Trabajo de Fin de Grado

**Grado en Ingeniería en Tecnologías Industriales (GETI)**

# Determinación de características de pronunciación de letras mediante el análisis frecuencial de fonemas

# MEMORIA

21 de junio de 2020


**Autor:** Abdulrahim Oukar Tatari

**Director:** Carlos Ocampo-Martinez


**Convocatòria:** 07/2020



Escola Tècnica Superior
d'Enginyeria Industrial de Barcelona


# Resumen

En este proyecto se crea una herramienta capaz de discernir entre tres consonantes fricativas del árabe, dos de ellas consideradas sonidos comunes y una de ellas considerada particular del árabe. Esta herramienta consiste en un modelo creado con código Python en la aplicación web Jupyter Notebook a partir del algoritmo de inteligencia artificial conocido como *Análisis de Discriminante Lineal*. Cuando este modelo recibe la información relativa a una de estas consonantes, el sistema devuelve el fonema correspondiente según lo que ha aprendido a partir de muestras anteriores.

La información relativa a cada consonante viene dada por un vector de valores de parámetros acústicos seleccionados a partir de investigaciones de vanguardia anteriores y la propia verificación sobre la relevancia de estos parámetros a la hora de diferenciar entre las consonantes realizada en este proyecto mediante el método estadístico de ANOVA en Minitab. Para extraer los parámetros de las grabaciones de sonido de cada consonante se utiliza el programa de análisis fonético del habla conocido como Praat.

Adicionalmente, se realiza un estudio con 15 individuos con tres idiomas nativos distintos (catalán, castellano e italiano), que no hablan el árabe, para comprobar la capacidad de pronunciar correctamente sonidos ajenos y la correlación de esta con el idioma nativo. El modelo resultó ser eficaz y veraz, no obstante, no se pudo establecer una correlación entre el idioma nativo y la correcta pronunciación de los sonidos desconocidos.

# Agradecimientos

Me gustaría agradecer a mi tutor Carlos Ocampo-Martinez por su constante y dedicado apoyo durante toda la elaboración de este proyecto, quien, con sus indicaciones ha conseguido enseñarme mucho sobre la correcta realización de trabajos de ámbito científico.

También, quiero agradecer especialmente a todas las personas, de las cuales la mayoría son familiares y amigos, que han participado en este proyecto prestando sus voces y enviando las grabaciones, sin las cuales, este proyecto no podría haberse llevado a cabo.

# Índice general

# Índice de figuras

# Índice de cuadros

# Capítulo 1

# Introducción

## 1.1. Origen y motivación del proyecto

El grado de ingeniería en tecnologías industriales me ha supuesto durante estos años una serie ininterrumpida de retos durante los cuales, he podido observar como mis aptitudes y habilidades mejoraban al enfrentarme a cada uno de ellos, expandiendo así mis límites. Por tanto, siguiendo esta misma tendencia, decidí que el trabajo de fin de grado debía surgir como una idea propia, la cual debería resolver usando el ingenio y la capacidad resolutiva adquirida en este tiempo.

Nacido en Valencia pero de origen sirio, aprendí a hablar el dialecto árabe levantino dentro del hogar y durante mis viajes a Siria con el resto de mi familia. No obstante, existen ciertos sonidos que no conseguí discernir o pronunciar correctamente. Es por esto, que me he interesado en hacer un proyecto relativo a la diferenciación acústica de algunas consonantes del árabe, buscando crear una herramienta más técnica, desde un punto de vista ingenieril, que pudiese replicar el *feedback* del cerebro referente a la correcta pronunciación.

Tras proponer esta idea a mi tutor Carlos Ocampo-Martinez del departamento de Control Automático, y estar él de acuerdo, procedimos a elaborarla más detalladamente y comenzar así, con el proyecto.

## 1.2. Objetivos del proyecto

El objetivo principal de este proyecto es diseñar una herramienta que sea capaz de discernir entre ciertas consonantes del árabe similares entre sí. Concretamente, la idea consiste en que, cuando esta herramienta reciba como señal de entrada una letra pronunciada por un individuo, el sistema devuelva como respuesta la información sobre si se ha pronunciado correctamente dicha letra o si, en cambio, se ha asemejado más a otra. Esta herramienta puede no ser necesariamente una pieza única y, consistir, por contra, en una composición de distintas herramientas con funciones aisladas que se combinen para lograr el objetivo.

Para cumplir con el objetivo principal, se ha de pasar por un objetivo intermedio consistente en averiguar que parámetros acústicos característicos son los necesarios para poder discernir, en conjunto, todos y cada uno de los sonidos involucrados en el proyecto. Hay que descubrir, por tanto, de que manera tratar las muestras y con que herramientas extraer los parámetros.

Por último, otro objetivo secundario que se ha considerado interesante es realizar un estudio sobre si ciertas características de la lingüística, como son el idioma nativo o el poliglotismo, influyen en la correcta pronunciación de sonidos ajenos. Se usará, por tanto, la herramienta creada en este proyecto para comprobar si individuos que no han escuchado un sonido anteriormente son capaces de reproducirlo con exactitud o si tienden, por contra, a pronunciar otro más semejante a uno conocido. Esto permitía, a su vez, comprobar la veracidad del trabajo.

Todos estos objetivos sirven, además, al propósito de usar y aplicar correctamente todos los conocimientos, métodos y herramientas aprendidas durante la carrera, así como utilizar con rigor los nuevos que se aprendan y apliquen durante el desarrollo de este. Por último, se busca aplicar el ingenio en la búsqueda y creación de soluciones a posibles complicaciones técnicas, que aunque no sirvan en todos los contextos, si optimicen los resultados del objetivo final.

## 1.3. Alcance del proyecto

Un factor importante a tener en cuenta en este trabajo, son las limitaciones que se tienen a la hora de realizar ciertas partes y que por tanto, afectan al alcance del proyecto en conjunto. La principal limitación a la que se enfrenta este proyecto es el diseño del experimento. Al tratarse de un proyecto de acústica, las muestras que se analizan son grabaciones de sonido realizadas por terceras personas. Esto supone un problema en lo referente la obtención de todas las muestras necesarias (tanto de cantidad como de tipo), y en el supuesto de la necesidad de un rediseño del experimento.

Además, a esto se ha de sumar las circunstancias extraordinarias en el marco temporal en el que se está realizando este proyecto, que supone un problema en la homogeneidad de todas las muestras. Otros factores que afectan al diseño del experimento son la falta de conocimiento y experiencia previa en el campo de la acústica y, la falta de información más relevante durante la primera etapa del trabajo, que sumado a la limitación del tiempo, hacen que el diseño del experimento inicial no sea el más óptimo para el futuro del trabajo.

## 1.4. Requerimientos previos

Para el correcto seguimiento y entendimiento de este trabajo no se requiere ninguna preparación especial previa, simplemente estar familiarizado con los conceptos empleados en este, cuya mayoría se estudian durante el grado de ingeniería en tecnologías industriales.

Concretamente, los conceptos más importantes utilizados son relativos a la inteligencia artificial aplicada con código Python mediante Jupyter Notebook y al análisis estadístico de comparación de tratamientos, como son el p-valor y el análisis ANOVA, mediante Minitab. También se hace una ligera mención a la turbulencia, concepto de la mecánica de fluidos.

Respecto a los conceptos que no se estudian durante la carrera de una manera específica (teoría acústica 2), o las herramientas nuevas tampoco utilizadas (Praat 3.1), se hace siempre una explicación previa.

## 1.5. Estado del concepto

Existen muchos estudios en la literatura reciente sobre los parámetros característicos de las consonantes que permiten discernir entre los distintos fonemas. Debido a sus propiedades, son especialmente interesantes los que examinan las consonantes fricativas. Razón por la cual en este estudio se ha optado por esta dirección. Aunque mayoritariamente han estado centrados en las fricativas inglesas, también se han hecho exhaustivos estudios sobre las fricativas árabes y de otros idiomas. En general, estudios que han pretendido clasificar los sonidos fricativos, han investigado parámetros comunes (explicados más detenidamente en la Sección 2.3),que pueden agruparse según el tipo: amplitudinales, temporales, formánticos y espectrales.

El primer estudio consultado estaba centrado en las fricativas del inglés [Jongman et al, 2000]. En este se estudian diez parámetros: el pico espectral, los cuatro momentos espectrales, la amplitud relativa, la amplitud normalizada, la duración, la ecuación de locus y la F2 de transición. Para comprobar el efecto de cada parámetro en la diferenciación de las distintas consonantes, se aplico el método ANOVA, de uno a varios factores según el caso, teniendo en cuenta el punto de articulación, la sonoridad, la vocal adyacente y el género.

En un estudio posterior [Maniwa et al, 2009], también relativo a las fricativas inglesas, esta vez centrado en clara pronunciación de estas, se añadieron cuatro parámetros más: pendientes espectrales superiores e inferiores a picos de frecuencia, media de la frecuencia fundamental de la vocal adyacente, la relación harmónico ruido (HNR) y la energía inferior a 500 Hz. Esta vez, se descartó la ecuación de locus. Se utilizaron los mismos métodos de análisis.

Aunque estos estudios son bastante exhaustivos, para este trabajo se necesita información más precisa sobre las fricativas árabes, puesto que muchas de ellas no están presentes en el inglés u otros idiomas. El siguiente estudio se centra en analizar y definir cada uno de los parámetros que se habían utilizado en la literatura anterior y en decidir cuales suponen una diferenciación significativa entre las 13 fricativas del árabe, todo ello teniendo en cuenta la influencia de las vocales [Al-Khairy, 2005]. El método de análisis vuelve a ser ANOVA, de uno o varios factores, según el caso. Una diferencia notable es que en este estudio, no se consideran la influencia del género, puesto que solo participan hombres. En este caso, también se creo un modelo de entrenamiento y predicción a partir de los resultados obtenidos.

Además de analizar las características de las fricativas, es interesante también ver que estudios se han hecho sobre otras particularidades del árabe, como las consonantes enfáticas (ver Sección 2.1.2). Un primer estudio investiga los tres primeros formantes en el inicio, mitad y final de las todas las posibles vocales adyacentes a las consonantes simples y sus contrapartes enfáticas [Al-Masri et al, 2007]. Otro trabajo posterior, relacionado con el anterior, se focalizaba en en el efecto del género en esta misma diferenciación, [Abudalbuh, 2011]. Ambos estudios no se centran solo en las fricativas, si no en las oclusivas también, por tanto también se investigaron otros parámetros concretos de cada tipo, sobre todo en el segundo trabajo.

Este trabajo no se centrará en el estudio exhaustivo completo de todos los parámetros necesarios para clasificar toda la familia de sonidos fricativos, sino que, basándose en la literatura anterior se seleccionarán aquellos parámetros que sean útiles para separar las consonantes involucradas en este proyecto, que a su vez, se habrán elegido en función de las herramientas de las cuales se disponga para la extracción de dichos parámetros y los primeros resultados de estos. Además, se añadirá la creación del modelo que permitirá discernir entre consonantes mediante las técnicas de inteligencia artificial estudiadas en la carrera. Esta última parte permitirá, por tanto, cumplir con el objetivo principal del proyecto y elabora el estudio posterior.

## 1.6. Estructura del documento

Este documento se divide en siete capítulos principales:

- El Capítulo 1 corresponde a la *Introducción*. En este, se explica el origen y la motivación de este proyecto a la vez que se exponen los objetivos de este, y se menciona el alcance del proyecto y los requerimientos previos para su seguimiento. También se hace un repaso de la literatura más relevante para la realización del trabajo.

- El Capítulo 2 se denomina *Teoría acústica* y en él, se hace una introducción a la fonología y fonética del árabe, explicando sus características más particulares. También se analizan las características (producción y parámetros acústicos característicos) de los sonidos que involucran este trabajo: las consonantes fricativas.

- El Capítulo 3, que recibe el nombre de *Extracción de parámetros*, consiste en analizar si los parámetros acústicos seleccionados para discernir entre las consonantes son estadísticamente significativos para dicha tarea, así como explicar las herramientas y los métodos empleados. También se detalla como se obtuvieron las muestras.

- En el Capítulo 4, llamado *Clasificación y estudio*, se realiza un modelo mediante un algoritmo de inteligencia artificial, del cual se explica su funcionamiento y la herramienta utilizada para su implementación, que es capaz de clasificar y predecir las consonantes involucradas en el proyecto. A partir de este modelo, se realiza un pequeño estudio sobre la capacidad de pronunciar sonidos ajenos por parte de hablantes de otros idiomas distintos al árabe.

- Los Capítulos 5 y 6 corresponden a la elaboración del presupuesto de este proyecto y a su impacto ambiental, respectivamente.

- Finalmente, en el Capítulo 7 se exponen las *Conclusiones* de este trabajo.

# Capítulo 2

# Teoría acústica

Este capítulo hace una introducción a la fonología y fonética del idioma árabe, mientras clasifica en un contexto acústico sus sonidos y comenta sus características más particulares. Además, se explican conceptos relacionados con las características de la producción fricativa y la parametrización del subconjunto de sonidos en los cuales se centrará este trabajo. Todo esto es necesario para facilitar el seguimiento del proyecto, pues muchos conceptos pueden ser desconocidos para el lector potencial del presente documento.

## 2.1. Fonología y fonética del árabe

El análisis lingüístico se puede dividir en dos ramas principales: la fonología, que estudia las unidades lingüísticas invariantes de carácter distintivo codificadas en las ondas sonoras, conocidas como fonemas, y la fonética, que estudia la variación acústica y articulatoria de los sonidos de este mismo habla [Llisterri, 2020]. Los fonemas pueden dividirse en dos tipos: consonánticos y vocálicos.

El idioma árabe actual desciende del árabe clásico o coránico del siglo VI d.C, tanto el idioma literario, conocido árabe moderno estándar (MSA), como las distintas variedades dialectales habladas. En árabe existe una distinción entre estos dos tipos. Por un lado, el MSA se reconoce como la única versión oficial del árabe y es usado en todas las situaciones de carácter formal, tanto de forma escrita (documentos, libros, artículos, entre otros), como de forma hablada (e.g., conferencias, noticias, discursos). Por otro lado, las variaciones dialectales en el lenguaje hablado, que difieren significativamente a lo largo de Oriente próximo y el norte de África, no se consideran como idiomas distintos [Javed, 2013]. El alfabeto árabe (alifato) consta de 28 letras (más algunas variantes y grafemas auxiliares), y su escritura se caracteriza por realizarse de derecha a izquierda con una traza cursiva.

### 2.1.1. Clasificación de los sonidos del árabe

Los sonidos del habla pueden clasificarse acústicamente de la siguiente forma [Llisterri, 2020]:

- Las vocales, consonantes nasales y consonantes laterales son sonidos periódicos compuestos producidos por la vibración de los pliegues vocales (frecuencia fundamental) y resonancia (armónicos) en el tracto vocal.
- Las consonantes oclusivas son sonidos aperiódicos impulsionales producidos por el cierre y explosión en el tracto vocal.
- Las consonantes fricativas son sonidos aperiódicos continuos producidos por la constricción en el tracto vocal.

El árabe moderno estándar presenta seis vocales puras: tres vocales cortas /a/, /i/, /u/ y sus respectivas versiones largas /i:/, /u:/, /a:/, respectivamente. También presenta dos diptongos con la vocal /a/: /w/ y /j/. Mientras que las vocales largas sí tienen una letra correspondiente, las vocales cortas se representan con grafías auxiliares que muchas veces se omiten por agilidad. Cabe mencionar que algunas letras pueden representar sonidos distintos según su localización en la palabra, por ejemplo /u:/ y /w/, /i:/ y /j/ o /a:/ y la primera letra del alifato, cuyo sonido es variante [Javed, 2013].

Respecto a las consonantes, el árabe presenta dos nasales (/m/, /n/), ocho oclusivas (/t, tˤ, k, q, b, d, dˤ, P/) 13 fricativas (/f, T, s, S, sˤ, X, è, h, ð, z, ðˤ, K, Q/) , dos aproximantes(/w, j/), una lateral aproximante (/l/) y una líquida (/r/). Hay una letra que se considera fricativa (/Z/) o africante (/Ã/) según el dialecto, siendo este último el caso mayoritario [Martínez, 2020] [Newman, 2020].

Para diferenciar las consonantes dentro de cada tipo, existen dos factores:

- **Punto de articulación en el tracto vocal**: Este puede ser bilabial, labio-dental, dental, alveolar, post-alveolar, palatal, velar, labial-velar, uvular, glotal o faríngeo. Este último es característica del árabe. Otra particularidad del árabe son las contrapartes enfáticas de algunos puntos de articulación anteriores.
- **Sonoridad**: Las consonantes oclusivas y las consonantes fricativas pueden ser sordas o sonoras. La principal diferencia es que la sonoridad implica la vibración de las cuerdas vocales.

El Cuadro 2.1 aglutina la clasificación previamente expuesta.

Tal y como se mencionó en la Sección 1.5, este trabajo se focalizará en las consonantes fricativas debido a sus características y a la amplia literatura que existe sobre estas. Inicialmente se pensó trabajar con ocho consonantes fricativas (/s, S, sˤ, X, è, h, z, Z/) debido a su similitud entre grupos e interés en relación a algunas letras únicas en el árabe de cara al estudio posterior. No obstante, por motivos que se explican en la Sección 3.3, se acabaron descartando cinco de ellas, quedando /s/, /S/ y /sˤ/.

Cuadro 2.1: *Clasificación de las consonantes del MSA*

| | | Bilabial | labio-dental | dental | alveolar | post-alveolar | palatal | labial-velar | velar | uvular | faríngea | glotal |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nasal | | /m/ | | | /n/ | | | | | | | |
| Oclusiva | sorda | | | /t/-/tˤ/ | | | | | /k/ | /q/ | | /ʔ/ |
| | sonora | /b/ | | /d/-/dˤ/ | | | | | | | | |
| Fricativa | sorda | | /f/ | /T/ | /s/-/sˤ/ | /S/ | | | | /X/ | /ħ/ | h |
| | sonora | | | /ð/-/ðˤ/ | /z/ | /Z/ | | | | /K/ | /Q/ | |
| Africante | | | | | | /Ã/ | | | | | | |
| Aproximante | | | | | | | /j/ | /w/ | | | | |
| Lateral aproximante | | | | | /l/ | | | | | | | |
| Líquida | | | | | /r/ | | | | | | | |

### 2.1.2. Faringealización

Una de las características fonéticas más particulares del árabe es la faringealización. La faringealización es un tipo de articulación que involucra la constricción de la faringe. Existen dos tipos, las consonantes faríngeas y las consonantes faringealizadas o enfáticas. Las consonantes faríngeas (/ħ/ y /Q/) tienen una constricción primaria, mientras que en las consonantes enfáticas (/sˤ/, /tˤ/,/dˤ/ y /ðˤ/), esta constricción es secundaria. Estas últimas pueden producirse velarizando el sonido [Hermes et al, 2017].

## 2.2. Producción fricativa

La producción de los sonidos fricativos del habla está estrechamente relacionada con el concepto de la turbulencia de un flujo. Es por esto que en este apartado se repasarán los conceptos de turbulencia en el ámbito de la mecánica de fluidos y cuál es su papel en la producción fricativa del habla. También se introducirá el modelo fricativo que intenta reproducir esta producción.

### 2.2.1. Turbulencia

En el campo de la mecánica de fluidos [1], donde se trata al fluido como un medio continuo, se define un flujo turbulento como aquel que se caracteriza por tener un movimiento con fuertes fluctuaciones que varía irregularmente y cuyas propiedades presentan variaciones aleatorias en el tiempo y el espacio (descritos por promedios estadísticos). El parámetro adimensional que permite decidir cuándo un flujo es turbulento es el número de Reynolds. Cuando este supera un valor límite, aproximadamente del orden de $10^6$, se puede considerar que el flujo es turbulento. El número de Reynolds, que se define como

$$Re = \frac{cL}{\nu} = \frac{\rho c L}{\mu}, \tag{2.1}$$

donde $c$ [m/s] es la velocidad del flujo, $L$ [m] es la longitud característica del medio u objeto por donde circula el flujo, $\rho$ [kg/m$^3$] es la densidad del fluido y $\nu$ [m$^2$/s] y $\mu$ [Pa/s] son las viscosidades cinemática y dinámica del fluido, respectivamente, mide la importancia relativa entre las fuerzas de inercia y las fuerzas viscosas. Es interesante observar que la velocidad de

[1] Apuntes de Mécanica de Fluidos. GETI, ETSEIB

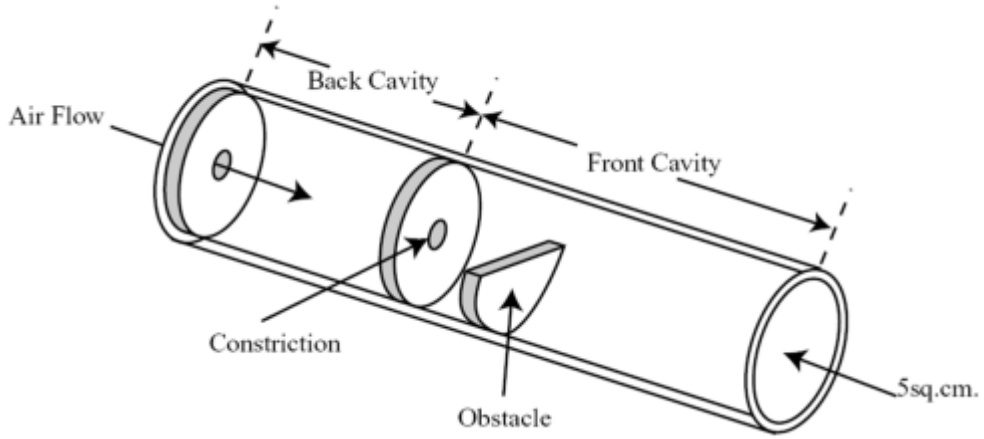


Figura 2.1: *Modelo de Shadle [Huckvale, 2017]*

flujo es directamente proporcional al número de Reynolds, por lo que se puede deducir, aunque dependiendo también de propiedades del fluido y el lugar por donde circula el flujo, una velocidad alta favorece la turbulencia.

### 2.2.2. Modelo de Shadle

La fuente de excitación, ocurrida en el tracto vocal, que produce los sonidos fricativos, es la turbulencia. Son las variaciones de presión aleatorias generadas en el aire provocadas por el flujo turbulento las que se escuchan en forma de una onda aperiodica o de ruido. La turbulencia en el tracto vocal puede llegar de dos formas: cuando el flujo se vuelve turbulento al pasar por una construcción, o cuando una constricción genera un chorro de aire a alta velocidad que impacta con un obstáculo estacionario de bordes afilados [Huckvale, 2017].

El modelo fricativo de Shadle (1990), el cual se puede observar en la Figura 2.1, es un modelo general de la producción fricativa, capaz de recrear la mayoría de los aspectos de la acústica fricativa. En este, el flujo de aire se vuelve turbulento tanto en una constricción como en un obstáculo.

## 2.3. Parámetros acústicos de las fricativas árabes

Dado que para llevar a cabo el objetivo del proyecto se ha de realizar una comparativa entre consonantes, es necesario poder definir dichos fonemas de una manera cuantificable y medible. Para poder encontrar esta firma característica de cada consonante y posteriormente realizar su clasificación, es necesario encontrar sus parámetros más representativos.

A continuación, se hace una recopilación de los parámetros más importantes en la diferenciación de los sonidos fricativos que se estudiaron en la literatura y a los que se hace referencia en la Sección 1.5 con el fin de detallar y explicar cómo se define cada uno. Esta recopilación se centrará especialmente en la investigación de Al-Khairy [Al-Khairy, 2005], puesto que es la semejante y relevante para este proyecto, aunque también se tendrán en cuenta los demás estudios relativos

a las consonantes enfáticas.

Cabe mencionar que este estudio no mostrará las conclusiones sobre la influencia de cada parámetro en la diferenciación de las consonantes ni los métodos empleados. Este análisis se detallará más adelante, en la Sección 3.3, y se centrará únicamente en las consonantes seleccionadas para el proyecto, con el fin de seleccionar los parámetros útiles para este.

### 2.3.1. Amplitud y duración

En esta subsección se definen los parámetros temporales (ms) y de amplitud (dB).

**Amplitud normalizada RMS:** Diferencia entre la amplitud del RMS (valor eficaz o valor cuadrático medio) del sonido fricativo y la amplitud media del RMS de tres períodos de tono consecutivos en el punto de máxima amplitud vocal.

**Amplitud relativa:** Diferencia entre la amplitud de una frecuencia específica (dependiente de la consonante) medida en el punto medio del sonido fricativo, y la amplitud de la correspondiente frecuencia medida en el inicio de la vocal.

**Duración absoluta:** Simplemente la duración absoluta del sonido fricativo.

**Duración normalizada:** Ratio entre la duración del sonido fricativo y la duración de la palabra entera en la que se encuentra.

### 2.3.2. Medidas espectrales

En este parte se introducirán los parámetros espectrales. Estos parámetros contemplan el pico espectral y los cuatro primeros momentos espectrales (centro de gravedad, desviación estándar, asimetría y curtosis).

**Pico espectral:** Frecuencia (Hz) asociada a la mayor densidad de energía del sonido fricativo.

**Centro de gravedad:** Frecuencia (Hz) que divide el espectro en dos mitades, tales que, la cantidad de energía en la mitad superior (frecuencias altas), es igual a la cantidad de energía en la mitad inferior (frecuencias bajas).

**Desviación estándar:** Medida de cuánto pueden desviarse del centro de gravedad las frecuencias de un espectro.

**Asimetría:** Medida de cuánto difiere la forma del espectro por debajo del centro de gravedad de la forma por encima de la frecuencia media.

**Curtosis:** Medida de cuánto difiere la forma del espectro alrededor del centro de gravedad de la distribución de Gauss.

### 2.3.3. Transiciones formánticas

En este apartado se comentarán aquellos parámetros relacionados con los formantes. Los formantes son picos de intensidad en el espectro de un sonido.

**F2 de transición:** Valor del segundo formante en la zona de transición entre el sonido fricativo y la vocal.

**Ecuación de locus:** Coeficientes de las ecuaciones de regresión de cada consonante aplicadas a las medidas del segundo formante a través del contexto vocálico, siendo el eje de ordenadas el valor de F2 en el inicio de la vocal y el eje de coordenadas el valor de F2 en la mitad de la vocal.

Otro parámetro relacionado con los formantes es el F1 de transición. Este parámetro, aunque no sirve para diferenciar significativamente las fricativas del inglés, si parece ser útil a la hora de diferenciar entre el sonido uvular /X/ y el sonido faríngeo /ѐ/, característico del árabe. Esto es mencionado por el propio Al-Khairy [Al-Khairy, 2005] en su repaso de la literatura. También, respecto a este contexto formántico, otros trabajos como los de de Al-Masri [Al-Masri et al, 2007] o Abudalbuh [Abudalbuh, 2011] amplían el estudio a más formantes y más zonas con el fin de determinar su influencia en la faringealización.

No obstante, teniendo en cuenta los resultados de los citados trabajos, este proyecto solo se ha centrado en la investigación de Al-Khairy [Al-Khairy, 2005] ya que sus parámetros se consideran suficientes para discernir entre las consonantes involucradas en este trabajo.

## 2.4. Resumen

En este capítulo se ha visto como el árabe es un idioma con distintos dialectos y cuya fonética es extensa y variada, teniendo características únicas como la faringealización. También se ha hecho una clasificación acústica de las consonantes del árabe, procediendo a hacer un énfasis en las que involucra este proyecto, las fricativas, y su relación con el concepto de turbulencia.

Además, se hace un resumen de los parámetros influyentes en la diferenciación de las consonantes árabes a partir del trabajo realizado por Al-Khairy [Al-Khairy, 2005]. Con esta información, en el Capítulo 3 se seleccionarán qué parámetros de los expuestos son los necesarios para discernir únicamente entre las consonantes involucradas en este proyecto y si, en efecto, son estadísticamente significativos en este trabajo, teniendo en cuenta además el género del emisor de la voz.

# Capítulo 3

# Extracción de parámetros

En el presente capítulo se analizan qué parámetros se han seleccionado como determinantes para definir y diferenciar entre las distintas consonantes involucradas en este proyecto, especificando los métodos y criterios empleados. Además, se detallan los métodos y las herramientas empleadas para la extracción de dichos parámetros. Por último, también se explica cómo se ha realizado la recolección de datos, es decir, la forma en la que se han grabado y enviado las letras pronunciadas de las cuales se extraen los parámetros.

## 3.1. Praat

La herramienta utilizada para extraer los parámetros característicos de las distintas consonantes es Praat, un programa gratuito para el análisis, sintetizado y manipulado fonético del habla. Fue desarrollado en 1992 por Paul Boersma y David Weenink en el Instituto de Ciencias Fonéticas de la Universidad de Ámsterdam [Boersma et al, 2001]. Su descarga está disponible, junto a la documentación para su uso, en la página web del programa (Praat) en distintas versiones para los diferentes sistemas operativos (Macintosh, Windows, Linux, entre otros).

Para el análisis, Praat permite grabar los sonidos desde el propio programa o importar el archivo en formato WAV. Una vez el audio esté en el programa, se puede seleccionar para su visualización. En la ventana superior se puede observar el oscilograma (forma de la onda) que viene representado como la amplitud en función del tiempo. La ventana inferior permite observar varios análisis acústicos: el espectrograma (representación en escala de grises de la cantidad de altas y bajas frecuencias de la señal), los formantes (representados por puntos rojos), la curva melódica (representada en color cían) y la curva de intensidad (representada en color amarillo) [Llisterri, 2020]. En la Figura 3.1 se muestra un ejemplo de una ventana de análisis completa de la letra /sīn/.

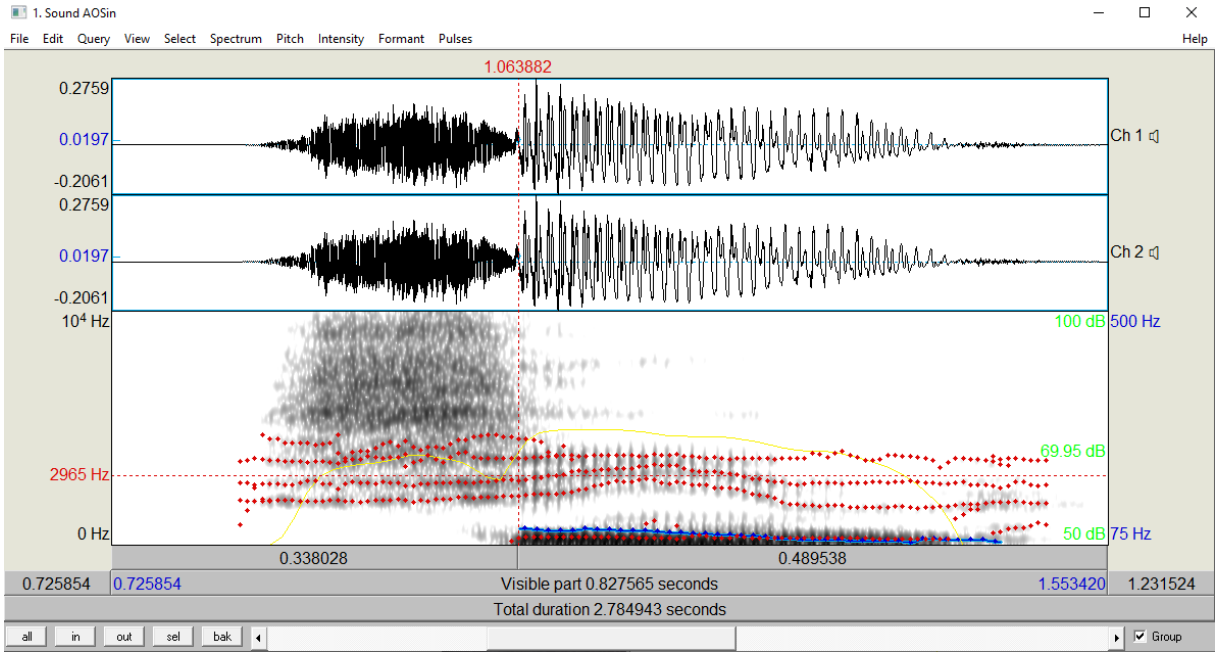


Figura 3.1: *Ventana de análisis en Praat*

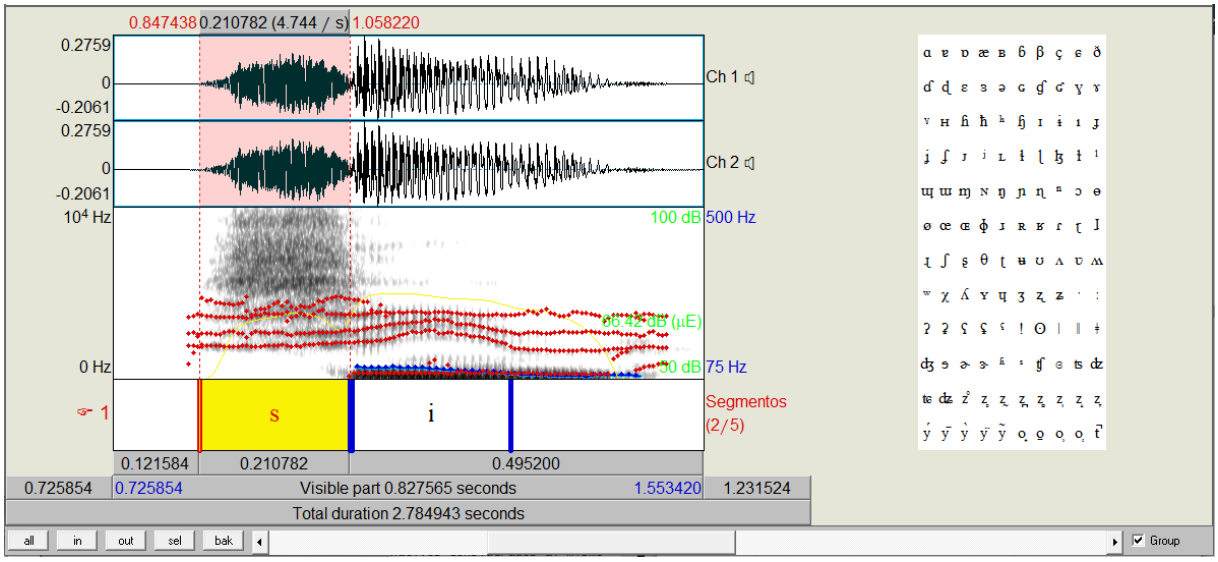


Figura 3.2: *Ejemplo de TextGrid en Praat*

### 3.1.1. TextGrids y Tiers

En ocasiones es necesario analizar una o varias unidades de un sonido en vez de la señal completa. Al proceso por el cual se delimitan estas unidades se le denomina *segmentación*. A estas unidades delimitadas también se les puede realizar un etiquetado con el símbolo apropiado para su descripción. En Praat, estos dos procesos se pueden realizar creando un documento de texto conocido como TextGrid. Dentro de un mismo TextGrid, se pueden hacer segmentaciones diferentes en distintos niveles llamados Tiers. En la Figura 3.2 se observa un ejemplo de segmentación en un TextGrid del mismo sonido /sīn/.

### 3.1.2. Scripts de Praat

Para optimizar un proceso de análisis, en el que muchas veces se requiere obtener parámetros de distintas muestras de forma reiterativa, es conveniente automatizarlo. En Praat esto es posible gracias a los Scripts, rutinas ejecutables dentro del mismo programa escritos en el lenguaje propio de programación de Praat. Son documentos de texto que se guardan en el disco duro del ordenador con extensión `.praat` que posteriormente se pueden abrir y ejecutar desde el programa.

## 3.2. Diseño del experimento

Una de las partes más importantes en los trabajos de investigación es la correcta recolección de datos. Este proyecto, al involucrar grabaciones de sonido como fuente de datos, requería que estas fuesen grabadas de la manera más homogénea y óptima posible (desde un punto de vista acústico). No obstante, debido a las circunstancias extraordinarias del marco temporal en el que se realiza este proyecto, las muestras tuvieron que ser grabadas a distancia por cada individuo con su propio dispositivo distinto. Otro problema adicional fue la falta de conocimiento clave sobre la materia en el inicio del proyecto, que implicó que no se recogiera información útil como la combinación de las consonantes con todas las vocales.

### 3.2.1. Participantes

En este proyecto, la participación se puede dividir en dos tipos: por un lado, se tienen a los participantes en la clasificación, también conocida como entrenamiento del modelo, conformados por los individuos de idioma nativo árabe, y por el otro lado, a los participantes en la fase del estudio, conformados por individuos cuyo idioma nativo es el catalán, el castellano o el italiano.

Para la parte de clasificación, participaron 12 hombres y 12 mujeres de 24 y 31 años de media, respectivamente. Respecto a los dialectos involucrados, 17 eran sirios (71 %), cuatro jordanos (17 %), dos iraquíes (8 %) y un yemení (4 %). Aunque es cierto que puede haber ligeras variaciones de pronunciación entre dialectos, en este proyecto no se han considerado relevantes.

Para la fase del estudio, participaron 15 personas de los tres tipos distintos de idiomas nativos: cinco personas con el catalán y castellano como idiomas nativos considerados al mismo nivel (dos hombres y tres mujeres), cinco con idioma nativo castellano, pero con idioma secundario el catalán de dialecto valenciano (tres hombres y dos mujeres) y, por último, cinco personas con idioma nativo italiano (tres hombres y dos mujeres). Se considera idioma secundario cuando no se usa regularmente en el día a día.

### 3.2.2. Método de grabación y envío

Para las grabaciones, se realizaron unas instrucciones que orientaban a los participantes en cómo realizarlas y enviarlas. En estas, se detallaba que se grabasen pronunciando cada letra en audios separados con un ordenador o un dispositivo móvil, a una distancia que considerasen normal y

en un lugar lo más silencioso posible. Por último, se pedía que se enviasen los audios a un correo electrónico creado específicamente para este trabajo. Para la petición de muestras se usaron distintas aplicaciones de mensajería y redes sociales.

Para los participantes en la parte clasificación, inicialmente se pidió que se grabasen diciendo el nombre de ocho consonantes fricativas de forma normal, sin alargar ni el sonido de la consonante ni la vocal. De estas ocho, se descartaron rápidamente tres (ver Sección 3.3). A los participantes en la fase del estudio, se les pidió que grabasen las cinco consonantes restantes intentando reproducir lo más fielmente posible el sonido de las muestras originales que se les había enviado. Estas muestras pertenecen a un participante en la fase de clasificación.

### 3.2.3. Homogeneización de las muestras

Un punto muy importante a tener en cuenta respecto al tratado de muestras es que sean homogéneas, sobre todo teniendo en cuenta el distinto origen de cada muestra en este proyecto. Como cada grabación se hizo con un dispositivo diferente, la calidad podía diferir, lo que se traduce en distintas frecuencias de muestreo. Por esto, el primer proceso que se aplicó a las muestras fue el remuestreo.

El teorema del muestreo o de Shanon[1] define que si una señal (e.g., una onda sonora compleja) no contiene frecuencias superiores a una frecuencia determinada $f_c$ [Hz], esta quedará completamente determinada por sus valores medidos en instantes de tiempo separados si

$$f_s \geq 2f_c, \tag{3.1}$$

donde $f_s$ es la frecuencia de muestreo. Por tanto, asumiendo que las frecuencias de muestreo de las grabaciones cumplen el teorema de Shanon, lo único que queda por verificar es que todas estén muestreadas a la misma frecuencia $f_s$. En este caso, esta frecuencia es de 44,1 kHz. Las muestras que ya se encontraban a esta frecuencia no se modificaron, mientras que las que se encontraban a una frecuencia mayor, concretamente a 48 kHz, se redujeron mediante Praat.

El segundo proceso de homogeneización que se aplicó fue la reducción de ruido. Esto fue realizado, además, como método para mejorar la calidad de las muestras. Concretamente, se aplicó un filtro pasa banda de 80 Hz a 11 kHz de frecuencia y se redujo el ruido en 20 dB. Esto fue implemento también en el programa Praat. Previamente a estos dos procesos, se tuvieron que convertir todos los archivos a formato WAV mediante un conversor online puesto que es el formato que Praat admite.

## 3.3. Selección de consonantes y análisis de parámetros

Para la selección de parámetros, el principal objetivo que se tuvo en cuenta fue que estos fuesen capaces de separar, de la manera más óptima posible, las consonantes involucradas en el trabajo. Para esto, se utilizaron dos criterios: el primero, consideraba una preselección de parámetros basado en los resultados obtenidos por Al-Khairy [Al-Khairy, 2005] y, el segundo, en un análisis

[1] Apuntes de Control Automático. GETI, ETSEIB.

de la varianza (ANOVA) de cada uno de estos parámetros para comprobar si, en efecto, se cumplían dichos resultados en las muestras poblacionales más reducidas de este estudio, es decir, verificar la significancia de los parámetros.

Inicialmente, el proyecto incluía el estudio de ocho consonantes fricativas (/s, S, sQ, X, è, h, z, Z/). No obstante, durante la extracción de los primeros parámetros, se descartaron /z/, /Z/, y /h/ puesto que la diferenciación con la vocal adyacente en la segmentación resultaba prácticamente imposible, por lo que no se podían estudiar de una manera certera dichos parámetros. Adicionalmente, durante los estudios de clasificación preliminares, se observó que, aunque /è/ se diferenciaba correctamente de las demás consonantes, las muestras de /X/ eran muy dispersas, confundiéndose con todas las consonantes. Debido a esto, aunque hubiese sido interesante incluir /è/ en el estudio posterior ya que es una consonante característica y única del árabe, se decidió eliminar /è/ y /X/ del proyecto.

Las consonantes fricativas del proyecto son, por tanto: la alveolar /s/, presente en todos los idiomas involucrados en el trabajo (y en la mayoría de idiomas del mundo), su contraparte enfática /sQ/, únicamente presente en el árabe, y la post-alveolar /S/, ausente en el castellano, pero presente en catalán e italiano.

Para la extracción de los parámetros se utilizaron dos Scripts de Praat ya creados y de libre uso. El primero de ellos es *zero-crossings-and-spectral-moments*[2]. Este analiza todos los ficheros de una carpeta que contienen los archivos de audio y los TextGrids con el segmento del sonido que se quiere analizar, el cual se debe especificar cuando se ejecuta. Este Script extrae y almacena en un archivo 12 parámetros: el inicio, final y duración del intervalo (ms), los cruces por cero dB en los primeros 30 ms, los cruces por cero dB en todo el intervalo y los cruces por cero dB por 10 y entre la duración del segmento, la intensidad, los cuatro momentos espectrales y el momento central. Además, permite la opción de aplicar un filtro pasa banda recomendado de 1 a 11 kHz. En lo referente a este proyecto, dado que ya se habían atenuado las frecuencias por encima de 11 kHz, este filtro solo afectó a las bajas frecuencias.

El segundo de ellos es *extracción-de-formantes-en-tabla*[3]. Este Script obtiene un archivo con los tres primeros formantes en el punto medio de cada segmento del Tier que se especifique. También se han de especificar la carpeta en la que se encuentran los audios, en la que se encuentran los TextGrids, donde se quiere que se guarde el archivo y el nombre. Por último, para el parámetro *valor máximo del formante más alto* hay que utilizar el valor 5000 Hz para las voces masculinas y 5500 Hz para las femeninas, según señala el autor[3] siguiendo la recomendación de Praat.

### 3.3.1. ANOVA y Minitab

El análisis de la varianza, también conocido como ANOVA [4], es un método empleado para la comparación de medias de distintos grupos de muestras. El modelo supone que todos los datos de cada grupo proceden de una distribución Normal con la misma varianza constante $\sigma^2$. El método consiste en estimar esta varianza $\sigma^2$ de dos modos distintos, a partir de la variabilidad

[2]Wendy Elvira-Garcia (2014). Laboratorio de fonética, Universidad de Barcelona.
[3]Joaquim Llisterri (2016). Laboratorio de fonética, Universidad Autónoma de Barcelona.
[4]Apuntes de Estadística. GETI, ETSEIB.

dentro y entre los grupos. Suponiendo que se tienen *k* grupos donde cada grupo *t* tiene $n_t$ muestras, se considera que los datos de los grupos responden al modelo

$$y_{ti} = \mu + \tau_t + \epsilon_{ti}, \tag{3.2}$$

donde $y_{ti}$ es el valor de la muestra *i* del grupo *t*, $\nu$ es la media general de todos los datos, $\tau_t$ es la variación de la respuesta del grupo *t*respecto a $\mu$ (cuyo sumatorio desde el primer hasta el *k* grupo es 0 para los modelos de *efectos fijos*), $\epsilon_{ti}$ es la variabilidad aleatoria asociada a cada muestra que se supone Normal de varianza $\sigma$ con todos los valores independientes entre sí. La estimación $\sigma^2$ de variabilidad dentro de los grupos consiste en calcular $S_R^2$, cuya expresión es

$$S_R^2 = \frac{\sum_{t=1}^{k}\sum_{i=1}^{n_t}(y_{ti} - \bar{y}_t)}{N-k}, \tag{3.3}$$

donde $N$ es el número total de muestras, mientras que la estimación mediante variabilidad entre grupos, asumiendo que todos los grupos tienen el mismo número de muestras *n*, pasa por el cálculo de $S_T^2$ que se define por la expresión

$$S_T^2 = \frac{n\sum_{t=1}^{k}(\bar{y}_t - \bar{y})^2}{k-1}. \tag{3.4}$$

El contraste de hipótesis sobre la comparación de las medias, en el que la hipotesis nula $H_0$ consiste en que todas las medias sean iguales y la alternativa $H_0$ en que exista alguna media distinta, se realiza a través de la comparación de las dos estimaciones distintas de $\sigma^2$. Si los valores de $S_R^2$ y $S_T^2$ son similares, no se puede descartar la hipótesis nula $H_0$, mientras que, si son distintas, concretamente si $S_T^2$ tiene un valor considerablemente mayor a $S_R^2$, se tiene que sospechar su incumplimiento. El test de las dos varianzas se realiza mediante la distribución F de Snedecor.

Para este proyecto, interesa conocer también la influencia del tipo de voz según el género del emisor en los valores de los parámetros en cada consonante y su interacción con estas. Por tanto, se realizará un análisis de la varianza de dos factores (voz y consonante), también conocido como *two-way* ANOVA. El modelo es similar al caso anterior, solo hay que añadir la influencia de cada factor.

La herramienta utilizada para realizar el análisis es Minitab, un programa informático diseñado para ejecutar funciones estadísticas básicas y avanzadas. En Minitab, el análisis ANOVA entrega una tabla con el p-valor de cada factor. Si el p-valor resulta menor que el nivel de significación $\alpha$=0.05, se rechaza la hipótesis nula $H_0$ y se considera que el factor tiene una influencia significativa y diferenciadora.

Previamente a aplicar el método ANOVA, hay que verificar el cumplimiento de las hipótesis del modelo. En Minitab esto se puede realizar obteniendo e interpretando cuatro gráficas de residuos:

- Gráfica de probabilidad Normal para comprobar que no hay indicios de que los datos no siguen una distribución Normal.

- Histograma para comprobar la Normalidad de los residuos.
- Gráfica de residuos contra valores previstos para verificar que la variabilidad no aumenta con el nivel de respuesta. De ser así, se incumpliría la hipótesis de varianza constante.
- Gráfica de residuos frente al orden de recogida de los datos para verificar la independencia de los residuos.

Por tanto, para cumplir las hipótesis del modelo, el primer gráfico tiene que mostrar como los puntos siguen la línea de normalidad, el histograma ha de tener una forma que se asemeje a la distribución Normal, el tercer gráfico no ha un aumento de la variabilidad frente al nivel de la respuesta, y por último, el cuarto gráfico no tiene que presentar ninguna tendencia o patrón reconocible de los residuos frente al orden.

### 3.3.2. Pico espectral

El cálculo del pico espectral se centró en todo sonido fricativo. Aunque los estudios previos sugerían un estudio en dos zonas diferenciadas (mitad y final del sonido fricativo), Al-Khairy [Al-Khairy, 2005] no observó diferencias significativas entre las dos zonas, por lo que solo reportó los resultados de la zona media. Debido a esto, y a la optimización de tiempo, puesto que la segmentación fue manual, se decidió hacer el análisis en todo el sonido fricativo. En su trabajo, este parámetro mostró diferencias significativas entre /s/ y /ʃ/ pero no entre /s/ y su contraparte enfática /sˤ/.

Para su extracción, se creó un segmento de la consonante de cada letra mediante TextGrids y se utilizó el Script *zero-crossing-and-spectral-moments* en la carpeta correspondiente especificando el nombre del segmento. Esta parte del Script de Praat convierte el sonido en un objeto del tipo Ltas (abreviatura inglesa para *Espectro promedio a largo plazo*), el cual representa la densidad espectral de potencia logarítmica en función de la frecuencia, con un ancho de banda de 150 Hz. Se obtiene el pico espectral (frecuencia asociada al pico de amplitud) con una aproximación cúbica.

Realizando un primer análisis exploratorio de datos (Figura 3.3), se pueden extraer algunas conclusiones a priori. En las muestras de voz femenina, la dispersión de valores dentro de cada consonante varía bastante, siendo /ʃ/ la de menor variación y /sˤ/ la de mayor. Además, el rango de valores de /s/ se encuentra incluido en el rango de valores de /sˤ/, por lo que no hay diferencias entre estas en lo referente a este parámetro. No obstante, sí que hay diferencias con /ʃ/, que presenta valores claramente inferiores. En cambio, las voces masculinas muestran una dispersión más constante dentro de cada consonante y, aunque sí parece haber cierta confusión entre consonantes, se observa una cierta transición más escalonada en el rango de valores. Las medianas de las voces masculinas son inferiores a las femeninas.

Ejecutando el análisis *two-way* ANOVA (Figura 3.4) se obtiene en la tabla de resultados que el p-valor de la voz es de 0.004 y el de la consonante de 0.000. Por tanto, esto quiere decir que se rechazan las hipótesis nulas, lo que implica que el pico espectral es un buen parámetro para diferenciar, al menos, una de las consonantes del resto y, que el tipo de voz (masculina o femenina) es influyente en el resultado. La interacción, con un p-valor igual a 0.180, no es significativa. Este análisis es válido puesto que, siguiendo los criterios establecidos en la Sección 3.3.1, las gráficas

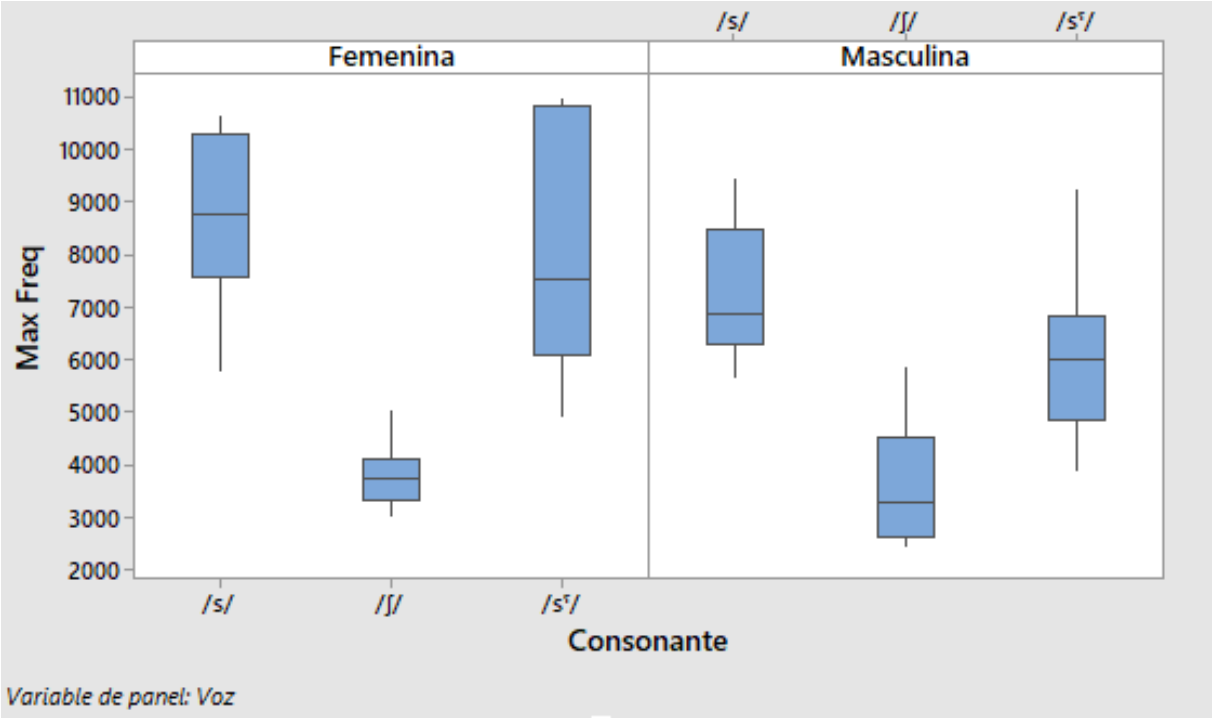


Figura 3.3: Boxplot del Pico Espectral

**Análisis de Varianza**

| Fuente | GL | SC Sec. | Contribución | SC Ajust. | MC Ajust. | Valor F | Valor p |
|---|---|---|---|---|---|---|---|
| Voz | 1 | 21880215 | 5,03% | 21880215 | 21880215 | 9,16 | 0,004 |
| Consonante | 2 | 246614739 | 56,75% | 246614739 | 123307370 | 51,62 | 0,000 |
| Voz*Consonante | 2 | 8421374 | 1,94% | 8421374 | 4210687 | 1,76 | 0,180 |
| Error | 66 | 157653486 | 36,28% | 157653486 | 2388689 | | |
| Total | 71 | 434569814 | 100,00% | | | | |

Figura 3.4: 2-way ANOVA del Pico Espectral

de los residuos, visibles en el Anexo 1 (Figura 7.1), no muestran indicios de un incumplimiento de las hipótesis del modelo.

Por último, para ver mejor el comportamiento de este parámetro, se pueden extraer gráficos de interacción (Figura 3.5). En estos gráficos se puede observar que, en ambos géneros, los valores de las medias de /s/ y /sᵠ/ son similares mientras la de /S/ es claramente inferior. En cuanto a las diferencias en el género, ambas consonantes alveolares bajan el valor de la media con una pendiente similar, mientras que la post-alveolar /S/ es prácticamente igual para ambos tipos de voz.

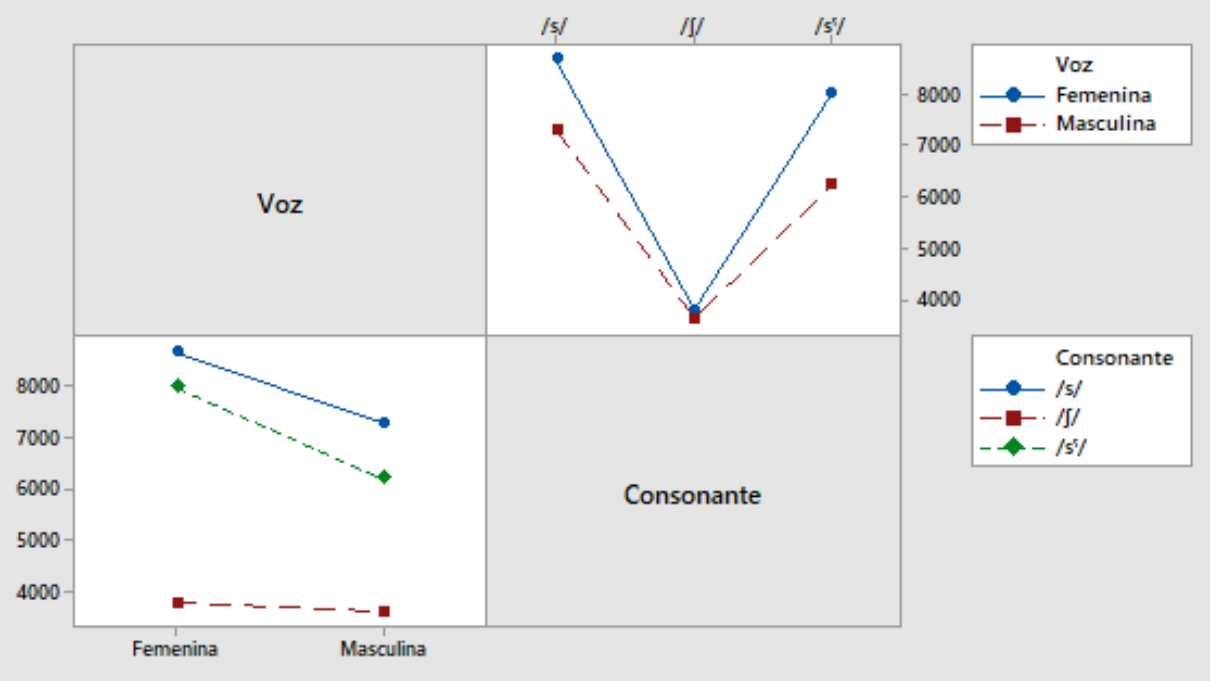


Figura 3.5: Interacción Voz-Consonante del Pico Espectral

### 3.3.3. Centro de gravedad

Al igual que el caso anterior, se obtuvo el centro de gravedad de todo el sonido fricativo. Contrariamente al pico espectral, Al-Khairy [Al-Khairy, 2005], siguiendo la literatura anterior, sí encontró diferencias significativas según las distintas zonas (inicio, mitad, final del sonido fricativo y transición con la vocal adyacente). No obstante, en este trabajo, tras observar que los valores de este parámetro dentro de una misma consonante tenían una tendencia clara sin grandes fluctuaciones y, por optimización del tiempo, puesto que segmentar manualmente todas las consonantes en cuatro unidades iguales habría llevado mucho tiempo, se decidió hacer este estudio sobre el sonido fricativo completo. Al igual que el pico espectral, este parámetro le sirvió para diferenciar entre /s/ y /ʃ/ pero no entre /s/ y /sˢ/.

Para su extracción, también se utilizó el Script *zero-crossing-and-spectral-moments* y los TextGrids anteriores en la misma carpeta, obteniendo los resultados junto al pico espectral y el resto de parámetros que entrega el Script. Praat calcula el centro de gravedad siguiendo la expresión

$$\frac{\int_0^\infty f|S(f)|^p\,df}{\int_0^\infty |S(f)|^p\,df}, \tag{3.5}$$

donde $S(f)$ es el espectro complejo, $f$ es la frecuencia [Hz] y el denominador es la energía [J]. En el Script, la cantidad $p$ se establece en 2, lo que implica que la frecuencia media es ponderada por la potencia del espectro en vez del espectro absoluto.

Realizando el análisis exploratorio inicial (Figura 3.6), se puede observar una tendencia parecida a la del parámetro anterior (pico espectral). En las voces femeninas, los valores de /s/ y /sˢ/ no muestran diferencia, pero en este caso, los valores de /s/ abarcan un rango tan elevado que llegan a confundirse con valores de /ʃ/, en principio muy alejados e inferiores. Para las muestras

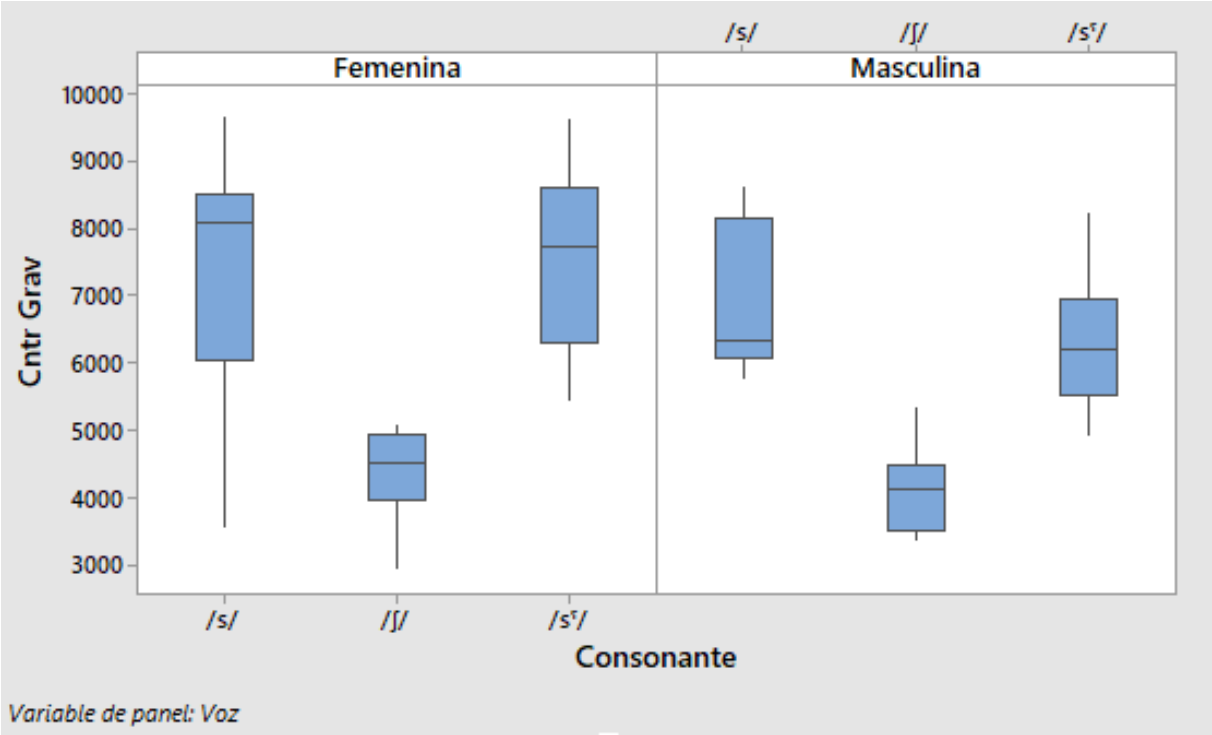


Figura 3.6: Boxplot del Centro de Gravedad

**Análisis de Varianza**

| Fuente | GL | SC Sec. | Contribución | SC Ajust. | MC Ajust. | Valor F | Valor p |
|---|---|---|---|---|---|---|---|
| Voz | 1 | 8640041 | 3,83% | 8640041 | 8640041 | 6,48 | 0,013 |
| Consonante | 2 | 125358261 | 55,54% | 125358261 | 62679131 | 47,04 | 0,000 |
| Voz*Consonante | 2 | 3777712 | 1,67% | 3777712 | 1888856 | 1,42 | 0,250 |
| Error | 66 | 87938561 | 38,96% | 87938561 | 1332402 | | |
| Total | 71 | 225714575 | 100,00% | | | | |

Figura 3.7: 2-way ANOVA del Centro de Gravedad

de voz masculina, las conclusiones son las mismas que antes, solo que en este caso, /S/ parece ligeramente más aislada. Esta vez, a excepción de la /s/ femenina, la dispersión de valores es más constante. De nuevo, las medianas de las voces masculinas son claramente inferiores a las femeninas. Esta diferencia es menos notable en la consonante /S/.

El análisis *two-way* ANOVA (Figura 3.8) muestra un p-valor igual a 0.013 y 0.000 para la voz y la consonante, respectivamente. Esto quiere decir de nuevo, que este parámetro es útil para diferenciar, al menos una de las consonantes del resto y que el género influye en los valores. El p-valor de la interacción de ambos factores es igual a 0.250, por lo que se deduce que la interacción vuelve a no ser relevante. Al igual antes, es correcto realizar este análisis ya que las gráficas de los residuos, que se observan en el Anexo 2 (Figura 7.2), cumplen con los requisitos establecidos en la Sección 3.3.1 y, por tanto, el modelo cumple con las hipótesis supuestas.

Esta vez, al crear los gráficos de interacción (Figura 3.7), se puede observar que la consonante

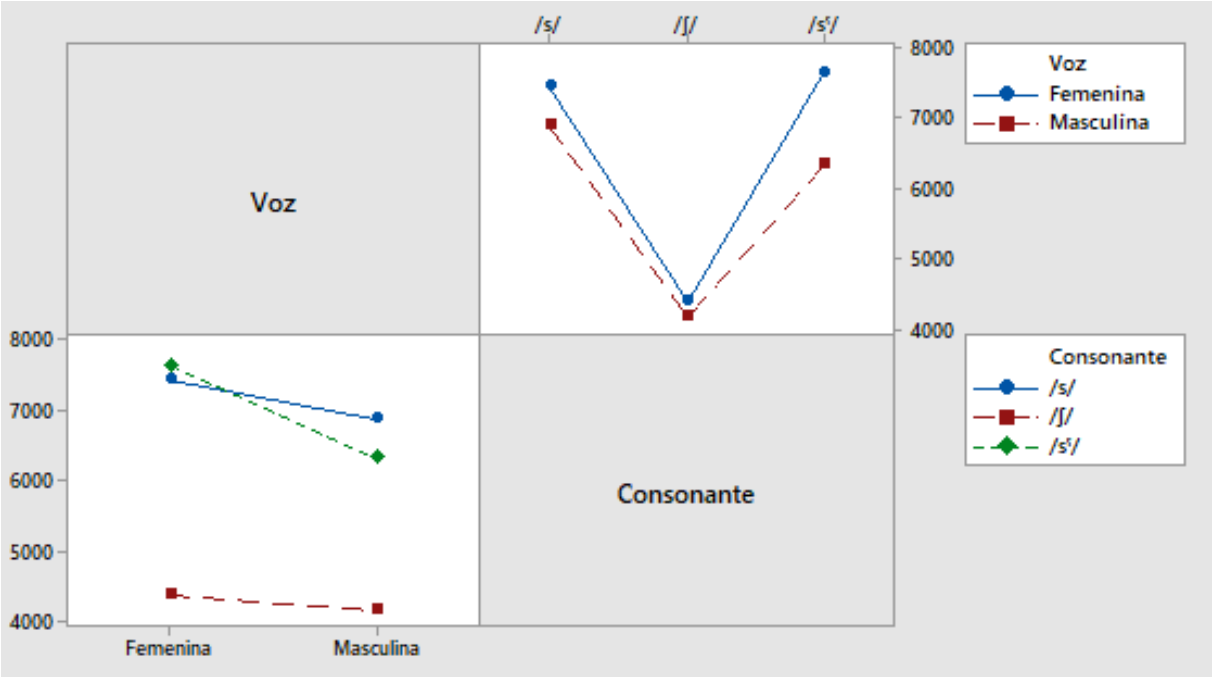


Figura 3.8: Interacción Voz-Consonante Centro de Gravedad

post-alveolar tiene claramente una media mucho inferior a las otras dos que presentan valores indiferenciados. De nuevo, las medias de /S/ son bastante similares para hombres y mujeres y las otras dos disminuyen su valor en hombres. No obstante, en este caso la pendiente es mucho más pronunciada en la alveolar enfática, llegando a tener valores inferiores a su contraparte simple en voces masculinas, al contrario de lo que pasa en las voces femeninas, aunque en estas, la diferencia es bastante inferior.

### 3.3.4. Cruces por cero dB por 10 entre duración del intervalo

El número de cruces por cero dB multiplicado por 10 y entre la duración del intervalo en milisegundos fue el único parámetro que no se seleccionó teniendo en cuenta la literatura anterior sino la experimentación. Al tratarse de un parámetro que entrega por defecto el Script *zero-crossing-and-spectral-moments*, se observó que los valores dentro de las mismas consonantes tendían a un valor sin muchas fluctuaciones y, que la consonante /S/ parecía tener valores inferiores a las otras dos. Por tanto, se decidió incluir este parámetro y comprobar si, en efecto, servía para diferenciar las consonantes.

El análisis a priori de los datos utilizando Boxplots (Figura 3.9) muestra unos resultados cualitativamente similares a los parámetros anteriores: las consonantes alveolares muestran valores semejantes mientras los valores de la post-alveolar son menores independientemente del género de la voz. Las medianas de los hombres también parecen inferiores.

Al realizar el análisis *two-way* ANOVA se puede observar de nuevo en la tabla (Figura 3.10), que se ha de rechazar la hipótesis nula tanto para la voz como para la consonante, con valores de p-valor iguales a 0.002 y 0.000 respectivamente, por lo que se concluye que ambos factores son significativos. La interacción, con un p-valor igual a 0.445, al igual que los casos anteriores,

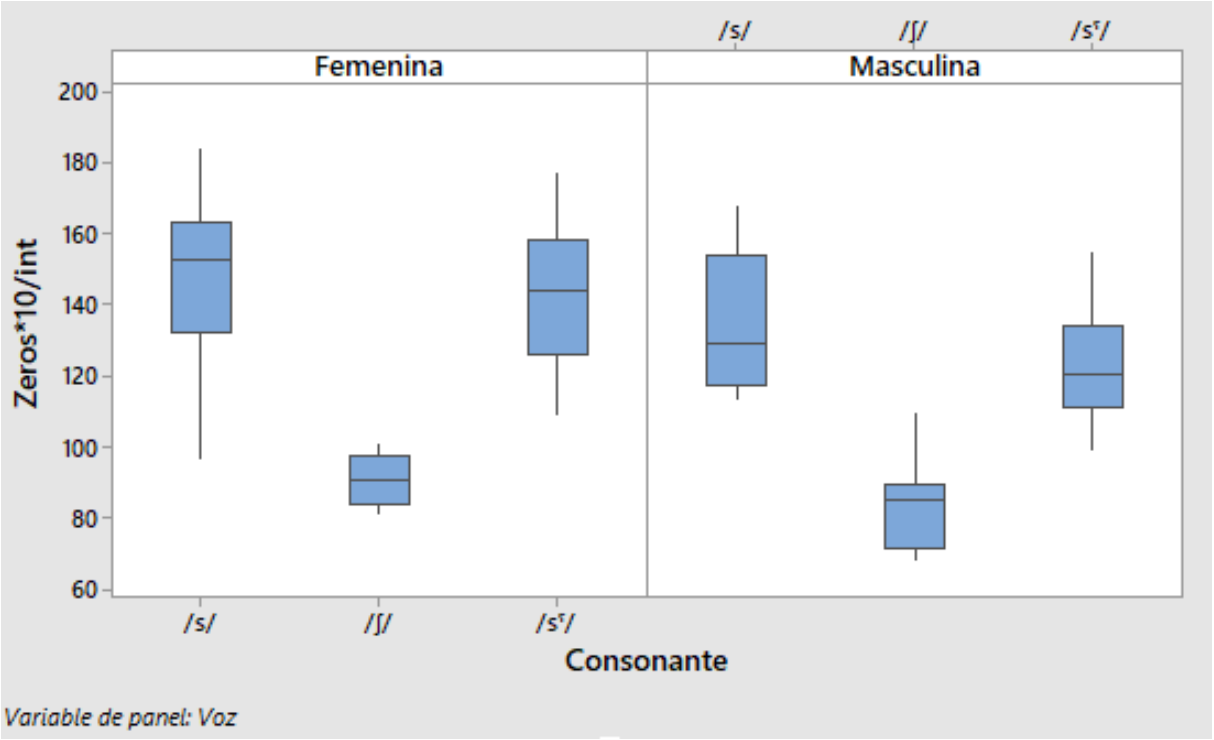


Figura 3.9: Boxplot de Cruces por Cero

**Análisis de Varianza**

| Fuente | GL | SC Sec. | Contribución | SC Ajust. | MC Ajust. | Valor F | Valor p |
|---|---|---|---|---|---|---|---|
| Voz | 1 | 3188,3 | 4,98% | 3188,3 | 3188,3 | 10,08 | 0,002 |
| Consonante | 2 | 39459,1 | 61,62% | 39459,1 | 19729,5 | 62,40 | 0,000 |
| Voz*Consonante | 2 | 518,8 | 0,81% | 518,8 | 259,4 | 0,82 | 0,445 |
| Error | 66 | 20868,3 | 32,59% | 20868,3 | 316,2 | | |
| Total | 71 | 64034,4 | 100,00% | | | | |

Figura 3.10: 2-way ANOVA de Cruces por Cero

tampoco es relevante. Previamente a realizar este análisis hay que comprobar las hipótesis del modelo siguiendo la interpretación gráfica explicada en la Sección 3.3.1. De nuevo, las gráficas de los residuos, que se encuentran en el Anexo 3 (Figura 7.3), sugieren el cumplimiento de estas hipótesis.

Por último, los gráficos de intersección (Figura 3.11) vuelven a mostrar la tendencia aditiva de los parámetros anteriores, con los tres valores descendiendo al pasar de voces femeninas a masculinas y siendo la más diferenciada con valores inferiores la post-alveolar y las otras dos las más similares entre sí, aunque /s$^{Q}$/ es ligeramente superior. Queda claro, por tanto, que este parámetro es efectivo en la separación de la consonante /$/, así que, se incluye en la clasificación aunque no salga en estudios anteriores.

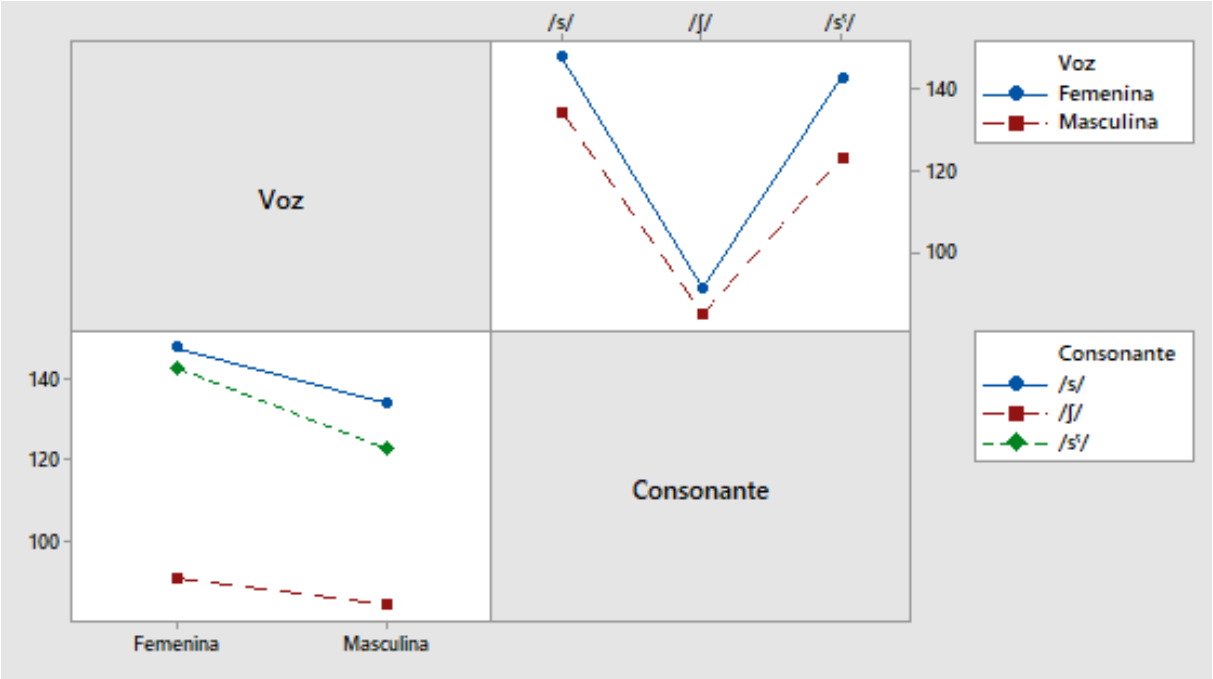


Figura 3.11: Intersección Voz-Consonante de Cruces por Cero

### 3.3.5. Amplitud relativa

Como se mencionó en la Sección 2.3.1, la amplitud relativa se define como la diferencia entre la amplitud de una frecuencia específica en el punto medio de la consonante fricativa y la amplitud de la correspondiente frecuencia en el inicio de la vocal. Esta frecuencia corresponde a F2 para la consonante /sˠ/ y a F3 para las consonantes /s/ y /ʃ/.

El proceso desarrollado para calcular este parámetro siguió los siguientes pasos:

1. Obtener los formantes en el punto medio de la consonante usando el Script *extracción-de-formantes-en-tabla* mediante los TextGrids utilizados para la extracción de los parámetros anteriores y, posteriormente, eliminar del archivo los correspondientes al punto medio de la vocal puesto que en el mismo Tier se habían segmentado la consonante y la vocal.

2. Obtener los formantes con el mismo Script usando TextGrids en los cuales se segmentaba la zona correspondiente al primer periodo en el que la vocal comenzaba a tener una apariencia de onda periódica constante.

3. Extraer los segmentos de las consonantes y las vocales de las grabaciones completas, obteniendo nuevos audios separados de cada segmento.

4. Transformar esos segmentos a Ltas y obtener las amplitudes correspondientes a cada segmento utilizando los formantes obtenidos anteriormente.

5. Restar las amplitudes correspondientes al inicio de la vocal a las amplitudes correspondientes a la mitad de la consonante.

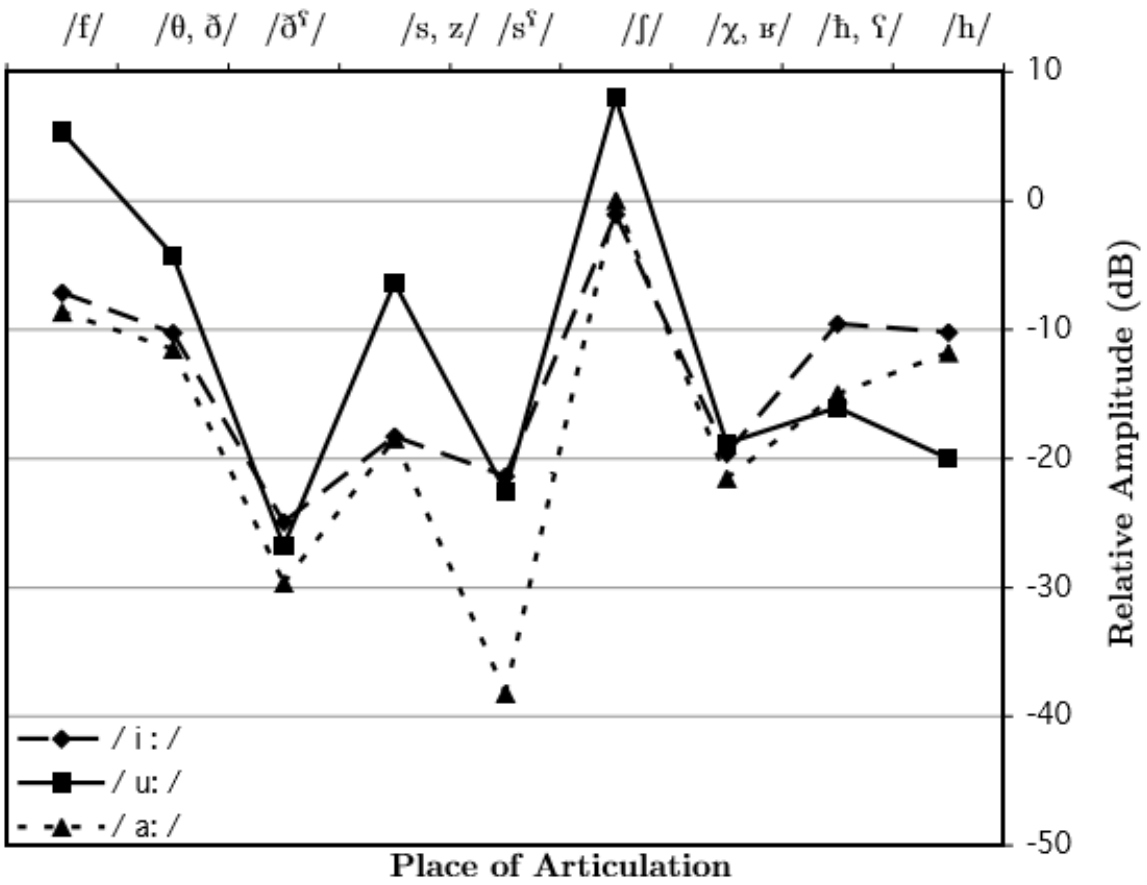


Figura 3.12: Amplitud relativa en función de la consonante y la vocal larga [Al-Khairy, 2005]

Uno de los puntos más cruciales a tener en cuenta en el uso e interpretación de este parámetro es la influencia de la vocal adyacente. Se ha de recordar que este trabajo no contempló la relevancia de las vocales en la fase del diseño del experimento, por lo que cada consonante va acompañada de la vocal del nombre original de la letra. En lo referente a este trabajo, esto se traduce en que las consonantes /s/ y /S/ van acompañadas de la vocal /i:/ y la consonante /sˤ/ de la vocal /a:/.

En principio, este problema deshabilitaría el uso de este parámetro puesto que la comparación entre consonantes no sería adecuada ya que los valores en cada fonema difieren según la vocal adyacente, lo que supone una confusión de efectos. No obstante, los resultados obtenidos por Al-Khairy [Al-Khairy, 2005] ayudan a demostrar que, en este caso particular, teniendo en cuenta el objetivo del proyecto, este problema supone, en realidad, una ventaja. Cabe recordar que este objetivo pasa por conseguir un modelo certero que separe y prediga con precisión únicamente las tres consonantes con las que se trabaja. De esta forma, el estudio posterior será verídico.

En la Figura 3.12, se puede observar que el valor de la amplitud relativa de las tres consonantes acompañadas de /a:/ difiere significativamente entre ellas. La diferencia es de aproximadamente 20 dB entre cada una de ellas, siendo las medias de: 0 dB para /S/, -20 dB para /s/ y -40 dB para /sˤ/. Los valores de la media con la vocal /i:/ se mantienen igual que con la vocal /a:/ para las consonantes /s/ y /S/. No obstante, la media de /sˤ/ con la /i:/ es prácticamente igual a las de /s/ con la /a:/ o la /i:/.

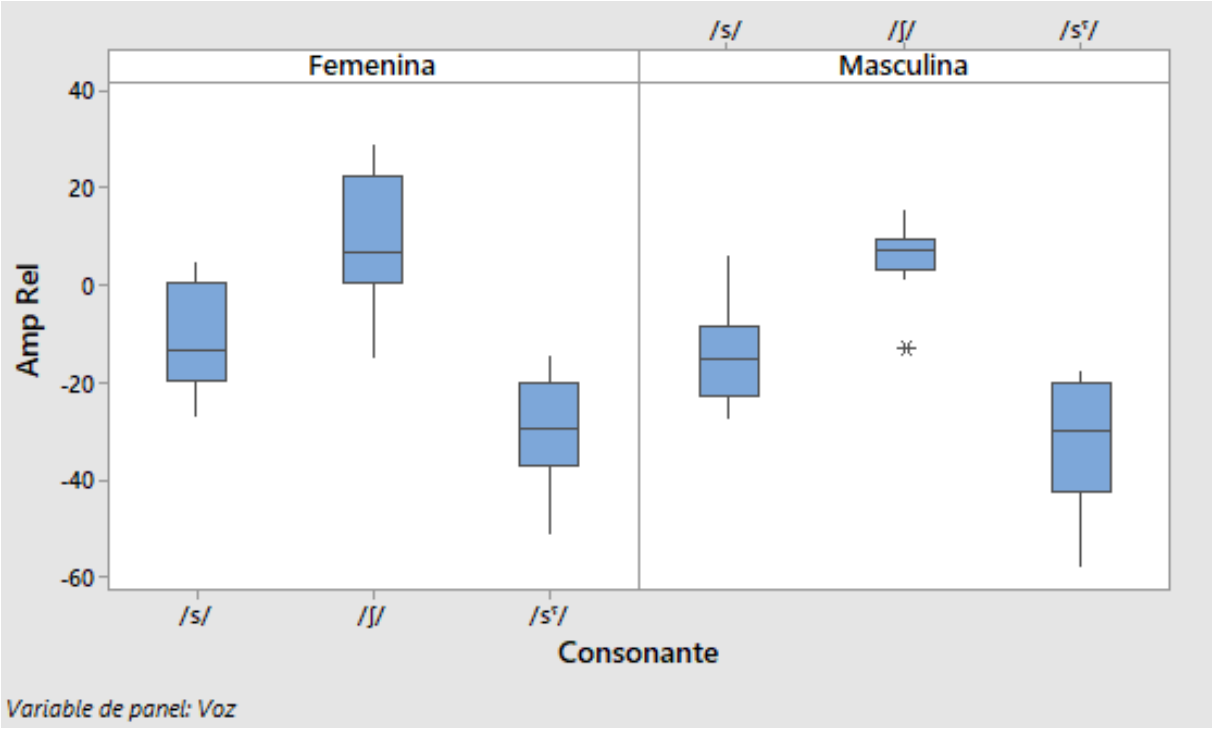


Figura 3.13: Boxplot de la Amplitud Relativa

Esto significa que, si el modelo es eficaz, si alguien pronuncia correctamente /sˤ/, que va acompañada de la /a:/ en en nombre de la letra, el acierto será seguro. No obstante, si el individuo la pronuncia incorrectamente, el modelo predecirá correctamente el error con la consonante correcta puesto que los valores de las otras dos consonantes presentan los mismos valores para /a:/ que para /i:/, siendo esta última la vocal que acompaña al nombre natural de las dos letras correspondientes a las consonantes.

Adicionalmente, en el improbable caso de que alguien del estudio posterior pronuncie /s/ o /ʃ/ como la consonante enfática, la cual no se encuentra en sus idiomas nativos, el modelo debería predecir que se ha dicho /s/ ya que el valor de /sˤ/ con la /i:/ es similar teóricamente a /s/ con /a:/. Con lo cual, el único error en esta suposición es que si alguien pronuncia la consonante /sˤ/ intentando reproducir /ʃ/, el modelo fallará y la confundirá con /s/, pero acertará diciendo que no se ha conseguido decir la consonante correctamente, por lo que el objetivo del proyecto no queda comprometido. No se contempla el caso en que alguien pronuncie la letra alveolar simple como enfática, pues /s/ es común para todos los idiomas de este trabajo.

Una vez asumido esto, se puede realizar el análisis exploratorio previo de los datos. En este caso, se puede observar en la Figura 3.13, cómo hay una tendencia escalada en el rango de datos, tanto en hombres como en mujeres, siendo los valores más pequeños los de la consonante alveolar enfática, seguida de la alveolar simple, y por último, con los valores más elevados la post-alveolar. En general, se observa que la dispersión es similar en todos los casos, a excepción de /ʃ/, donde es considerablemente menor y tiene la particularidad de tener un valor «outlier». No parece haber una diferencia entre las medianas según la voz.

Al ejecutar el análisis *two-way* ANOVA se puede observar en la tabla (Figura 3.14) que, en efecto, este parámetro es eficiente a la hora de diferenciar entre las consonantes puesto que se obtiene

## Análisis de Varianza

| Fuente | GL | SC Sec. | Contribución | SC Ajust. | MC Ajust. | Valor F | Valor p |
|---|---|---|---|---|---|---|---|
| Voz | 1 | 106,9 | 0,40% | 106,9 | 106,90 | 0,85 | 0,360 |
| Consonante | 2 | 18056,1 | 68,18% | 18056,1 | 9028,07 | 71,64 | 0,000 |
| Voz*Consonante | 2 | 1,1 | 0,00% | 1,1 | 0,55 | 0,00 | 0,996 |
| Error | 66 | 8317,5 | 31,41% | 8317,5 | 126,02 | | |
| Total | 71 | 26481,7 | 100,00% | | | | |

Figura 3.14: 2-way ANOVA de la Amplitud Relativa

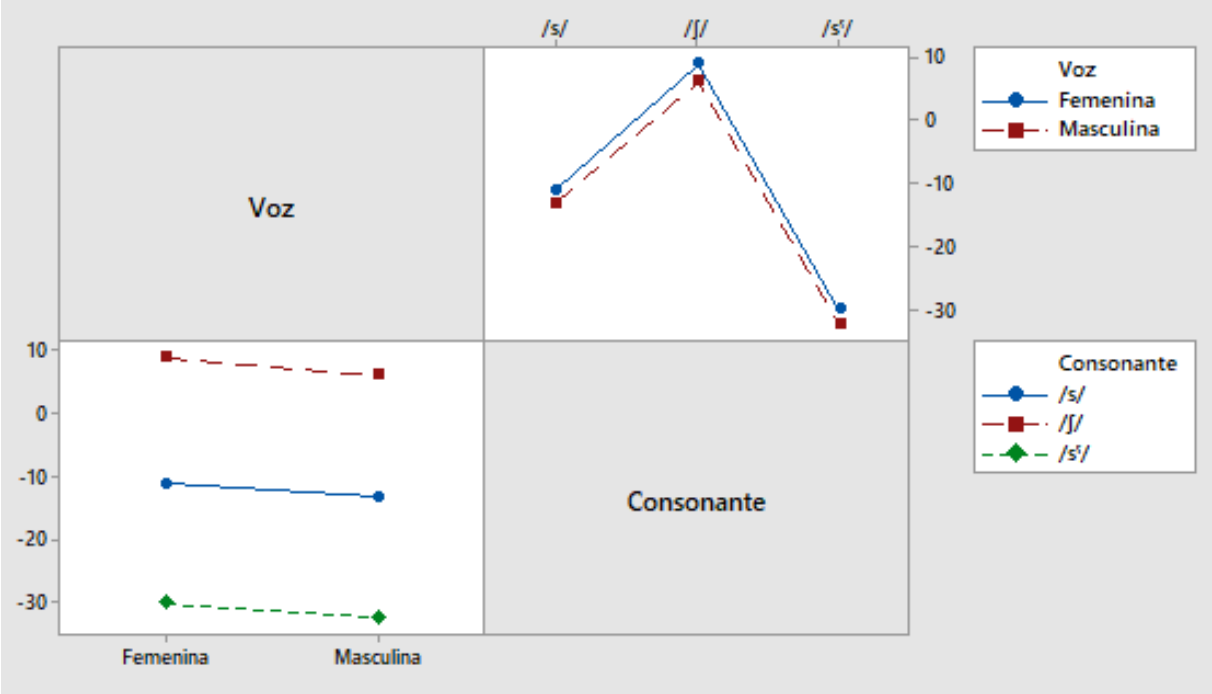


Figura 3.15: Interacción Voz-Consonante de la Amplitud Relativa

un p-valor para este factor de 0.000. No obstante, tal como apuntaba el análisis preliminar, a diferencia de los demás casos, esta vez la diferenciación por voz no es significativa ya que el p-valor es igual a 0.360. Asimismo, la interacción vuelve a ser no relevante con un p-valor igual a 0.996. De nuevo, previamente se comprobó que el modelo cumplía las hipótesis supuestas usando las gráficas de los residuos tal y como se indica en la Sección 3.3.1. Los resultados son visibles en el Anexo 4 (Figura 7.4).

Mediante los gráficos de interacción (Figura 3.15) se puede observar en efecto que el género no tiene ninguna influencia. Las medias de todas las consonantes son prácticamente idénticas. También se hace notable la buena diferenciación entre las tres consonantes, puesto que tienen medias significativamente distintas, siguiendo los resultados esperados.

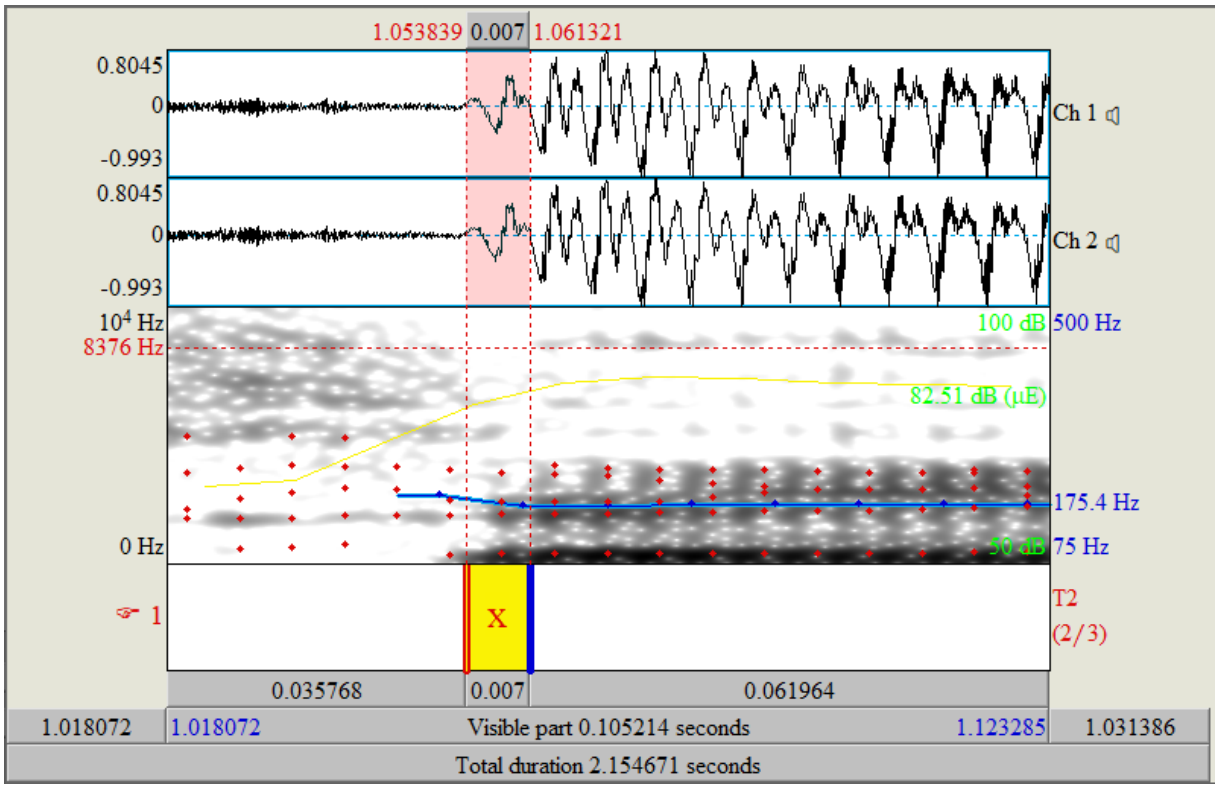


Figura 3.16: Ejemplo de zona de Transición para el cálculo de F2

### 3.3.6. F2 de transición

El segundo formante en la zona de transición se obtuvo con el Script *extracción-de-formantes-en-tabla* especificando la carpeta donde se encontraban los archivos de sonido, los TextGrids, la carpeta de destino con el nombre del fichero con los formantes (F1, F2 Y F3), el nombre del Tier (en este caso solo había uno) y el valor del formante máximo (5000 Hz para hombres y 5500 Hz para mujeres). El resto de parámetros se mantuvieron en su valor por defecto. Los textGrids creados para cada consonante tenían como segmento el primer periodo de onda en el que la onda aleatoria correspondiente al ruido fricativo comenzaba a transformarse en la oscilación periódica de la vocal. Un ejemplo de este TextGrid se puede ver en la Figura 3.16.

Para este parámetro también se tuvo que tener en cuenta la variación de los valores según la vocal adyacente. En este caso, debido a que los valores de la consonante alveolar enfática son inferiores a las otras dos que presentan valores similares, a la vez que los valores generales de las consonantes con la /a:/, tal y como se puede observar en la Figura 3.17, son inferiores a los de las acompañadas con la /i:/, se tuvo que aplicar un factor corrector para eliminar la confusión de los efectos y tener una comparación correcta.

Para realizar esto, se pidió a tres individuos de idioma nativo árabe que participaron en la parte de clasificación (dos hombres y una mujer) que grabasen las dos letras correspondientes a las consonantes no enfáticas con la /a:/ (y el resto del nombre de la letra que representa el fonema /s$^{Q}$/). Dividiendo los valores del segundo formante de transición de las dos consonantes simples con /a:/ entre sus respectivos valores con la /i:/ para cada individuo y haciendo la media de todas, se obtuvo un factor corrector de 0.834. Este factor se aplicó a las dos consonantes con la

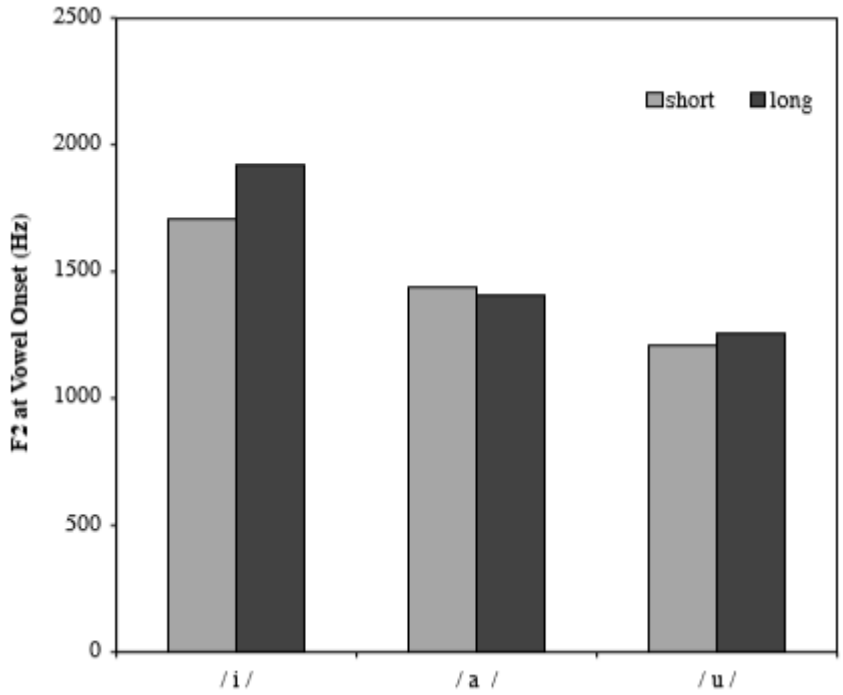


Figura 3.17: Valores de F2 de transición según vocal [Al-Khairy, 2005]

/i:/, con lo que todas las consonantes representaban entonces los valores con la /a:/.

El análisis exploratorio usando los Boxplots (Figura 3.18) muestra que los valores de las consonantes simples /s/ y /ʃ/ son muy similares mientras que las de la consonante enfática /sˤ/ son claramente inferiores. Esto sucede para ambas voces, pero la voz masculina presenta valores inferiores.

La tabla del análisis *two-way* ANOVA (Figura 3.19) muestra que ambos factores son significativos, teniendo los dos un p-valor de 0.000. Además, al igual que todos los demás parámetros, la interacción no es significativa, teniendo esta, en este caso, un p-valor igual a 0.407. Por última vez, previamente a calcular esta tabla, se calcularon los gráficos de residuos que se muestran en el Anexo 5 (Figura 7.5) y, siguiendo el criterio explicado en la Sección 3.3.1, se dictaminó que no había indicios de incumplimiento de las hipótesis del modelo.

Por último, la interacción entre factores (Figura 3.20) muestra que el paso de la voz femenina a la masculina tiene una pendiente descendiente muy similar entre todas las consonantes, creando un efecto proporcional entre las medias. De nuevo se hace evidente que este parámetro es muy útil para diferenciar la consonante enfática /sˤ/.

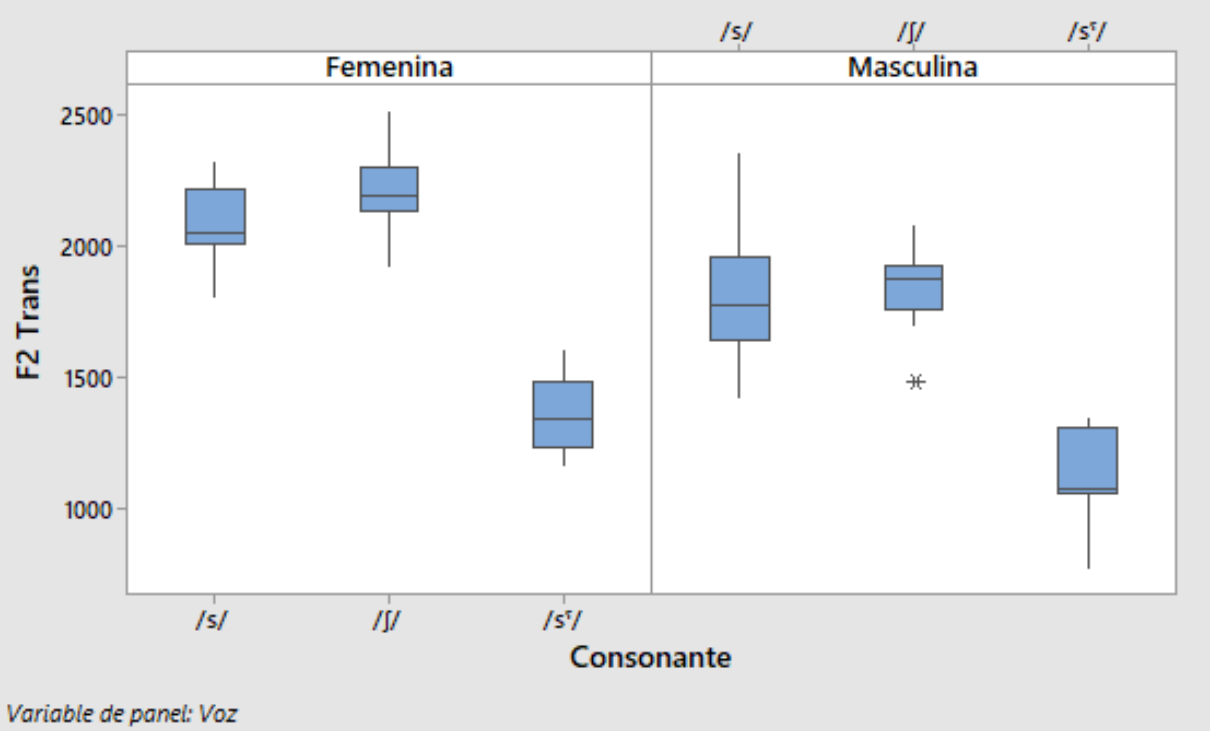


Figura 3.18: Boxplots de F2 de Transición

## Análisis de Varianza

| Fuente | GL | SC Sec. | Contribución | SC Ajust. | MC Ajust. | Valor F | Valor p |
|---|---|---|---|---|---|---|---|
| Voz | 1 | 1534773 | 12,32% | 1534773 | 1534773 | 51,04 | 0,000 |
| Consonante | 2 | 8882022 | 71,31% | 8882022 | 4441011 | 147,69 | 0,000 |
| Voz*Consonante | 2 | 54850 | 0,44% | 54850 | 27425 | 0,91 | 0,407 |
| Error | 66 | 1984564 | 15,93% | 1984564 | 30069 | | |
| Total | 71 | 12456209 | 100,00% | | | | |

Figura 3.19: 2-way ANOVA de F2 de Transición

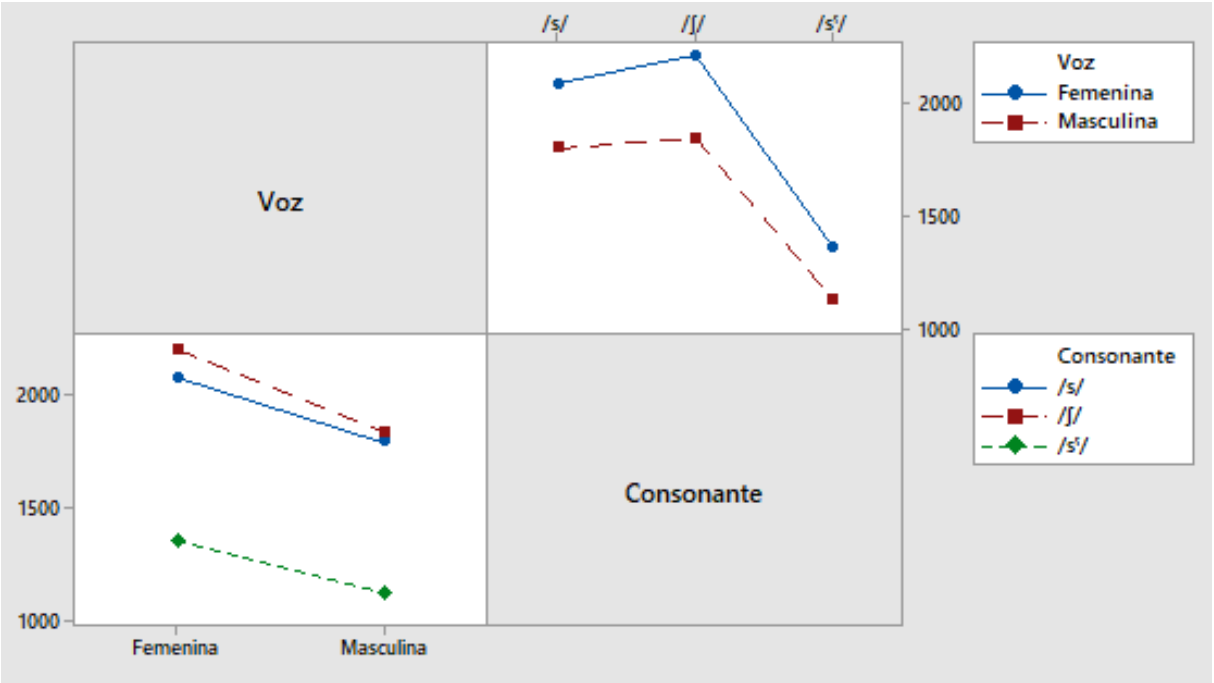


Figura 3.20: Interacción Voz-Consonante de F2 de Transición

## 3.4. Resumen

En este capítulo, se ha explicado el funcionamiento general de Praat. Se ha concretado el diseño del experimento y todo el preprocesado al que se han sometido las muestras de sonido. También se han detallado los Scripts de Praat utilizados, así como los procesos de extracción de todos los parámetros de las consonantes seleccionadas.

Se ha explicado el fundamento teórico del método ANOVA, el cual se ha utilizado para determinar si los parámetros extraídos eran significativos a la hora de diferenciar entre las consonantes seleccionadas. A lo largo de este análisis exhaustivo se han añadido gráficos que ayudaban a entender el comportamiento de los datos. También se ha estudiado el efecto del género del emisor de la voz. Previamente al análisis de cada parámetro se ha justificado su selección y uso.

Se determinó que los cinco parámetros seleccionados (pico espectral, centro de gravedad, cruces por cero dB por 10 entre la duración del intervalo, amplitud relativa y F2 de transición) eran relevantes para diferenciar entre las tres consonantes del proyecto (/s/, /S/ y /sˠ/). Además, el género del emisor de la voz resulta significativa en todos los parámetros a excepción de la amplitud relativa.

A partir de estos resultados, en el Capítulo 4 se utilizarán los parámetros extraídos para entrenar y construir un modelo para clasificar y predecir las tres consonantes utilizando técnicas de inteligencia artificial. Se utilizará este modelo para realizar, también, el estudio posterior relativo a los hablantes no nativos.

# Capítulo 4

# Clasificación y estudio

En este capítulo se detalla el código de programación utilizado, así como sus extensiones y herramientas particulares utilizadas para implementar el método de *machine learning*, del cual también se explican sus fundamentos teóricos, utilizado para crear el modelo que permite discernir entre las tres consonantes involucradas en el proyecto a partir de los parámetros extraídos en el Capítulo 3. Posteriormente, se realiza el estudio sobre la capacidad de pronunciar los sonidos ajenos por parte de los individuos que no hablan árabe. Por último, también se comprueba la veracidad del modelo.

## 4.1. Python y Anaconda

Python es un lenguaje de programación de alto nivel, interpretado, orientado a objetos y con semántica dinámica. Su sintaxis es simple y fácil de aprender, lo que enfatiza la legibilidad y por tanto, reduce el coste de mantenimiento del programa. Soporta módulos y paquetes, con lo que se fomenta la modularidad del programa y la reutilización del código. Es un lenguaje muy atractivo debido al aumento de productividad que proporciona gracias a la carencia de un paso de compilación, hecho que implica que el ciclo de edición-prueba-depuración sea muy rápido. El interprete de Python y la extensa biblioteca estándar están disponibles en forma de código fuente o binaria sin cargo para todas las plataformas principales, y pueden ser distribuidos libremente [Python Software Foundation, 2020].

Anaconda es una distribución libre y abierta del lenguaje Python desarrollada por Continuum Analytics y disponible para Windows, Mac OS X y Linux. Es utilizada para el análisis de datos y el aprendizaje automático, e incluye aplicaciones y paquetes que se utilizarán en este proyecto: Jupyter Notebook, NumPy, Pandas, SciKit-learn y Matplotlib [Weston et Bjornson , 2016].

### 4.1.1. Jupyter Notebook

Jupyter Notebook es una aplicación web de código abierto que permite crear y compartir documentos que contienen código en vivo, ecuaciones, visualizaciones y texto narrativo. Sus usos incluyen: limpieza y transformación de datos, simulación numérica, modelado estadístico, vi-

sualización de datos, aprendizaje automático y mucho más. En definitiva, amplía el enfoque basado en la consola de la informática interactiva en una dirección cualitativamente nueva, proporcionando una aplicación basada en la web adecuada para capturar todo el proceso de computación: desarrollo, documentación y ejecución del código, así como la comunicación de los resultados [Jupyter Team, 2015].

### 4.1.2. Bibliotecas y extensiones

**Pandas** es una paquete de Python que proporciona estructuras de datos rápidas, flexibles y expresivas diseñadas para hacer que el trabajo con datos que presentan relaciones o etiquetas sea fácil e intuitivo. Se adapta a distintos tipos de datos como: datos tabulares con columnas heterogéneas, datos de series de tiempo ordenados y no ordenados, datos de matrices arbitrarias o cualquier otra forma de conjuntos de datos de estadísticos. Las dos estructuras principales de **Pandas**, *Series* (unidimensional) y *DataFrame* (multidimensional), sirven para ocuparse de la mayoría de casos de ámbito ingenieril [The pandas development team, 2014].

**NumPy** es una biblioteca de Python que proporciona un objeto de matriz multidimensional, varios objetos derivados, y un surtido de rutinas para operaciones rápidas en matrices. El núcleo del paquete **NumPy** es el objeto *ndarray*. Este encapsula matrices n-dimensionales de tipos de datos homogéneos, con muchas operaciones que se realizan en código compilado para su ejecución. Las matrices **Numpy** facilitan operaciones matemáticas avanzadas y otras operaciones con un gran número de datos. Tienen un tamaño fijo y los elementos que la componen tienen que ser del mismo tipo de datos [The SciPy community, 2020].

**Scikit-learn** es un módulo de Python que integra una amplia gama de algoritmos de aprendizaje automático de última generación para problemas supervisados y no supervisados. Resalta la facilidad de uso, el rendimiento, la documentación y la coherencia de la API. Tiene dependencias mínimas y se distribuye bajo la licencia simplificada BSD, fomentando su uso en entornos académicos [Pedregosa et al, 2011].

**Matplotlib** es una biblioteca para crear visualizaciones estáticas, animadas e interactivas en Python. Matplotlib produce figuras de calidad en una variedad de formatos impresos y entornos interactivos a través de plataformas [Python Software Foundation, 2020].

## 4.2. Análisis de Disciminante Lineal (LDA)

El análisis de discriminante lineal [Kanaan et al, 2013], conocido como LDA por sus siglas en inglés, es un método de *machine learning* que sirve para determinar qué variables o combinación de estas en un conjunto de datos discriminan entre dos o más clases predefinidas. Es un método, por tanto, supervisado. Se utiliza como clasificador lineal y/o como reductor de variables. A continuación, se hará una descripción de la técnica del método aplicada al caso de dos clases puesto que las expresiones matemáticas quedan simplificadas. No obstante, el método es extensible a casos con más clases, como este proyecto.

Al igual que la mayoría de técnicas de inteligencia artificial, el LDA consta de una parte de entrenamiento y una parte de predicción. En el entrenamiento se parte de un conjunto de $m$

muestras del que se tiene previo conocimiento de la asignación de cada una de estas a su clase correspondiente, y se genera un modelo lineal de las *n* variables que permita discriminar entre las clases. En la fase de predicción, se utiliza el modelo lineal para predecir las clases de las muestras restantes no utilizadas en el entrenamiento.

La formulación matemática del LDA consiste en modelizar de una manera probabilística la distribución de datos en cada una de las clases, que en este caso dicotómico son $\alpha_1$ y $\alpha_2$. El proceso de entrenamiento consiste en una estimación estadística de la cual se obtienen parámetros de una determinada densidad de probabilidad.

La versión más simple de este análisis supone una distribución de probabilidad Normal de los datos de las clases. Esto se expresa, para el caso particular de $\alpha_1$ y $\alpha_2$, mediante la densidad de probabilidad condicionada de que un dato $x$ pertenezca a una de las clases, lo que se formula como

$$p(x|\alpha_i) = \frac{1}{\sigma_i(2\pi)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma_i^2}(x-\mu_i)^T(x-\mu_i)\right], \quad i = 1, 2, \tag{4.1}$$

donde $N$ es el número de variables, y $\mu_i$ y $\sigma_i$ son la media y la desviación estándar de la distribución de datos de la clase $\alpha_i$, respectivamente.

El siguiente paso es introducir la función discriminante $g(x)$, la cual permite definir la frontera de asignación de los datos a sus respectivas clases y que, en el caso dicotómico, se define como

$$g(x) = g_1(x) - g_2(x), \tag{4.2}$$

donde $g_i(x)$ es la función de asignación a la clase $i$ cuya expresión es

$$g_i(x) = \ln p(x|\alpha_i) + \ln P(\alpha_i), \quad i = 1, 2, \tag{4.3}$$

donde $p(x|\alpha_i)$ es la densidad de probabilidad definida en (4.1) y $P(\alpha_i)$ es la probabilidad con la que se asignaría a priori un dato a la clase $\alpha_i$. Introduciendo (4.1) en (4.3), y esta a su vez en (4.2), la función discriminante queda finalmente definida como

$$g(x) = -\frac{1}{\sigma_1^2}(x-\mu_1)^T(x-\mu_1) + \frac{1}{\sigma_2^2}(x-\mu_2)^T(x-\mu_2) - N\ln\frac{\sigma_1}{\sigma_2} + \ln P(\alpha_1) - \ln P(\alpha_2). \tag{4.4}$$

Cuando se desconoce la pertenencia a priori a una de las dos clases, se tiene que $P(\alpha_1) = P(\alpha_2) = \frac{1}{2}$ y la expresión anterior queda simplificada tal que

$$g(x) = -\frac{1}{\sigma_1^2}(x-\mu_1)^T(x-\mu_1) + \frac{1}{\sigma_2^2}(x-\mu_2)^T(x-\mu_2) - N\ln\frac{\sigma_1}{\sigma_2}. \tag{4.5}$$

A (4.5), se le puede hacer una interpretación geométrica por la cual la asignación de un dato $x$ a una de las clases pasa por calcular la distancia de este al centroide de cada clase $i$. Concretamente, se trata de la *distancia de Mahalanobis*, que viene dada por la expresión

$$d_i(x) = \frac{1}{\sigma_i^2}(x-m_i)^T(x-m_i), \tag{4.6}$$

donde $m_i$ es el centroide de la clase $i$. La *distancia de Mahalanobis* constituye una métrica ponderada del dato $x$ al centroide $m_i$ utilizando la varianza $\sigma_i^2$. La función discriminante está compuesta

por las distancias de Mahalanobis del dato $x$ a cada una de las clases presentes y se le añade la corrección $N \ln \frac{\sigma_1}{\sigma_2}$ debido a la posible disparidad entre las varianzas de los datos de cada clase.

El criterio de asignación de un punto $x_0$ a una clase consiste en: si la función discriminante es positiva, se asigna el punto a la clase $\alpha_1$, mientras que si es negativa, se asigna a $\alpha_2$. En caso de neutralidad ($g(x_0) = 0$), el punto $x_0$ es igualmente asignable a ambas clases, es decir, se encuentra en la frontera de decisión. Entonces, por ejemplo, para asignar un dato a la clase $\alpha_1$ significa que $g(x_0) > 0$ y, por tanto, $g_1(x_0) > g_2(x_0)$ lo que, entendiendo la igualdad entre la media y el centroide, implica que

$$\frac{1}{\sigma_1^2}(x - m_1)^T(x - m_1) + N \ln \sigma_1 < \frac{1}{\sigma_2^2}(x - m_2)^T(x - m_2) + N \ln \sigma_2. \tag{4.7}$$

De (4.7) se deduce que, cuando las varianzas son iguales, el LDA asigna el dato $x_0$ a aquella clase cuyo centroide esté más cerca de este en distancia Mahalanobis. En caso contrario ($\sigma_1 \neq \sigma_2$), en la decisión se penaliza a la clase que presente una mayor dispersión estadística mediante el término $N \ln \sigma_i$. La interpretación estadística se entiende como la asignación de un punto a la clase cuya pertenencia resulta más estadísticamente verosímil.

## 4.3. Preprocesado de datos

Como se observó en la Sección 3.3, el tipo de voz según el género del emisor presenta, en cuatro de los cinco parámetros, una diferencia significativa. Esto implica que, a priori, no se puede realizar un modelo completo con todas las muestras. Con el fin de solucionar esto, existen dos alternativas para el análisis: crear un modelo para cada género o neutralizar el efecto de la voz.

En este proyecto se ha optado por neutralizar el efecto de la voz. Esto es debido especialmente a que los modelos de inteligencia artificial quedan ajustados a medida que aumenta el número de muestras y, por tanto, las conclusiones son más veraces. Para realizar esta neutralización, se decidió multiplicar las muestras masculinas por el factor equivalente a la media de valores femeninos entre la media de valores masculinos para cada uno de los parámetros.

La razón del uso de este factor es homogeneizar los valores de las muestras de los hombres y mujeres, lo que se puede traducir en asemejar lo máximo posible las medias masculinas y femeninas para cada consonante en cada parámetro. Este factor se aplicó a los parámetros de pico espectral, centro de gravedad, cruces por cero y F2 de transición, con valores de 1.1169, 1.1932, 1.1195 y 1.1839, respectivamente. Aunque es interesante observar la similitud de los valores, no se hizo una media con estas, aunque implicase una simplificación en los cálculos, ya que no se tienen indicios de una relación entre estos parámetros que justificase esta semejanza. No se aplicó esta corrección para la amplitud relativa puesto que la voz no era significativa desde un inicio.

Cabe mencionar que, al no ser la diferencia entre las medias por voz proporcionales a lo largo de las consonantes, este factor obtenido con las medias generales de las tres afectará en distinta medida a cada una de ellas. Esto quiere decir que mientras algunas medias acabarán asemejándose más, otras cuya diferencia no sea significativa se verán afectadas negativamente, haciendo que las medias se separen. Un ejemplo se puede observar en la Figura 3.4, donde claramente

la consonante /ʃ/ empeorará mientras las otras mejorarán. Este fenómeno no se ha considerado relevante puesto que, aunque si empeorará ligeramente en una parte local, el resultado general será favorable, eliminando la relevancia de la voz.

Para analizar los datos, se creó un archivo `.csv` a partir de un *Google Datashet*, donde las primeras cinco columnas corresponden a los parámetros y una sexta columna etiqueta cada fila de parámetros con el símbolo del fonema consonántico correspondiente. Se creó un documento en Jupyter Notebook y se subió el archivo de datos. En este documento, se importaron las librerías de pandas, NumPy, Matplotlib y las herramientas necesarias de Scikit-learn.

El conjunto de datos creado consistía en un *DataFrame* con los valores de los parámetros $(72, 5)$ y un *Series* con las etiquetas de las consonantes asociadas a cada fila de parámetros. A su vez, estos se dividieron en otros dos conjuntos de datos con el mismo formato, uno de entrenamiento con el $67\,\%$ de los datos $(48$ muestras$)$ seleccionados aleatoriamente, y uno de predicción con el resto de las muestras.

En *machine learning*, es común la necesidad de realizar una normalización de los datos previa a la aplicación de métodos de reducción de variables (e.g, PCA) o de clasificación (e.g, K-NN). No obstante, en este proyecto no es necesario puesto que el único método utilizado es el LDA, el cual internamente ya centra los datos y compensa las proporciones entre varianzas.

## 4.4. Reducción de variables

El primer paso para descubrir el comportamiento del modelo es crear un gráfico 2D que represente las agrupaciones de las consonantes, comprobando así la dispersión de cada una de ellas y la separación relativa entre ellas. Dado que se tienen cinco parámetros, es decir, cinco dimensiones, estas no se pueden representar directamente sino que se tiene que reducir la dimensionalidad de las variables a dos, manteniendo a su vez toda la información esencial del modelo original.

El método LDA también se puede usar como un reductor de dimensionalidad puesto que, durante la fase de entrenamiento, el modelo aprende los ejes más discriminantes entre clases, y estos se pueden utilizar para definir un hiperplano para proyectar los datos. Esta técnica destaca por su óptima separación entre todas las clases [Géron, 2017].

Aplicando la reducción por LDA al *DataFrame* completo de este proyecto, y representando los puntos en un gráfico 2D, se obtiene la Figura 4.1. Con este gráfico se pueden extraer algunas conclusiones de cómo se comporta el modelo. Las tres consonantes están bien diferenciadas, aun siendo la frontera entre /s/ y /ʃ/ algo difusa, pues se confunden un par de valores. La dispersión parece ser menor en /sˠ/. Las dos consonantes alveolares presentan cada una de ellas un valor aislado de sus respectivos agrupamientos. Este gráfico permite intuir que el modelo es un buen clasificador y que es capaz de discernir entre las tres consonantes.

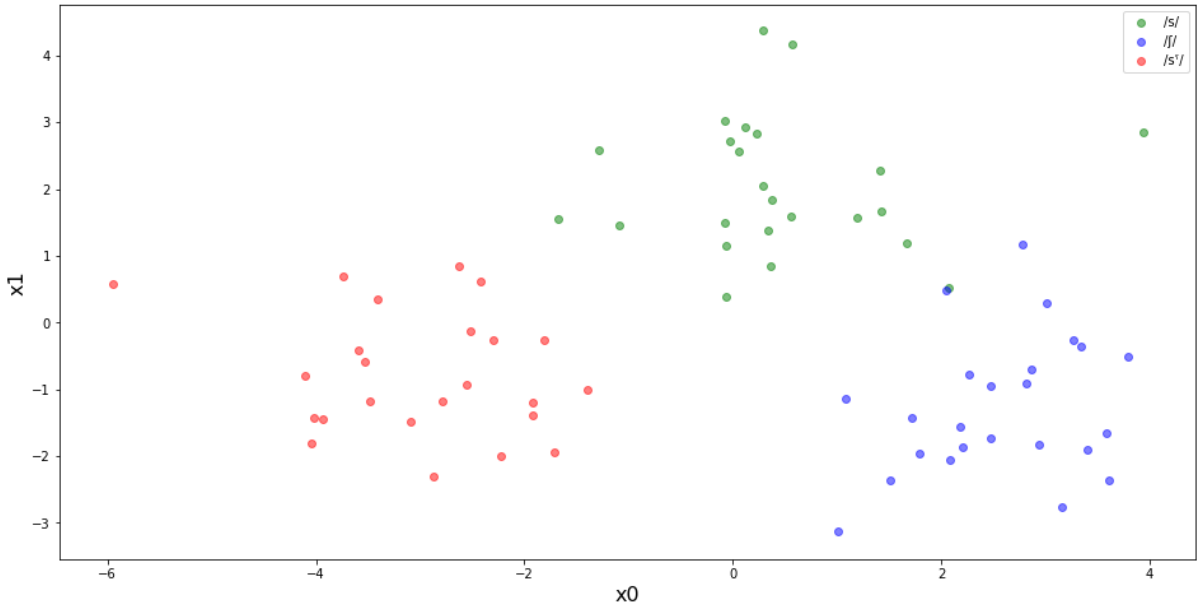


Figura 4.1: Representación 2D del modelo por LDA

| Predicted | /s/ | /sˤ/ | /ʃ/ |
|---|---|---|---|
| Real | | | |
| /s/ | 8 | 0 | 1 |
| /sˤ/ | 0 | 8 | 0 |
| /ʃ/ | 0 | 0 | 7 |

Figura 4.2: Matriz de confusión. Precisión=96 %

## 4.5. Entrenamiento y precisión del modelo

Para comprobar de una manera más técnica la eficacia del modelo, se construyó una función que entrenaba el modelo, mediante el método LDA, con el conjunto de datos de entrenamiento, y devolvía la precisión del modelo a la hora de acertar con el conjunto de predicción. La función genera dos *Series* correspondientes a las etiquetas del conjunto real de predicción y a las predichas por el modelo, a partir de las cuales se puede obtener la precisión y, posteriormente, la matriz de confusión. En la Figura 4.2 se puede observar un ejemplo de matriz de confusión en el que el modelo predice una letra como /S/ cuando, en realidad, es una /s/. La precisión es de 96 %.

No obstante, debido al relativamente reducido número de muestras, según sea el conjunto aleatorio de muestras seleccionadas para el entrenamiento, la precisión del modelo variará en cada caso. Por esto, obtener únicamente una matriz de confusión con su respectiva precisión no es suficiente para analizar correctamente el modelo.

Por tanto, para tener una visión más extensa del comportamiento del modelo, se realizó una

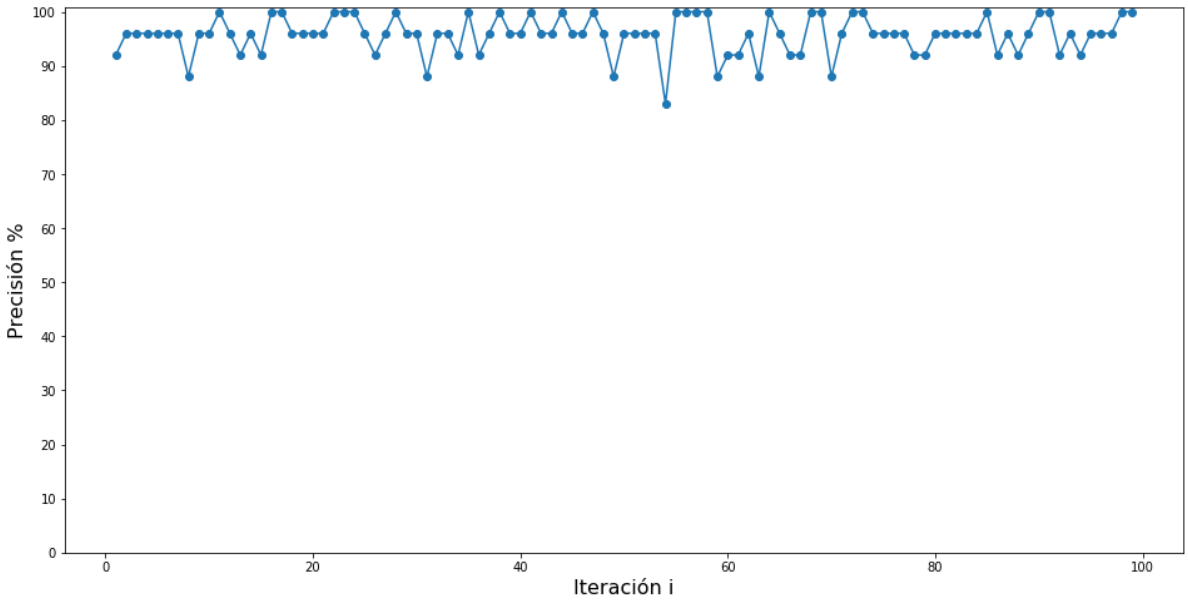


Figura 4.3: Precisión del modelo en 100 iteraciones

función en bucle de 1000 iteraciones que, a cada una de estas, calculaba la precisión del modelo con un conjunto de entrenamiento aleatorio distinto y la añadía a una lista inicialmente vacía. En la Figura 4.3 se puede observar la representación gráfica de las precisiones correspondientes a las 100 primeras iteraciones.

Dentro de este bucle, también se generaba una lista donde cada elemento correspondía a otra lista con la precisión de cada iteración como primer elemento, y una sublista de tuples como segundo elemento. Estos tuples correspondían a las confusiones de cada iteración, siendo el primer valor de cada tuple la consonante real y el segundo la consonante predicha.

A partir de estas dos listas se obtuvo la información que permite extraer conclusiones más detalladas y esclarecedoras del comportamiento y eficacia del modelo. Por un lado, a partir de la lista de 1000 precisiones se contabilizó el número de veces que se repetía una misma precisión. Por otro lado, a partir de la segunda lista, se obtuvo el número de veces que se confundían dos parejas de consonantes en el total de 1000 iteraciones.

Se ha de tener en cuenta, de nuevo, que debido al reducido número de muestras, los resultados varían a cada ejecución. Por este motivo, las conclusiones se extrajeron tras realizar varias ejecuciones y observar la tendencia aproximada. Esta técnica se consideró asumible y robusta puesto que al tener un número elevado de iteraciones, los resultados eran relativamente constantes.

La precisión mostró únicamente cinco valores posibles: 79 %, 83 %, 88 %, 92 %, 96 % y 100 % correspondientes a cinco, cuatro, tres, dos, un y cero fallos, respectivamente. Alrededor del 40 % de los casos corresponden un fallo, dos y ningún fallo tienen un porcentaje similar de ocasiones del 25 % aproximadamente, tres fallos ocurren en un 7.5 % de las ocasiones, cuatro fallos en un 2 % y por último, los cinco fallos ocurren muy anómalamente en menos del 0.5 % de las veces.

Las consonantes /s/ y /S/ se confunden entre 900 y 1000 veces en total mientras que /s/ y /sˠ/ se confunden solo entre 100 y 200 veces en total. Las consonantes /S/ y /sˠ/ no se confunden

nunca. Es decir, las confusiones entre la consontante alveolar y post-alveolar representan el 90 % mientras el 10 % restante es debido a la confusión entre alveolares.

Todo esto va en concordancia con lo expuesto en la Sección 4.4 sobre la ligera difusión en la frontera entre las consonantes /s/ y /ʃ/. También se demuestra que la consonante /sˠ/ está más aislada del resto. Se puede concluir que es un modelo eficaz, con una preción entre el 90 % y 100 %.

## 4.6. Estudio y resultados

Para la parte del estudio se hizo un procesado de datos análogo al conjunto de clasificación del modelo de muestras de hablantes nativos de árabe. Se creó un `.csv` con las 15 muestras de no hablantes de árabe. La única diferencia es que esta vez se añadió una sexta columna con las etiquetas del idioma nativo correspondientes a cada fila de parámetros. Al igual que el caso anterior, se escalaron los valores provenientes de emisores de voz masculinos en los parámetros necesarios con los mismos factores para eliminar la influencia de la voz.

Una vez creado el *DataFrame* del conjunto de estudio completo, se ordenó según el idioma nativo para facilitar la división según este factor. Al final, se poseían tres conjuntos de datos en formato *DataFrame* según el idioma nativo con sus respectivos *Series* con las etiquetas de las consonantes.

Aprovechando la estructura de la función de predicción anterior, se hizo otra para predecir los subconjuntos de datos de cada idioma. Esta vez, la parte de entrenamiento con LDA se realizaba con todo el conjunto de datos de la clasificación, incluyendo las usadas para determinar la precisión del modelo anterior.

Los resultados fueron los siguientes:

- El conjunto de idioma nativo catalán obtuvo un acierto del 73 %, confundiendo una /ʃ/ con una /s/, el mismo fallo a la inversa y dos /sˠ/ con /s/.
- El grupo con idioma nativo castellano acertó todas las letras.
- El bloque de idioma nativo italiano acertó en un 80 % de las veces, fallando dos letras /s/ por /ʃ/ y una /sˠ/ por /ʃ/.

Teniendo en cuenta la débil frontera entre /ʃ/ y /s/, y el reducido número de fallos de /sˠ/, no se puede establecer una correlación entre el idioma nativo y la correcta pronunciación o no de las consonantes. En principio se esperaba un posible mayor número de fallos en la letra /ʃ/ por parte del conjunto de idioma nativo castellano y un mayor número de fallos en general para la letra /sˠ/.

Los aciertos de la letra post-alveolar en los hablantes con idioma nativo castellano son razonables puesto que el sonido no es ajeno ya que se encuentra en el valenciano e inglés, idiomas que tenían como secundarios. También influye que la consonante es pronunciada de forma simple, se tendría que ver el efecto cuando se pronuncia dentro de palabras y en distintas zonas.

Respecto al alto número de aciertos de /sˠ/, se considera que, o bien efectivamente la mayoría de

individuos han pronunciado correctamente la consonante ajena, o bien el modelo no se corresponde con la realidad. Este segundo caso podría deberse a que la suposición sobre la amplitud relativa (ver Sección 3.3.5) no fuese asumible y/o a que el factor corrector aplicado al parámetro F2 de transición (ver Sección 3.3.6) no fuese certero.

## 4.7. Veracidad del modelo

Para probar la veracidad del modelo se utilizaron dos métodos: utilizar el criterio de oyentes nativos como clasificador que determinase qué individuos pronunciaron correctamente la consonante /sˤ/ e insertar como entrada al modelo consonantes /s/ acompañadas esta vez de la vocal /a:/ para comprobar si se predecía correctamente. Estas consonantes se obtuvieron de los mismos tres individuos de los que se extrajo el factor corrector del F2 de Transición obtenido a partir de las grabaciones de las consonantes no enfáticas acompañadas de la vocal /i:/, tal como se explica en la Sección 3.3.6.

Para el primer método se dividieron las 15 muestras no árabes de /sˤ/ en tres bloques según el idioma nativo y se añadió a cada uno de ellos una de las consonantes /s/ con /a:/ a modo de cebo para verificar el buen criterio de los oyentes nativos. Cada bloque se entregó a un oyente diferente a quien se le pidió que decidiese si lo que oían se trataba de una consonante /s/ o una /sˤ/. Por último, una oyente adicional analizó de una manera más detallada todas las muestras, señalando si había algunas de estas que no eran completamente una consonante u otra.

Para el bloque de idioma nativo catalán se señaló, por parte del oyente al que únicamente le correspondía ese bloque, que había tres consonantes /s/ y dos /sˤ/. La oyente de todas las muestras especificó que había una /s/, dos /sˤ/ y las otras dos se encontraban en un punto intermedio, más próximo a /sˤ/. Ambos clasificaron correctamente el cebo.

Respecto al bloque de idioma nativo castellano, el oyente al que sólo le correspondía este bloque interpretó que en todos los audios se había pronunciado, aunque de una manera más suave, el sonido /sˤ/. No obstante, el oyente falló el cebo, señalando que seguía la tendencia del bloque entero. La oyente de todos los audios determinó que tres de las consonantes eran /sˤ/, dos estaban en un punto intermedio entre /s/ y /sˤ/. Esta oyente si acertó el cebo, señalándolo como /s/.

Por último, en cuanto al conjunto italiano, el oyente de bloque único dedujo que en todas las muestras se pronunciaba /sˤ/, señalando únicamente y correctamente el cebo como /s/. La oyente que se encargaba de todas las muestras dictaminó que cuatro eran /sˤ/ y la restante generaba confusión. También acertó el cebo.

Para el segundo método, se creo un pequeño `.cvs` de la misma forma que antes con las tres consonantes /s/ acompañadas de /a:/ y, con una función análoga a las anteriores se predijeron. El modelo clasificó dos de las tres como, efectivamente, /s/, mientras que la otra la predijo como /sˤ/. Esta última coincide con el cebo introducido en el bloque de idioma nativo castellano anterior, en el cual un oyente lo calificó como /sˤ/ suave. Adicionalmente, se pregunto de una manera informal a otro de los oyentes sobre este audio y mencionó que lo calificaría como /s/, aunque no estaba completamente seguro.

Los resultados están en concordancia con lo establecido por el modelo, habiendo predicho este

correctamente las consonantes en las que había consenso entre los dos oyentes de cada bloque. En aquellas en las que no lo había, y que se consideraron, por tanto, intermedias entre las dos consonantes, el modelo las asignó indiferentemente a cada una de ellas. Además el modelo acertó dos de las tres consonantes empleadas en el segundo método, fallando, de hecho, aquella que generaba duda a los oyentes nativos. Por todo esto, se puede concluir que el modelo se corresponde con la realidad y que es, además, muy eficaz y certero.

El único error llamativo del modelo fue la confusión de una de las consonantes /sˤ/ con /ʃ/ (conjunto italiano), puesto que ninguno de los oyentes señaló que alguna de las consonantes sonase como la consonante post-alveolar, además, no se tienen precedentes de confusión entre estas dos consonantes en el modelo inicial de entrenamiento con el conjunto árabe. No obstante, este error se atribuye a la débil frontera entre las dos consonantes simples y, por tanto, se considera asumible.

## 4.8. Resumen

En este capítulo se ha hecho una introducción al código Python y a las bibliotecas y extensiones utilizadas para implementar el modelo de inteligencia artificial en la aplicación web Jupyter Notebook. También se explican los fundamentos teóricos del método LDA, utilizado tanto para la clasificación (entrenamiento y predicción) del modelo como para la reducción de variables, que permite visualizar en un gráfico 2D el comportamiento del modelo.

Se ha explicado el procesado de datos previo a entrenar el modelo, como es la neutralización del efecto de la voz según el género del emisor en los parámetros, la creación de los `.cvs` con los datos o las posteriores divisiones del conjunto de datos necesarios para cada análisis. También se especifican todas las funciones creadas para obtener los resultados.

Se ha obtenido que el modelo es muy preciso, fallando entre cero y tres consonantes como máximo en el 90 % de las veces según el conjunto aleatorio de datos seleccionados para el entrenamiento. También se ha obtenido que, en la mayoría de fallos, se confunden las consonantes alveolar y post-alveolar y que, sólo en un porcentaje mucho menor, se llega a confundir la consonante alveolar enfática con su contraparte simple.

Se determina que el modelo se corresponde con la realidad. No obstante, debido al elevado número de aciertos por parte de los individuos no hablantes de árabe en el sonido desconocido, no se pudo determinar una correlación entre el idioma nativo y la correcta pronunciación o no de los sonidos ajenos.

# Capítulo 5

# Presupuesto

El presupuesto para realizar este proyecto contempla dos partes. Por un lado, se ha de tener en cuenta el coste de las horas empleadas en la realización del mismo y, por otro lado, se ha de considerar el coste del material necesario para elaborar el proyecto, así como las licencias de los programas o aplicaciones utilizadas.

## 5.1. Coste de horas empleadas

Este trabajo de fin de grado esta valorado dentro del grado de ingeniería en tecnologías industriales en 12 créditos ECTS. Teniendo en cuenta que cada crédito equivale a 25 horas de dedicación, el tiempo necesario para realizar este trabajo se estima en 300 horas. Asignando la retribución mínima de 8 €/h establecida en la *Normativa de Prácticas Externas* de la ETSEIB para los alumnos de este grado, la cual se ha tomado como referencia, el coste total de horas equivale a 2400 €.

## 5.2. Coste de material y licencias

Para la realización de este proyecto es necesario poseer un ordenador que pueda procesar los programas utilizados: Praat, las aplicaciones y paquetes de Anaconda y Minitab. En este caso se adquirió un ordenador Lenovo IdeaPad S340 de procesador i5 de 10ª generación valorado en 600 €.

En cuanto a las licencias, tanto Praat como los paquetes y aplicaciones de Anaconda son gratuitos, por lo que solo se tiene en cuenta el coste de la licencia de Minitab, cuyo precio es de 1330 €.

# Capítulo 6

# Impacto ambiental

El impacto ambiental de este proyecto se considera nulo puesto que la totalidad de la realización de este tuvo lugar en un espacio virtual. Todas las muestras fueron grabadas por todos los participantes en sus dispositivos electrónicos y enviados electrónicamente. El resto del proyecto fue realizado en un solo ordenador. Por tanto, lo único que podría considerarse es el consumo eléctrico, pero de hecho, se considera insignificante y no se contempla su evaluación.

# Capítulo 7

# Conclusiones

En este proyecto, se ha creado una herramienta informática mediante inteligencia artificial y *machine learning* capaz de emular, para tres consonantes fricativas del árabe, el *feedback* que cerebro humano genera al oír un sonido y decidir de cuál se trata. Es decir, se ha conseguido elaborar un instrumento virtual que permite discernir entre tres sonidos considerados similares a partir de muestras de audio de un perfil específico tras elaborarles un procesado previo (extracción de parámetros característicos), cumpliendo así con el objetivo principal del trabajo.

Este trabajo se ha realizado en un contexto donde la literatura de vanguardia se había centrado en el exhaustivo estudio sobre los parámetros acústicos más relevantes en la diferenciación de las consonantes fricativas del árabe. Es a partir de estos, y junto a la experimentación propia, que se han seleccionado los parámetros necesarios para discernir entre las consonantes involucradas en este proyecto y se ha comprobado con éxito significación estadística a la hora de discernir entre estas, descubriendo a su vez la influencia del género del emisor de la voz.

La herramienta ha resultado ser altamente eficiente además de veraz. Además, el marco temporal en el que se realiza este trabajo ha conseguido que fuese robusta, pues es capaz de realizar su función a partir de audios grabados en distintos dispositivos electrónicos, difiriendo así de los trabajos anteriores en los cuales el experimento se diseñaba de una manera más restrictiva y controlada. Otra novedad aportada ha sido la independencia del género, pues a partir de la información anterior se ha neutralizado el efecto del tipo de voz según el género del emisor.

Adicionalmente, se ha realizado un estudio para comprobar la correlación entre el idioma nativo en la correcta pronunciación o no de sonidos ajenos, del cual debido al diseño de este, no se obtuvieron las conclusiones esperadas, acertando la mayoría de individuos los sonidos desconocidos independientemente de su idioma nativo. No obstante, este sí ha servido para mostrar la ajustada precisión de la herramienta.

Entre las razones de los resultados del estudio, se cree que la enfatizada manera de pronunciar la vocal adyacente a la consonante desconocida por parte del individuo del que se extrajo la muestra que los involucrados en el estudio tenían que reproducir, forzó la correcta articulación a la hora de pronunciar este sonido fricativo.

A partir de todo esto, se abre la posibilidad de automatizar todo este proceso en futuros proyectos llegando a la posibilidad de crear una aplicación realmente útil e interesante para el apren-

dizaje y mejora de la pronunciación en un contexto sociolingüístico. Además, se recomienda optimizar ciertos factores en los futuros trabajos. Entre las posibles mejoras se propone aumentar el número de muestras de los hablantes nativos, incluir todas las combinaciones con el resto de vocales y extender el proyecto a más fonemas e idiomas.

En cuanto al estudio sobre la correlación de la correcta pronunciación y el idioma nativo, resultaría interesante observar el efecto de la pronunciación en distintas zonas dentro de una palabra completa, puesto que en este trabajo solo se ha considerado la consonante en el inicio del nombre de la letra. También se sugiere estudiar el efecto del poliglotismo, el cual aunque se contempló en un inicio para este trabajo, se descartó debido al homogéneo número de idiomas que conocían los individuos involucrados en esta parte.

Todo esto, sumado al correcto uso de las herramientas y métodos y, a la ingeniosa resolución de problemas relativos a defectos en el diseño del experimento, permiten concluir que este trabajo ha cumplido con las expectativas de forma exitosa.

# Anexo

Este Anexo incluye los cuatro gráficos de residuos correspondientes a los cinco parámetros estudiados para la diferenciación de las consonantes, a partir de los cuales, interpretándolos según se indica en la Sección 3.3.1, se puede determinar si cada parámetro cumple con las hipótesis del bajo las cuales se puede aplicar el método ANOVA.

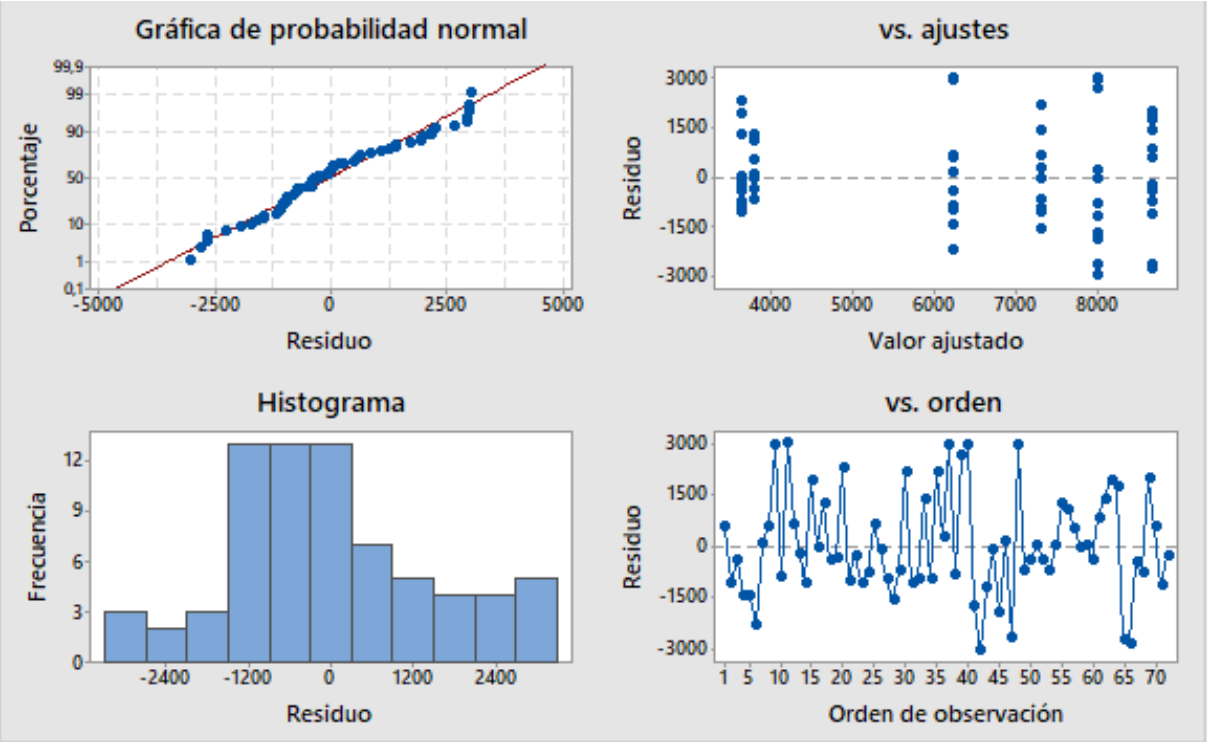


Figura 7.1: Anexo 1. Gráfica de residuos para el Pico Espectral

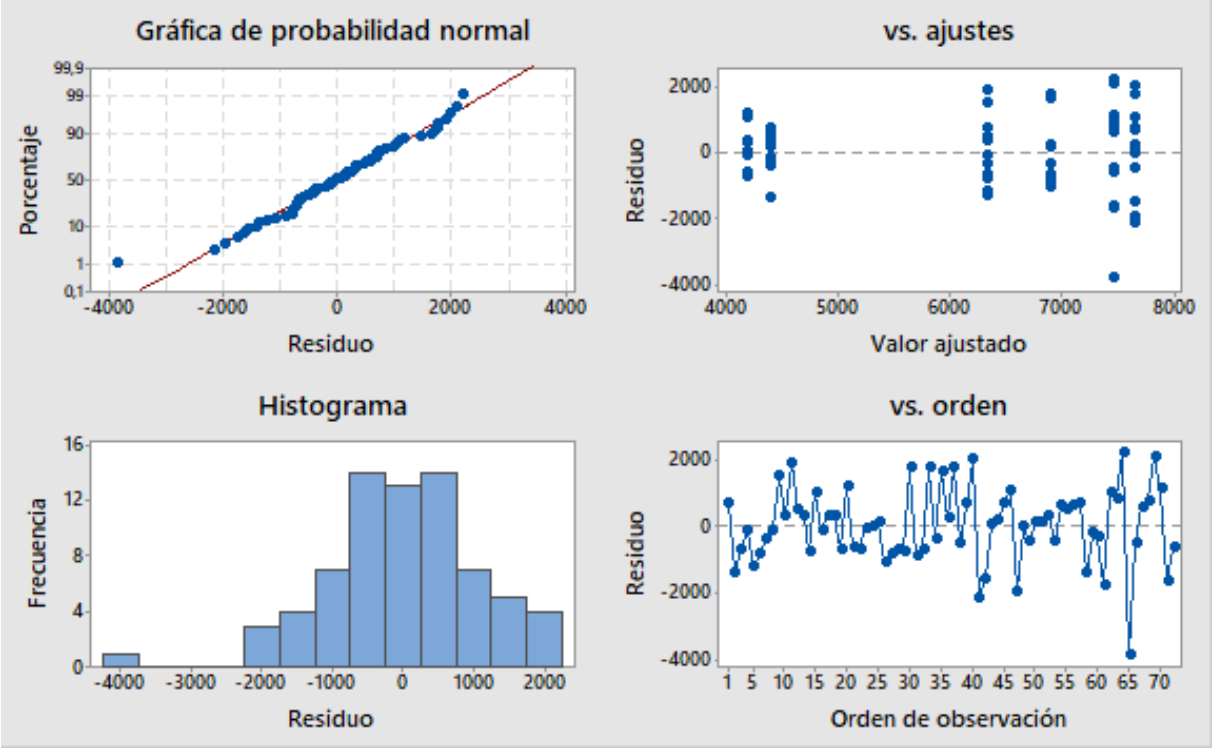


Figura 7.2: Anexo 2. Gráfica de residuos para el Centro de Gravedad

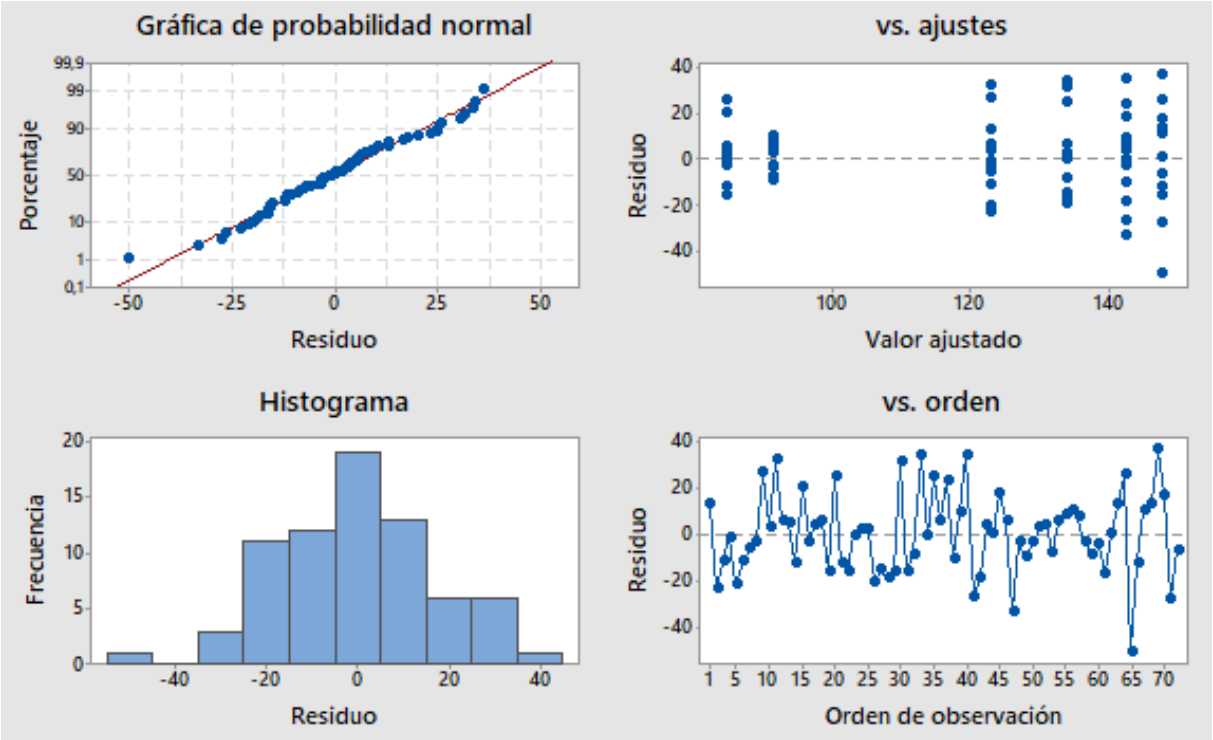


Figura 7.3: Anexo 3. Gráfica de residuos para los Cruces por Cero

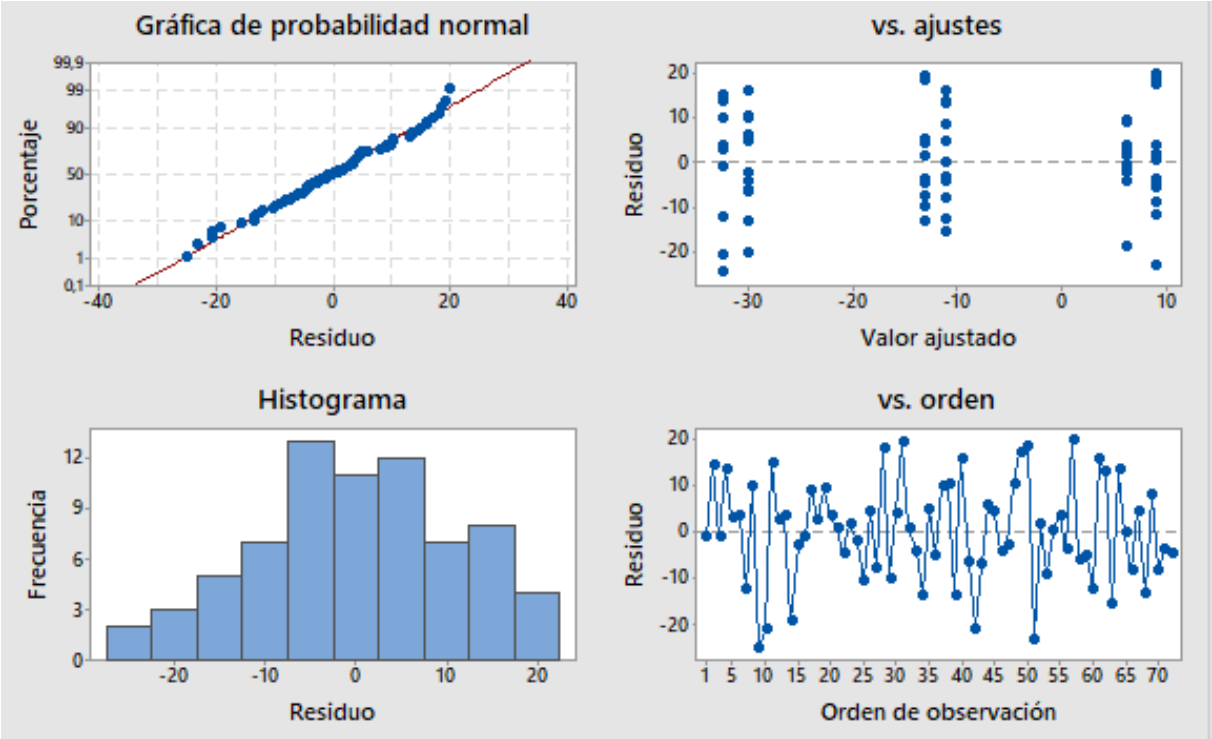


Figura 7.4: Anexo 4. Gráfica de residuos para la Amplitud Relativa

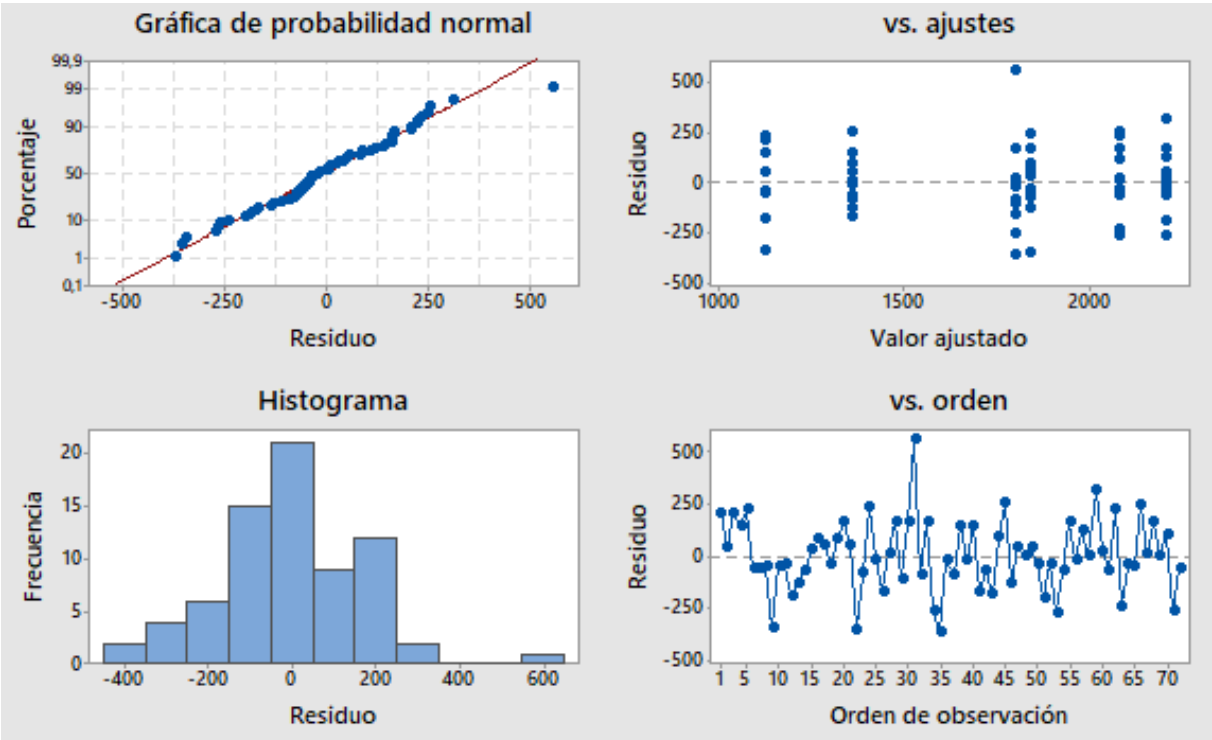


Figura 7.5: Anexo 5. Gráfica de residuos para F2 de Transición

# Bibliografía


[Jongman et al, 2000] Allard Jongman, Ratree Wayland, Serena Wong *Acoustic characteristics of English fricatives*, J. Acoust. Soc. Am., Vol. 108, No. 3, Sep 2009
https://kuscholarworks.ku.edu/bitstream/handle/1808/13393/Jongman%20et%20al.%20JASA2000.pdf?sequence=1&isAllowed=y

[Maniwa et al, 2009] Kazumi Maniwa, Allard Jongman, Travis Wade *Acoustic characteristics of clearly spoken English fricatives*, J. Acoust. Soc. Am., Vol. 125, No. 6, June 2009
https://kuscholarworks.ku.edu/bitstream/handle/1808/13399/Maniwa_Jongman_Wade%20JASA%202009.pdf?sequence=1

[Al-Khairy, 2005] Mohamed Ali Al-Khairy *Acoustic characteristics of arabic fricatives*, Doctoral dissertation, University of Florida, 2005
https://ufdcimages.uflib.ufl.edu/UF/E0/01/13/99/00001/alkhairy_m.pdf

[Al-Masri et al, 2007] Mohammad Al-Masri, Allard Jongman, Wendy Herd *Acoustic correlates of emphasis in Arabic*, ICPhS XVI, ID 1235, Aug 2007
https://eis.hu.edu.jo/deanshipfiles/pub10619416.pdf

[Abudalbuh, 2011] Mujdey Abudalbuh *E ects of Gender on the Production of Emphasis in Jordanian Arabic: A Sociophonetic Study*, Kansas Working Papers in Linguistics, Vol. 32 (2011), 20-47 https://kuscholarworks.ku.edu/bitstream/handle/1808/8096/KWPL-32-Abudalbuh.pdf?sequence=1

[Llisterri, 2020] Joaquim Llisterri, *Fonética y fonología*, Departament de Filologia Espanyola, Universitat Autònoma de Barcelona, Material en línea, última actualización Mzo 2020.
http://liceu.uab.es/~joaquim/phonetics/fon_def_ambits/fonetica_fonologia.html#Fonética_y_fonología_1

[Javed, 2013] Farheen Javed, *Arabic and English Phonetics: A Comparative Study*, The Criterion An International Journal in English, Vol. 4, Issue-IV, Aug 2013
http://www.the-criterion.com/V4/n4/Javed.pdf

[Llisterri, 2020] Joaquim Llisterri, *La clasificación acústica de los sonidos del habla*, Departament de Filologia Espanyola, Universitat Autònoma de Barcelona. Material en línea, última actualización Mzo 2020.
liceu.uab.es/~joaquim/phonetics/fon_anal_acus/fon_acust.html#La_clasificación_acústica_de_los_sonidos_del_habla

[Martínez, 2020] Juan B. Martínez Guevara, *Lengua árabe: Fonología y fonética*, Taawilaalkitaaba, Sitio web.
https://sites.google.com/site/taawilaalkitaaba/arabe/fonetica

[Newman, 2020] Daniel L.Newman *The phonetics of arabic*, Arabic Phonetics: Sound Descriptions, Durham University, Online Resources
http://community.dur.ac.uk/daniel.newman/phon5.pdf

[Hermes et al, 2017] Zainab Hermes, Marissa Barlaz, Ryan Shosted, Zhi-Pei Liang, Brad Sutton *Phonetic Correlates of Pharyngeal and Pharyngealized Consonants in Saudi, Lebanese, and Jordanian Arabic: an rt-MRI Study*, INTERSPEECH 2017, Stockholm, Sweden.
https://pdfs.semanticscholar.org/f9f4/c903762e1102ad98f44444d6f5d9923da1fc.pdf

[Huckvale, 2017] Mark Huckvale *6. Measuring Syllables*, SPEECH, HEARING & PHONETIC SCIENCES, UCL Division of Psychology and Language Sciences, Online Resources, last modified Feb 2017.
https://www.phon.ucl.ac.uk/courses/spsci/expphon/week6.php

[Boersma et al, 2001] Paul Boersma, Vincent van Heuven,Rob Goedemans *Speak and unSpeak with PRAAT*, Glot International Vol. 5, No. 9/10, November/December 2001
http://www.fon.hum.uva.nl/paul/papers/speakUnspeakPraat_glot2001.pdf

[Llisterri, 2020] Joaquim Llisterri *Análisis acústico del habla mediante Praat* Departament de Filologia Espanyola, Universitat Autònoma de Barcelona. Material en línea, última actualización Mzo 2020.
http://liceu.uab.es/~joaquim/phonetics/fon_Praat/Praat.html

[Python Software Foundation, 2020] Python Software Foundation *What is Python? Executive Summary* Python Website.
https://www.python.org/doc/essays/blurb/

[Weston et Bjornson , 2016] Stephen Weston and Robert Bjornson *Introduction to Anaconda* Yale Center for Research Computing Yale University.
https://research.computing.yale.edu/sites/default/files/files/anaconda.pdf

[Jupyter Team, 2015] Jupyter Team *The Jupyter Notebook* Jupyter Website.
https://jupyter-notebook.readthedocs.io/en/stable/notebook.html

[The pandas development team, 2014] The pandas development team *Package overview* pandas Website.
https://pandas.pydata.org/pandas-docs/stable/getting_started/overview.html

[The SciPy community, 2020] The SciPy community *What is NumPy?* NumPy Website.
https://numpy.org/devdocs/user/whatisnumpy.html

[Pedregosa et al, 2011] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion *Scikit-learn: Machine Learning in Python* Journal of Machine Learning Research 12 (2011) 2825-2830.
http://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

[Python Software Foundation, 2020] Python Software Foundation *Project description* matplotlib 3.2.1 Website.
https://pypi.org/project/matplotlib/

[Kanaan et al, 2013] Samir Kanaan, Gerard Escudero, Raúl Benítez *Inteligencia artificial avanzada* PID 00174137 UOC.
https://www.exabyteinformatica.com/uoc/Inteligencia_artificial/Inteligencia_artificial_avanzada/Inteligencia_artificial_avanzada_(Modulo_1).pdf

[Géron, 2017] Aurélien Géron *Hands-On Machine Learning with Scikit-Learn & TensorFlow* O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.
https://www.lpsm.paris/pageperso/has/source/Hand-on-ML.pdf