

**ANÁLISIS DE LA RELACIÓN EXISTENTE ENTRE LA  
LONGITUD DEL TRACTO VOCAL, OBTENIDA A PARTIR  
DE IMÁGENES POR RESONANCIA MAGNÉTICA, Y  
PARÁMETROS ACÚSTICOS DE LA VOZ.**

JULIAN ORLANDO REYES MORENO  
PAULA ANDREA VASQUEZ SERRANO



UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS  
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE  
TELECOMUNICACIONES  
BUCARAMANGA

2018

ANÁLISIS DE LA RELACIÓN EXISTENTE ENTRE LA  
LONGITUD DEL TRACTO VOCAL, OBTENIDA A PARTIR  
DE IMÁGENES POR RESONANCIA MAGNÉTICA, Y  
PARÁMETROS ACÚSTICOS DE LA VOZ.

JULIAN ORLANDO REYES MORENO  
PAULA ANDREA VASQUEZ SERRANO

*Trabajo de grado para optar al título de  
Ingeniero Electrónico*

Director  
FRANKLIN ALEXANDER SEPÚLVEDA SEPÚLVEDA  
PhD Ingeniería Automática

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS  
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE  
TELECOMUNICACIONES  
BUCARAMANGA  
2018

## *Agradecimientos*

*A mi familia por brindarme su apoyo incondicional, especialmente a mis padres que con sus enseñanzas, amor y confianza me acompañaron por este camino y me dieron fuerzas para alcanzar esta meta.*

*A mi abuela, mi estrella en el cielo, por su eterno amor, por ser ejemplo de valentía, lucha y fortaleza, y por enseñarme que a pesar de las dificultades y obstáculos que haya en el camino, siempre se puede salir adelante.*

*Paula Andrea Vasquez Serrano*

## Índice general

<b>INTRODUCCIÓN</b>	<b>18</b>
<b>1 MÉTODO</b>	<b>20</b>
1.1 Producción de la voz y el tracto vocal . . . . .	20
1.2 Base de datos . . . . .	21
1.3 Conjunto de datos para la realización de experimentos . . . . .	22
1.4 Representación paramétrica de la señal de voz . . . . .	23
1.4.1 Pre-procesamiento . . . . .	23
1.4.2 Estimación de formantes con LPC . . . . .	24
1.4.3 MFCC . . . . .	26
1.5 Algoritmo de medición del valor de referencia longitud de tracto vocal . . . . .	28
1.6 Pre-procesamiento de datos . . . . .	32
1.6.1 Remoción de silencios . . . . .	32
1.6.2 Detección de valores atípicos . . . . .	33
1.7 Análisis de regresión . . . . .	34
1.7.1 Estimación de coeficientes . . . . .	35
1.7.2 Modelos . . . . .	36
1.8 Evaluación del modelo . . . . .	37
<b>2 PRUEBAS Y ANÁLISIS DE RESULTADOS</b>	<b>39</b>
2.1 Pruebas de análisis de regresión . . . . .	39
2.1.1 Experimento intra-hablante . . . . .	40
2.1.2 Experimento entre-hablantes . . . . .	46
2.2 Resultados pruebas <i>k-fold</i> . . . . .	48
2.2.1 Experimento intra-hablante . . . . .	49
2.2.2 Experimento entre-hablantes . . . . .	54
<b>ANEXOS</b>	<b>62</b>

	8
A Frases de la base de datos	62
B Banco de filtros de mel	66
C Resultados error cuadrático medio para análisis de regresión	67
D Coeficientes modelos de regresión lineal, para experimento intrahablante	69
E Coeficientes modelos de regresión lineal, para experimento entrehablantes	74
BIBLIOGRAFÍA	84

## Índice de figuras

Figura 1.1	Tres puntos de referencia (puntos rojos) y línea central para la construcción del gridline (línea amarilla) . . . . .	31
Figura 1.2	Estimación de la región de las vías respiratorias . . . . .	31
Figura 2.1	Comparación modelos de regresión empleados en hablantes mujeres, utilizando formantes calculados con LPC como variables predictoras. . . . .	41
Figura 2.2	Comparación modelos de regresión empleados en hablantes hombres, utilizando formantes calculados con LPC como variables predictoras. . . . .	42
Figura 2.3	Comparación modelos de regresión empleados en hablantes mujeres, utilizando los MFCC como variables predictoras. . . . .	43
Figura 2.4	Comparación modelos de regresión empleados en hablantes hombres, utilizando los MFCC como variables predictoras. . . . .	44
Figura 2.5	Comparación de modelos de regresión usando como variables predictoras los formantes calculados con LPC, con los modelos usando los MFCC. Experimento intra-hablante, en mujeres. . . . .	45
Figura 2.6	Comparación de modelos de regresión usando como variables predictoras los formantes calculados con LPC, con los modelos usando los MFCC. Experimento intra-hablante, en hombres. . . . .	45
Figura 2.7	Comparación de modelos de regresión usando como variables predictoras los formantes calculados con LPC, con los modelos usando los MFCC. Experimento entre-hablantes. . . . .	48
Figura 2.8	Comparación error de <i>k-fold</i> en hablantes mujeres, utilizando formantes calculados con LPC como variables predictoras. . . . .	50
Figura 2.9	Comparación error de <i>k-fold</i> en hablantes hombres, utilizando formantes calculados con LPC como variables predictoras. . . . .	50
Figura 2.10	Comparación error de <i>k-fold</i> en hablantes mujeres, utilizando MFCC como variables predictoras. . . . .	52

Figura 2.11 Comparación error de <i>k-fold</i> en hablantes hombres, utilizando MFCC como variables predictoras. . . . .	52
Figura 2.12 Comparación error de <i>k-fold</i> entre las pruebas realizadas usando como predictores los formantes calculados con LPC, y los MFCC. Experimento intra-hablante, en mujeres. . . . .	54
Figura 2.13 Comparación error de <i>k-fold</i> entre las pruebas realizadas usando como predictores los formantes calculados con LPC, y los MFCC. Experimento intra-hablante, en hombres. . . . .	55
Figura 2.14 Comparación error de <i>k-fold</i> entre las pruebas realizadas usando como predictores los formantes calculados con LPC, y los MFCC. Experimento entre-hablantes. . . . .	57

## Índice de tablas

Tabla 1.1	Información de los hablantes usados en los análisis estadísticos . . .	23
Tabla 1.2	Rangos y Longitudes promedio del tracto vocal de los hablantes utilizados . . . . .	32
Tabla 2.1	Promedio error cuadrático medio de la regresión de cada modelo, utilizando como variables predictoras los formantes calculados con LPC. Experimento intra-hablante. . . . .	41
Tabla 2.2	Promedio error cuadrático medio de la regresión para cada modelo, utilizando como variables predictoras los coeficientes MFCC. Experimen- to intra-hablante. . . . .	43
Tabla 2.3	Promedio error cuadrático medio de cada modelo, utilizando for- mantes como variables predictoras. Experimento entre-hablantes. . . . .	47
Tabla 2.4	Promedio error cuadrático medio de cada modelo, utilizando los MFCC como variables predictoras. Experimento entre-hablantes. . . . .	47
Tabla 2.5	Error de <i>k-fold</i> de cada modelo, utilizando los formantes halla- dos con LPC como variables predictoras. Experimento intra-hablante, en mujeres. . . . .	51
Tabla 2.6	Error de <i>k-fold</i> de cada modelo, utilizando formantes hallados con LPC como predictores. Experimento intra-hablante, en hombres. . . . .	51
Tabla 2.7	Error de <i>k-fold</i> de cada modelo, utilizando MFCC como predicto- res. Experimento intra-hablante, en mujeres. . . . .	53
Tabla 2.8	Error de <i>k-fold</i> de cada modelo, utilizando MFCC como predicto- res. Experimento intra-hablante, en hombres. . . . .	53
Tabla 2.9	Error de <i>k-fold</i> para experimento entre-hablantes, usando forman- tes como variables predictoras . . . . .	56
Tabla 2.10	Error de <i>k-fold</i> para experimento entre-hablantes, usando coefi- cientes de MFCC como predictores . . . . .	56

Tabla A.1	Frases de la base de datos MOCHA-TIMIT utilizadas en el proyecto	65
Tabla B.1	Banco de filtros MFCC . . . . .	66
Tabla C.1	Error cuadrático medio de validación de cada modelo utilizado, para los hablantes mujeres, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante. . . . .	67
Tabla C.2	Error cuadrático medio de validación de cada modelo utilizado, para los hablantes hombres, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante. . . . .	68
Tabla C.3	Error cuadrático medio de validación de cada modelo utilizado, para los hablantes mujeres, con MFCC. . . . .	68
Tabla C.4	Error cuadrático medio de validación de cada modelo utilizado, para los hablantes hombres, con MFCC. . . . .	68
Tabla D.1	Coefficientes de la regresión para el modelo <i>lineal</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en mujeres . . . . .	69
Tabla D.2	Coefficientes de la regresión para el modelo <i>lineal</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en hombres . . . . .	70
Tabla D.3	Coefficientes de la regresión para el modelo <i>interacciones</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en mujeres . . . . .	70
Tabla D.4	Coefficientes de la regresión para el modelo <i>interacciones</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en hombres . . . . .	71
Tabla D.5	Coefficientes de la regresión para el modelo <i>cuadrático con interacciones</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en mujeres . . . . .	71
Tabla D.6	Coefficientes de la regresión para el modelo <i>cuadrático con interacciones</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en hombres . . . . .	72
Tabla D.7	Coefficientes de la regresión para el modelo <i>cuadrático</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en mujeres . . . . .	72

Tabla D.8	Coeficientes de la regresión para el modelo <i>cuadrático</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en hombres . . . . .	73
Tabla D.9	Coeficientes de la regresión para el modelo <i>puramente cuadrático</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en mujeres . . . . .	73
Tabla D.10	Coeficientes de la regresión para el modelo <i>puramente cuadrático</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en hombres . . . . .	73
Tabla E.1	Coeficientes de la regresión para el modelo <i>lineal</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes mujeres . . . . .	74
Tabla E.2	Coeficientes de la regresión para el modelo <i>lineal</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres . . . . .	75
Tabla E.3	Coeficientes de la regresión para el modelo <i>lineal</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres y mujeres juntos . .	75
Tabla E.4	Coeficientes de la regresión para el modelo <i>interacciones</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes mujeres . . . . .	76
Tabla E.5	Coeficientes de la regresión para el modelo <i>interacciones</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres . . . . .	76
Tabla E.6	Coeficientes de la regresión para el modelo <i>interacciones</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres y mujeres juntos . .	77
Tabla E.7	Coeficientes de la regresión para el modelo <i>cuadrático con interacciones</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes mujeres .	78
Tabla E.8	Coeficientes de la regresión para el modelo <i>cuadrático con interacciones</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres	79

Tabla E.9 Coeficientes de la regresión para el modelo <i>cuadrático con interacciones</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres y mujeres juntos . . . . .	80
Tabla E.10 Coeficientes de la regresión para el modelo <i>cuadrático</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes mujeres . . . . .	81
Tabla E.11 Coeficientes de la regresión para el modelo <i>cuadrático</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres . . . . .	81
Tabla E.12 Coeficientes de la regresión para el modelo <i>cuadrático</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres y mujeres juntos . . . . .	82
Tabla E.13 Coeficientes de la regresión para el modelo <i>puramente cuadrático</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes mujeres . . . . .	82
Tabla E.14 Coeficientes de la regresión para el modelo <i>puramente cuadrático</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres . . . . .	83
Tabla E.15 Coeficientes de la regresión para el modelo <i>puramente cuadrático</i> , utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres y mujeres juntos . . . . .	83

## RESUMEN

**Título:** Análisis de la relación existente entre la longitud del tracto vocal, obtenida a partir de imágenes por resonancia magnética, y parámetros acústicos de voz.<sup>1</sup>

**Autores:**

Paula Andrea Vasquez Serrano<sup>2</sup>

Julian Orlando Reyes Moreno<sup>2</sup>

**Palabras Clave:** Tracto vocal, Formantes, Coeficientes cepstrales en las frecuencias de mel, codificación predictiva lineal, Validación cruzada, Imágenes por resonancia magnética.

### DESCRIPCIÓN

En este proyecto se presenta la evaluación de 5 diferentes modelos estadísticos de tipo lineal, a los cuales se les realizó el análisis de regresión para la estimación de la longitud del tracto de vocal, a partir de parámetros obtenidos de la voz humana. Para ello se hizo uso de la base de datos USC-TIMIT, la cual cuenta con una serie de audios e imágenes en tiempo real por resonancia magnética (rt-MRI) de 10 individuos, a partir de los cuales se extraen los parámetros necesarios para el ajuste y validación de los modelos. De esta forma, para la parametrización de la señal de la voz se utilizaron dos métodos diferentes: la codificación predictiva lineal (LPC) y los coeficientes cepstrales en las frecuencias de mel (MFCC). Los resultados obtenidos con cada uno de estos métodos fueron comparados, mostrando mejores resultados, en términos del error cuadrático medido, MFCC. Para la obtención de los valores de referencia de longitud del tracto vocal, se hizo uso de las imágenes por resonancia magnética, la cuales fueron procesadas con un algoritmo ya implementado en MATLAB. Se realizaron pruebas conformando diferentes grupos de datos, con el fin de observar el comportamiento de los modelos y la relación entre las variables. Además, se realizaron pruebas de validación cruzada de *k-fold* para la evaluación y comparación de los modelos.

---

<sup>1</sup>Trabajo de Grado

<sup>2</sup>Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Franklin Alexander Sepúlveda Sepúlveda, PhD.

## ABSTRACT

**Title:** Analysis of the relation between the vocal tract length, obtained from magnetic resonance imaging, and acoustic parameters of the voice.<sup>1</sup>

**Authors:**

Paula Andrea Vasquez Serrano<sup>2</sup>

Julian Orlando Reyes Moreno<sup>2</sup>

**Key Words:** Vocal tract, formants, Mel frequency cepstral coefficients, Linear predictive coding, Cross validation, Magnetic resonance imaging.

### DESCRIPTION

The following document presents the evaluation of five different statistical models of linear type, to which the regression analysis were applied for the estimation of the vocal tract length from parameters of the human voice. For this purpose, the USC-TIMIT database was used, this database has a set of audios and real-time magnetic resonance images (rt-MRI) of 10 speakers, from which the necessary parameters for the fit and validation of the models were extracted. Two different methods were used for the parameterization of the speech signal: the linear predictive coding (LPC), and the mel frequency cepstral coefficients (MFCC). The results obtained with these methods were compared, showing better results, in terms of the mean square error, the MFCC. The vocal tract length reference values were taken from the magnetic resonance images, this images were processed with an algorithm implemented in MATLAB. Tests were made forming different data sets, in order to analyze the behavior of the models, and the relation between the variables. In addition, *k-fold* cross validation tests were made for the evaluation and comparison of the models.

---

<sup>1</sup>Bachelor Thesis

<sup>2</sup>Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Franklin Alexander Sepúlveda Sepúlveda, PhD.

## **OBJETIVO GENERAL**

Desarrollar y evaluar algoritmos de estimación de la longitud del tracto vocal en humanos a partir de parámetros acústicos de la voz, estimados sobre señales de voz, mediante el uso de técnicas de análisis estadístico y regresión.

## **OBJETIVOS ESPECÍFICOS**

- ❖ Identificar mediante consulta del estado del arte, parámetros de la voz relacionados con la longitud del tracto vocal y estimarlos mediante métodos de procesamiento de señales.
- ❖ Estimar la longitud del tracto vocal a partir del procesamiento de imágenes por resonancia magnética en tiempo real del tracto vocal (base de datos USC-TIMIT) mediante métodos de procesamiento de imágenes.
- ❖ Desarrollar un análisis basado en técnicas de análisis estadístico y de regresión para la estimación del tracto vocal, a partir de parámetros acústicos de la voz.

## *INTRODUCCIÓN*

La longitud del tracto vocal es una de las principales características morfológicas del aparato fonador que más ejerce influencia sobre el comportamiento de la señal acústica de la voz del individuo. La importancia de la estimación de la longitud del tracto vocal radica en que contribuye a la descripción del aparato fonador, lo cual es útil para propósitos de modelado de la producción de la señal acústica y en sistemas que incluyen síntesis articulatoria. De otra parte, algunos autores han manifestado una posible relación entre esta medida y características propias del individuo tales como estatura [14] y género, lo cual podría ser de gran ayuda en propósitos de identificación de personas [2, 19]. Finalmente, la estimación de la longitud del tracto vocal podría constituirse en una herramienta útil para mejorar los sistemas de inversión articulatoria [43], los cuales a su vez tienen potencial para el estudio y diagnóstico de enfermedades relacionadas con el aparato fonador.

La longitud del tracto vocal puede ser obtenida de forma directa mediante el procesamiento de imágenes de resonancia magnética (MRI); sin embargo, dado el costo, tiempo y equipos necesarios para realizar dicho procedimiento, este no constituye en un método eficaz y práctico para obtener dicho parámetro, y por ende no resultaría de utilidad en aplicaciones prácticas.

En consecuencia, teniendo en cuenta la relación existente entre la longitud del tracto vocal y la señal de la voz, y dada la diversidad de herramientas asequibles que existen actualmente para obtener esta señal acústica y sus parámetros, resulta notablemente útil desarrollar un estimador de la longitud del tracto vocal a partir de la información contenida en la voz.

Aunque se han desarrollado algoritmos de estimación de la longitud del tracto vocal [13, 26, 46], muchos de estos no han sido debidamente evaluados en conjuntos de datos reales de voz humana, o presentan errores de estimación que no son aceptables para su posterior uso en aplicaciones. Es por esto que con este proyecto de grado se busca explorar el desarrollo de algoritmos para la estimación de la longitud del tracto vocal a partir de la señal acústica del habla.

En particular, en [46] se estima la longitud del tracto vocal a partir de la frecuencia promedio del segundo formante, el cual a su vez es determinado por las frecuencias de los tres primeros formantes. El otro método propuesto, también expuesto en [46], se basa en la diferencia entre las frecuencias de formantes adyacentes de orden alto. Según los mismos autores, el primer método mencionado presenta un error absoluto promedio de 15.6 %, mientras que el segundo presenta un error de 7.1 %. Sin embargo, la base de datos utilizada por los autores para evaluar estos estimadores era considerablemente pequeña, y solo contaba con datos simulados obtenidos a partir de un sintetizador articulatorio de voz. Adicionalmente, en [13] se desarrolló un método basado en la distancia promedio entre cada par adyacente de formantes. A este método se le conoce como dispersión de frecuencia y ha sido utilizado para predecir la estatura y peso en humanos [36] y algunas especies de animales [13, 20, 35, 37]. Particularmente en animales, este estimador ha demostrado tener una buena correlación con las características mencionadas. Finalmente, Lammert & Narayanan en [26] proponen un modelo lineal, considerando que cada frecuencia de formante puede ser tomada como una característica que puede brindar información sobre la longitud del tracto vocal. Este estimador fue evaluado tanto en datos simulados como en datos reales acústicos y de MRI. El valor de error cuadrático medio (MSE, *mean square error*) que se obtuvo en el mencionado trabajo corresponde a 0.631cm y 1.277cm, para señales sintetizadas y señales reales, respectivamente. Para el caso de datos tomados del habla humana se utilizaron cinco personas.

En el presente trabajo se plantea realizar un análisis de la relación existente entre la longitud del tracto vocal y la señal acústica de voz. Para ello se usarán datos reales provenientes de la base de datos USC-TIMIT [32] y modelos de regresión, con lo cuales se busca utilizar modelos más flexibles. Adicionalmente, el presente método no se limita al uso de formantes, y en su lugar incluye parámetros MFCC (*mel frequency cepstral coefficients*).

# 1. MÉTODO

## 1.1. Producción de la voz y el tracto vocal

La voz es una señal acústica resultante de los sonidos generados por el aparato fonador. El sonido en su frecuencia fundamental y armónicos, es originado en las cuerdas vocales, estas ondas sonoras se propagan a través del tracto vocal, el cual actúa como resonador y filtro del sonido [29].

El tracto vocal es una cavidad comprendida entre los labios y la glotis. En los humanos está conformado por la cavidad oral, la cavidad nasal, la faringe y la laringe. Cada una de estas partes también está compuesta por otras más pequeñas, dentro de las que se incluyen los labios, mandíbula, dientes, lengua, paladar duro, velo del paladar (paladar blando), como partes de la cavidad oral; y las cuerdas vocales y la glotis como partes de la laringe, entre otras [6, 18]. Las posiciones y formas que tomen los órganos articuladores (paladar, lengua, dientes, labios, glotis) definen la forma del tracto vocal, el cual actúa como filtro acústico que modifica el espectro del sonido originado por las cuerdas vocales, determinando así los diversos formantes de cada unidad segmental del habla [12].

Los formantes son zonas del espectro acústico en torno a las cuales se concentra la energía acústica, que a su vez corresponden a picos de la respuesta en frecuencia de la cavidad resonadora (tracto vocal) [4]. Los formantes se caracterizan por tres parámetros importantes: centro de frecuencia, ancho de banda y energía. Cuando se modifica la forma del tracto vocal, estos tres parámetros también se modifican, y por lo tanto la función de transferencia del filtro y el sonido obtenido también cambian [18]. Estas modificaciones en la forma de las partes que conforman el tracto vocal se ven reflejadas principalmente en dos aspectos: a) como cambios en la longitud del tracto vocal,

medida desde la glotis hasta los labios; y b) en los diferentes diámetros transversales a lo largo de esta longitud. Dependiendo del género y la edad (niñez ó etapa adulta), el largo del tracto vocal difiere [14, 47], por lo que los valores de los formantes para un fonema en específico también serán diferentes de acuerdo a estas características de la persona [7]. De esta forma el largo del tracto vocal y los formantes intervienen tanto en la caracterización del hablante, como en la caracterización de los diferentes sonidos y fonemas del habla en humanos, convirtiéndose en un elemento fundamental para el estudio de la voz [4].

Para el desarrollo de este proyecto, se partió del modelo lineal de producción de la voz, también conocido como modelo fuente-filtro. Según este modelo, la producción de la voz involucra la acción de un filtro, que corresponde al tracto vocal, y una fuente del sonido, que corresponde a la glotis y las cuerdas vocales. La función de transferencia del tracto vocal es modelada como un filtro lineal, cuya entrada es la señal originada en la fuente, y la salida es la señal de la voz [47]. La base fundamental de este modelo es que asume que la fuente y el filtro son independientes uno del otro, lo que implica que cada uno realiza contribuciones independientes a las características y parámetros de la voz. Aunque esta suposición no es estrictamente cierta, el modelo fuente-filtro es considerado un modelo útil y acertado para la síntesis y análisis de la voz [5, 38].

## 1.2. Base de datos

Los registros fuente utilizados para realizar el presente proyecto provienen de vídeos de resonancia magnética que incluyen además la señal acústica del habla sincronizada con el vídeo, los cuales fueron tomados de la base de datos USC-TIMIT [8]. Esta es una base de datos que muestra la producción de habla mediante imágenes en tiempo real por resonancia magnética (rt-MRI) de 5 hombres y 5 mujeres adultos, con edades entre los 20 y 46 años, hablantes nativos de inglés Americano y, ninguno de los cuales presentaba patología de escucha o habla.

Para cada uno de ellos utilizaron las mismas oraciones incluidas en la base de datos del MOCHA-TIMIT [50]. De esta forma, la base de datos USC-TIMIT contiene un conjunto de 460 frases diseñadas para obtener todos los fonemas del inglés americano en una amplia gama de contextos fonológicos [32]. Los datos están organizados de tal

forma que en cada vídeo hay 5 frases, separadas por espacios de silencio entre cada una.

Los vídeos de resonancia magnética fueron adquiridos en Los Angeles County Hospital con un escáner Signa Excite HD 1.5T (GE Healthcare, Waukesha, WI). La reconstrucción de las imágenes fue realizada usando MATLAB (Mathworks, South Natick, MA), obteniéndose finalmente vídeos de 23,18 fotogramas por segundo, con una resolución de imagen de 68x68 píxeles (2,9x2,9 mm por píxel).

Los audios fueron grabados simultáneamente dentro del escáner MRI con una frecuencia de muestreo de 20kHz. La sincronización del audio con la señal de vídeo fue controlada con el master clock del escáner. Se utilizó un micrófono de fibra óptica (Optoacoustics Ltd., Moshav Mazor, Israel) con el fin de minimizar la influencia del ruido electromagnético sobre el micrófono. Aún así, los autores de la base de datos utilizaron un algoritmo de procesamiento de señal adaptativo que tuvo en cuenta la estructura periódica del ruido generado por el escáner de MRI para la eliminación de este ruido. Para una descripción mas detallada de la base de datos ver [32].

### 1.3. Conjunto de datos para la realización de experimentos

De la base de datos USC-TIMIT [8] ya descrita, se tomaron 18 audios con sus respectivos vídeos de RMI para cada uno de los 10 hablantes (5 hombres y 5 mujeres), por lo que en total, se extrajeron datos de 180 audios y 180 vídeos. Esto equivale en promedio a 4576 muestras por cada hablante. Cabe aclarar que para cada hablante se utilizaron las mismas 90 frases (ver apéndice A). De los audios se extrajeron parámetros de la voz, y de los vídeos, el valor de referencia para la longitud del tracto vocal.

En la tabla 1.1 se muestra información sobre los hablantes y los datos utilizados para la realización de pruebas.

Hablante	Género	Edad	Número de vídeos y audios	Total frases	Total de observaciones (*)
F1	Femenino	23	18	90	4534
F2	Femenino	32	18	90	4505
F3	Femenino	20	18	90	4456
F4	Femenino	46	18	90	4629
F5	Femenino	25	18	90	4731
M1	Masculino	29	18	90	4891
M2	Masculino	33	18	90	4581
M3	Masculino	26	18	90	4143
M4	Masculino	26	18	90	4625
M5	Masculino	27	18	90	4671

(\*) En el número total de observaciones no están incluidos los espacios de silencio

Tabla 1.1: Información de los hablantes usados en los análisis estadísticos

## 1.4. Representación paramétrica de la señal de voz

La parametrización de la voz busca convertir pequeños segmentos de la señal acústica, en conjuntos de parámetros que describan las propiedades y características de cada uno de estos fragmentos [15], de forma tal que se elimine la información que no contribuye al análisis de la voz, y se enfatizen aquellas características representativas. Para el desarrollo de este proyecto se emplearon dos tipos diferentes de representaciones paramétricas de la voz, la codificación predictiva lineal (LPC) y el método por coeficientes cepstrales en las frecuencias de mel (MFCC).

**1.4.1 Pre-procesamiento** Antes de realizar el proceso de parametrización, la señal de la voz pasa por una etapa de pre-procesamiento, en la cual se prepara la señal de forma tal que tenga un formato que facilite su análisis. El pre-procesamiento consiste básicamente de dos pasos, el pre-énfasis y el enventanado.

- ❖ **Pre-énfasis:** La señal acústica primeramente es filtrada para atenuar las bajas frecuencias de la voz, y enfatizar las altas. El filtro de pre-énfasis es un filtro FIR de primer orden cuya función de transferencia está dada por la ecuación 1.1,

donde  $a$  es un coeficiente cuyo valor normalmente se encuentra entre 0,95 - 0,98. Para el presente trabajo se estableció un valor de 0,97.

$$H(z) = 1 - az^{-1} \quad (1.1)$$

- ❖ **Enventanado:** La señal de voz es dividida en cortos segmentos de tiempo, normalmente de entre 20-60 ms, en donde la señal puede ser considerada como casi-estacionaria [39], sin perder sus características e información. El proceso de enventanado consiste en multiplicar la señal de la voz por una ventana en el tiempo, obteniéndose estos segmentos o tramos de la señal. Para el presente estudio se utilizó una ventana de tipo Hamming, y la señal fue dividida en tramos de 43,14 ms.

**1.4.2 Estimación de formantes con LPC** La codificación predictiva lineal (LPC) es una herramienta ampliamente usada en el procesamiento de voz, y el análisis de formantes [9]. Se basa en la teoría de producción de la voz fuente-filtro, modelando el tracto vocal como un sistema lineal e invariante en el tiempo, a partir de un filtro todopolo (*all-pole filter*, ecuación 1.3), en el que cada par conjugado de polos, equivale a la frecuencia de un formante, y cada muestra de la señal de la voz es estimada basada en una combinación lineal de sus muestras anteriores [16], como se muestra en la ecuación 1.2. Donde  $s(n)$  es una muestra de la señal de la voz en un tiempo  $n$ , y  $p$  es el numero total de muestras de la señal.

$$s(n) \approx a_1s(n-1) + a_2s(n-2) + \dots + a_ps(n-p) \rightarrow \tilde{s} = \sum_{i=1}^p a_i s(n-i) \quad (1.2)$$

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} = \frac{1}{A(z)} \quad (1.3)$$

En la ecuación 1.3,  $U(z)$  corresponde a la señal generada en la fuente, escalada por una ganancia  $G$ , y  $S(z)$  es la señal de la voz resultante. Esta ecuación también puede ser expresada en el dominio del tiempo discreto como se muestra en la ecuación 1.4.

$$s(n) = \sum_{i=1}^p a_i s(n-i) + Gu(n) \quad (1.4)$$

Finalmente, el análisis LPC busca encontrar los coeficientes  $a_i$  de tal forma que se minimice el error de estimación  $e(n)$  (ecuación 1.5).

$$e(n) = s(n) - \tilde{s}(n) \quad (1.5)$$

La estimación de las frecuencias de formantes a partir de señales acústicas, mediante la técnica de LPC, se realizó con *Wavesurfer*, el cual es un software libre para la edición de audio. *Wavesurfer* tiene una interfaz de usuario simple que provee funcionalidad de forma intuitiva para que se pueda adaptar a diferentes tareas. Sus aplicaciones más típicas son en análisis de sonido y habla [44]. Esta herramienta presenta un buen desempeño para mostrar las ondas de presión de sonido, secciones espectrales, espectrogramas, seguimiento de tono y transcripciones. Además, puede leer un amplio número de formatos de archivo, incluyendo TIMIT.

Además de realizar el proceso de detección de formantes con LPC, *Wavesurfer* también realiza el pre-procesamiento de la señal. Antes de realizar la extracción de estos valores se deben ajustar algunos parámetros en la configuración de *Wavesurfer*, estos son:

- ❖ Coeficiente del filtro de pre-énfasis.
- ❖ Ancho del tramo (frame).
- ❖ Tipo de ventana de análisis.
- ❖ Número de formantes.
- ❖ Orden del LPC.

*Wavesurfer* utiliza el filtro de pre-énfasis descrito en ecuación 1.1, como se había mencionado anteriormente, el coeficiente  $a$  es ajustado en 0,97. Para poder obtener valores de formantes para cada valor de longitud del tracto vocal, se necesita que el ancho de cada tramo en el que se va a dividir la señal de los audios sea igual al de los vídeos por resonancia magnética; teniendo en cuenta que los vídeos tienen una tasa de 23,18 fotogramas por segundo, el ancho de frame usado fue de 0,04314 segundos. La ventana de análisis utilizada es de tipo Hamming.

El orden del LPC está asociado al número de formantes, y se debe cumplir con la siguiente regla para un correcto funcionamiento:

$$\text{Numero de formantes} = \frac{\text{LPC Order} - 4}{2} \quad (1.6)$$

Dado que se decidió utilizar 4 formantes, siguiendo la ecuación 1.6 el orden del LPC debe ser de 12. Se eligió este número de formantes ya que esta suele ser la configuración más confiable. Esta afirmación se puede sustentar en el artículo “*Formant Estimation for Speech Recognition*” de Lutz Welling and Hermann Ney, Member, IEEE [49], donde se compara el resultado del reconocimiento de voz con 3, 4 y 5 formantes, presentando un menor error la configuración con sólo los 4 primeros formantes [49].

Ya habiendo configurado estos parámetros, se pueden extraer los valores de formantes necesarios para el experimento correspondiente, solo es necesario ingresar cada archivo de audio de nuestro conjunto de datos, y automáticamente *Wavesurfer* imprime en un archivo de texto la lista de formantes (en kHz) para cada uno de ellos.

**1.4.3 MFCC** Los coeficientes cepstrales en las frecuencias de mel (MFCC) son una técnica de extracción de parámetros de la voz basada en la forma en la que el sistema auditivo humano percibe el habla. Con esta técnica se busca extraer características de la señal de audio que sean relevantes para el análisis de la voz, desenfatiando otros tipos de información que poco aporten al proceso de análisis y reconocimiento de voz [9].

MFCC ha sido ampliamente usado en sistemas de reconocimiento de voz [25, 30, 40, 41, 48], debido a sus esquemas de computación eficientes y a su robustez en presencia de diferentes tipos de ruido [15], lo cual lo hace que brinde mejores resultados en el reconocimiento de voz que otras técnicas de parametrización como LFCC (coeficientes cepstrales de frecuencia lineal) y LPCC (coeficientes cepstrales derivados de la predicción lineal) [15, 17].

Los coeficientes MFCC son el resultado de la transformada de coseno, del logaritmo real del espectro de energía de tiempo corto, expresados en la escala de frecuencia de

mel [40]. La escala de mel es una escala perceptual que proporciona una relación entre la frecuencia percibida de un tono puro, como una función de su frecuencia acústica. Esta escala es lineal por debajo de 1 kHz y logarítmica por encima de este umbral [15]; puede ser calculada a partir de valores de frecuencia (en Hz) como se muestra en la ecuación 1.7.

$$Mel = 2595 \log_{10} \left( 1 + \frac{f}{700} \right) \quad (1.7)$$

Para calcular los coeficientes cepstrales en las frecuencias de mel, la señal de la voz primeramente debe haber sido pre-procesada; a cada tramo en el que fue dividida la señal, se le aplica la transformada de Fourier discreta, como se muestra en la ecuación 1.8, donde  $s(n)$  corresponde a cada uno los segmentos de señal de voz y  $N$  equivale al número de muestras contenidas en cada tramo.

$$S(k) = \sum_{n=0}^{N-1} s(n) \cdot \exp \left( \frac{-j2\pi nk}{N} \right), k = 0, 1, \dots, N - 1 \quad (1.8)$$

El resultado obtenido de realizar la transformada discreta de Fourier, es utilizado para calcular la amplitud del espectro de la señal  $|S(k)|$ . Esta información pasa por un banco de filtros de mel ( $H_i$ ), y se calcula el logaritmo de la energía de salida del banco de filtros, según se muestra en la ecuación 1.9

$$S_i = \log_{10} \left( \sum_{k=0}^{N-1} |S(k)| \cdot H_i(k) \right), i = 1, 2, \dots, M \quad (1.9)$$

siendo  $M = \text{número de filtros del banco de filtros de mel}$ .

Finalmente se aplica la transformada de coseno discreta a estos logaritmos (ecuación 1.10), obteniéndose los coeficientes cepstrales.

$$MFCC(r) = \sqrt{\frac{2}{M}} \sum_{i=0}^{M-1} S_{i+1} \cdot \cos \left( \frac{r(i + 0,5)\pi}{M} \right), r = 0, 1, \dots, R - 1 \quad (1.10)$$

donde  $R \leq M$ , siendo  $R$  el número de coeficientes que se desean obtener.

El banco de filtros usado corresponde al propuesto en el Auditory Toolbox [45], está compuesto por 40 filtros triangulares medio solapados, de igual área, que cubren un rango de frecuencia de 133,33 a 6855 Hz. Las frecuencias centrales de los primeros 13 filtros están linealmente espaciadas en un rango de 200 a 1000 Hz, con un paso de aproximadamente 66,7Hz; y las frecuencias centrales de los siguientes 27 filtros están logarítmicamente espaciadas en un rango de 1071 a 6400 Hz. Los límites de las bandas de frecuencia de cada filtro se encuentran en el apéndice B.

La extracción de los coeficientes de MFCC se realizó en MATLAB con la función *mfcc*, la cual calcula automáticamente los coeficientes cepstrales en las frecuencias de mel, haciendo uso del banco de filtros de mel ya descrito. Antes de hacer uso de esta función, los audios son multiplicados por el filtro de pre-énfasis previamente indicado en la ecuación 1.1, este filtro también es implementado en MATLAB. Por su parte, el proceso de inventariado ya está incluido en la función *mfcc*.

Para usar esta función se debe especificar: el número de coeficientes, que en este caso es de 13, dado que es el valor comúnmente empleado; y el número de muestras en la ventana de análisis, para el cual se escogió un valor de 862, de forma tal que la señal de audio quedara dividida en el mismo número de segmentos de los videos por resonancia magnética. La función también permite calcular el logaritmo de la energía dentro de los frames, y los delta y delta delta de los coeficientes, que equivalen a la velocidad y aceleración con la que cambian los tramos en los que ha dividido la señal. Estas características de los MFCC no van a ser utilizadas en nuestro estudio, por lo que los valores de configuración relacionados con ellas son ajustados en 0. El primer coeficiente no es utilizado dado que equivale a la energía de la señal; para realizar el análisis de regresión se utilizan los coeficientes del 2 al 13.

## 1.5. Algoritmo de medición del valor de referencia longitud de tracto vocal

Para la obtención del valor de longitud de tracto vocal usado como referencia, se utilizó un software de MATLAB llamado *rtmri\_seg*, el cual es una implementación del algoritmo descrito en [24]. El software es de uso libre y está disponible en [23]. Se utilizó la versión 4 puesto que es la más reciente hasta el momento. Este software

contiene códigos para la extracción de parámetros del tracto vocal y permite el análisis automático y sistemático de imágenes de las vías respiratorias superiores. El software realiza lo siguiente:

- ❖ Mejoramiento de las imágenes, que corresponde a la corrección de la intensidad del pixel de las imágenes por resonancia magnética, reducción del ruido, incremento del contraste entre los tejidos y la vía respiratoria.
- ❖ Construcción semi-automática del gridline basado en una línea trazada manualmente por el centro de los límites del tejido de las vías respiratorias, y 3 puntos de referencia también seleccionados manualmente por el usuario.
- ❖ Detección del borde frontal de los labios y el tope de la laringe.
- ❖ Segmentación automática de los límites del tejido de las vías respiratorias en el tracto vocal y estimación de un camino de la vía respiratoria faríngea dentro de este.
- ❖ Medición de la distancia entre los límites superior e inferior de las vías respiratorias.
- ❖ Cálculo del vector para la forma de la vía respiratoria superior.
- ❖ Cálculo de la longitud del tracto vocal para cada imagen por resonancia magnética.

La trayectoria de las vías respiratorias así como la ubicación de los labios y la laringe son encontradas usando el algoritmo de Viterbi. Los límites de los tejidos de las vías respiratorias son encontrados para cada gridline buscando el pixel más cercano de mayor intensidad [24]. El algoritmo implementado en el software `rtmri_seg` permite tener una segmentación robusta de los tejidos, el cual reporta un error cuadrático medio de 0,71 pixeles y 0,93 pixeles para la estimación de los límites interiores y exteriores de los tejidos de las vías respiratorias, respectivamente. Para una descripción más detallada de este algoritmo, ver [24].

Dado que los vídeos de imágenes por resonancia magnética están en un formato `.avi`, antes de realizar la extracción de parámetros se deben obtener todos los fotogramas de cada vídeo y convertirlos a una matriz de datos. El software `rtmri_seg_v4` [23]

ya contiene un código para efectuar esto, puesto que con la matriz de datos realiza el procesamiento de las imágenes. Para la construcción del gridline y la extracción de parámetros del tracto vocal, inicialmente se toma una imagen de referencia para cada vídeo, y en ella se debe trazar a pulso una línea que siga el camino del tracto vocal, desde los labios hasta la glotis. Posteriormente se deben ubicar manualmente los tres puntos de referencia, estos son: el punto más bajo de labio superior, el paladar duro, y por último, la glotis. Además, se ingresan los rangos de tiempo donde hay ausencia del habla, para que el software al momento de realizar la extracción de parámetros elimine los datos registrados en estos segmentos de tiempo. Ya teniendo estos puntos determinados, el software procesa todos los fotogramas del vídeo, construye el gridline, e identifica los límites de los tejidos de las vías respiratorias de manera automática. A partir de esto, el software extrae los parámetros y calcula el valor de la longitud del tracto vocal para cada uno de los fotogramas del vídeo. Los valores de longitud entregados por el software están dados en pixeles, por lo que posteriormente son transformamos a unidades de longitud métricas (centímetros), para ello se tiene en cuenta que en las imágenes por resonancia magnética utilizadas un pixel equivale a 2,9mm.

Para una imagen inicial de MRI, las figuras 1.1, ?? y 1.2, son ejemplos de la ubicación de los puntos de referencia ingresados manualmente, el gridline y la estimación de la trayectoria de las vías respiratorias del software, respectivamente. En la tabla 1.2 se muestran los rangos de longitud del tracto vocal encontrados por cada hablante, y la longitud promedio.

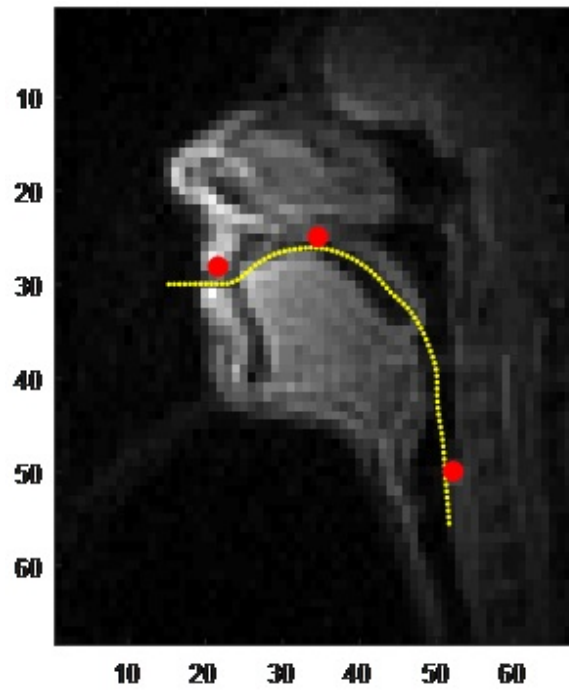


Figura 1.1: Tres puntos de referencia (puntos rojos) y línea central para la construcción del gridline (línea amarilla)

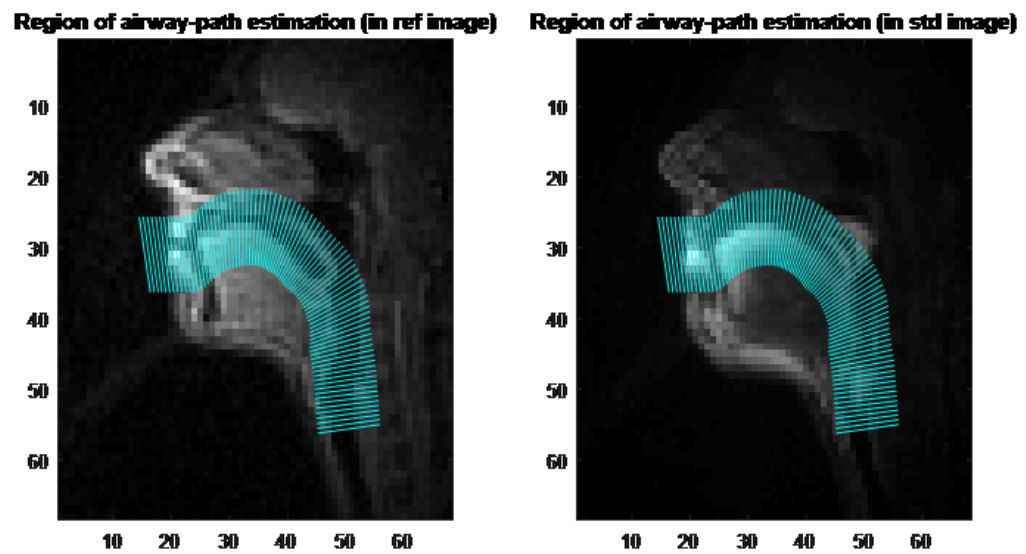


Figura 1.2: Estimación de la región de las vías respiratorias

Hablante	Longitud tracto vocal [cm]	
	Rango	Promedio
F1	11,07 - 15,24	13,09
F2	13,47 - 17,69	15,55
F3	12,26 - 16,27	14,28
F4	11,19 - 17,39	14,14
F5	12,92 - 16,30	14,57
M1	14,66 - 17,76	16,21
M2	13,97 - 17,48	15,74
M3	14,24 - 17,34	15,74
M4	13,52 - 17,22	15,31
M5	15,31 - 19,31	17,31

Tabla 1.2: Rangos y Longitudes promedio del tracto vocal de los hablantes utilizados

## 1.6. Pre-procesamiento de datos

El pre-procesamiento de datos incluye dos pasos. Con el primer paso se busca remover aquella información acústica no relevante, en este caso los silencios; y con el segundo, se remueven datos anómalos antes de realizar el procedimiento de regresión.

**1.6.1 Remoción de silencios** Como ya se había mencionado, los audios de la base de datos contienen cinco frases, con espacios de silencio (ausencia de voz) entre cada una de ellas. La información contenida en estos espacios, así como de otros, no debe tenerse en cuenta al momento de realizar la estimación del modelo estadístico. Por tanto, se deben detectar dichos segmentos y eliminarlos antes de la etapa de estimación de parámetros y análisis de regresión. Para la identificación de los tiempos en los que hay ausencia de habla se utiliza VAD (*Voice Activity Detector*), una técnica utilizada en el procesamiento de voz, que se encarga de detectar la presencia o ausencia de la voz humana. Este procedimiento es comúnmente usado en reconocimiento de voz [10].

Existen diversos algoritmos de VAD, pero para el desarrollo del presente proyecto se utilizó el *Google WebRTC Voice Activity Detector* (WebRTC-VAD) en su versión “webrtcvad 2.0.10”, el cual está desarrollado para Python. Para utilizar este VAD es necesario instalar el modulo webrtcvad, y descargar y configurar el archivo correspondiente de la página de paquetes de Python <https://pypi.python.org/pypi/webrtcvad>.

Previamente los archivos de audio que se deseen trabajar con dicho software deben estar o ser convertidos a una frecuencia de muestreo de 8 kHz, 16 kHz, o 32 kHz. Para el presente caso los audios se convirtieron a una frecuencia de 32 kHz.

El resultado obtenido de dicho procedimiento es un vector binario, el cual asocia un “1” cuando encuentra voz y “0” cuando no encuentra voz. Este vector es procesado por un código que registra la cantidad de unos y ceros, y extrae los rangos de tiempos.

**1.6.2 Detección de valores atípicos** Ya teniendo los valores de longitud de tracto vocal con sus respectivos formantes, se decide eliminar los valores atípicos (*outliers*) de cada variable. Es decir, aquellos valores que están estadísticamente alejados de los demás valores del conjunto de datos. Los valores atípicos pueden alterar los resultados del modelo obtenido a partir de los datos, especialmente cuando se utiliza el criterio de error cuadrático medio para la estimación de parámetros [22]. Por tanto resulta pertinente eliminar estos valores antes de generar un modelo estadístico. Los *outliers* podrían ser el resultado de errores al momento de realizar las mediciones y toma de datos, o durante el procesamiento de estos. En el caso que nos atañe, p.e. si se utilizan formantes, no se espera que estos cambien de una manera abrupta debido a que están relacionados con las frecuencias de resonancia del tracto vocal, y este por ser un sistema biológico no se espera que asuma configuraciones extremas.

La media y la desviación estándar son fuertemente influenciadas por la presencia de *outliers* [21, 27], sin embargo, la mediana no es muy sensible a estos valores dado que tiene un punto de quiebre de 50% [27]. El punto de quiebre es una medida de prueba de la robustez de un procedimiento estadístico. Para un estimador estadístico equivale a la fracción de datos a los que se les puede dar valores arbitrarios, sin hacer que el estimador proporcione un resultado incorrecto [42].

Teniendo en cuenta lo anterior, para el presente caso se decidió emplear un método de detección de valores atípicos que utilice la mediana, más específicamente, uno basado en la desviación mediana absoluta (MAD). La MAD es una medida robusta de la dispersión alrededor de la mediana y, al igual que esta, no se ve significativamente alterada por los valores atípicos [1].

Para una variable  $X$ , con  $n$  observaciones, la desviación mediana absoluta está definida como

$$MAD = b \times mediana(|x_i - mediana(x)|) \quad (1.11)$$

donde  $i = 1, 2, \dots, N$  y  $b = 1, 4826$ .

En el desarrollo del presente proyecto, se consideraron como valores atípicos, aquellos cuya desviación absoluta con respecto a la mediana de los datos estuviera 3 veces arriba de la MAD. De esta forma, un valor es un *outlier* sí:

$$\frac{|x_i - mediana(x)|}{MAD} > 3 \quad (1.12)$$

Previamente a la realización del modelado estadístico, se identificaron y eliminaron los valores atípicos de cada una de las variables a emplear. Esto se hizo para el total de las muestras de cada hablante (caso intra-hablante), y de hombres y mujeres por separado (caso entre-hablantes).

## 1.7. Análisis de regresión

La etapa de regresión requiere de dos fases: la selección de la estructura del modelo y el ajuste de los parámetros del modelo. En lo que se refiere a la estructura del modelo se inicia probando con modelos lineales sencillos, en consonancia con lo encontrado en trabajos como [11, 13, 26, 28, 33, 46]. Se realizaron pruebas con diferentes modelos estadísticos, pero primero los *outliers* fueron eliminados.

El análisis de regresión es un proceso estadístico que permite estudiar y estimar la relación entre una variable dependiente, y una o más variables independientes, o predictores [22]. De esta forma la regresión, además de permitir generar un modelo con fines de predicción de la longitud del tracto vocal, también ayuda a entender cómo se ve afectada esta variable, al presentarse cambios en los valores de las variables independientes [3]. Para el caso tratado en este proyecto, la variable dependiente equivale a la longitud del tracto vocal y los predictores son, por un lado, los valores de frecuencia de los formantes 1, 2, 3 y 4 hallados con LPC; y por otro, 12 coeficientes MFCC de la señal acústica.

Los modelos de regresión pueden ser de tipo lineal o no lineal. Dada la naturaleza y relación de las variables independientes (en consonancia con lo reportado en trabajos previos), se decidió usar modelos de regresión lineal como primera opción. De forma general, el modelo de regresión lineal es expresado como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (1.13)$$

Donde  $y$  es la variable dependiente, y  $x_1, x_2, \dots, x_p$ , son las variables independientes o predictores.  $\beta_0$  es la intersección o término constante.  $\beta_1, \beta_2, \dots, \beta_p$ , son los parámetros del modelo o coeficientes, siendo  $p$  el número de variables independientes de la regresión. Finalmente,  $\varepsilon$  es un término de error aleatorio.

Cabe aclarar que los modelos de tipo lineal predicen el valor de una variable a partir de relaciones lineales de los coeficientes del modelo, sin embargo, las variables predictoras no necesariamente deben estar relacionadas en forma lineal en la ecuación [34].

**1.7.1 Estimación de coeficientes** En el modelo de regresión lineal descrito en la ecuación 1.13, los coeficientes  $\beta_k$  inicialmente son desconocidos, por lo que el objetivo es estimar estos coeficientes de forma tal que el modelo logre ajustarse bien a los datos disponibles. Aunque existen diversos métodos para la estimación los coeficientes, el más común, y el que va a ser utilizado en el presente caso, es el basado en el criterio de mínimos cuadrados [22]. Este método se describe a continuación.

Siendo  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$  las estimaciones de los coeficientes, se pueden hacer predicciones de la respuesta del modelo, a partir de la ecuación 1.14.

$$\hat{y} = \hat{\beta}_0 x_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p \quad (1.14)$$

Donde  $\hat{y}$  es una predicción del valor de  $y$ , basada en los valores reales de las variables independientes  $x_k$ . Teniendo en cuenta el criterio de mínimos cuadrados, se seleccionan los valores de coeficientes de tal forma que minimicen la suma de cuadrados de los

residuos, esto es:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1.15)$$

**1.7.2 Modelos** A continuación se describen los modelos empleados para realizar el análisis de regresión. Las ecuaciones de ejemplo de cada uno de estos modelos corresponden al análisis realizado tomando como variables predictoras las frecuencias de formantes.

- ❖ Lineal: Este modelo es la forma básica de un modelo de regresión lineal múltiple.

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \quad (1.16)$$

- ❖ Interacciones: Este modelo presenta un término constante, los términos lineales de las variables predictoras y todos los productos de pares de estas, sin incluir términos al cuadrado.

$$\begin{aligned} y = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1 x_2 \\ & + \beta_6 x_1 x_3 + \beta_7 x_1 x_4 + \beta_8 x_2 x_3 + \beta_9 x_2 x_4 + \beta_{10} x_3 x_4 \end{aligned} \quad (1.17)$$

- ❖ Cuadrático con interacciones: Este modelo tiene un término constante y los términos lineales, interacciones y cuadrados de las variables predictoras.

$$\begin{aligned} y = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1 x_2 + \beta_6 x_1 x_3 + \beta_7 x_1 x_4 \\ & + \beta_8 x_2 x_3 + \beta_9 x_2 x_4 + \beta_{10} x_3 x_4 + \beta_{11} x_1^2 + \beta_{12} x_2^2 + \beta_{13} x_3^2 + \beta_{14} x_4^2 \end{aligned} \quad (1.18)$$

- ❖ Cuadrático: Este modelo contiene un término constante y los términos lineales y cuadrados de los predictores.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1^2 + \beta_6 x_2^2 + \beta_7 x_3^2 + \beta_8 x_4^2 \quad (1.19)$$

- ❖ Puramente cuadrático: En este modelo se incluye un término constante y los cuadrados de los predictores.

$$y = \beta_0 + \beta_1 x_1^2 + \beta_2 x_2^2 + \beta_3 x_3^2 + \beta_4 x_4^2 \quad (1.20)$$

## 1.8. Evaluación del modelo

Existen diferentes técnicas para validar el análisis por regresión, una de ellas es la validación cruzada, que es un método estadístico de re-muestreo utilizado para evaluar y comparar modelos, que permite que los resultados de la aplicación de algún modelo estadístico sean menos dependientes de alguna división particular entre los datos de entrenamiento y prueba [31]. La validación cruzada puede ser usada para estimar el error esperado asociado a un modelo estadístico, además de facilitar la selección del nivel apropiado de flexibilidad del modelo [22].

En el presente proyecto para realizar la evaluación de los modelos, se utilizó específicamente la validación cruzada de  $k$  iteraciones, o *k-fold*; en este tipo de validación cruzada el conjunto de observaciones es dividido aleatoriamente en  $K$  grupos, de aproximadamente el mismo tamaño. Uno de los grupos es tratado como conjunto de validación y los demás ( $k-1$ ) son utilizados para entrenar el modelo. Este proceso se repite durante  $k$  iteraciones, y en cada una de ellas un grupo diferente de observaciones es tratado como conjunto de validación [31].

En cada una de las iteraciones se calcula el error cuadrático medio, obteniéndose  $k$  estimaciones del error de validación. El error final (error de *k-fold*) se halla calculando la media aritmética de estos valores, como se muestra en la ecuación 1.21 [22].

$$E = \frac{1}{k} \sum_{i=1}^k MSE_i \quad (1.21)$$

Debido al costo computacional, no se suele usar un número de iteraciones muy grande, normalmente se utiliza un valor de  $k$  entre 5 y 10. Aunque el valor depende del tamaño del conjunto de datos. Teniendo en cuenta el conjunto de datos usado en el presente proyecto y los modelos a ajustar, se escogió  $k = 20$ .

Además de presentar una ventaja computacional con otras técnicas de validación cruzada y re-muestreo, como la validación cruzada dejando uno fuera y la aleatoria, *k-fold* también permite tener una variabilidad pequeña del error de validación, dando estimaciones más acertadas de este. Esto debido a que el promedio de las salidas de los  $k$  modelos ajustados están menos correlacionados entre sí, puesto que el solapamiento entre los conjuntos de entrenamiento es pequeño [22].

## 2. PRUEBAS Y ANÁLISIS DE RESULTADOS

### 2.1. Pruebas de análisis de regresión

El análisis de regresión descrito en la sección 1.7 fue realizado con los modelos lineales ya especificados, teniendo en cuenta que existen dos grupos de variables predictoras: las frecuencias de formantes calculadas con LPC (en Hertz), y los coeficientes MFCC. Este análisis se hizo para cada hablante (experimento intra-hablante), y también con el total de datos provenientes tanto de hombres como de mujeres (experimento entre-hablantes).

Primeramente se estimaron los valores de los coeficientes de cada modelo, obteniéndose el ajuste del modelo; a los datos utilizados para realizar este proceso se les denomina conjunto de datos de entrenamiento. Posteriormente se procedió a evaluar la exactitud del ajuste, calculando el error cuadrático medio de validación; a los datos utilizados para la evaluación del modelo se les llama conjunto de datos de validación.

El proceso de ajuste de modelos fue realizado con MATLAB, haciendo uso de la función *fitlm*, puesto que ella permite construir un ajuste de mínimos cuadrados de un modelo lineal. Para ello se debe:

- ❖ Ingresar los datos de entrenamiento.
- ❖ Especificar la variable dependiente, que corresponde a la longitud del tracto vocal en centímetros.
- ❖ Especificar los predictores.
- ❖ Ingresar la ecuación del tipo de modelo a utilizar.

La validación del modelo y el cálculo del error cuadrático medio también fueron realizados en MATLAB. Los coeficientes estimados para las experimentos intra-hablantes y entre-hablantes se encuentran en el apéndice D y E, respectivamente.

**2.1.1 Experimento intra-hablante** La estimación de coeficientes de la regresión, para el caso intra-hablante, se realizó con los parámetros obtenidos de 2/3 de los audios y vídeos de MRI de cada hablante del conjunto de datos de este proyecto, esto equivale a 12 audios y 12 vídeos por cada hablante. Estos datos conforman el conjunto de entrenamiento. Las frases utilizadas en este conjunto de datos corresponden a las 60 primeras registradas en el apéndice A. Para el conjunto de validación se tomaron los datos restantes, es decir, los parámetros obtenidos de los 6 audios y 6 vídeos de MRI que no fueron utilizados en el conjunto de entrenamiento. A este conjunto de datos le corresponden las frases 61 a la 90 del apéndice A.

Para cada modelo, los valores del MSE de validación de cada hablante son promediados (tablas 2.1 y 2.2). En los experimentos intra-hablante, los resultados de los hablantes hombres y mujeres, se trabajan y analizan por separado, teniendo en cuenta que las características del tracto vocal y los rangos de frecuencia, difieren respecto al género. Los resultados del error cuadrático medio de validación de la regresión de cada uno de los modelos empleados, para cada hablante, se encuentran en el apéndice C.

**Resultados utilizando frecuencias de formantes.** En las pruebas de regresión intra-hablante en las que se utilizaron como variables predictoras las frecuencias de los formantes halladas con LPC, el modelo que registró el promedio de MSE más bajo en las mujeres fue el cuadrático, con un valor de 0,6426 cm. El MSE de este modelo para los diferentes hablantes mujeres registró valores entre 0,4476-1,2909 cm. El hablante F5 presentó valores de error más altos.

Por otra parte, los hablantes M4 y M5 registraron los valores de MSE más altos en los hablantes hombres. El modelo cuadrático con interacciones y el cuadrático presentan los promedios de MSE mas bajos, con valores de 0,3385 y 0,3386 cm, respectivamente. A su vez, sus rangos de error son de 0,2409-0,4530 cm, para el modelo cuadrático con interacciones, y 0,24-0,4531 cm para el cuadrático.

Teniendo en cuenta estos resultados, se puede concluir que para las pruebas intra-hablante usando frecuencias de formantes calculadas con LPC, el modelo cuadrático es el que mejor se ajusta, tanto en hombres como en mujeres. Aunque en los hombres, el modelo cuadrático con interacciones fue el que presentó menor valor promedio de MSE, la diferencia con el modelo cuadrático es mínima (0,0001 cm), mientras que el número de coeficientes a calcular es casi el doble, siendo 9 para el cuadrático, y 15 para el cuadrático con interacciones. Cabe resaltar que en general, todos los modelos presentaron valores de MSE cercanos. Las figuras 2.1, y 2.2, son una representación gráfica de los valores obtenidos, para efectos de comparación de los modelos.

Método Genero	Error cuadrático medio [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
Mujeres	0,6448	0,6432	0,6457	0,6426	0,6441
Hombres	0,3436	0,3427	0,3385	0,3386	0,3438

Tabla 2.1: Promedio error cuadrático medio de la regresión de cada modelo, utilizando como variables predictoras los formantes calculados con LPC. Experimento intra-hablante.

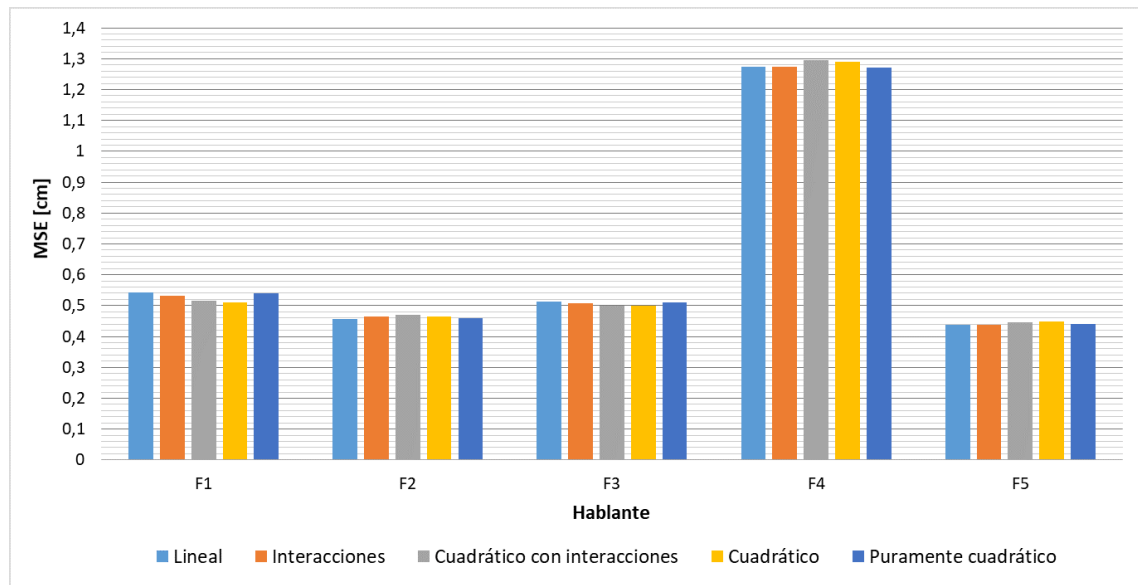


Figura 2.1: Comparación modelos de regresión empleados en hablantes mujeres, utilizando formantes calculados con LPC como variables predictoras.

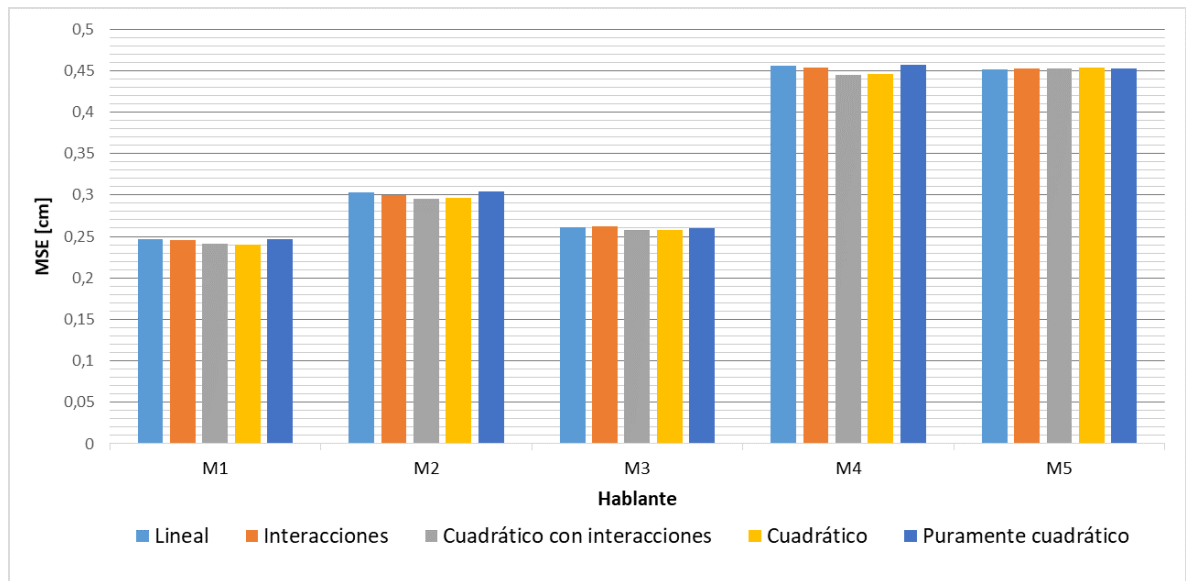


Figura 2.2: Comparación modelos de regresión empleados en hablantes hombres, utilizando formantes calculados con LPC como variables predictoras.

**Resultados utilizando MFCC.** En las pruebas de regresión en las que se utilizaron como variables predictoras los MFCC, el modelo que presentó menor promedio de MSE en los hablantes mujeres, fue el cuadrático, seguido por el lineal, con valores de 0,6289 y 0,6307 cm respectivamente. Los MSE de estos modelos para los diferentes hablantes están en un rango de 0,4261-1,2081 cm, para el cuadrático, y 0,4261-1,2243 cm en el lineal. En esta prueba, el hablante F4 registró los valores de error más altos entre las mujeres.

De otra parte, en los hombres, fueron los hablantes M4 y M5 quienes registraron mayores valores de MSE. Adicionalmente, el modelo que presentó el promedio de MSE más bajo fue el de interacciones (0,3225 cm) junto con el cuadrático (0,3228 cm). Los valores de error cuadrático medio para los diferentes hablantes hombres están entre 0,2399-0,4503 cm, para el modelo interacciones, y 0,2410-0,4413 cm, para el cuadrático. La diferencia entre el promedio de MSE de estos dos modelos es de apenas 0,0003 cm, pero el modelo cuadrático requiere de 6 coeficientes menos que el de interacciones.

En general, se puede decir que para las pruebas intra-hablante, usando los MFCC como variables predictoras, el modelo que mejor se ajusta, tanto en hablantes hombres como en mujeres, es el cuadrático. Sin embargo, los diferentes modelos empleados

presentan valores de error bastante cercanos.

Las figuras 2.3 y 2.4 son una representación gráfica de los valores obtenidos, para efectos de comparación de los modelos.

Método Genero	Error cuadrático medio [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
Mujeres	0,6307	0,6314	0,6341	0,6289	0,6457
Hombres	0,3257	0,3225	0,3234	0,3228	0,3375

Tabla 2.2: Promedio error cuadrático medio de la regresión para cada modelo, utilizando como variables predictoras los coeficientes MFCC. Experimento intra-hablante.

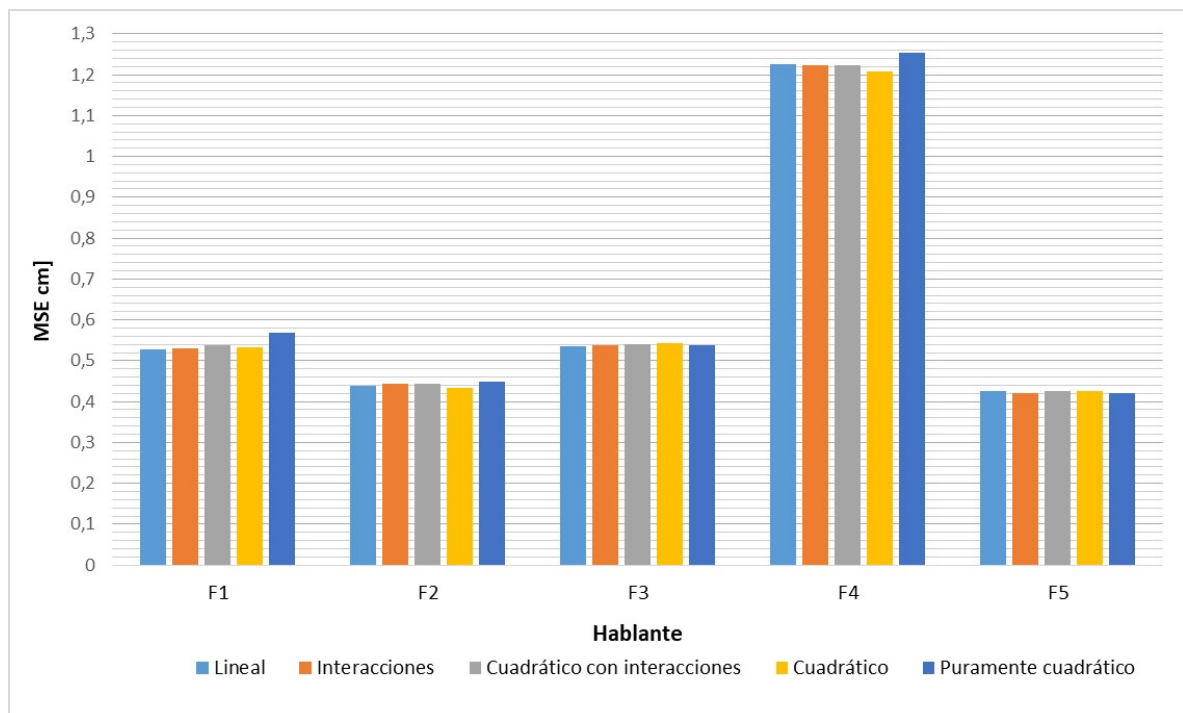


Figura 2.3: Comparación modelos de regresión empleados en hablantes mujeres, utilizando los MFCC como variables predictoras.

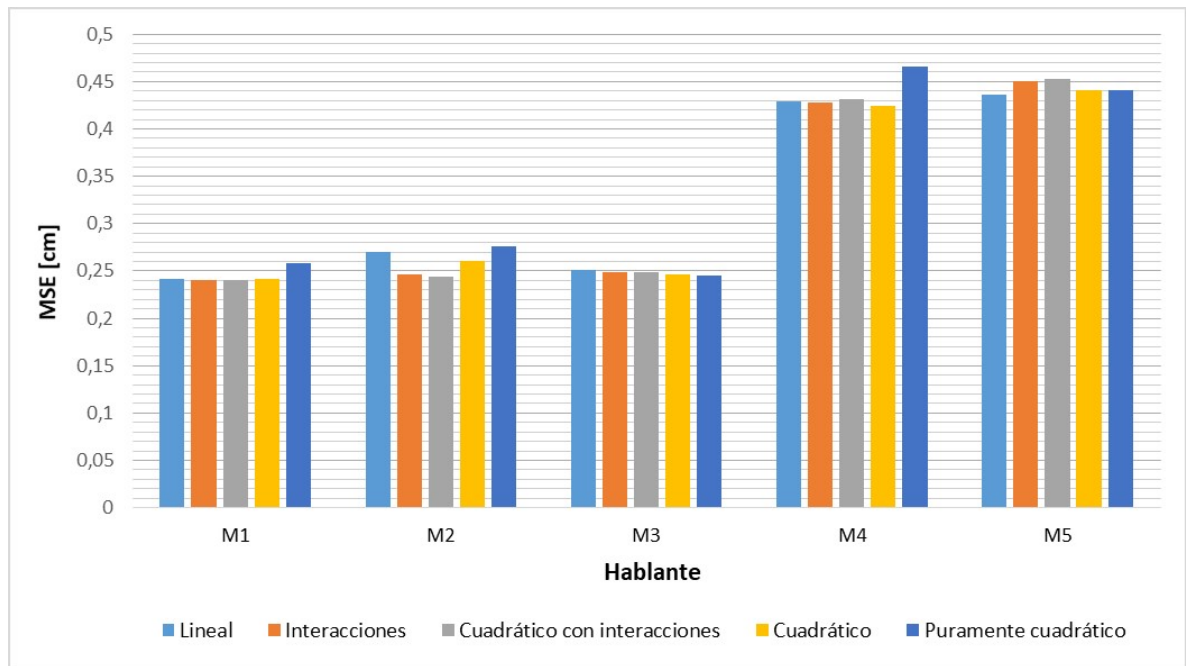


Figura 2.4: Comparación modelos de regresión empleados en hablantes hombres, utilizando los MFCC como variables predictoras.

**Comparación LPC con MFCC.** En las figuras 2.5 y 2.6 se muestra el promedio del error cuadrático medio de validación de los diferentes modelos, en hombres y mujeres; esto para efectos comparativos entre el análisis de regresión realizado tomando como variables predictoras los formantes calculados con LPC, y el realizado con los MFCC.

Tanto en hablantes hombres como en mujeres, los resultados usando MFCC son mejores, exceptuando en el modelo puramente cuadrático en las mujeres, en el que LPC registra un menor promedio de error que MFCC. En promedio, los valores del MSE usando MFCC se ven reducidos en un 1,54 % con respecto a los registrados usando LPC, para los hablantes mujeres. En los hombres disminuyen en un 4,42 %. Aunque hay una reducción del error cuadrático medio, se debe tener en cuenta que al utilizar los MFCC se necesitan estimar muchos más parámetros en cada modelo, dado que se tienen 12 variables predictoras, mientras que con LPC solo se tienen 4.

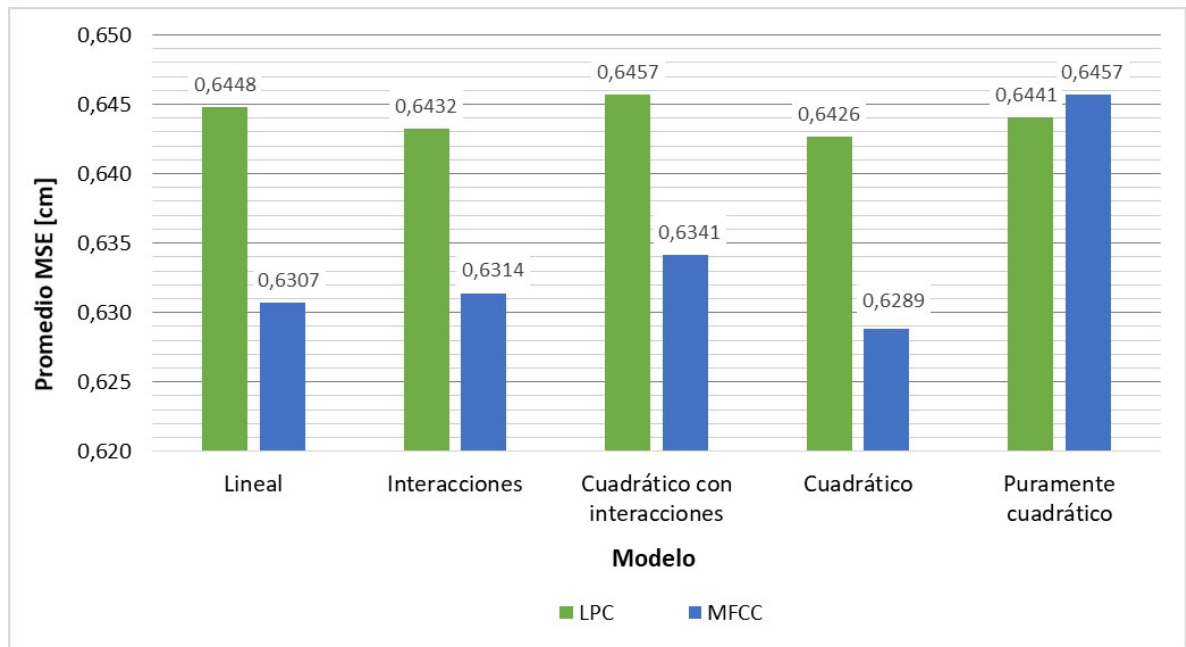


Figura 2.5: Comparación de modelos de regresión usando como variables predictoras los formantes calculados con LPC, con los modelos usando los MFCC. Experimento intra-hablante, en mujeres.

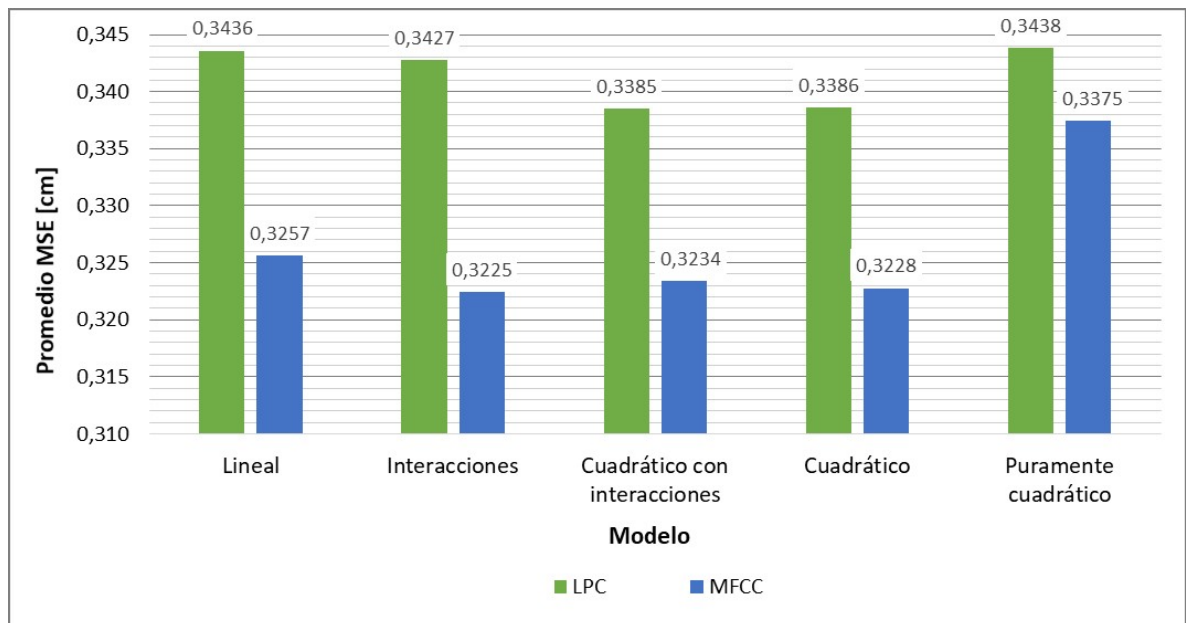


Figura 2.6: Comparación de modelos de regresión usando como variables predictoras los formantes calculados con LPC, con los modelos usando los MFCC. Experimento intra-hablante, en hombres.

**2.1.2 Experimento entre-hablantes** Para el experimento entre-hablantes se tienen 2 casos, para uno se trabajan los datos de hombres y mujeres por separado, de esta forma, para cada uno de estos grupos, el conjunto de entrenamiento está conformado con los datos de 4 de los hablantes, y el conjunto de validación está conformado con los datos del hablante restante. El hablante utilizado para la validación se va rotando, realizándose en total 5 análisis de regresión por cada grupo de hablantes (hombres y mujeres). En los dos grupos, los MSE de validación de los análisis de regresión de cada modelo se promedian.

En el otro caso, se unen los datos de hablantes hombres y mujeres; el conjunto de entrenamiento está conformado por 4 hablantes hombres y 4 mujeres, mientras que el conjunto de validación lo va a conformar por un lado, el hablante mujer que no hace parte del conjunto de entrenamiento, y en otro, el hablante hombre que tampoco hace parte de este conjunto. De esta forma se realizan en total 10 análisis de regresión por cada modelo, y se promedian los valores de MSE de validación.

Al realizar los experimentos entre-hablantes, se presentan algunos errores de validación considerablemente altos, para ciertos hablantes, esto probablemente se deba a que ninguno de los datos del hablante utilizado para la validación, hacen parte de conjunto de entrenamiento y, teniendo en cuenta que cada hablante tiene características diferentes, así sean del mismo género, algunos parámetros no se ajustan bien al modelo.

**Resultados utilizando formantes calculados con LPC.** En las pruebas entre-hablantes en las que se utilizaron los formantes como predictores, el modelo interacciones presentó el menor MSE promedio, entre las mujeres, con un valor de 1,4241 cm; mientras que para los hombres, y para el caso de hablantes hombres y mujeres juntos, el modelo lineal presentó menor promedio de MSE, con valores de 1,1054 y 1,6283 cm, respectivamente. Cabe resaltar que todos los modelos presentaron valores de error cercanos, con diferencias de no más de 0,15 cm para el caso de las mujeres, 0,09 cm para el caso de los hombres, y 0,04 cm para el caso de hombres y mujeres juntos. Los resultados del promedio del MSE de cada modelo para estas pruebas se encuentran en la tabla 2.3.

Método Género		Error cuadrático medio [cm]				
		Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
Mujeres		1,4487	1,4241	1,5599	1,5724	1,4532
Hombres		1,1054	1,1362	1,1983	1,1617	1,1057
Ambos		1,6283	1,6373	1,6713	1,6587	1,6303

Tabla 2.3: Promedio error cuadrático medio de cada modelo, utilizando formantes como variables predictoras. Experimento entre-hablantes.

**Resultados utilizando MFCC.** Para el experimento entre-hablantes, usando los MFCC como variables predictoras, el modelo que presentó el menor promedio de MSE en las mujeres fue el cuadrático con interacciones, seguido por el de interacciones, con valores de 1,2568 y 1,2578 cm respectivamente. En los hombres, el modelo que registró el menor valor promedio de error fue el puramente cuadrático, con un valor de 0,5041 cm; mientras que para el caso de hombres y mujeres juntos, fue el modelo interacciones, con un valor de 1,1877 cm. Las diferencias mas grandes entre los errores de los modelos son de 0,47 cm para el caso de las mujeres, 0,18 cm para el caso de los hombres, y 0,30 cm para el caso de hombres y mujeres juntos. Los resultados del promedio del MSE para estas pruebas se encuentran en la tabla 2.4.

Método Género		Error cuadrático medio [cm]				
		Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
Mujeres		1,3341	1,2578	1,2568	1,3439	1,7259
Hombres		0,5938	0,5590	0,6194	0,6823	0,5041
Ambos		1,3410	1,1877	1,3420	1,4897	1,4401

Tabla 2.4: Promedio error cuadrático medio de cada modelo, utilizando los MFCC como variables predictoras. Experimento entre-hablantes.

**Comparación LPC con MFCC.** En términos del error cuadrático medio, se observa una mejora al utilizar los MFCC como predictores, esto aplica para todos los modelos, tanto para el caso de hombres y mujeres por separado, como juntos, exceptuando el modelo puramente cuadrático en el caso de las mujeres, dónde el análisis realizado

usando LPC presenta un menor error que con MFCC. En las mujeres la reducción del error cuadrático medio al utilizar MFCC es de aproximadamente 14,47 %, con respecto a LPC; en los hombres la reducción es del 48,21 %, y para el caso de hombres y mujeres juntos es de 17,33 %. Nuevamente cabe resaltar, que aunque se presenten mejoras en términos del error cuadrático medio, al utilizar MFCC los modelos requieren de un mayor número de coeficientes a estimar, que al utilizar LPC.

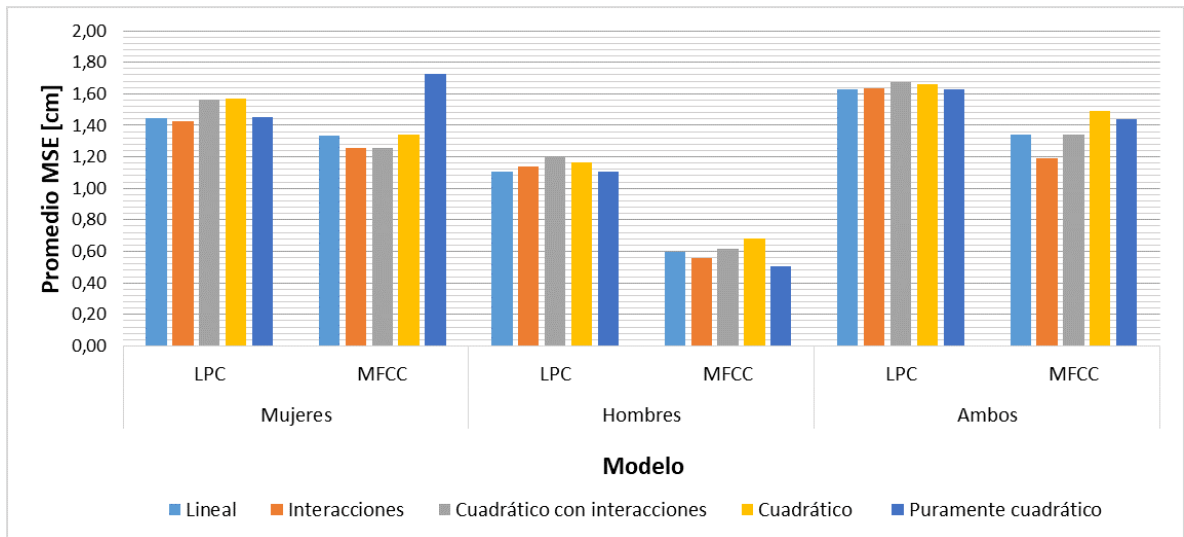


Figura 2.7: Comparación de modelos de regresión usando como variables predictoras los formantes calculados con LPC, con los modelos usando los MFCC. Experimento entre-hablantes.

## 2.2. Resultados pruebas *k-fold*

La validación cruzada de  $k$ -iteraciones se realizó en MATLAB, haciendo uso de la función *cvpartition*, la cual divide las observaciones del conjunto de datos en los  $k$  subgrupos (20) de manera aleatoria. Ya teniendo los subgrupos se continúa con el procedimiento de *k-fold* descrito en la sección 1.8, realizando el ajuste del modelo mediante el criterio de mínimos cuadrados de la ecuación 1.15. Como variables predictoras se usaron los formantes y los coeficientes de MFCC. Esto se hizo con cada uno de los modelos planteados, tanto para el caso intra-hablante, como para el entre-hablantes.

**2.2.1 Experimento intra-hablante** Para el experimento intra-hablante, se realizó una prueba de *k-fold* por cada hablante, utilizando el total de los datos de cada uno, después de haber eliminado los *outliers*. Los errores de *k-fold* obtenidos utilizando como variables predictoras tanto los formantes calculados con LPC, como los MFCC, se muestran en las tablas 2.5, 2.6, 2.7 y 2.8. La comparación de estos errores se observa en las figuras 2.8, 2.9, 2.10 y 2.11. En los resultados se diferencian los hablantes hombres de las mujeres.

**Resultados usando formantes calculados con LPC.** En las mujeres, el modelo cuadrático con interacciones registra los errores de *k-fold* más pequeños, exceptuando para el hablante F2, sin embargo la diferencia del valor obtenido con este modelo en comparación con el que presentó menor error, es muy pequeña (0,0028 cm). El error de *k-fold* del modelo cuadrático con interacciones está entre 0,3715-0,5234 cm. F4 continúa registrando errores de *k-fold* más altos entre los hablantes mujeres.

En los hombres, los errores de *k-fold* más pequeños los registran los modelos cuadrático y cuadrático con interacciones, con valores entre 0,2334-0,4509 cm y 0,2348-0,4505 cm, respectivamente. Los hablantes M4 y M5 son los que presentan errores de *k-fold* más altos.

En general, se puede concluir que el modelo cuadrático con interacciones fue el que mejores resultados presentó en estas pruebas, en términos del error cuadrático medio, sin embargo, este modelo tiene la desventaja de presentar el mayor número de coeficientes a estimar, mientras que los demás modelos, con un menor número de coeficientes, registran errores igualmente válidos y cercanos entre sí. Por poner un caso, el modelo cuadrático presenta el menor promedio de error de *k-fold*, después del cuadrático con interacciones, con una diferencia de apenas 0,0003 cm en las mujeres, y 0,0009 cm en los hombres, y emplea 6 coeficientes menos que el modelo cuadrático con interacciones.

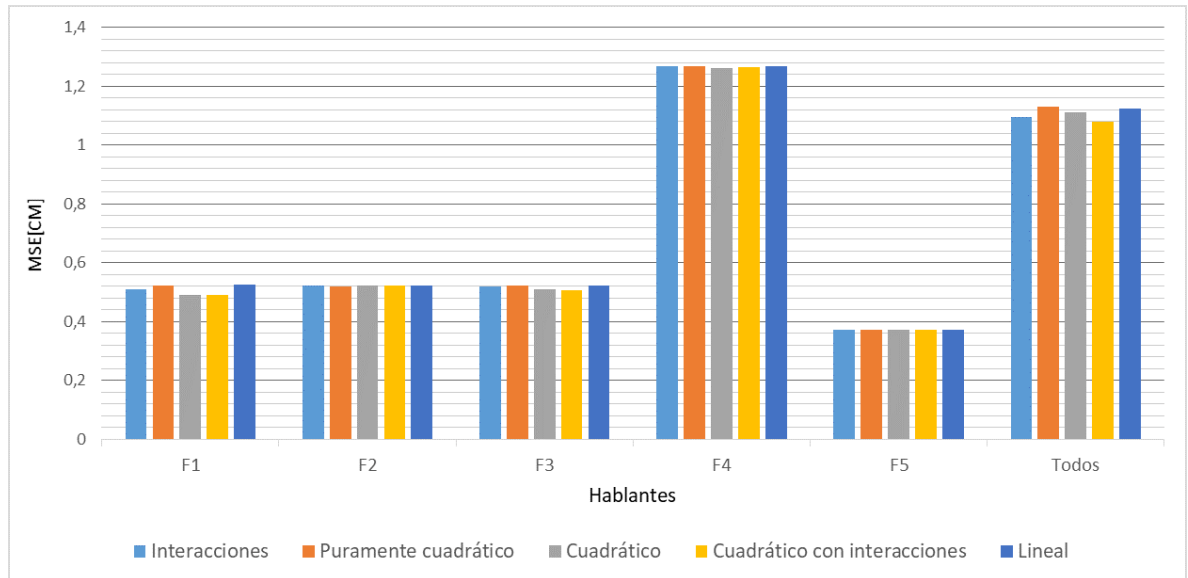


Figura 2.8: Comparación error de  $k$ -fold en hablantes mujeres, utilizando formantes calculados con LPC como variables predictoras.

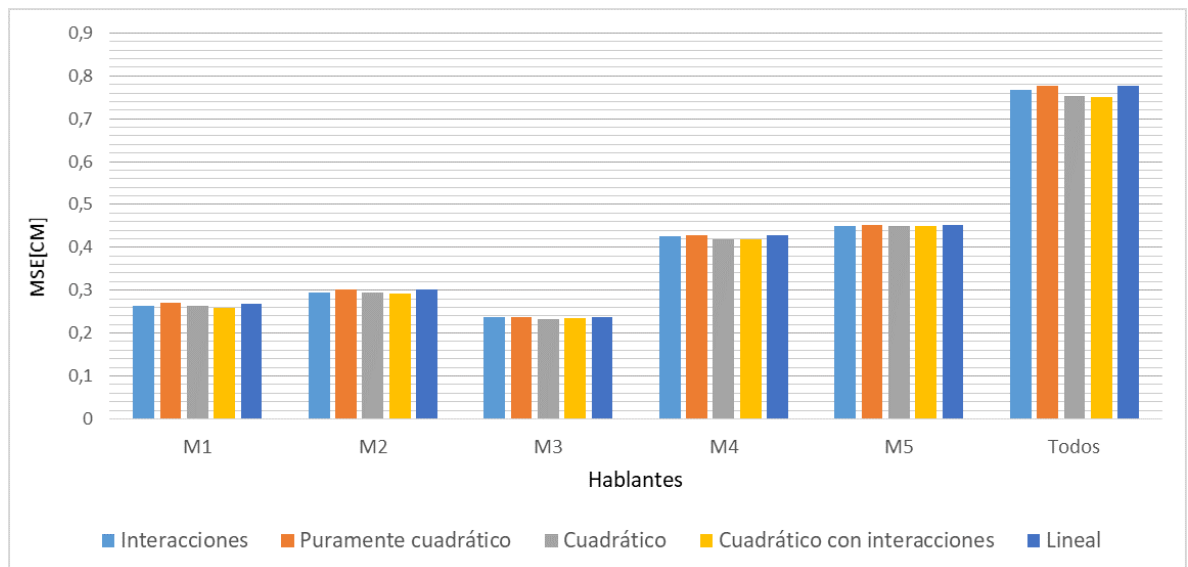


Figura 2.9: Comparación error de  $k$ -fold en hablantes hombres, utilizando formantes calculados con LPC como variables predictoras.

Método Hablaante	Error cuadrático medio [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
F1	0,5244	0,5099	0,4903	0,4906	0,5237
F2	0,5215	0,5231	0,5234	0,5233	0,5206
F3	0,5229	0,5196	0,5077	0,5086	0,5231
F4	1,2686	1,2672	1,2624	1,2626	1,2686
F5	0,3720	0,3722	0,3715	0,3719	0,3721

Tabla 2.5: Error de *k-fold* de cada modelo, utilizando los formantes hallados con LPC como variables predictoras. Experimento intra-hablante, en mujeres.

Método Hablaante	Error cuadrático medio [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
M1	0,2693	0,2640	0,2595	0,2623	0,2700
M2	0,3023	0,2951	0,2922	0,2953	0,3025
M3	0,2375	0,2380	0,2348	0,2334	0,2370
M4	0,4279	0,4264	0,4185	0,4180	0,4286
M5	0,4516	0,4508	,4505	0,4509	0,4519

Tabla 2.6: Error de *k-fold* de cada modelo, utilizando formantes hallados con LPC como predictores. Experimento intra-hablante, en hombres.

**Resultados utilizando MFCC.** Al realizar la prueba de *k-fold* utilizando los coeficientes de MFCC como predictores, tanto los hablantes hombres como mujeres presentaron un menor error de *k-fold* con el modelo cuadrático con interacciones. En las mujeres, el error está entre 0,3443 y 0,5013 cm mientras que para los hombres está entre 0,2392 y 0,3958 cm. Sin embargo, los valores de error de *k-fold* entre los modelos, son bastante cercanos, y requieren de menos coeficientes que el modelo cuadrático con interacciones.

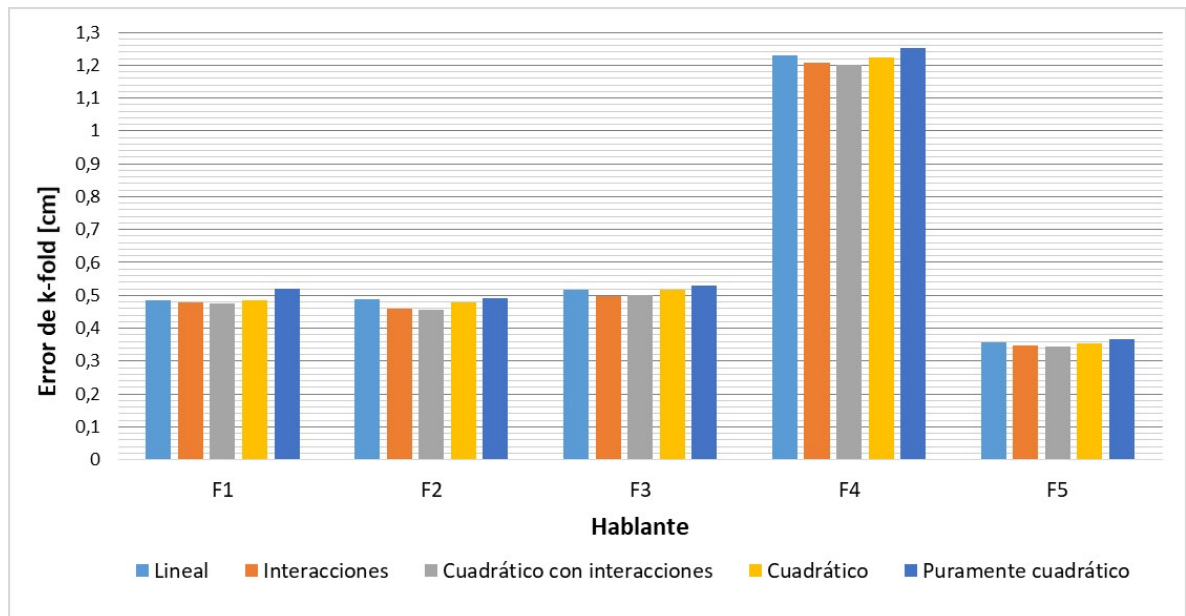


Figura 2.10: Comparación error de  $k$ -fold en hablantes mujeres, utilizando MFCC como variables predictoras.

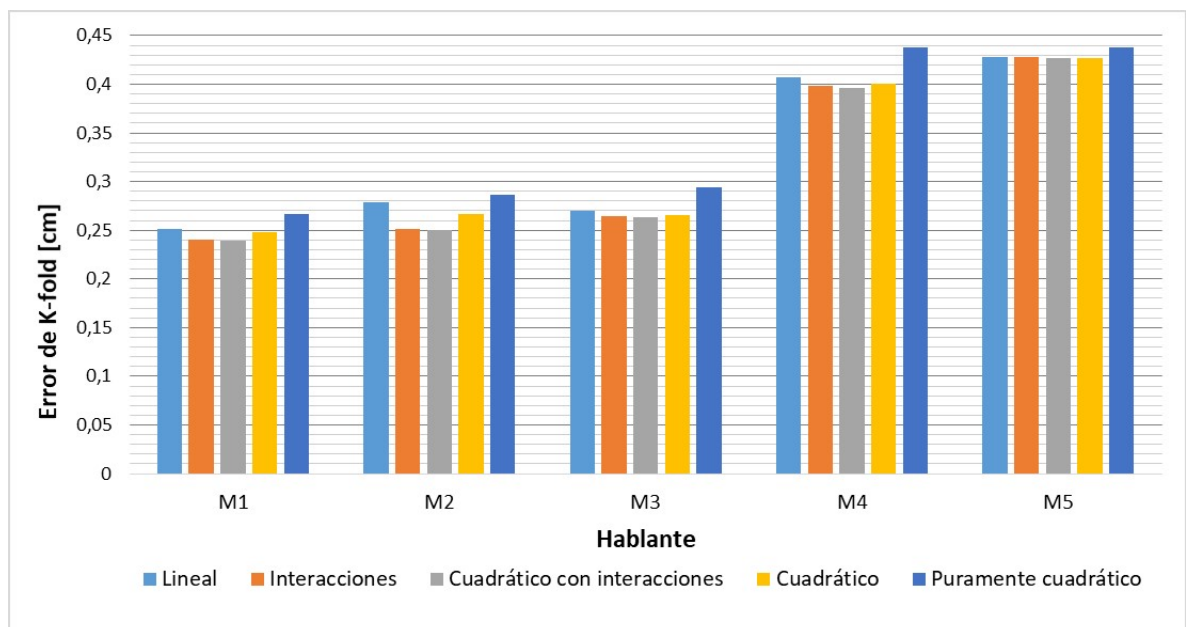


Figura 2.11: Comparación error de  $k$ -fold en hablantes hombres, utilizando MFCC como variables predictoras.

Método Hablaante	Error cuadrático medio [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
F1	0,4841	0,4787	0,4766	0,4845	0,5214
F2	0,4894	0,4594	0,4575	0,4776	0,4911
F3	0,5182	0,4981	0,5013	0,5153	0,5296
F4	1,2308	1,2067	1,2004	1,2226	1,2524
F5	0,3584	0,3461	0,3443	0,3533	0,3681

Tabla 2.7: Error de *k-fold* de cada modelo, utilizando MFCC como predictores. Experimento intra-hablante, en mujeres.

Método Hablaante	Error cuadrático medio [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
M1	0,2513	0,2406	0,2392	0,2480	0,2670
M2	0,2788	0,2516	0,2499	0,2668	0,2861
M3	0,2699	0,2648	0,2636	0,2651	0,2936
M4	0,4068	0,3980	0,3958	0,4010	0,4375
M5	0,4283	0,4275	0,4271	0,4272	0,4380

Tabla 2.8: Error de *k-fold* de cada modelo, utilizando MFCC como predictores. Experimento intra-hablante, en hombres.

**Comparación LPC con MFCC** Para efectos de comparación entre los resultados obtenidos utilizando como predictores los formantes calculados con LPC, y los MFCC, se promediaron los valores de error de *k-fold* de los hablantes para cada modelo, manteniendo la diferenciación de género. Los resultados se pueden ver en las gráficas 2.12 y 2.13.

Se puede observar que tanto en los hablantes hombres como en mujeres, los valores obtenidos con LPC y MFCC son bastante cercanos, presentando mejores resultados MFCC. En los hablantes mujeres, el modelo de interacciones presenta la diferencia más grande entre estos dos métodos, con una disminución del promedio del error de *k-fold* del 6,36 % utilizando MFCC, respecto a LPC. Para el caso cuadrático con interacciones la disminución es del 5,56 %, para el lineal del 4,01 % y para el puramente cuadrático del 1,42 %.

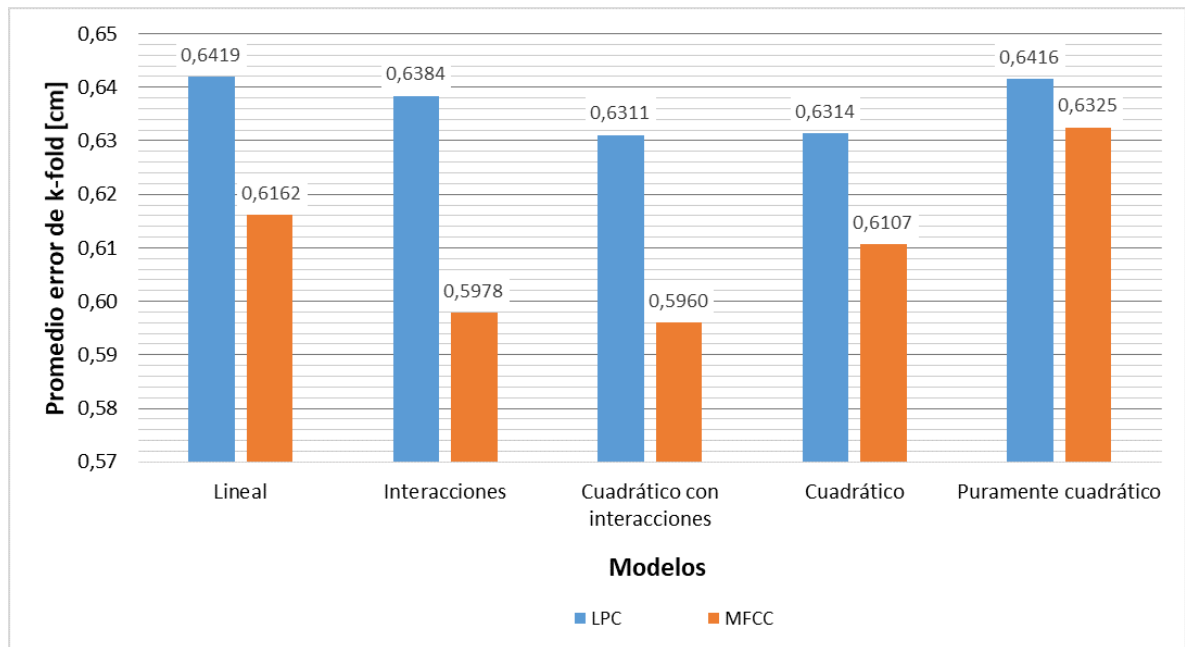


Figura 2.12: Comparación error de  $k$ -fold entre las pruebas realizadas usando como predictores los formantes calculados con LPC, y los MFCC. Experimento intra-hablante, en mujeres.

En los hablantes hombres, el modelo puramente cuadrático es el único en el que LPC presenta errores de  $k$ -fold más pequeños que utilizando MFCC, siendo el incremento del 1,91 % utilizando MFCC con respecto a LPC. En los demás modelos, se mantiene la mejoría de MFCC con respecto a LPC, siendo la disminución promedio del error de  $k$ -fold de 3,17 % para el lineal, 5,48 % para el de interacciones, 4,83 % para el cuadrático con interacciones, y 3,12 % para el cuadrático. A pesar de la reducción del error de  $k$ -fold, se debe tener presente que se necesitan estimar más coeficientes para cada modelo, esto sumado al número de iteraciones que se realizan con  $k$ -fold, hacen que los análisis y pruebas realizados con MFCC sean más complejos.

**2.2.2 Experimento entre-hablantes** Para el experimento entre-hablantes, se realizó una prueba de  $k$ -fold con los datos de todos los hablantes mujeres, y una con los datos de todos los hombres. También se realiza una prueba de  $k$ -fold uniendo el total de los datos de hombres y mujeres.

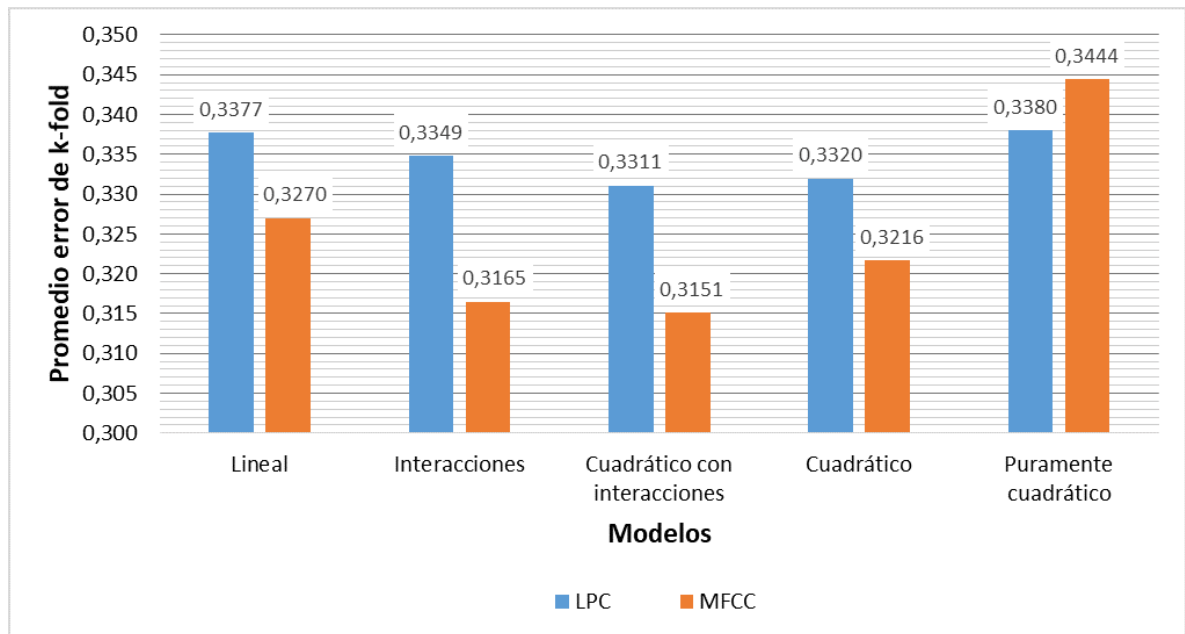


Figura 2.13: Comparación error de  $k$ -fold entre las pruebas realizadas usando como predictores los formantes calculados con LPC, y los MFCC. Experimento intra-hablante, en hombres.

**Resultados utilizando formantes calculados con LPC.** Para el experimento entre-hablantes, en el que se utilizaron formantes como variables predictoras, el modelo cuadrático con interacciones presentó los mejores resultados en cuanto al error de  $k$ -fold, tanto en hombres como en mujeres, registrando valores de 0,7493 cm y 1,0801 cm, respectivamente. Para el caso en el que se utilizaron los datos de hombres y mujeres juntos, este modelo también registró el error de  $k$ -fold más bajo, siendo este de 1,3409 cm. Cabe resaltar que todos los modelos presentan errores de  $k$ -fold cercanos entre sí, para cada caso.

Al realizar la prueba de  $k$ -fold con los datos de hombres y mujeres juntos, el error de las pruebas de  $k$ -fold se ve incrementado en un promedio de 79,40 % con respecto a los resultados de solo los hablantes hombres, y con respecto los resultados de solo hablantes mujeres, presenta un incremento de 23,86 %. Estos resultados se encuentran en la tabla 2.9.

Método Genero	Error de k-fold [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
Mujeres	1,1233	1,0936	1,0801	1,1116	1,1288
Hombres	0,7765	0,7683	0,7493	0,7517	0,7775
Ambos	1,3936	1,3538	1,3409	1,3674	1,4030

Tabla 2.9: Error de *k-fold* para experimento entre-hablantes, usando formantes como variables predictoras

**Resultados utilizando MFCC.** Utilizando los MFCC como predictores, para hombres y mujeres por separado, así como con los datos de estos dos grupos juntos, el modelo cuadrático con interacciones presentó nuevamente un menor error de *k-fold*, con valores de 0,6241, 0,8144 y 1,0908 cm, respectivamente. El modelo interacciones presentó errores cercanos a los generados con el modelo cuadrático con interacciones. Con una diferencia de 0,0084 cm en las mujeres, 0,0079 cm en los hombres, y 0,0256 cm en hombres y mujeres juntos; además de necesitar de menos coeficientes.

Al usar los datos de los hablantes hombres y mujeres juntos, se incrementa el error de *k-fold* con respecto al generado con cada grupo por separado. Este incremento en promedio es de 44,88 % con respecto a las mujeres y de 89,12 % con respecto a los hombres. Los resultados del error de *k-fold* en esta prueba se encuentran en la tabla 2.10.

Método Hablaante	Error de k-fold [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
Mujeres	0,8839	0,8228	0,8144	0,8727	1,0409
Hombres	0,6928	0,6320	0,6241	0,6798	0,7660
Ambos	1,3462	1,1165	1,0908	1,2933	1,6057

Tabla 2.10: Error de *k-fold* para experimento entre-hablantes, usando coeficientes de MFCC como predictores

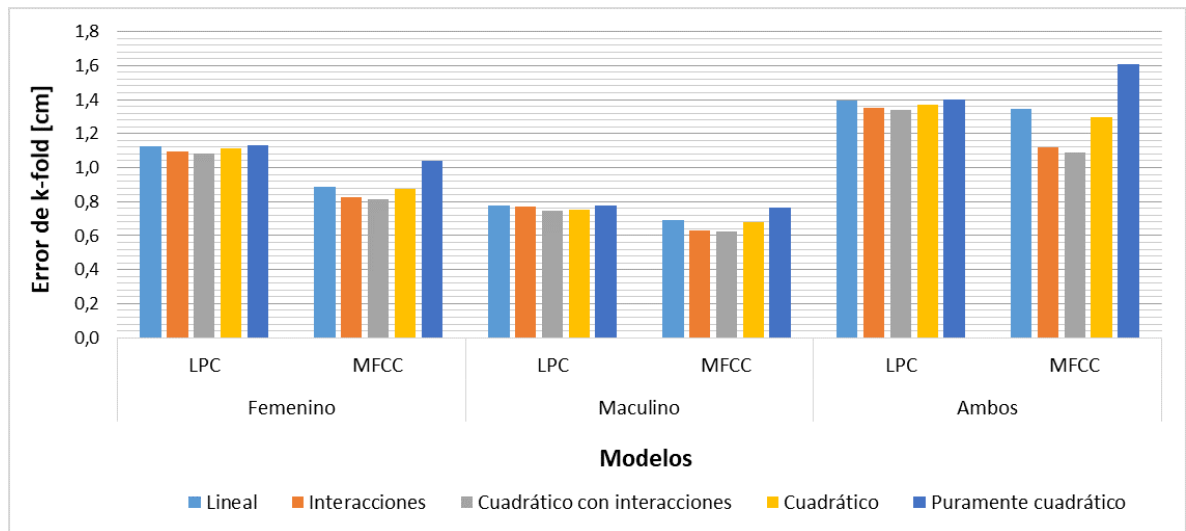


Figura 2.14: Comparación error de  $k$ -fold entre las pruebas realizadas usando como predictores los formantes calculados con LPC, y los MFCC. Experimento entre-hablantes.

**Comparación LPC con MFCC.** Se observa una disminución del error de  $k$ -fold al utilizar los MFCC como variables predictoras, para todos los casos, exceptuando particularmente el modelo puramente cuadrático para el caso de hablantes hombres y mujeres juntos. En las mujeres, la reducción del error de  $k$ -fold, en promedio, es de 19,99 % con respecto a LPC, en los hombres de 11,26 % , y para el caso de hombres y mujeres juntos es de 6,11 %. Nuevamente cabe resaltar que al utilizar MFCC se tiene una mayor complejidad computacional, debido al número de coeficientes que se deben estimar en cada modelo, junto con el número de iteraciones que se realizan en las pruebas de  $k$ -fold.

En la figura 2.14 se muestra la comparación entre los resultados obtenidos utilizando LPC, y los obtenidos con MFCC.

## *CONCLUSIONES*

En el presente proyecto de grado se evaluaron diferentes modelos estadísticos de tipo lineal (5 en total) para la estimación de la longitud del tracto vocal, a partir de parámetros obtenidos de la señal de la voz, para lo cual se realizaron diferentes pruebas y formas de evaluación, partiendo del análisis de regresión.

Se optó por realizar pruebas de ajuste y validación separando los hablantes por género, al igual que tomándolos juntos, teniendo en cuenta que las características morfológicas del tracto vocal difieren entre hombres y mujeres. Al realizar las pruebas esto se ve evidenciado, puesto que se obtienen mejores resultados analizando los datos en grupos separados por género.

En general, los diferentes modelos reportaron valores de MSE aceptables, y bastante cercanos entre sí, por lo que la principal diferencia radica en la complejidad de los modelos, y el número de coeficientes a estimar, dado que no conviene utilizar un modelo que tenga el doble, o más, coeficientes que otro, cuando su disminución del error cuadrático medio sea apenas mínima.

La representación de parámetros de la voz se hizo con formantes (calculados usando LPC) y MFCCs, teniendo en cuenta que estos dos métodos han sido ampliamente usados en procesos de análisis y reconocimiento de voz; además, estos parámetros presentan facilidad para su implementación, pudiéndose realizar de forma semi-automática con programas como *Wavesufer* y MATLAB. En general, al usar MFCCs, los resultados son consistentemente mejores que al usar los formantes, pero se tiene un costo computacional más alto debido al número de variables predictoras del tipo MFCC (12), que por ende hacen que el número de coeficientes a estimar para cada modelo sea considerablemente mayor. Esto se ve reflejado especialmente en los modelos interacciones

y cuadrático con interacciones, para los cuales se necesita estimar 79 y 91 coeficientes respectivamente, mientras que al usar formantes se emplean solo 11 y 15 coeficientes. Sin embargo, desde el punto de vista de representación de la señal de la voz, es más recomendable usar MFCC, dado que computacionalmente son más fáciles de hallar, y presentan mayor robustez.

En particular, al utilizar diferentes hablantes para la etapa de entrenamiento y validación (experimento entre-hablantes), y emplear los MFCC, se obtienen valores de MSE de 1,3341 cm y 0,5938 cm en mujeres y hombres respectivamente, para el modelo lineal. Para el caso en el que se utiliza el modelo interacciones, se obtienen valores de 1,2578 cm en las mujeres, y 0,5590 cm en los hombres. Al utilizar el modelo cuadrático con interacciones los valores de MSE son 1,2568 cm en las mujeres, y 0,6194 cm en los hombres. Para el caso del modelo cuadrático, el MSE en las mujeres es de 1,3439 cm, y en los hombres de 0,6823 cm. Y, en el caso del modelo puramente cuadrático, el MSE tiene un valor de 1,7259 cm para las mujeres, y 0,5041 cm para los hombres.

En contraste, utilizando formantes como variables predictoras, se obtienen valores de MSE de 1,4487 cm y 1,1054 cm, para mujeres y hombres respectivamente, al utilizar el modelo lineal. En el modelo interacciones se presentan errores de 1,4241 cm en las mujeres, y 1,1362 cm en los hombres. Para el caso del modelo cuadrático con interacciones, se obtienen valores de 1,5599 cm en las mujeres y 1,1983 en los hombres. Para el caso del modelo cuadrático el MSE es de 1,5724 cm en mujeres, y 1,1617 cm en hombres. Por último, para el caso en el que se utiliza el modelo puramente cuadrático, se tienen errores de 1,4532 cm en mujeres y 1,1057 cm en hombres.

Estos resultados son más confiables que aquellos reportados en trabajos tales como [32], en el cual se utilizan 5 hablantes (3 mujeres y 2 hombres), y solo 5 frases por cada uno de ellos, para desarrollar y evaluar el modelo propuesto en su investigación. En tanto, en el presente trabajo se utilizan 10 hablantes (5 hombres y 5 mujeres) y 90 frases por cada hablante. Cabe resaltar que la extracción de parámetros para esta cantidad de recursos, fue un proceso que implicó un cantidad significativa de tiempo, dado que los programas empleados realizan los procesos de forma semi-automática, y fue necesario modificarlos y ajustarlos para los procesos necesarios a desarrollar en este proyecto.

## *TRABAJOS FUTUROS*

Los modelos utilizados en el presente proyecto son de tipo lineal, y son ajustados a partir de datos provenientes de 10 hablantes. Incluir más hablantes es una tarea difícil a corto plazo teniendo en cuenta lo costoso que es obtener vídeos de resonancia magnética del tracto vocal. Sin embargo, a modo de trabajo a corto plazo se plantea el probar con modelos no lineales, y analizar modelos del tracto vocal más complejos, que involucren también otros parámetros. Además, se plantea analizar simultáneamente la relación que tiene la longitud del tracto vocal, con características de la persona, como la estatura.

# APÉNDICES

## *Apéndice A*

### *Frases de la base de datos*

A continuación se muestran las frases tomadas de la base de datos del MOCHA-TIMIT que fueron utilizadas en el desarrollo del proyecto.

Número	Frase
1	This was easy for us.
2	Jane may earn more money by working hard.
3	She is thinner than I am.
4	Bright sunshine shimmers on the ocean.
5	Nothing is as offensive as innocence.
6	Why yell or worry over silly items.
7	Where were you while we were away?
8	Are your grades higher or lower than Nancy's.
9	He will allow a rare lie.
10	Will robin wear a yellow lily.
11	Swing your arm as high as you can.
12	Before Thursday's exam review every formula.
13	The museum hires musicians every evening.
14	A roll of wire lay near the wall.
15	Carl lives in a lively home.
16	Alimony harms a divorced man's wealth.
17	She wore warm fleecy woolen overalls.
18	Alfalfa is healthy for you.
19	When all else fails use force.

Sigue en la página siguiente.

Número	Frase
20	Although always alone we survive.
21	Only lawyers love millionaires.
22	Did dad do academic bidding?
23	Help greg to pick a peck of potatoes.
24	A good attitude is unbeatable.
25	Coconut cream pie makes a nice dessert.
26	Don't do Charlie's dirty dishes.
27	Help celebrate your brother's success.
28	Only the most accomplished artists obtain popularity.
29	Critical equipment needs proper maintenance.
30	Young people participate in athletic activities.
31	Barb's gold bracelet was a graduation present.
32	Stimulating discussions keep students' attention.
33	Etiquette mandates compliance with existing regulations.
34	Biblical scholars argue history.
35	Elderly people are often excluded.
36	Basketball can be an entertaining sport.
37	Addition and subtraction are learned skills.
38	Grandmother outgrew her upbringing in petticoats.
39	At twilight on the twelfth day we'll have chablis.
40	Catastrophic economic cutbacks neglect the poor.
41	Ambidextrous pickpockets accomplish more.
42	Even a simple vocabulary contains symbols.
43	The eastern coast is a place for pure pleasure and excitement.
44	The lack of heat compounded the tenant's grievances.
45	Academic aptitude guarantees your diploma.
46	The prowler wore a ski mask for disguise.
47	We experience distress and frustration obtaining our degrees.
48	The singer's finger had a splinter.
49	The legislature met to judge the state of public education.
50	Chocolate and roses never fail as a romantic gift.
51	Any contributions will be greatly appreciated.

Segue en la página siguiente.

Número	Frase
52	Continental drift is a geological theory.
53	Regular attendance is seldom required.
54	Challenge each general's intelligence.
55	We got drenched from the uninterrupted rain.
56	Last year's gas shortage caused steep price increases.
57	Upgrade your status to reflect your wealth.
58	Eat your raisins outdoors on the porch steps.
59	Porcupines resemble sea urchins.
60	Spring street is straight ahead.
61	Cliff's display was misplaced on the screen.
62	An official deadline cannot be postponed.
63	Fill that canteen with fresh spring water.
64	Gently place Jim's foam sculpture in the box.
65	Bagpipes and bongos are musical instruments.
66	Doctors prescribe drugs too freely.
67	Will you please describe the idiotic predicament.
68	It's impossible to deal with bureaucracy.
69	Good service should be rewarded by big tips.
70	My instructions desperately need updating.
71	Cooperation along with understanding alleviate dispute.
72	Cement is measured in cubic yards.
73	Primitive tribes have an upbeat attitude.
74	Flying standby can be practical if you want to save money.
75	It's hard to tell an original from a forgery.
76	'The thinker' is a famous sculpture.
77	The misprint provoked an immediate disclaimer.
78	A large household needs lots of appliances.
79	Cut a small corner off each edge.
80	Iguanas and alligators are tropical reptiles.
81	Masquerade parties tax one's imagination.
82	Penguins live near the icy antarctic.
83	Guess the question from the answer.

Sigue en la página siguiente.

Número	Frase
84	Medieval society was based on hierarchies.
85	Project development was proceeding too slowly.
86	Kindergarten children decorate their classrooms for all holidays.
87	Special task forces rescue hostages from kidnappers.
88	Call an ambulance for medical assistance.
89	He stole a dime from a beggar.
90	You must explicitly delete files.

Tabla A.1: Frases de la base de datos MOCHA-TIMIT utilizadas en el proyecto

## *Apéndice B*

### *Banco de filtros de mel*

Se muestra el los límites de frecuencia de los filtros de mel, utilizados para hallar los coeficientes cepstrales en las frecuencias de mel.

Filtro No.	Límites del filtro [Hz]	Filtro No.	Límites del filtro [Hz]
1	[133 267]	21	[1618 1856]
2	[200 333]	22	[1733 1988]
3	[267 400]	23	[1856 2130]
4	[333 467]	24	[1988 2281]
5	[400 533]	25	[2130 2444]
6	[467 600]	26	[2281 2618]
7	[533 667]	27	[2444 2804]
8	[600 733]	28	[2618 3004]
9	[667 800]	29	[2804 3217]
10	[733 867]	30	[3004 3446]
11	[800 933]	31	[3217 3692]
12	[867 999]	32	[3446 3954]
13	[933 1071]	33	[3692 4236]
14	[999 1147]	34	[3954 4537]
15	[1071 1229]	35	[4236 4860]
16	[1147 1316]	36	[4537 5206]
17	[1229 1410]	37	[4860 5577]
18	[1316 1510]	38	[5206 5973]
19	[1410 1618]	39	[5577 6399]
20	[1510 1733]	40	[5973 6854]

Tabla B.1: Banco de filtros MFCC

## *Apéndice C*

### *Resultados error cuadrático medio para análisis de regresión*

Se muestran los resultados del error cuadrático medio obtenido al realizar el análisis de regresión con cada uno de los modelos planteados, utilizando tanto los formantes calculados con LPC (tablas C.1 y C.2), como los MFCC (tablas C.3 y C.4), como variables predictoras. Esto para el experimento intrahablante.

Método Hablante	Error cuadrático medio [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
F1	0,5414	0,5315	0,5161	0,5100	0,5399
F2	0,4575	0,4642	0,4710	0,4658	0,4589
F3	0,5125	0,5084	0,4999	0,4989	0,5091
F4	1,2737	1,2744	1,2952	1,2909	1,2725
F5	0,4389	0,4375	0,4461	0,4476	0,4399

Tabla C.1: Error cuadrático medio de validación de cada modelo utilizado, para los hablantes mujeres, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante.

Método Hablante	Error cuadrático medio [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
M1	0,2464	0,2457	0,2409	0,2400	0,2466
M2	0,3032	0,2996	0,2955	0,2962	0,3040
M3	0,2610	0,2626	0,2582	0,2579	0,2598
M4	0,4563	0,4537	0,4449	0,4458	0,4565
M5	0,4510	0,4521	0,4530	0,4531	0,4522

Tabla C.2: Error cuadrático medio de validación de cada modelo utilizado, para los hablantes hombres, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante.

Método Hablante	Error cuadrático medio [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
F1	0,5276	0,5311	0,5371	0,5325	0,5690
F2	0,4394	0,4434	0,4445	0,4347	0,4480
F3	0,5362	0,5370	0,5409	0,5429	0,5381
F4	1,2243	1,2236	1,2223	1,2081	1,2531
F5	0,4261	0,4219	0,4258	0,4261	0,4202

Tabla C.3: Error cuadrático medio de validación de cada modelo utilizado, para los hablantes mujeres, con MFCC.

Método Hablante	Error cuadrático medio [cm]				
	Lineal	Interacciones	Cuadrático con interacciones	Cuadrático	Puramente Cuadrático
M1	0,2421	0,2399	0,2409	0,2410	0,2587
M2	0,2706	0,2460	0,2443	0,2606	0,2766
M3	0,2505	0,2486	0,2483	0,2463	0,2457
M4	0,4285	0,4275	0,4310	0,4247	0,4659
M5	0,4366	0,4503	0,4526	0,4413	0,4405

Tabla C.4: Error cuadrático medio de validación de cada modelo utilizado, para los hablantes hombres, con MFCC.

*Apéndice D*

*Coefficientes modelos de regresión lineal, para  
experimento intrahablante*

A continuación se muestran valores estimados de los coeficientes al realizar el ajuste de regresión, utilizando formantes como variables predictoras, para el experimento intrahablante.

Hb \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
F1	14,34680	-0,06490	-0,21443	-0,58532	0,21543
F2	18,26232	-0,67288	-0,12562	-0,68782	-0,06066
F3	16,20636	-0,54332	-0,48149	-0,54903	0,17312
F4	16,16091	-1,30457	-0,42887	-0,31751	0,09930
F5	16,33274	-0,67602	-0,07957	-0,24140	-0,14737

(Hb = Hablante) (Cf = Coeficiente)

Tabla D.1: Coeficientes de la regresión para el modelo *lineal*, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en mujeres

Hb \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
M1	17,22485	-1,17453	-0,20867	0,02204	-0,03476
M2	16,94561	-1,43808	0,01433	-0,13684	-0,04752
M3	17,89460	-0,94348	-0,64968	-0,22856	0,03308
M4	16,83021	-0,69849	-0,39591	-0,12530	-0,00679
M5	18,81030	-0,77114	-0,01823	0,00540	-0,27757

(Hb = Hablante) (Cf = Coeficiente)

Tabla D.2: Coeficientes de la regresión para el modelo *lineal*, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en hombres

Cf \ Hb	F1	F2	F3	F4	F5
$\beta_0$	26,4888	15,7268	11,3252	21,0306	15,7543
$\beta_1$	0,15532	-6,19383	0,65717	-4,23048	-3,21757
$\beta_2$	-3,82627	-0,31065	-1,81999	-2,78435	-1,67226
$\beta_3$	-4,50556	-0,00322	1,23614	-3,08938	0,64425
$\beta_4$	-1,93925	1,68558	1,95036	1,11885	0,67435
$\beta_5$	0,25607	-0,00154	0,13336	-0,07597	-0,31361
$\beta_6$	-0,04149	1,49054	-0,25053	1,71841	0,75352
$\beta_7$	-0,21594	0,37496	-0,22528	-0,41318	0,24528
$\beta_8$	0,91648	0,47522	0,46367	1,22620	0,31103
$\beta_9$	0,28024	-0,27430	-0,01414	-0,25241	0,22400
$\beta_{10}$	0,66920	-0,58191	-0,61799	-0,12427	-0,49064

(Hb = Hablante) (Cf = Coeficiente)

Tabla D.3: Coeficientes de la regresión para el modelo *interacciones*, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en mujeres

Cf \ Hb	M1	M2	M3	M4	M5
$\beta_0$	24,90140	19,34884	31,80407	20,80274	21,28465
$\beta_1$	-3,08891	-1,14795	-5,22485	-3,25580	0,91513
$\beta_2$	-3,32939	-0,37608	-4,84724	-3,03975	-2,18198
$\beta_3$	-2,62866	-1,05481	-2,48778	-1,88898	-1,98576
$\beta_4$	-1,03771	-0,68714	-3,57638	0,61362	0,44184
$\beta_5$	0,12593	-0,22949	-0,37067	0,41727	0,51328
$\beta_6$	0,01951	-0,05631	0,07446	0,74729	0,72008
$\beta_7$	0,48803	0,07406	1,28680	-0,00427	-1,24924
$\beta_8$	1,16332	0,13339	0,27328	1,19770	0,83170
$\beta_9$	0,04391	0,05593	1,02966	-0,21917	-0,09192
$\beta_{10}$	0,28851	0,21260	0,49437	-0,12163	0,07786

(Hb = Hablante) (Cf = Coeficiente)

Tabla D.4: Coeficientes de la regresión para el modelo *interacciones*, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en hombres

Cf \ Hb	F1	F2	F3	F4	F5
$\beta_0$	28,98379	13,04088	12,58790	23,39171	14,39447
$\beta_1$	0,59860	-0,80791	7,55659	-0,64167	0,76509
$\beta_2$	-5,02966	1,15461	-2,14869	-2,52842	-1,24672
$\beta_3$	0,65085	0,14810	-2,52818	-4,94607	0,98641
$\beta_4$	-6,99761	1,47992	2,87609	0,48553	0,30083
$\beta_5$	0,59345	-0,84099	-0,67912	0,48919	-0,25325
$\beta_6$	-0,44651	1,59818	-0,60203	1,60517	0,62285
$\beta_7$	-0,04171	0,16192	0,25493	-0,85386	0,03592
$\beta_8$	0,53085	0,57305	0,38110	0,29061	0,19577
$\beta_9$	0,17447	-0,31587	0,08585	-0,79162	-0,07322
$\beta_{10}$	-0,73999	-0,52078	-0,25668	-0,19656	-0,42763
$\beta_{11}$	-0,37560	-2,61286	-4,55416	-1,86363	-2,23440
$\beta_{12}$	0,60175	-0,29908	0,26529	1,05692	0,28775
$\beta_{13}$	0,16981	-0,11640	0,51787	0,68406	-0,06268
$\beta_{14}$	1,27901	0,02811	-0,33286	0,28591	0,11164

(Hb = Hablante) (Cf = Coeficiente)

Tabla D.5: Coeficientes de la regresión para el modelo *cuadrático con interacciones*, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en mujeres

Cf \ Hb	M1	M2	M3	M4	M5
$\beta_0$	28,027325	21,505908	31,139231	19,704168	21,898964
$\beta_1$	-5,20492	-3,04611	-1,65428	-2,95994	2,04364
$\beta_2$	-4,91609	-0,34992	-4,30957	-4,37239	-2,14823
$\beta_3$	-2,54193	-1,37447	-4,12852	-2,79952	-1,07211
$\beta_4$	-1,88435	-1,50234	-2,97035	2,51006	-0,80743
$\beta_5$	1,39205	0,16954	-0,84356	1,21672	0,24946
$\beta_6$	-0,62140	-0,88192	-0,34859	0,18484	0,70014
$\beta_7$	0,56702	0,62450	1,05897	-0,00170	-1,13296
$\beta_8$	0,91485	-0,17700	0,30633	0,43566	0,88043
$\beta_9$	-0,10657	0,15865	0,92295	-0,01650	-0,09788
$\beta_{10}$	0,11297	-0,16707	-0,27062	-0,17545	0,11557
$\beta_{11}$	1,48355	1,44084	-0,68960	0,02090	-0,87736
$\beta_{12}$	0,69443	0,08543	0,04146	0,73527	0,01065
$\beta_{13}$	0,25146	0,52337	0,91868	0,49921	-0,21708
$\beta_{14}$	0,19775	0,18007	0,22222	-0,31778	0,15839

(Hb = Hablante) (Cf = Coeficiente)

Tabla D.6: Coeficientes de la regresión para el modelo *cuadrático con interacciones*, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en hombres

Cf \ Hb	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
F1	28,295	0,7700	-3,3001	0,4539	-7,4317	-0,5441	0,7901	-0,1728	1,1131
F2	13,665	2,6052	0,5749	0,5574	0,6545	-2,5490	-0,2039	-0,2345	-0,0932
F3	12,195	5,2252	-1,4985	-2,7088	3,3636	-4,1593	0,3494	0,4092	-0,4490
F4	24,270	1,7790	-3,5113	-4,2140	-0,4470	-2,1808	0,8202	0,6951	0,0771
F5	16,286	2,1319	-1,2240	0,3826	-0,4619	-2,1617	0,3140	-0,1229	0,0416

(Hb = Hablante) (Cf = Coeficiente)

Tabla D.7: Coeficientes de la regresión para el modelo *cuadrático*, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en mujeres

Hb \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
M1	23,3726	-0,7783	-2,2868	-1,6609	-1,5755	-0,2844	0,7034	0,3499	0,2081
M2	21,115	-2,1653	-0,1697	-2,0694	-0,9817	0,7911	0,0920	0,3992	0,1272
M3	26,148	0,0043	-1,9373	-4,6273	-1,1884	-0,6654	0,4518	0,8665	0,1727
M4	17,331	0,1731	-2,5201	-2,8514	2,5729	-0,6054	0,7134	0,5384	-0,3966
M5	19,360	0,7426	-0,1932	0,5136	-1,1638	-1,2549	0,0645	-0,1002	0,1287

(Hb = Hablante) (Cf = Coeficiente)

Tabla D.8: Coeficientes de la regresión para el modelo *cuadrático*, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en hombres

Hb \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
F1	13,67575	-0,03496	-0,05017	-0,11686	0,03680
F2	16,97651	-0,55588	-0,04043	-0,13085	-0,00884
F3	15,22322	-0,43331	-0,14508	-0,09728	0,02181
F4	15,07624	-0,96519	-0,10062	-0,05832	0,01184
F5	15,44843	-0,54922	-0,01653	-0,04401	-0,02006

(Hb = Hablante) (Cf = Coeficiente)

Tabla D.9: Coeficientes de la regresión para el modelo *puramente cuadrático*, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en mujeres

Hb \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
M1	16,65893	-1,05919	-0,05046	0,00706	-0,00440
M2	16,31781	-1,38049	0,00855	-0,02288	-0,00720
M3	16,86131	-0,75294	-0,20844	-0,03835	0,00300
M4	16,10792	-0,57608	-0,11265	-0,01834	-0,00366
M5	18,09212	-0,65036	-0,00230	0,00104	-0,03970

(Hb = Hablante) (Cf = Coeficiente)

Tabla D.10: Coeficientes de la regresión para el modelo *puramente cuadrático*, utilizando formantes calculados con LPC como variables predictoras. Experimento intrahablante, en hombres

***Apéndice E***  
***Coefficientes modelos de regresión lineal, para***  
***experimento entre-hablantes***

A continuación se muestran valores estimados de los coeficientes al realizar el ajuste de regresión, utilizando formantes como variables predictoras, para el experimento entre-hablantes.

Hv \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
F1	17,12321	-1,54463	-0,35122	-0,78513	0,32528
F2	15,01480	-1,71850	-0,63025	-0,29834	0,58233
F3	16,75503	-2,56859	-0,87217	-0,72401	0,75448
F4	16,69677	-2,43307	-0,77612	-0,81486	0,76985
F5	17,27723	-2,66127	-0,83327	-0,94674	0,75339

(Hv = Hablante utilizado para realizar la validación) (Cf = Coeficiente)

Tabla E.1: Coeficientes de la regresión para el modelo *lineal*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes mujeres

Hv \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
M1	15,29410	0,33345	0,19392	-0,20547	0,23185
M2	15,94062	-0,31095	-0,02993	-0,14251	0,24325
M3	15,27136	0,63985	0,24663	-0,29946	0,26454
M4	15,80320	0,28352	0,31553	0,07987	-0,10334
M5	17,09345	-1,21516	-0,33275	-0,23950	0,14173

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.2: Coeficientes de la regresión para el modelo *lineal*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres

Hv \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
F1	19,95484	-2,23818	-0,90690	-0,72303	0,02857
F2	20,98860	-4,00869	-1,67499	-0,77681	0,39718
F3	20,31815	-3,67323	-1,47146	-0,65087	0,35126
F4	20,50708	-3,52040	-1,49713	-0,73087	0,35604
F5	20,01839	-3,85964	-1,42304	-0,64585	0,37949
M1	19,95484	-2,23818	-0,90690	-0,72303	0,02857
M2	20,98860	-4,00869	-1,67499	-0,77681	0,39718
M3	20,31815	-3,67323	-1,47146	-0,65087	0,35126
M4	20,50708	-3,52040	-1,49713	-0,73087	0,35604
M5	20,01839	-3,85964	-1,42304	-0,64585	0,37949

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.3: Coeficientes de la regresión para el modelo *lineal*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres y mujeres juntos

Cf \ Hv	F1	F2	F3	F4	F5
$\beta_0$	24,53020	21,96750	32,43335	29,27439	30,35041
$\beta_1$	-5,14546	-4,25214	-7,56668	-7,21561	-6,74266
$\beta_2$	0,29349	-5,96014	-3,65368	-4,33189	-3,46833
$\beta_3$	-4,00633	-0,57778	-4,65434	-3,15238	-4,31810
$\beta_4$	-1,12870	0,08991	-2,86563	-1,88649	-2,05957
$\beta_5$	-1,47084	0,14351	-1,37544	-1,51696	-1,21307
$\beta_6$	1,19238	0,24343	0,89083	0,79894	0,85010
$\beta_7$	0,81430	0,45767	1,46602	1,52431	1,15908
$\beta_8$	0,34184	1,00233	0,55256	0,70850	0,61175
$\beta_9$	-0,17853	0,64751	0,57862	0,67986	0,47142
$\beta_{10}$	0,49895	-0,40260	0,65053	0,17681	0,48065

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.4: Coeficientes de la regresión para el modelo *interacciones*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes mujeres

Cf \ Hv	M1	M2	M3	M4	M5
$\beta_0$	15,55233	21,55377	16,09565	14,01745	23,01247
$\beta_1$	-0,47225	-1,81850	0,36736	1,57612	-2,69753
$\beta_2$	-1,63908	-1,77636	-2,03635	-1,51085	-2,50457
$\beta_3$	0,65579	-2,19224	1,65422	0,48887	-1,71179
$\beta_4$	0,40658	-0,54080	-0,58891	1,21097	-1,06152
$\beta_5$	-1,34257	-1,10185	-0,97922	-1,21335	-0,48246
$\beta_6$	-0,86117	0,12664	-2,69660	-0,94104	0,31563
$\beta_7$	1,42046	0,84077	2,45686	0,79054	0,42236
$\beta_8$	0,82770	0,96251	0,57880	1,35049	0,46734
$\beta_9$	0,10761	-0,03404	0,36123	-0,29362	0,35437
$\beta_{10}$	-0,44900	0,15641	-0,38787	-0,51820	0,17873

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.5: Coeficientes de la regresión para el modelo *interacciones*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres

Hv \ Cf	F1	F2	F3	F4	F5	M1	M2	M3	M4	M5
$\beta_0$	10,4726	19,6272	18,4998	16,4103	23,5476	10,4726	19,6272	18,4998	16,4103	23,5476
$\beta_1$	2,0251	-6,2667	-2,4359	-2,7216	-4,2891	2,0251	-6,2667	-2,4359	-2,7216	-4,2891
$\beta_2$	-0,3662	-6,8526	-3,4517	-4,4306	-3,9681	-0,3662	-6,8526	-3,4517	-4,4306	-3,9681
$\beta_3$	2,7479	2,5069	3,1075	5,0090	0,8316	2,7479	2,5069	3,1075	5,0090	0,8316
$\beta_4$	1,9076	1,6080	-0,5952	-0,2076	-1,4190	1,9076	1,6080	-0,5952	-0,2076	-1,4190
$\beta_5$	-1,6251	-0,6292	-2,1226	-2,1715	-1,0107	-1,6251	-0,6292	-2,1226	-2,1715	-1,0107
$\beta_6$	-1,4202	-0,3928	-2,2020	-2,2911	-1,0473	-1,4202	-0,3928	-2,2020	-2,2911	-1,0473
$\beta_7$	0,5715	1,2508	2,2979	2,4775	1,3851	0,5715	1,2508	2,2979	2,4775	1,3851
$\beta_8$	0,2493	1,0596	0,0423	-0,0807	-0,0651	0,2493	1,0596	0,0423	-0,0807	-0,0651
$\beta_9$	-0,0754	0,7371	0,8585	1,2074	0,9130	-0,0754	0,7371	0,8585	1,2074	0,9130
$\beta_{10}$	-0,8273	-1,2823	-0,7148	-1,1552	-0,2074	-0,8273	-1,2823	-0,7148	-1,1552	-0,2074

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.6: Coeficientes de la regresión para el modelo *interacciones*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres y mujeres juntos

Cf \ Hv	F1	F2	F3	F4	F5
$\beta_0$	26,32985	20,88909	35,64358	31,95259	33,76784
$\beta_1$	-3,80233	-3,40570	-11,75830	-11,58543	-11,95634
$\beta_2$	-0,10311	-4,81788	-3,89409	-4,50668	-4,64970
$\beta_3$	-0,66069	-3,68583	-1,47925	-0,67739	-0,06344
$\beta_4$	-4,89300	2,39752	-6,38632	-4,49279	-5,80500
$\beta_5$	-1,47692	-0,04541	-1,46158	-1,54487	-0,97316
$\beta_6$	0,58735	0,41994	0,45353	0,49015	0,12586
$\beta_7$	0,91134	0,22748	2,07615	1,98829	1,86632
$\beta_8$	0,31425	0,95789	0,77058	0,99803	0,76729
$\beta_9$	-0,11706	0,69927	0,72931	0,74779	0,58213
$\beta_{10}$	-0,62942	0,20159	-0,10849	-0,31341	-0,29003
$\beta_{11}$	0,19404	-0,12512	2,58167	2,71473	3,21548
$\beta_{12}$	0,08113	-0,27803	-0,24045	-0,24963	0,04165
$\beta_{13}$	0,27580	0,15343	-0,04970	-0,15005	-0,19960
$\beta_{14}$	0,89643	-0,54527	0,67490	0,48145	0,71262

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.7: Coeficientes de la regresión para el modelo *cuadrático con interacciones*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes mujeres

Cf \ Hv	M1	M2	M3	M4	M5
$\beta_0$	4,07604	16,83977	10,80678	10,26613	25,09344
$\beta_1$	0,26050	-3,69444	-3,14884	1,44964	-4,31675
$\beta_2$	-2,26893	-4,04780	-3,64936	-2,76379	-2,95897
$\beta_3$	-2,85330	-3,24085	0,45170	-1,36419	-2,15636
$\beta_4$	9,99285	4,35726	4,72803	5,30674	-1,51023
$\beta_5$	-0,02981	0,62559	0,54164	-0,33663	0,01065
$\beta_6$	0,10563	0,52488	-2,23641	-0,61337	-0,32012
$\beta_7$	-0,61549	-0,42351	1,18405	0,02650	0,71935
$\beta_8$	0,53648	0,55877	0,33296	1,05044	0,21608
$\beta_9$	-0,09042	-0,14949	0,32093	-0,26330	0,40485
$\beta_{10}$	0,50514	0,72711	0,23168	-0,24520	-0,09380
$\beta_{11}$	1,25611	1,98142	3,66635	0,36681	1,32225
$\beta_{12}$	0,42344	0,87536	0,48397	0,46666	0,22210
$\beta_{13}$	-0,02370	-0,15005	-0,20598	0,20491	0,42553
$\beta_{14}$	-1,50708	-0,76189	-0,85595	-0,61271	0,12196

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.8: Coeficientes de la regresión para el modelo *cuadrático con interacciones*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres

Hv \ Cf	F1	F2	F3	F4	F5	M1	M2	M3	M4	M5
$\beta_0$	7,8484	13,6872	14,3492	11,3435	25,3103	7,8484	13,6872	14,3492	11,3435	25,3103
$\beta_1$	2,2520	-6,8098	-2,0460	-3,5899	-6,1737	2,2520	-6,8098	-2,0460	-3,5899	-6,1737
$\beta_2$	-0,1342	-6,0838	-2,4585	-3,3492	-3,7444	-0,1342	-6,0838	-2,4585	-3,3492	-3,7444
$\beta_3$	1,3609	-0,5571	2,0579	3,4630	0,6169	1,3609	-0,5571	2,0579	3,4630	0,6169
$\beta_4$	4,3379	7,1044	2,0548	3,5096	-2,1044	4,3379	7,1044	2,0548	3,5096	-2,1044
$\beta_5$	-1,5295	-0,2398	-2,1108	-2,1093	-0,9687	-1,5295	-0,2398	-2,1108	-2,1093	-0,9687
$\beta_6$	-1,2230	0,2312	-1,5943	-1,5179	-1,6540	-1,2230	0,2312	-1,5943	-1,5179	-1,6540
$\beta_7$	0,3742	0,0850	1,6772	1,6880	1,6879	0,3742	0,0850	1,6772	1,6880	1,6879
$\beta_8$	0,1179	1,0498	0,5126	0,4222	0,0946	0,1179	1,0498	0,5126	0,4222	0,0946
$\beta_9$	-0,0595	0,7211	0,8068	1,0517	0,9886	-0,0595	0,7211	0,8068	1,0517	0,9886
$\beta_{10}$	-0,3506	-0,3494	-0,1229	-0,2679	-0,4521	-0,3506	-0,3494	-0,1229	-0,2679	-0,4521
$\beta_{11}$	-0,2422	1,7476	0,0743	1,1352	1,8539	-0,2422	1,7476	0,0743	1,1352	1,8539
$\beta_{12}$	0,0038	-0,2533	-0,6046	-0,5574	-0,2765	0,0038	-0,2533	-0,6046	-0,5574	-0,2765
$\beta_{13}$	-0,0605	-0,1436	-0,4304	-0,5779	0,2478	-0,0605	-0,1436	-0,4304	-0,5779	0,2478
$\beta_{14}$	-0,4967	-1,0092	-0,5169	-0,7295	0,1386	-0,4967	-1,0092	-0,5169	-0,7295	0,1386

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.9: Coeficientes de la regresión para el modelo *cuadrático con interacciones*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres y mujeres juntos

Cf \ Hv	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
F1	27,4220	-2,4486	-1,0029	-1,0760	-4,9424	0,9865	0,1893	0,0719	0,7295
F2	12,5088	-2,1613	-0,7463	-2,1340	3,6112	0,3166	0,0437	0,3401	-0,4357
F3	26,0651	-5,9361	-1,2946	-1,9565	-2,9605	2,6792	0,1162	0,2502	0,5235
F4	22,3755	-6,5005	-1,3917	-1,2400	-1,1447	3,1160	0,1604	0,0872	0,2686
F5	25,8924	-6,9578	-2,2677	-0,5275	-3,0455	3,3481	0,3827	-0,0739	0,5354

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.10: Coeficientes de la regresión para el modelo *cuadrático*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes mujeres

Cf \ Hv	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
M1	3,2485	-1,5469	-1,4773	-1,7267	9,5663	1,1986	0,5155	0,2773	-1,3360
M2	12,9401	-2,4679	-2,7813	-1,5971	4,6556	1,6774	0,8841	0,2574	-0,6184
M3	7,6159	-2,8305	-1,5640	-0,7765	6,4870	2,7877	0,5574	0,0576	-0,8682
M4	10,4128	-0,4003	-1,6278	-1,0236	4,7845	0,3197	0,6198	0,1958	-0,6777
M5	21,8268	-2,3428	-1,3793	-2,4324	-0,4147	1,1083	0,3583	0,4353	0,0757

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.11: Coeficientes de la regresión para el modelo *cuadrático*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres

Hv \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
F1	10,6026	-2,3701	-1,2417	0,7979	4,5172	-0,0811	0,0922	-0,3046	-0,6280
F2	9,2634	-6,6504	-2,5398	-0,7551	8,0823	1,9668	0,2640	-0,0218	-1,0892
F3	9,3301	-3,2478	-0,6994	0,8395	5,2048	-0,4702	-0,2204	-0,3124	-0,6771
F4	5,5889	-4,6005	-1,1879	1,8881	7,0844	0,6777	-0,0880	-0,5248	-0,9376
F5	20,2912	-5,4291	-1,6267	-1,0502	0,8955	1,2430	0,0569	0,0774	-0,0729
M1	10,6026	-2,3701	-1,2417	0,7979	4,5172	-0,0811	0,0922	-0,3046	-0,6280
M2	9,2634	-6,6504	-2,5398	-0,7551	8,0823	1,9668	0,2640	-0,0218	-1,0892
M3	9,3301	-3,2478	-0,6994	0,8395	5,2048	-0,4702	-0,2204	-0,3124	-0,6771
M4	5,5889	-4,6005	-1,1879	1,8881	7,0844	0,6777	-0,0880	-0,5248	-0,9376
M5	20,2912	-5,4291	-1,6267	-1,0502	0,8955	1,2430	0,0569	0,0774	-0,0729

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.12: Coeficientes de la regresión para el modelo *cuadrático*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres y mujeres juntos

Hv \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
F1	15,77951	-1,12882	-0,08824	-0,14827	0,04882
F2	14,43773	-1,24288	-0,16387	-0,04818	0,08100
F3	15,36305	-1,83576	-0,22579	-0,12965	0,10872
F4	15,36044	-1,72267	-0,19931	-0,14832	0,10972
F5	15,57218	-1,84828	-0,21059	-0,17660	0,10976

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.13: Coeficientes de la regresión para el modelo *puramente cuadrático*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes mujeres

Hv \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
M1	15,68315	0,31437	0,07295	-0,03506	0,02684
M2	16,03325	-0,19909	0,00899	-0,02359	0,03002
M3	15,72039	0,62220	0,08906	-0,05626	0,03214
M4	16,04042	0,25561	0,11266	0,01777	-0,01799
M5	16,39339	-1,02378	-0,09596	-0,04357	0,02009

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.14: Coeficientes de la regresión para el modelo *puramente cuadrático*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres

Hv \ Cf	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
F1	17,63274	-1,78280	-0,24880	-0,13724	0,00161
F2	17,95669	-3,08114	-0,45614	-0,13926	0,05090
F3	17,74218	-2,93031	-0,41137	-0,12109	0,04692
F4	17,88330	-2,79421	-0,41867	-0,13691	0,04565
F5	17,41239	-3,00807	-0,39280	-0,12140	0,05379
M1	17,63274	-1,78280	-0,24880	-0,13724	0,00161
M2	17,95669	-3,08114	-0,45614	-0,13926	0,05090
M3	17,74218	-2,93031	-0,41137	-0,12109	0,04692
M4	17,88330	-2,79421	-0,41867	-0,13691	0,04565
M5	17,41239	-3,00807	-0,39280	-0,12140	0,05379

(Hv = Hablante utilizado para la validación) (Cf = Coeficiente)

Tabla E.15: Coeficientes de la regresión para el modelo *puramente cuadrático*, utilizando formantes calculados con LPC como variables predictoras. Experimento entre-hablantes, para el caso de hablantes hombres y mujeres juntos

## *Bibliografía*

- [1] IBM Knowledge Center - Mean and median absolute deviation.
- [2] ARSIKERE, H., LEUNG, G. K., LULICH, S. M., AND ALWAN, A. Automatic estimation of the first three subglottal resonances from adults' speech signals with application to speaker height estimation. *Speech Communication* 55, 1 (2013), 51 – 70.
- [3] CANAVOS, G. C. *Probabilidad y Estadística - Aplicaciones y Métodos*, 1 ed. MCGRAW-HILL/INTERAMERICANA DE MEXICO, S.A. DE C.V.
- [4] CARRANZA, R., AND ALESSANDRONI, N. *Dicción, Fonética y Técnica Vocal. El Entrenamiento Vocal en clave interdisciplinaria*. PhD thesis, 2013.
- [5] CASSIDY, S. Chapter 7. The Source Filter Model of Speech Production.
- [6] CASTAÑEDA, P. F. Explorando la producción sonora del habla. In *El lenguaje verbal del niño*. 1999, ch. 8, pp. 123–140.
- [7] CISTERNAS LILLO, P. F., AND DÍAZ APABLAZA, S. K. *Características acústicas de las vocales producidas por sujetos de habla hispana, Chilenos residentes en la ciudad de Santiago*. Tesis, UNIVERSIDAD ANDRÉS BELLO, 2012.
- [8] CSS TEMPLATES MARKET. span | USC-TIMIT database. Accedido en 26-04-2018 a <https://sail.usc.edu/span/usc-timit/index.html>, 2014.
- [9] DAVE, N. Feature Extraction Methods LPC , PLP and MFCC In Speech Recognition. *International Journal for Advance Research in Engineering and Technology* 1, Vi (2013), 1–5.

- [10] DOUGLAS, J. NELSON; SMITH, D. C., AND TOWNSEND, J. L. Voice Activity Detector, 2003.
- [11] DUSAN, S. Estimation of Speaker s Height and Vocal Tract Length from Speech Signal. *Interspeech* (2005), 1989–1992.
- [12] FAY, R. R., POPPER, A. N., NEW, S., HEIDELBERG, Y. B., KONG, H., MILAN, L., TOKYO, P., GREENBERG, S., AINSWORTH, W. A., AND PROGRAM, R. R. F. *Springer Handbook of Auditory Research Speech Processing in the Auditory System*.
- [13] FITCH, W. T. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America* 102, 2 (1997), 1213–1222.
- [14] FITCH, W. T., AND GIEDD, J. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America* 106, 3 (1999), 1511–1522.
- [15] GANCHEV, T. *Contemporary Methods for Speech Parameterization*. 2011.
- [16] GARGOURI, D., KAMMOUN, M. A. L. I., AND HAMIDA, A. B. E. N. A Comparative Study of Formant Frequencies Estimation Techniques. *Proceedings of the 5th WSEAS International Conference on Signal Processing, Istanbul, Turkey* (2006), 15–19.
- [17] GULZAR, T., SINGH, A., AND SHARMA, S. Comparative Analysis of LPCC , MFCC and BFCC for the Recognition of Hindi Words using Artificial Neural Networks. *International Journal of Computer Applications* 101, 12 (2014), 22–27.
- [18] GUZMAN, M. Acústica del tracto vocal \_ ORTOFON \_\_ Centro de Logopedia, Psicología Clínica y Psicopedagogía.
- [19] HANSEN, J. H. L., WILLIAMS, K., AND BOÑIL, H. Speaker height estimation from speech: Fusing spectral regression and statistical acoustic models. *The Journal of the Acoustical Society of America* 138, 2 (2015), 1052–1067.
- [20] HARRIS, T. R., FITCH, W. T., GOLDSTEIN, L. M., AND FASHING, P. J. Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass. *Ethology* 112, 9 (2006), 911–920.

- [21] HUBER, P. J. Robust statistics. *Statistical Science*, 1986 (2011), 1–11.
- [22] JAMES, G., WITTEN, D., HASTIE, T., AND TIBSHIRANI, R. *An introduction to Statistical Learning*, vol. 7. 2000.
- [23] KIM, J. Index of /old/software/rtmri\_seg. Accedido en 09-04-2018 a [http://sail.usc.edu/old/software/rtmri\\_seg/](http://sail.usc.edu/old/software/rtmri_seg/), 2015.
- [24] KIM, J., KUMAR, N., LEE, S., AND NARAYANAN, S. Enhanced airway-tissue boundary segmentation for real-time magnetic resonance imaging data. *Proceedings of the 10th International Seminar on Speech Production (ISSP)* (2014), 222–225.
- [25] KUMAR, C., REHMAN, F., KUMAR, S., MEHMOOD, A., AND SHABIR, G. Analysis of MFCC and BFCC in a Speaker Identification System. *2018 International Conference on Computing, Mathematics and Engineering Technologies* (2018).
- [26] LAMMERT, A. C., AND NARAYANAN, S. S. On short-time estimation of vocal tract length from formant frequencies. *PLoS ONE* 10, 7 (2015), 1–23.
- [27] LEYS, C., LEY, C., KLEIN, O., BERNARD, P., AND LICATA, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology* 49, 4 (2013), 764–766.
- [28] LYNN KIRLIN, R. A Posteriori Estimation of Vocal Tract Length. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26, 6 (1978), 571–574.
- [29] MIYARA, F. *La voz humana*, 2004.
- [30] MOHAN, B. J., AND RAMESH BABU, N. Speech recognition using MFCC and DTW. *2014 International Conference on Advances in Electrical Engineering, ICAEE 2014* (2014).
- [31] MUDRY, A., AND TJELLSTRÖM, A. Historical background of bone conduction hearing devices and bone conduction hearing aids. *Advances in Oto-Rhino-Laryngology* 71 (2011), 1–9.
- [32] NARAYANAN, S., TOUTIOS, A., RAMANARAYANAN, V., LAMMERT, A., KIM, J., LEE, S., NAYAK, K., KIM, Y.-C., ZHU, Y., GOLDSTEIN, L., BYRD, D., BRESCH,

- E., GHOSH, P., KATSAMANIS, A., AND PROCTOR, M. Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC). *The Journal of the Acoustical Society of America* 136, 3 (2014), 1307–1311.
- [33] PAIGE, A., AND ZUE, V. W. Calculation of Vocal Tract Length. *IEEE Transactions on Audio and Electroacoustics* 18, 3 (1970), 268–270.
- [34] PONTAQUE, F. C. *Modelos lineales*. E-UMAB. Publicacions i Edicions de la Universitat de Barcelona, 2005.
- [35] REBY, D., AND MCCOMB, K. Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour* 65, 3 (2003), 519–530.
- [36] RENDALL, D., KOLLIAS, S., NEY, C., AND LLOYD, P. Pitch ( $f_0$ ) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America* 117, 2 (2005), 944–955.
- [37] RIEDE, T., AND FITCH, T. Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*). *The Journal of experimental biology* 202 (1999), 2859–2867.
- [38] ROTHENBERG, M. The Source-Filter Model Lives ( if you are careful ). *Voice Foundation Annual Symposium* (2008), 1–9.
- [39] RUEDA, L. *Mejoras en reconocimiento del habla basadas en mejoras en la parametrización de la voz*. PhD thesis, 2011.
- [40] S. DAVIS AND P. MERMELSTEIN. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28, 4 (1980), 357–366.
- [41] SAHIDULLAH, M., AND SAHA, G. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. *Speech Communication* 54, 4 (2012), 543–565.
- [42] SAKATA, S., AND WHITE, H. *Breakdown Point*. American Cancer Society, 2006.

- [43] SEPULVEDA-SEPULVEDA, A., CASTELLANOS-DOMINGUEZ, G., AND GÓMEZ-VILDA, P. Subject-independent acoustic-to-articulatory mapping of fricative sounds by using vocal tract length normalization. *Revista Facultad de Ingeniería Universidad de Antioquia* (12 2015), 162 – 169.
- [44] SJÖLANDER, K., AND BESKOW, J. "TMH KTH :: Wavesurfer". Accedido en 09-04-2018 a <http://www.speech.kth.se/wavesurfer/>, 2012.
- [45] SLANEY, M. Auditory toolbox: A Matlab Toolbox for Auditory Modeling Work. Tech. rep., 1998.
- [46] SOROKIN, V. N., AND GERAS'KIN, I. V. Vocal-tract length estimation. *Journal of Communications Technology and Electronics* 58, 12 (2013), 1292–1301.
- [47] STEVENS, K. N. *Acoustic Phonetics*. Acoustic Phonetics. CogNet, 2000.
- [48] WEI HAN, CHEONG-FAT CHAN, CHIU-SING CHOY, AND KONG-PANG PUN. An efficient MFCC extraction method in speech recognition. *2006 IEEE International Symposium on Circuits and Systems* (2006), 4.
- [49] WELLING, L., AND NEY, H. Formant estimation for speech recognition. *IEEE Transactions on Speech and Audio Processing* 6, 1 (1998), 36–48.
- [50] WRENG, A. MOCHA-TIMIT. Accedido en 12-05-2018 a <http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html>, 1999.