la búsqueda de material teórico previo y un análisis personal realizado por los autores de este proyecto, en base a grabaciones hechas a distintas personas de diversos sexos y edades.

- Realizar un estudio a fondo de las características de los fonemas del castellano constituye un trabajo completamente extenso y complejo de realizar, el cual ha sido objeto de desarrollo de varios proyectos y trabajos alrededor del mundo (Massone, 1988; Fernández y Feijóo, 2004). En la elaboración del presente proyecto investigativo se pudo comprobar que a distintos investigadores y catedráticos de diversas universidades e institutos a nivel mundial, les ha costado mucho trabajo encontrar características determinantes en el comportamiento de formantes de los fonemas. Es así que para poder obtener mejores resultados, optaron por segmentar sus estudios a grupos específicos de estos. Debido a esto existen estudios por separado de consonantes nasales, oclusivas, vibrantes, etc., y cada uno de ellos ha sido desarrollado con suma complejidad.

A pesar de esto, en el presente trabajo se optó por abarcar el estudio de todas las consonantes y vocales del castellano, segmentando dicho estudio general en estudios más pequeños basados en las clasificaciones mencionadas en los capítulos correspondientes. Aunque los resultados obtenidos aquí no sean tan exactos como en los trabajos mencionados de otros investigadores, se puede concluir que las caracterizaciones acústicas presentadas comprenden un análisis suficiente para satisfacer las necesidades y los objetivos planteados a inicios del proyecto. Además, es satisfactorio saber que se ha llegado a conclusiones similares en la caracterización de los fonemas, a las de otros investigadores que contaban con mayor tiempo y recursos disponibles.

- Además de las diferencias propias de cada fonema debido a su generación por el lugar y modo de articulación, los formantes presentes en ellos constituyen una característica predominante para su diferenciación con respecto a los demás fonemas. Cada formante ofrece información distinta que permite a un oyente identificar de manera óptima los sonidos que se producen al hablar. Sin embargo, es necesario dejar constancia de que un único valor de formante no es válido para que se produzca cualquier tipo de identificación, ya sea de una persona o de algún algoritmo computacional programado en un ordenador.

La observación anterior conlleva a una conclusión fundamental en el análisis de formantes realizado. Esta consiste en que el comportamiento temporal de cada curva de formantes ofrece la información necesaria para caracterizar a un único sonido. Esto se lo pudo comprobar debido a que, a medida en que se consideraba una menor cantidad de muestras de formantes a lo largo del tiempo, la identificación realizada era más pobre.

Lo anterior es lógico de pensar, debido a que cada formante corresponde a un pico de resonancia producido por los resonadores internos del aparato fonatorio (cavidad oral y nasa); en donde las distintas configuraciones adoptadas por todos los órganos, van cambiando en función en que el sonido se va desarrollando en su totalidad, desde su comienzo hasta su fin (esto debido a que ambas etapas poseen un comportamiento de formantes distinto).

Es coherente pensar, por ejemplo, que el comienzo del sonido de una /p/ es más explosivo que su finalización, ya que el flujo de aire sale por completo en ese instante repentino; lo cual se pudo comprobar en el análisis correspondiente realizado en la sección 2.2.4.1. Sin embargo, la posible razón que hace obviar dicho pensamiento es que todos esos cambios se producen en instantes de tiempo muy cortos.

Por consiguiente, se finaliza que cada parte presente en una curva de formantes, ofrece información importante acerca de las resonancias producidas en el aparato fonatorio a lo largo del tiempo de generación de

cierto fonema, lo cual ayuda a que exista una correcta identificación del sonido.

La segunda sección de conclusiones comprende aquellas referentes a todos los procesos realizados por el conjunto de Redes Neuronales Artificiales en la identificación de los fonemas y la animación emitida correspondiente. Se debe recalcar que el principal objetivo de esta etapa del proyecto consistió en lograr una representación visual del movimiento de los labios diferenciada de acuerdo al fonema hablado, por lo que se trató de optimizar los procesos para obtener dicho fin en base a una identificación no tan exacta de la sílaba correspondiente. Esto debido a que existen algunas consonantes en las que se puede visualizar un movimiento labial similar o igual a otras.

En base a todos los procesos computacionales realizados a lo largo del desarrollo del presente proyecto de titulación, en donde se realizaron varios procedimientos experimentales buscando obtener el mejor desenvolvimiento posible, se pueden mencionar las siguientes conclusiones:

- Una de las principales variaciones realizadas a lo largo del desarrollo del proyecto consistió en la modificación del vector de entrada de las Redes Neuronales Artificiales. Esta modificación abarcó un cambio radical en la consideración de los formantes que servirían como vector de entrada para las ANN, ya que en primera instancia se pensaba realizar el cálculo de formantes directamente en el mismo software, lo cual conllevaría a muchas complicaciones innecesarias.

Al considerar que la información de formantes que ingrese a las Redes Neuronales Artificiales debía proceder de archivos de texto previamente obtenidos por el usuario, se logró facilitar en gran manera los procedimientos de identificación, ya que se pudo optimizar tanto las etapas de entrenamiento como de ejecución. Todo esto conllevó a asignar correctamente los recursos de trabajo a los procesos netamente necesarios en el desarrollo del presente proyecto investigativo, el cual no abarca ninguna etapa de cálculo de formantes sino que los usa como herramienta para la realización de los cálculos computacionales establecidos.

- Elaborar un proceso de identificación de archivos de texto con información de formantes de manera automatizada, constituye un gran avance en cuanto a las facilidades brindadas a los usuarios en cualquier tipo de programa computacional, ya sea experimental o comercial. Es por esto, que en el presente proyecto se buscó realizar dicha identificación, en base a la estructuración de los nombres de archivos con una nomenclatura lógica que describa su contenido. Al haberlo conseguido, este proceso se constituyó en un paso indispensable para facilitar la extracción de información, y poder así optimizar tareas que no debían ser realizadas manualmente por el usuario.
- En la etapa de entrenamiento de cada ANN se obtuvieron los mejores resultados referentes al aprendizaje adquirido por cada red. Esto significó que las posibilidades de lograr una correcta identificación aumenten en base a las curvas de desempeño obtenidas. Sin embargo, los resultados finales mostraron que las redes son muy dependientes de las muestras de entrenamiento que se les asignaron. Esto se debió principalmente al fenómeno de “overfitting” que se presentó en la etapa de entrenamiento, ya que el número de épocas (50) no fue elegido correctamente, y en lugar de brindar beneficios a la identificación, fue perjudicial, ya que hizo que los vectores de verificación y prueba tengan una separación muy grande con respecto a los valores de error del entrenamiento (ver figura 57).
Luego de hacer un análisis correspondiente a distintos factores, como: topología de cada red, conexiones, cantidad de neuronas, entrenamiento realizado, entre otros; se pudo concluir que la dependencia de las redes hacia los valores de entrada, también radica en el tipo de entrenamiento realizado, ya que por algunas complicaciones de tiempo y plazos establecidos, no fue posible proveer al software de un mayor número de muestras correspondientes a distintas personas, para un mejor aprendizaje de las tendencias acústicas de cada fonema.

- La utilización de la frecuencia de los formantes como el único parámetro acústico dentro del proceso de identificación, conllevó al reconocimiento de varios patrones de comportamiento basados en él; ya que en base a la frecuencia se pudieron analizar patrones como: comportamiento de las curvas de frecuencia de formantes a lo largo del tiempo, transiciones consonante-vocal, frecuencia del inicio de las transiciones, etc. A pesar de esto, se cree que la identificación correcta de sílabas habladas por distintas personas hubiera presentado mejores resultados en base a la inclusión de otro parámetro acústico, como lo es la amplitud.

Aunque la amplitud de los formantes estuvo inmersa en ciertos casos de análisis frecuencial debido a las razones ya expuestas en los capítulos dedicados, su inclusión como un parámetro independiente hubiera reflejado una mejor caracterización del fonema y por lo tanto derivado en la obtención de un mejor porcentaje de reconocimiento.

- Como se mencionó en el capítulo correspondiente, el porcentaje de reconocimiento en la vocal fue menor que en la consonante.

Si bien la teoría establece lo contrario, ya que la complejidad de reconocimiento de consonantes en términos generales es mayor; existe una razón muy determinante para que exista este acontecimiento. Esta razón consistió en la reducción de muestras de formantes consideradas para el entrenamiento, debido a las complejidades presentadas en la programación de MATLAB. Esta decisión radicó en la cantidad de elementos del vector de entrada permitida por el software, la cual debía ser igual en todos los casos. Es por esto que se decidió establecer el número de elementos en sesenta, debido a que dicho valor comprende la menor cantidad de muestras de formantes presentes en un archivo.

Debido a todo lo anterior, se concluye que la cantidad de muestras de formantes es determinante para obtener un mejor reconocimiento de un fonema, ya que mientras mayor sea la cantidad, mejores resultados se pueden obtener. Esto se basa en la información que aporta a un fonema

cada formante a lo largo del tiempo, debido a las configuraciones distintas que adopta el aparato fonatorio.

- La decisión de implementar un conjunto de seis Redes Neuronales Artificiales funcionando de manera simultánea, en lugar de una única red con todas las ochenta y cinco sílabas posibles; se basó en la necesidad de obtener un mejor funcionamiento y desempeño general en la identificación de todas las consonantes.

A pesar de esto, los resultados generales de todo este conjunto no fueron los esperados, debido entre otros factores, al fenómeno de “overfitting”. Es por esto que, a pesar de conocer esta causa de mal desempeño, no fue posible establecer las mejoras correspondientes, debido a que luego de realizar el análisis dedicado de las razones de este comportamiento, no se lograron identificar dichas conclusiones a tiempo para poder implementarlas.

- En términos generales hubieron algunas consideraciones que no pudieron ser optimizadas en la implementación de las ANN. Debido a limitaciones de tiempo no fue posible, por ejemplo, establecer el número óptimo de neuronas presentes en cada capa de las ANN. En base a este hecho se hubiera podido obtener una mejora en los resultados de desempeño así como optimizar recursos del computador.

La obtención del número óptimo de neuronas es un proceso que consiste en la variación de dicho número en base a los resultados y desempeño de una Red Neuronal Artificial. Dicha variación se realiza de acuerdo a cada red y no puede ser generalizada para todos los casos debido a que la funcionalidad de una red es distinta en base a los objetivos para los que fue diseñada.

Otra consideración que no pudo ser optimizada fue el tipo de algoritmo de ANN. Esto se basa en que en la actualidad existen varios tipos de Redes Neuronales Artificiales, los cuales presentan mejores resultados para ciertos fenómenos en lugar de otros; por lo que seguramente con mayor tiempo de

investigación se hubiera podido establecer el mejor tipo de algoritmo de ANN para obtener los mejores resultados de reconocimiento de fonemas.

- En la actualidad existen varios experimentos realizados con Redes Neuronales Artificiales acerca de reconocimiento de patrones del habla continua y segmentada. Es necesario mencionar que la mayoría de estos experimentos presentan mejores resultados cuando se usan métodos estadísticos de análisis de información. Una de las principales conclusiones del presente proyecto ciertamente consiste en que a diferencia de los experimentos mencionados, en este se trató de obtener resultados satisfactorios en base a un análisis acústico de cada fonema, basándose únicamente en los formantes y su comportamiento. Al ser un proyecto enfocado en una carrera acústica, se decidió no incluir métodos estadísticos para poder enfocar al habla como lo que en realidad es, un fenómeno netamente sonoro.

Sin embargo, es completamente necesario mencionar que al estar involucrados un gran número de factores que no corresponden a un concepto acústico, sino más bien estadístico debido a la inclusión de probabilidades; el enfoque acústico debe ser complementado con la implementación de métodos probabilísticos que modelen comportamientos inestables en términos de tiempo (variaciones en la velocidad del habla) y frecuencia (diferencias de entonación).

Adicionalmente a todas las conclusiones mencionadas anteriormente, se cree necesario también mencionar algunas recomendaciones acerca de ciertas observaciones realizadas a lo largo del proyecto. Estas recomendaciones, se basan en la posibilidad de desarrollar futuros programas computacionales acerca del reconocimiento del habla, tomando como punto de partida el presente proyecto:

- En los casos de habla continua y normal (no considerados en el presente proyecto), la complejidad de realizar un análisis a fondo y una identificación acertada, aumenta en gran manera. La coarticulación entre fonemas no solo

hace que las características acústicas dependan del sonido hablado y las posibles combinaciones de letras, sílabas y palabras; sino que aparecen nuevas variables en este fenómeno que aumentan la complejidad en base a factores determinantes como particularidades propias de cada hablante, variaciones en la velocidad del habla, silencios entre frases, diferencias en la entonación, etc.

En base a lo anterior se puede establecer que en las condiciones mencionadas, la información de formantes no es suficiente para que se produzca una identificación adecuada. En estas condiciones, el conocimiento de las características de cada fonema así como de su información y comportamiento de formantes, ayuda a tener ideas más claras de los fenómenos producidos por la coarticulación, para poder así plantear nuevos métodos de reconocimiento.

Los métodos estadísticos basados en el estudio de probabilidades, son muy empleados en la actualidad en el desarrollo de aplicaciones con este fin. Uno de ellos corresponde a la implementación de los denominados “Modelos Ocultos de Markov” (HMM por sus siglas en inglés).

Un futuro desarrollo del software planteado podría considerar casos de habla continua como los mencionados anteriormente. En estos se deberían considerar las variables especificadas y para ello se podría implementar modelos estadísticos basados en los HMM, además de un complemento necesario de Redes Neuronales Artificiales, por su capacidad de adaptarse a las condiciones naturales de un fenómeno específico.

- A pesar de conocer que hubiera sido beneficioso utilizar otro parámetro de identificación a parte de la frecuencia, es necesario mencionar que no fue posible incluir a la amplitud como un parámetro acústico independiente en el entrenamiento de las Redes Neuronales Artificiales y en la posterior etapa de reconocimiento. Esto se debió a que en la actualidad existe una complejidad en las herramientas de cálculo de formantes, para establecer un valor específico de su amplitud. Es por esto que aunque se intentó en varias oportunidades y con distintos métodos poder establecer un valor

tangible a la representación energética visible en los espectrogramas presentados en este proyecto, no fue posible obtener dicho fin por lo que se optó por no incluirlo en el desarrollo de este trabajo, aunque no debe descartarse poder incluirlo (junto con otros factores más) en posibles futuros desarrollos del proyecto.

- Teniendo el conocimiento de las razones del “sobre-entrenamiento” de las redes neuronales, sería necesario en un futuro establecer de manera óptima la cantidad de épocas para el proceso de entrenamiento. Esto conllevaría a poder tener mejores resultados de generalización de las tendencias de los formantes, y lograr así que el porcentaje de error se reduzca en comparación a los valores obtenidos en el presente proyecto, acerca de muestras de otras personas distintas a la que proporcionó los formantes de entrenamiento de las RNA.

Referencias

Basogain, X. (2000). *Redes neuronales artificiales y sus aplicaciones*, Escuela Superior de Ingeniería de Bilbao, España. Recuperado el 25 de Junio de 2012 de <http://ocw.ehu.es/enseñanzas-tecnicas/redes-neuronales-artificiales-y-sus-aplicaciones/contenidos/pdf/libro-del-curso>, pp.1-34.

Bobadilla, J. Gómez, P. y Bernal, J. (1999). *Posición y evolución de los formantes del habla. Estado del arte*. Recuperado el 25 de Junio de 2012 de <http://www.raco.cat/index.php/EFE/article/viewFile/144487/256867>, pp. 14-33.

Boersma, P. y Weenink, D. (2012). *PRAAT's manual*. Institute of Phonetics Sciences of the University of Amsterdam, Netherlands.

Childers, D. (1978). *Modern spectrum analysis*, IEEE Press. Recuperado el 25 de Junio de 2012 de http://www.mpi-hd.mpg.de/astrophysik/HEA/internal/Numerical_Recipes/f13-7.pdf, pp. 265-269.

Demuth, H. Beale, M. Hagan, M. (2010). *Neural Network Toolbox. User's Guide*. The MathWorks, Inc. Natick, Massachusetts, U.S.A.

Fernández, S. y Feijóo, S. (2004). *Modelos acústicos de sílabas consonante-vocal para el reconocimiento de fricativas*, Acústica 2004, Guimarães, Portugal, pp. 1-6.

Frías, X. (2001). *Introducción a la fonética y fonología del español*. Recuperado el 21 de Enero de 2012 de <http://www.romaniaminor.net/ianua/sup/sup04.pdf>, pp. 3-23.

Guzmán, M. (2010). *Acústica del tracto vocal*. Recuperado el 21 de Enero de 2012 de <http://lavoz.unsl.edu.ar/users/users/ACUSTICA%20EL%20TRACTO%20VOCAL.pdf>, pp. 1-3.

Holmes, J. y Holmes, W. (2001). *Speech synthesis and recognition, second edition*. Taylor & Francis, pp. 1-57, 159-218.

Little, J. y Moler, C. (2009), *MATLAB User's Guide*. The MathWorks, Inc. Natick, Massachusetts, U.S.A.

Massone, M. (1988). *Estudio acústico y perceptivo de las consonantes nasales y líquidas del español*. Facultad de Filología, Universidad de Barcelona, España, pp. 15-30.

Miyara, F. (2004). *La voz humana*. Recuperado el 10 de Enero de 2012 de <http://www.fceia.unr.edu.ar/acustica/biblio/fonatori.pdf>, pp. 1-10.

Quilis, A. (2000). *Fonética Acústica de la lengua Española, primera edición*. Gredos. pp. 1-70.

Rosero, J. (2009). *Oralidad y competencia comunicativa*. Recuperado el 12 de Mayo de 2012 de <http://es.scribd.com/doc/42043039/La-Oralidad-y-cia-Comunicativa>, pp. 102-113.

Tebelskis, J. (1995). *Speech recognition using neural networks*. School of Computer Science. Carnegie Mellon University, Pennsylvania, U.S.A., pp. 1-56.

Zañartu, M. (2003). Aplicaciones del análisis acústico en los estudios de la voz humana. *Seminario Internacional de Acústica 2003*, Chile, Santiago, pp. 1-9.

ANEXOS

Anexo 1.- Funciones y comandos de MATLAB R2009a usados en la etapa de programación

- **abs:** determina el valor absoluto de un número x en el caso de ser real, y la magnitud compleja si es imaginario.

abs (X)

- **cell (n):** crea una matriz de celdas vacías de “n” filas por “n” columnas.

c=cell(n)

- **char(N):** convierte al arreglo de matriz N, que contiene números enteros, en una matriz con caracteres de MATLAB.

S=char(N)

- **clc:** borra todos los datos de la ventana “Command Window” de MATLAB.

clc;

- **clear:** libera memoria borrando los datos de las variables de base y de funciones.

clear

- **close all:** borra todos los datos de los handles que estén ocultos y de los que no lo están. Los handles son valores de MATLAB que llaman a una función de manera indirecta.

close all

- **disp:** muestra una determinada matriz o texto.

disp (X)

- **divideind:** separa a los vectores de la red en tres puntos específicos que son: entrenamiento, validación y prueba, los cuales se los denota como trainInd, valInd y testInd.

[trainP, valP, testV] = divideind(p, trainInd, valInd, testInd)

- **elseif:** ejecuta una sentencia adicional si es que la sentencia de if es verdadera.

elseif X == n

end

- **figure:** crea una ventana con el gráfico del objeto.

figure(h)

- **floor:** redondea un valor o conjunto de valores al número entero menor en caso de ser positivo, y hacia el número entero mayor en caso de ser negativo.

floor(n)

- **fopen(filename):** abre archivos especificados en su argumentación y lee la información que posee.

fid=fopen(filename)

- **for:** orden de repetición hasta un límite específico.

for variable = initval:endval

statement

...

statement

end

- **fscanf:** comando que lee la información de un archivo, y convierte su formato al que se especifica en su argumento.

A = fscanf(fid, format)

- **hold off:** reinicia las propiedades de los ejes de los gráficos a su valor por defecto antes de que se creen nuevos gráficos.

hold off

- **hold on:** conserva el esquema del gráfico actual con las características de los ejes, para que los comandos de una nueva gráfica se añadan a la anterior.

hold on

- **if:** es una sentencia a ejecutar.

if expression

statements

end

- **load:** carga las variables que se encuentran grabadas en disco duro.

load filename

- **max:** determina el número máximo de entre los elementos.

C=max(A)

- **mean**: saca el promedio de los elementos.

$$M = \text{mean}(A)$$

- **newff**: crea una Red Neuronal Artificial del tipo “Feed Forward Back Propagation” de MATLAB.

$$\text{net} = \text{newff}(P, T, [S1 S2 \dots])$$

- **objArray = javaArray(PackageName.ClassName,x1,...,xn)**: es un arreglo de Java que elabora una matriz vacía para los objetos de la clase Java definida. Los argumentos de entrada pueden ser el String, que detalla la clase Java o las dimensiones del arreglo, y como argumento de salida se obtiene el arreglo de Java de una determinada dimensión.

- **pack**: libera espacio para que se pueda re-organizar los datos, utilizando la mínima memoria.

$$\text{pack}(\text{'filename'})$$

- **plot(x,y)**: dibuja el eje “x” versus el eje “y”.

$$\text{plot}(X1, Y1, \dots)$$

- **save**: guarda variables en disco duro.

$$\text{save filename}$$

- **sim(net,P)**: comando que permite la simulación de la Red Neuronal Artificial, el cual incorpora en sus argumentos a la red y al vector de entrada P.

$$Y = \text{sim}(\text{net}, P)$$

- **size(n)**: mide el tamaño del archivo.

$$d = \text{size}(X)$$

- **sprintf**: transforma los datos en formato String.

$$[s, \text{errmsg}] = \text{sprintf}(\text{format}, A, \dots)$$

- **String**: es un vector compuesto por componentes, los cuales son códigos numéricos que forman los caracteres, los primeros 127 códigos son ASCII.

- **xlim ([xmin xmax])**: determinan los límites, mínimo y máximo, del eje “x”.

$$\text{xlim}([x_{\min} x_{\max}])$$

- ***ylim* ([*ymin ymax*])**: determinan los límites, mínimo y máximo, del eje “y”.

ylim([*ymin ymax*])

- ***zeros*(*n,m*)**: crea una matriz de ceros de “n” filas por “m” columnas”.

B = *zeros*(*n*)

- ***A*=[*x,y*]**: matriz de “x” filas por “y” columnas.
- ***x*=1:*n***; vector que va desde uno hasta “n”.
- **<=**: condición de menor igual para establecer una comparación generalmente en un condicional *if* o *elseif*.
- **>=**: condición de mayor igual para establecer una comparación generalmente en un condicional *if* o *elseif*.
- **==**: condición de igual para establecer una comparación.

Anexo 2.- Descripción de la programación realizada

- **Programación de la identificación de información, nomenclatura y extracción de archivos**

La programación descrita corresponde a la RNA específica de las consonantes “F”, “S”, “J”. Este procedimiento se repite para las demás redes neuronales.

```
close all;
clear all;
pack;
clc;
nop = 1;
strArray = java_array('java.lang.String', nop);
strArray = {'X'};
opciones = cell(strArray);
clear strArray;
ncons = 3;
strArray = java_array('java.lang.String', ncons);
strArray = {'F','S','J'};
consonants = cell(strArray);
clear strArray;
nvoca = 5;
strArray = java_array('java.lang.String', nvoca);
strArray = {'A', 'E', 'I', 'O', 'U'};
vocals = cell(strArray);
clear strArray;
npers = 1;
strArray = java_array('java.lang.String', npers);
strArray = {'p'};
personas = cell(strArray);
clear strArray;
nvelo = 1;
strArray = java_array('java.lang.String', nvelo);
strArray = {'n'};
```

```

velocidades = cell(strArray);
clear strArray;
narchivos = nop*ncons*nvoca*nvelo*npers;
strArray = java_array('java.lang.String', narchivos);
cont = 1;
for e = 1:nop
for a = 1:ncons
for b = 1:nvoca
for c = 1:nvelo
for d = 1:npers
strArray(cont)=java.lang.String(sprintf('SC%s%s%s%s%s.txt',char(cell2mat(opci
ones(e))),char(cell2mat(consonantes(a))),char(cell2mat(vocales(b))),char(cell2
mat(velocidades(c))),char(cell2mat(personas(d))))));
cont = cont+1;
end
end
end
end
end
filename = cell(strArray);
clear strArray;
clear strArray;
clear cont;
clear a;
clear b;
clear c;
clear d;
narchivoslab = ncons+nvoca;
strArray = java_array('java.lang.String', narchivoslab);
cont = 1;
for a = 1:ncons

```

```
strArray(cont)=java.lang.String(sprintf('CC%s.txt',char(cell2mat(consonantes(a))
)));
cont = cont+1;
end
for b = 1:nvoca
strArray(cont)=java.lang.String(sprintf('CV%s.txt',char(cell2mat(vocales(b)))));
cont = cont+1;
end
filename1 = cell(strArray);
clear strArray;
clear a;
clear b;
clear cont;
clear ncons;
clear nvoca;
clear npers;
clear nvelo;
clear consonantes;
clear vocales;
clear velocidades;
clear personas;
cont1 = 1;
cont2 = 1;
c = 1;
for a = 1:narchivos
fid = fopen(char(cell2mat(filename(a))), 'rt');
datemp = fscanf(fid, '%10g\n', [1, inf]);
st = fclose(fid);
tama(a) = size(datemp, 2);
mtam(a) = floor(tama(a)/2);

fid = fopen(char(cell2mat(filename1(c))), 'rt');
```



```

datemp1 = fscanf(fid,'%3g\t%3g\t%3g\n',[1,inf]);
st = fclose(fid);
fid = fopen(char(cell2mat(filename1(cont2+2))), 'rt');
datemp2 = fscanf(fid,'%3g\t%3g\t%3g\n',[1,inf]);
st = fclose(fid);
d = 40;
for b = 1:d
datempr = datemp+rand(size(datemp))*10;
P((a-1)*d+b,:) = [datempr(1:60) datempr(mtam(a)+1:mtam(a)+60)];
T((a-1)*d+b,:) = [datemp1(1,1) datemp1(1,2) datemp1(1,3) datemp2(1,1)
datemp2(1,2) datemp2(1,3)];
end
cont2 = cont2+1;
if cont2>5
c = c+1;
cont2 = 1;
end
end
itre = round(d*0.7);
iver = round(d*0.15);
ipru = round(d*0.15);
atre = 1:1:itre*narchivos;
aver = 1:1:iver*narchivos;
apru = 1:1:ipru*narchivos;
cont = 1;
ctre = 1;
cver = 1;
cpru = 1;
for k = 1:narchivos*d
if(cont<=itre)
atre(ctre) = k;
ctre = ctre+1;

```

```
end
if(cont>itre && cont<=(itre+ipru))
    apru(cpru) = k;
    cpru = cpru+1;
end
if(cont>(itre+ipru) && cont<=(itre+ipru+iver))
    aver(cver) = k;
    cver = cver+1;
end
cont = cont+1;
if cont>(itre+ipru+iver)
    cont = 1;
end
end
P = P';
T = T';
net = newff(P,T,[15],{'tansig','purelin'},'trainlm');
net.divideFcn = 'divideind';
net.divideparam.trainInd = atre;
net.divideparam.valInd = aver;
net.divideparam.testInd = apru;
Y = sim(net,P);
net.trainParam.show = 50;
net.trainParam.lr = 0.05;
net.trainParam.epochs = 1000;
net.trainParam.goal = 1e-9;
net.trainParam.max_fail = 100;
[net,tr] = train(net,P,T);
Y = sim(net,P);
save ('fricativasorda','net');
```

- **Programación de la reducción de datos erróneos de formantes**

```
function dc = desviacion(d)
    n=size(d,2);
    m = mean(d);
    ds = std(d);
    dife = zeros(size(d));
    for a = 1:n-1
        dife(a) = abs(d(a+1) - d(a));
    end
    mdif = mean(dife);
    dsdif = std(dife);
    for a = 2:n-1
        if dife(a) >= mdif+dsdif
            d(a) = (d(a-1)+d(a+1))/2;
        end
    end
    if dife(1) >= mdif+dsdif
        d(1) = (d(2)+d(3))/2;
    end
    if dife(n) >= mdif+dsdif
        d(n) = (d(n-1)+d(n-2))/2;
    end
    dc = d;
```

- **Programación del reconocimiento de la sílaba que se obtenga de las seis redes neuronales**

```
function nt = pruebafinal(n)
    networkname = 'nasaes';
    load (networkname,'net');
    tama = size(n,2);
    mtam = floor(tama/2);
```

```

Pn(1,:) = [n(1:60) n(mtam+1:mtam+60)];
Pn = Pn';
Yn = sim(net,Pn);
networkname = 'vibrante';
load (networkname,'net');
tama1 = size(n,2);
mtam1 = floor(tama1/2);
Pv(1,:) = [n(1:60) n(mtam1+1:mtam1+60)];
Pv = Pv';
Yv = sim(net,Pv);
networkname = 'latefrica';
load (networkname,'net');
tama2 = size(n,2);
mtam2 = floor(tama2/2);
Pl(1,:) = [n(1:60) n(mtam2+1:mtam2+60)];
Pl = Pl';
Yl = sim(net,Pl);
networkname = 'fricativasorda';
load (networkname,'net');
tama3 = size(n,2);
mtam3 = floor(tama3/2);
Pf(1,:) = [n(1:60) n(mtam3+1:mtam3+60)];
Pf = Pf';
Yf = sim(net,Pf);
networkname = 'oclusivasonora';
load (networkname,'net');
tama4 = size(n,2);
mtam4 = floor(tama4/2);
Po(1,:) = [n(1:60) n(mtam4+1:mtam4+60)];
Po = Po';
Yo = sim(net,Po);
networkname = 'oclusivasorda';

```

```
load (networkname,'net');
tama5 = size(n,2);
mtam5 = floor(tama5/2);
Pos(1,:) = [n(1:60) n(mtam5+1:mtam5+60)];
Pos = Pos';
Yos = sim(net,Pos);
clear tama;
clear tama1;
clear tama2;
clear tama3;
clear tama4;
clear tama5;
clear mtama
clear mtama1
clear mtama2
clear mtama3
clear mtama4
clear mtama5
Comp = abs(Yn-[0; 0; 0; 0.8; 0.6; 4.9]);
mMA = mean(Comp);
Comp1 = abs(Yn-[0; 0; 0; 0.8; 0.5; 4.6]);
mME = mean(Comp1);
Comp2 = abs(Yn-[0; 0; 0; 0.6; 0.5; 3.7]);
mMI = mean(Comp2);
Comp3 = abs(Yn-[0; 0; 0; 0.5; 0.3; 2.9]);
mMO = mean(Comp3);
Comp4 = abs(Yn-[0; 0; 0; 0.5; 0.2; 2.1]);
mMU = mean(Comp4);
Comp5 = abs(Yn-[0.5; 0.4; 3.5; 0.8; 0.6; 4.9]);
mNA = mean(Comp5);
Comp6 = abs(Yn-[0.5; 0.4; 3.5; 0.8; 0.5; 4.6]);
mNE = mean(Comp6);
```

Comp7 = abs(Yn-[0.5; 0.4; 3.5; 0.6; 0.5; 3.7]);
mNI = mean(Comp7);
Comp8 = abs(Yn-[0.5; 0.4; 3.5; 0.5; 0.3; 2.9]);
mNO = mean(Comp8);
Comp9 = abs(Yn-[0.5; 0.4; 3.5; 0.5; 0.2; 2.1]);
mNU = mean(Comp9);
Comp10 = abs(Yn-[0.5; 0.4; 3.5; 0.8; 0.6; 4.9]);
mNIA = mean(Comp10);
Comp11 = abs(Yn-[0.5; 0.4; 3.5; 0.8; 0.5; 4.6]);
mNIE = mean(Comp11);
Comp12 = abs(Yn-[0.5; 0.4; 3.5; 0.6; 0.5; 3.7]);
mNII = mean(Comp12);
Comp13 = abs(Yn-[0.5; 0.4; 3.5; 0.5; 0.3; 2.9]);
mNIO = mean(Comp13);
Comp14 = abs(Yn-[0.5; 0.4; 3.5; 0.5; 0.2; 2.1]);
mNIU = mean(Comp14);
Comp15 = abs(Yv-[0.9; 0.6; 4; 0.8; 0.6; 4.9]);
mRA = mean(Comp15);
Comp16 = abs(Yv-[0.9; 0.6; 4; 0.8; 0.5; 4.6]);
mRE = mean(Comp16);
Comp17 = abs(Yv-[0.9; 0.6; 4; 0.6; 0.5; 3.7]);
mRI = mean(Comp17);
Comp18 = abs(Yv-[0.9; 0.6; 4; 0.5; 0.3; 2.9]);
mRO = mean(Comp18);
Comp19 = abs(Yv-[0.9; 0.6; 4; 0.5; 0.2; 2.1]);
mRU = mean(Comp19);
Comp20 = abs(Yv-[0.9; 0.8; 4.7; 0.8; 0.6; 4.9]);
mRRA = mean(Comp20);
Comp21 = abs(Yv-[0.9; 0.8; 4.7; 0.8; 0.5; 4.6]);
mRRE = mean(Comp21);
Comp22 = abs(Yv-[0.9; 0.8; 4.7; 0.6; 0.5; 3.7]);
mRRI = mean(Comp22);

Comp23 = abs(Yv-[0.9; 0.8; 4.7; 0.5; 0.3; 2.9]);
mRRO = mean(Comp23);
Comp24 = abs(Yv-[0.9; 0.8; 4.7; 0.5; 0.2; 2.1]);
mRRU = mean(Comp24);
Comp25 = abs(YI-[0.9; 0.8; 4.5; 0.8; 0.6; 4.9]);
mLA = mean(Comp25);
Comp26 = abs(YI-[0.9; 0.8; 4.5; 0.8; 0.5; 4.6]);
mLE = mean(Comp26);
Comp27 = abs(YI-[0.9; 0.8; 4.5; 0.6; 0.5; 3.7]);
mLI = mean(Comp27);
Comp28 = abs(YI-[0.9; 0.8; 4.5; 0.5; 0.3; 2.9]);
mLO = mean(Comp28);
Comp29 = abs(YI-[0.9; 0.8; 4.5; 0.5; 0.2; 2.1]);
mLU = mean(Comp29);
Comp30 = abs(YI-[0.5; 0.4; 3; 0.8; 0.6; 4.9]);
mYA = mean(Comp30);
Comp31 = abs(YI-[0.5; 0.4; 3; 0.8; 0.5; 4.6]);
mYE = mean(Comp31);
Comp32 = abs(YI-[0.5; 0.4; 3; 0.6; 0.5; 3.7]);
mYI = mean(Comp32);
Comp33 = abs(YI-[0.5; 0.4; 3; 0.5; 0.3; 2.9]);
mYO = mean(Comp33);
Comp34 = abs(YI-[0.5; 0.4; 3; 0.5; 0.2; 2.1]);
my = mean(Comp34);
Comp35 = abs(YI-[0.6; 0.5; 4.5; 0.8; 0.6; 4.9]);
mCHA = mean(Comp35);
Comp36 = abs(YI-[0.6; 0.5; 4.5; 0.8; 0.5; 4.6]);
mCHE = mean(Comp36);
Comp37 = abs(YI-[0.6; 0.5; 4.5; 0.6; 0.5; 3.7]);
mCHI = mean(Comp37);
Comp38 = abs(YI-[0.6; 0.5; 4.5; 0.5; 0.3; 2.9]);
mCHO = mean(Comp38);

```
Comp39 = abs(Yl-[0.6; 0.5; 4.5; 0.5; 0.2; 2.1]);  
mCHU = mean(Comp39);  
Comp40 = abs(Yf-[0.4; 0.3; 2; 0.8; 0.6; 4.9]);  
mFA = mean(Comp40);  
Comp41 = abs(Yf-[0.4; 0.3; 2; 0.8; 0.5; 4.6]);  
mFE = mean(Comp41);  
Comp42 = abs(Yf-[0.4; 0.3; 2; 0.6; 0.5; 3.7]);  
mFI = mean(Comp42);  
Comp43 = abs(Yf-[0.4; 0.3; 2; 0.5; 0.3; 2.9]);  
mFO = mean(Comp43);  
Comp44 = abs(Yf-[0.4; 0.3; 2; 0.5; 0.2; 2.1]);  
mFU = mean(Comp44);  
Comp45 = abs(Yf-[0.5; 0.3; 3.5; 0.8; 0.6; 4.9]);  
mSA = mean(Comp45);  
Comp46 = abs(Yf-[0.5; 0.3; 3.5; 0.8; 0.5; 4.6]);  
mSE = mean(Comp46);  
Comp47 = abs(Yf-[0.5; 0.3; 3.5; 0.6; 0.5; 3.7]);  
mSI = mean(Comp47);  
Comp48 = abs(Yf-[0.5; 0.3; 3.5; 0.5; 0.3; 2.9]);  
mSO = mean(Comp48);  
Comp49 = abs(Yf-[0.5; 0.3; 3.5; 0.5; 0.2; 2.1]);  
mSU = mean(Comp49);  
Comp50 = abs(Yf-[0.6; 0.5; 4; 0.8; 0.6; 4.9]);  
mJA = mean(Comp50);  
Comp51 = abs(Yf-[0.6; 0.5; 4; 0.8; 0.5; 4.6]);  
mJE = mean(Comp51);  
Comp52 = abs(Yf-[0.6; 0.5; 4; 0.6; 0.5; 3.7]);  
mJI = mean(Comp52);  
Comp53 = abs(Yf-[0.6; 0.5; 4; 0.5; 0.3; 2.9]);  
mJO = mean(Comp53);  
Comp54 = abs(Yf-[0.6; 0.5; 4; 0.5; 0.2; 2.1]);  
mJU = mean(Comp54);
```


Comp55 = abs(Yo-[0; 0; 0; 0.8; 0.6; 4.9]);
mBA = mean(Comp55);
Comp56 = abs(Yo-[0; 0; 0; 0.8; 0.5; 4.6]);
mBE = mean(Comp56);
Comp57 = abs(Yo-[0; 0; 0; 0.6; 0.5; 3.7]);
mBI = mean(Comp57);
Comp58 = abs(Yo-[0; 0; 0; 0.5; 0.3; 2.9]);
mBO = mean(Comp58);
Comp59 = abs(Yo-[0; 0; 0; 0.5; 0.2; 2.1]);
mBU = mean(Comp59);
Comp60 = abs(Yo-[0.6; 0.4; 4.4; 0.8; 0.6; 4.9]);
mDA = mean(Comp60);
Comp61 = abs(Yo-[0.6; 0.4; 4.4; 0.8; 0.5; 4.6]);
mDE = mean(Comp61);
Comp62 = abs(Yo-[0.6; 0.4; 4.4; 0.6; 0.5; 3.7]);
mDI = mean(Comp62);
Comp63 = abs(Yo-[0.6; 0.4; 4.4; 0.5; 0.3; 2.9]);
mDO = mean(Comp63);
Comp64 = abs(Yo-[0.6; 0.4; 4.4; 0.5; 0.2; 2.1]);
mDU = mean(Comp64);
Comp65 = abs(Yo-[0.8; 0.6; 4.5; 0.8; 0.6; 4.9]);
mGA = mean(Comp65);
Comp66 = abs(Yo-[0.8; 0.6; 4.5; 0.8; 0.5; 4.6]);
mGE = mean(Comp66);
Comp67 = abs(Yo-[0.8; 0.6; 4.5; 0.6; 0.5; 3.7]);
mGI = mean(Comp67);
Comp68 = abs(Yo-[0.8; 0.6; 4.5; 0.5; 0.3; 2.9]);
mGO = mean(Comp68);
Comp69 = abs(Yo-[0.8; 0.6; 4.5; 0.5; 0.2; 2.1]);
mGU = mean(Comp69);
Comp70 = abs(Yos-[0; 0; 0; 0.8; 0.6; 4.9]);
mPA = mean(Comp70);

```

Comp71 = abs(Yos-[0; 0; 0; 0.8; 0.5; 4.6]);
mPE = mean(Comp71);
Comp72 = abs(Yos-[0; 0; 0; 0.6; 0.5; 3.7]);
mPI = mean(Comp72);
Comp73 = abs(Yos-[0; 0; 0; 0.5; 0.3; 2.9]);
mPO = mean(Comp73);
Comp74 = abs(Yos-[0; 0; 0; 0.5; 0.2; 2.1]);
mPU = mean(Comp74);
Comp75 = abs(Yos-[0.7; 0.5; 4; 0.8; 0.6; 4.9]);
mTA = mean(Comp75);
Comp76 = abs(Yos-[0.7; 0.5; 4; 0.8; 0.5; 4.6]);
mTE = mean(Comp76);
Comp77 = abs(Yos-[0.7; 0.5; 4; 0.6; 0.5; 3.7]);
mTI = mean(Comp77);
Comp78 = abs(Yos-[0.7; 0.5; 4; 0.5; 0.3; 2.9]);
mTO = mean(Comp78);
Comp79 = abs(Yos-[0.7; 0.5; 4; 0.5; 0.2; 2.1]);
mTU = mean(Comp79);
Comp80 = abs(Yos-[0.6; 0.5; 4.5; 0.8; 0.6; 4.9]);
mKA = mean(Comp80);
Comp81 = abs(Yos-[0.6; 0.5; 4.5; 0.8; 0.5; 4.6]);
mKE = mean(Comp81);
Comp82 = abs(Yos-[0.6; 0.5; 4.5; 0.6; 0.5; 3.7]);
mKI = mean(Comp82);
Comp83 = abs(Yos-[0.6; 0.5; 4.5; 0.5; 0.3; 2.9]);
mKO = mean(Comp83);
Comp84 = abs(Yos-[0.6; 0.5; 4.5; 0.5; 0.2; 2.1]);
mKU = mean(Comp84);
if Comp <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp('MA')
elseif Comp1 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp('ME')

```

```
elseif Comp2 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp('MI')
elseif Comp3 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp('MO')
elseif Comp4 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('MU')
elseif Comp5 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('NA')
elseif Comp6 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('NE')
elseif Comp7 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('NI')
elseif Comp8 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('NO')
elseif Comp9 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('NU')
elseif Comp10 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('ñA')
elseif Comp11 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('ñE')
elseif Comp12 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('ñI')
elseif Comp13 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('ñO')
elseif Comp14 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('ñU')
end
if Comp15 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('RA')
elseif Comp16 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('RE')
elseif Comp17 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
```

```
disp ('RI')
elseif Comp18 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('RO')
elseif Comp19 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('RU')
elseif Comp20 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('RRA')
elseif Comp21 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('RRE')
elseif Comp22 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('RRI')
elseif Comp23 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('RRO')
elseif Comp24 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('RRU')
end
if Comp25 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('LA')
elseif Comp26 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('LE')
elseif Comp27 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('LI')
elseif Comp28 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('LO')
elseif Comp29 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('LU')
elseif Comp30 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('YA')
elseif Comp31 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('YE')
elseif Comp32 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('YI')
```

```
elseif Comp33 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('YO')
elseif Comp34 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('YU')
elseif Comp35 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('CHA')
elseif Comp36 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('CHE')
elseif Comp37 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('CHI')
elseif Comp38 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('CHO')
elseif Comp39 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('CHU')
end
if Comp40 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('FA')
elseif Comp41 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('FE')
elseif Comp42 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('FI')
elseif Comp43 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('FO')
elseif Comp44 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('FU')
elseif Comp45 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('SA')
elseif Comp46 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('SE')
elseif Comp47 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('SI')
elseif Comp48 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
```

```
disp ('SO')
elseif Comp49 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('SU')
elseif Comp50 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('JA')
elseif Comp51 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('JE')
elseif Comp52 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('JI')
elseif Comp53 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('JO')
elseif Comp54 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('JU')
end
if Comp55 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('BA')
elseif Comp56 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('BE')
elseif Comp57 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('BI')
elseif Comp58 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('BO')
elseif Comp59 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('BU')
elseif Comp60 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('DA')
elseif Comp61 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('DE')
elseif Comp62 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('DI')
elseif Comp63 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('DO')
```

```
elseif Comp64 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('DU')
elseif Comp65 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('GA')
elseif Comp66 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('GE')
elseif Comp67 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('GI')
elseif Comp68 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('GO')
elseif Comp69 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('GU')
end
if Comp70 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('PA')
elseif Comp71 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('PE')
elseif Comp72 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('PI')
elseif Comp73 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('PO')
elseif Comp74 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('PU')
elseif Comp75 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('TA')
elseif Comp76 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('TE')
elseif Comp77 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('TI')
elseif Comp78 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('TO')
elseif Comp79 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
```

```

disp ('TU')
elseif Comp80 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('KA')
elseif Comp81 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('KE')
elseif Comp82 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('KI')
elseif Comp83 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('KO')
elseif Comp84 <= [0.001; 0.001; 0.001; 0.001; 0.001; 0.001]
disp ('KU')
end
letras = 85;
strArray = java_array('java.lang.String', letras);
strArray = {'MA', 'ME', 'MI', 'MO', 'MU', 'NA', 'NE', 'NI', 'NO', 'NU', 'ÑA', 'ÑE', 'ÑI',
'ÑO', 'ÑU', 'RA', 'RE', 'RI', 'RO', 'RU', 'RRA', 'RRE', 'RRI', 'RRO', 'RRU', 'LA', 'LE',
'LI', 'LO', 'LU', 'YA', 'YE', 'YI', 'YO', 'YU', 'CHA', 'CHE', 'CHI', 'CHO', 'CHU', 'FA',
'FE', 'FI', 'FO', 'FU', 'SA', 'SE', 'SI', 'SO', 'SU', 'JA', 'JE', 'JI', 'JO', 'JU', 'BA', 'BE',
'BI', 'BO', 'BU', 'DA', 'DE', 'DI', 'DO', 'DU', 'GA', 'GE', 'GI', 'GO', 'GU', 'PA', 'PE',
'PI', 'PO', 'PU', 'TA', 'TE', 'TI', 'TO', 'TU', 'KA', 'KE', 'KI', 'KO', 'KU'};
silabas = cell(strArray);
compfinal = [mMA, mME, mMI, mMO, mMU, mNA, mNE, mNI, mNO, mNU,
mNIA, mNIE, mNII, mNIO, mNIU, mRA, mRE, mRI, mRO, mRU, mRRA, mRRE,
mRRI, mRRO, mRRU, mLA, mLE, mLI, mLO, mLU, mYA, mYE, mYI, mYO,
mYU, mCHA, mCHE, mCHI, mCHO, mCHU, mFA, mFE, mFI, mFO, mFU,
mSA, mSE, mSI, mSO, mSU, mJA, mJE, mJI, mJO, mJU, mBA, mBE, mBI,
mBO, mBU, mDA, mDE, mDI, mDO, mDU, mGA, mGE, mGI, mGO, mGU,
mPA, mPE, mPI, mPO, mPU, mTA, mTE, mTI, mTO, mTU, mKA, mKE, mKI,
mKO, mKU];
valorminglobal = min(compfinal);
for a = 1:85
if valorminglobal == compfinal(a);

```



```

vectfinal(a) = java.lang.String ((char(cell2mat(silabas(a)))));
nt = char(vectfinal);%Lared entrega la respuesta en formato string
end
end

```

- **Programación para la animación de las diferentes aberturas de la boca**

```

close all;
clear all;
clc;
fid = fopen('SCXRAnp.txt','rt');
n = fscanf(fid,'%10g\n',[1,inf]);
nt = pruebafinal(n);
st = fclose(fid);
letras = 85;
strArray = java_array('java.lang.String', letras);
strArray = {'MA', 'ME', 'MI', 'MO', 'MU', 'NA', 'NE', 'NI', 'NO', 'NU', 'ÑA', 'ÑE', 'ÑI',
'ÑO', 'ÑU', 'RA', 'RE', 'RI', 'RO', 'RU', 'RRA', 'RRE', 'RRI', 'RRO', 'RRU', 'LA', 'LE',
'LI', 'LO', 'LU', 'YA', 'YE', 'YI', 'YO', 'YU', 'CHA', 'CHE', 'CHI', 'CHO', 'CHU', 'FA',
'FE', 'FI', 'FO', 'FU', 'SA', 'SE', 'SI', 'SO', 'SU', 'JA', 'JE', 'JI', 'JO', 'JU', 'BA', 'BE',
'BI', 'BO', 'BU', 'DA', 'DE', 'DI', 'DO', 'DU', 'GA', 'GE', 'GI', 'GO', 'GU', 'PA', 'PE',
'PI', 'PO', 'PU', 'TA', 'TE', 'TI', 'TO', 'TU', 'KA', 'KE', 'KI', 'KO', 'KU'};
nsilabas = cell(strArray);
YV = size(nt)
YVV = max(YV)
resp = nt(YVV,1)
resp1 = nt(YVV,2)
if YV == [YVV 3]
resp2 = nt(YVV,3)
cons = [resp resp1 resp2]
else
cons = [resp resp1]
end
end

```

```
if YV == [YVV 2]
if cons == 'MA'
AE = [0 0 5 0.8 0.6 5];
elseif cons == 'ME'
AE = [0 0 5 0.8 0.5 5];
elseif cons == 'MI'
AE = [0 0 5 0.6 0.5 5];
elseif cons == 'MO'
AE = [0 0 5 0.5 0.3 5];
elseif cons == 'MU'
AE = [0 0 5 0.5 0.2 5];
elseif cons == 'NA'
AE = [0.5 0.4 5 0.8 0.6 5];
elseif cons == 'NE'
AE = [0.5 0.4 5 0.8 0.5 5];
elseif cons == 'NI'
AE = [0.5 0.4 5 0.6 0.5 5];
elseif cons == 'NO'
AE = [0.5 0.4 5 0.5 0.3 5];
elseif cons == 'NU'
AE = [0.5 0.4 5 0.5 0.2 5];
elseif cons == 'BA'
AE = [0 0 5 0.8 0.6 5];
elseif cons == 'BE'
AE = [0 0 5 0.8 0.5 5];
elseif cons == 'BI'
AE = [0 0 5 0.6 0.5 5];
elseif cons == 'BO'
AE = [0 0 5 0.5 0.3 5];
elseif cons == 'BU'
AE = [0 0 5 0.5 0.2 5];
elseif cons == 'DA'
```

```
AE = [0.6 0.4 5 0.8 0.6 5];  
elseif cons == 'DE'  
AE = [0.6 0.4 5 0.8 0.5 5];  
elseif cons == 'DI'  
AE = [0.6 0.4 5 0.6 0.5 5];  
elseif cons == 'DO'  
AE = [0.6 0.4 5 0.5 0.3 5];  
elseif cons == 'DU'  
AE = [0.6 0.4 5 0.5 0.2 5];  
elseif cons == 'FA'  
AE = [0.4 0.3 5 0.8 0.6 5];  
elseif cons == 'FE'  
AE = [0.4 0.3 5 0.8 0.5 5];  
elseif cons == 'FI'  
AE = [0.4 0.3 5 0.6 0.5 5];  
elseif cons == 'FO'  
AE = [0.4 0.3 5 0.5 0.3 5];  
elseif cons == 'FU'  
AE = [0.4 0.3 5 0.5 0.2 5];  
elseif cons == 'GA'  
AE = [0.8 0.6 5 0.8 0.6 5];  
elseif cons == 'GE'  
AE = [0.8 0.6 5 0.8 0.5 5];  
elseif cons == 'GI'  
AE = [0.8 0.6 5 0.6 0.5 5];  
elseif cons == 'GO'  
AE = [0.8 0.6 5 0.5 0.3 5];  
elseif cons == 'GU'  
AE = [0.8 0.6 5 0.5 0.2 5];  
elseif cons == 'JA'  
AE = [0.6 0.5 5 0.8 0.6 5];  
elseif cons == 'JE'
```

```
AE = [0.6 0.5 5 0.8 0.5 5];  
elseif cons == 'JI'  
AE = [0.6 0.5 5 0.6 0.5 5];  
elseif cons == 'JO'  
AE = [0.6 0.5 5 0.5 0.3 5];  
elseif cons == 'JU'  
AE = [0.6 0.5 5 0.5 0.2 5];  
elseif cons == 'KA'  
AE = [0.6 0.5 5 0.8 0.6 5];  
elseif cons == 'KE'  
AE = [0.6 0.5 5 0.8 0.5 5];  
elseif cons == 'KI'  
AE = [0.6 0.5 5 0.6 0.5 5];  
elseif cons == 'KO'  
AE = [0.6 0.5 5 0.5 0.3 5];  
elseif cons == 'KU'  
AE = [0.6 0.5 5 0.5 0.2 5];  
elseif cons == 'LA'  
AE = [0.9 0.8 5 0.8 0.6 5];  
elseif cons == 'LE'  
AE = [0.9 0.8 5 0.8 0.5 5];  
elseif cons == 'LI'  
AE = [0.9 0.8 5 0.6 0.5 5];  
elseif cons == 'LO'  
AE = [0.9 0.8 5 0.5 0.3 5];  
elseif cons == 'LU'  
AE = [0.9 0.8 5 0.5 0.2 5];  
elseif cons == 'PA'  
AE = [0 0 5 0.8 0.6 5];  
elseif cons == 'PE'  
AE = [0 0 5 0.8 0.5 5];  
elseif cons == 'PI'
```

```
AE = [0 0 5 0.6 0.5 5];  
elseif cons == 'PO'  
AE = [0 0 5 0.5 0.3 5];  
elseif cons == 'PU'  
AE = [0 0 5 0.5 0.2 5];  
elseif cons == 'RA'  
AE = [0.9 0.6 5 0.8 0.6 5];  
elseif cons == 'RE'  
AE = [0.9 0.6 5 0.8 0.5 5];  
elseif cons == 'RI'  
AE = [0.9 0.6 5 0.6 0.5 5];  
elseif cons == 'RO'  
AE = [0.9 0.6 5 0.5 0.3 5];  
elseif cons == 'RU'  
AE = [0.9 0.6 5 0.5 0.2 5];  
elseif cons == 'SA'  
AE = [0.5 0.3 5 0.8 0.6 5];  
elseif cons == 'SE'  
AE = [0.5 0.3 5 0.8 0.5 5];  
elseif cons == 'SI'  
AE = [0.5 0.3 5 0.6 0.5 5];  
elseif cons == 'SO'  
AE = [0.5 0.3 5 0.5 0.3 5];  
elseif cons == 'SU'  
AE = [0.5 0.3 5 0.5 0.2 5];  
elseif cons == 'TA'  
AE = [0.7 0.5 5 0.8 0.6 5];  
elseif cons == 'TE'  
AE = [0.7 0.5 5 0.8 0.5 5];  
elseif cons == 'TI'  
AE = [0.7 0.5 5 0.6 0.5 5];  
elseif cons == 'TO'
```

```
AE = [0.7 0.5 5 0.5 0.3 5];
elseif cons == 'TU'
AE = [0.7 0.5 5 0.5 0.2 5];
elseif cons == 'YA'
AE = [0.5 0.4 5 0.8 0.6 5];
elseif cons == 'YE'
AE = [0.5 0.4 5 0.8 0.5 5];
elseif cons == 'YI'
AE = [0.5 0.4 5 0.6 0.5 5];
elseif cons == 'YO'
AE = [0.5 0.4 5 0.5 0.3 5];
elseif cons == 'YU'
AE = [0.5 0.4 5 0.5 0.2 5];
elseif cons == 'ÑA'
AE = [0.5 0.4 5 0.8 0.6 5];
elseif cons == 'ÑE'
AE = [0.5 0.4 5 0.8 0.5 5];
elseif cons == 'ÑI'
AE = [0.5 0.4 5 0.6 0.5 5];
elseif cons == 'ÑO'
AE = [0.5 0.4 5 0.5 0.3 5];
elseif cons == 'ÑU'
AE = [0.5 0.4 5 0.5 0.2 5];
end
end
if YV == [YVV 3]
if cons == 'CHA'
AE = [0.6 0.5 5 0.8 0.6 5];
elseif cons == 'CHE'
AE = [0.6 0.5 5 0.8 0.5 5];
elseif cons == 'CHI'
AE = [0.6 0.5 5 0.6 0.5 5];
```

```

elseif cons == 'CHO'
AE = [0.6 0.5 5 0.5 0.3 5];
elseif cons == 'CHU'
AE = [0.6 0.5 5 0.5 0.2 5];
elseif cons == 'RRA'
AE = [0.9 0.8 5 0.8 0.6 5];
elseif cons == 'RRE'
AE = [0.9 0.8 5 0.8 0.5 5];
elseif cons == 'RRI'
AE = [0.9 0.8 5 0.6 0.5 5];
elseif cons == 'RRO'
AE = [0.9 0.8 5 0.5 0.3 5];
elseif cons == 'RRU'
AE = [0.9 0.8 5 0.5 0.2 5];
end
end
CC1 = AE(1,1);
CC2 = AE(1,2);
CC3 = AE(1,3);
CV1 = AE(1,4);
CV2 = AE(1,5);
CV3 = AE(1,6);
for n = 1:10
A(n) = (CV1/10)*n;
B(n) = (CV2/10)*n;
C(n) = (CC1/10)*n;
D(n) = (CC2/10)*n;
end
x = [0 CV3*0.25 CV3*0.5 CV3*0.75 CV3];
y = [0 0 0 0 0];
y1 = [0 D(1) C(1) D(1) 0];
y2 = [0 D(2) C(2) D(2) 0];

```

```

y3 = [0 D(3) C(3) D(3) 0];
y4 = [0 D(4) C(4) D(4) 0];
y5 = [0 D(5) C(5) D(5) 0];
y6 = [0 D(6) C(6) D(6) 0];
y7 = [0 D(7) C(7) D(7) 0];
y8 = [0 D(8) C(8) D(8) 0];
y9 = [0 D(9) C(9) D(9) 0];
y10 = [0 D(10) C(10) D(10) 0];
y11 = [0 -D(1) -C(1) -D(1) 0];
y12 = [0 -D(2) -C(2) -D(2) 0];
y13 = [0 -D(3) -C(3) -D(3) 0];
y14 = [0 -D(4) -C(4) -D(4) 0];
y15 = [0 -D(5) -C(5) -D(5) 0];
y16 = [0 -D(6) -C(6) -D(6) 0];
y17 = [0 -D(7) -C(7) -D(7) 0];
y18 = [0 -D(8) -C(8) -D(8) 0];
y19 = [0 -D(9) -C(9) -D(9) 0];
y20 = [0 -D(10) -C(10) -D(10) 0];
y25 = [0 B(5) A(5) B(5) 0];
y26 = [0 B(6) A(6) B(6) 0];
y27 = [0 B(7) A(7) B(7) 0];
y28 = [0 B(8) A(8) B(8) 0];
y29 = [0 B(9) A(9) B(9) 0];
y30 = [0 B(10) A(10) B(10) 0];
y36 = [0 -B(6) -A(6) -B(6) 0];
y37 = [0 -B(7) -A(7) -B(7) 0];
y38 = [0 -B(8) -A(8) -B(8) 0];
y39 = [0 -B(9) -A(9) -B(9) 0];
y40 = [0 -B(10) -A(10) -B(10) 0];
for u = 1:18
figure(u)
if u <= 3

```



```
plot(x,y)
elseif u==4
plot(x,y1,'-o')
hold on
plot(x,y11,'-o')
xlim([0 5])
ylim([-1 1])
elseif u==5
plot(x,y2,'-o')
hold on
plot(x,y12,'-o')
xlim([0 5])
ylim([-1 1])
elseif u==6
plot(x,y3,'-o')
hold on
plot(x,y13,'-o')
xlim([0 5])
ylim([-1 1])
elseif u==7
plot(x,y4,'-o')
hold on
plot(x,y14,'-o')
xlim([0 5])
ylim([-1 1])
elseif u==8
plot(x,y5,'-o')
hold on
plot(x,y15,'-o')
xlim([0 5])
ylim([-1 1])
elseif u==9
```

```
plot(x,y6,'-o')
hold on
plot(x,y16,'-o')
xlim([0 5])
ylim([-1 1])
elseif u==10
plot(x,y7,'-o')
hold on
plot(x,y17,'-o')
xlim([0 5])
ylim([-1 1])
elseif u==11
plot(x,y8,'-o')
hold on
plot(x,y18,'-o')
xlim([0 5])
ylim([-1 1])
elseif u==12
plot(x,y9,'-o')
hold on
plot(x,y19,'-o')
xlim([0 5])
ylim([-1 1])
elseif u==13
plot(x,y10,'-o')
hold on
plot(x,y20,'-o')
xlim([0 5])
ylim([-1 1])
hold off
elseif u==14
plot(x,y26,'-o')
```

```
hold on
plot(x,y36,'-o')
xlim([0 5])
ylim([-1 1])
hold off
elseif u==15
plot(x,y27,'-o')
hold on
plot(x,y37,'-o')
xlim([0 5])
ylim([-1 1])
hold off
elseif u==16
plot(x,y28,'-o')
hold on
plot(x,y38,'-o')
xlim([0 5])
ylim([-1 1])
hold off
elseif u==17
plot(x,y29,'-o')
hold on
plot(x,y39,'-o')
xlim([0 5])
ylim([-1 1])
hold off
elseif u==18
plot(x,y30,'-o')
hold on
plot(x,y40,'-o')
xlim([0 5])
ylim([-1 1])
```

hold off
end
end