

# Características acústicas de la vocales del español rioplatense

Aronson, Leonor (Depto Implante Coclear – Fundación Arauz y Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET )

Furmanski, Hilda M. (Fonoaudióloga)

Rufiner, Leonardo (Facultad de Bioingeniería. Universidad Nacional de Entre Ríos )

Estienne, Patricia (Departamento De Implante Coclear – Fundación Arauz )

## Abstract

El propósito de este estudio es el de calcular los valores de las formantes de las vocales aisladas del Español Rioplatense pronunciadas por adultos normoyentes. Se realizaron registros de 40 voces femeninas y 40 voces masculinas correspondientes a sujetos entre 18 y 35 años, nativos de habla español rioplatense, sin ningún tipo de patología vocal asociada. El análisis de los 3 primeros formantes y los respectivos anchos de banda, así como el 4º formante cuando fué posible, se realizó utilizando métodos espectrográficos. Los contornos formánticos se midieron mediante el método LPC (Linear Predictive Coding) utilizando un dispositivo desarrollado para este propósito.

El objetivo de este estudio es obtener patrones formánticos que van a ser utilizados como normativa en estudios de voces patológicas.

## Introducción

La frecuencia de cierre y apertura de la glotis debida a la presión del flujo de aire traqueal, caracteriza la emisión de sonidos sonoros (fonación). A esta frecuencia se conviene en llamar frecuencia glótica o fundamental, F0. El espectro del pulso glótico generado (espectro de la fuente glotal) será afectado por los filtros del tracto vocal, para producir sonidos vocálicos. Es decir: el tracto vocal es un filtro ó conjunto de filtros mecánicos que enfatiza algunas de las componentes del espectro glotal, especialmente las componentes que coinciden con sus propias frecuencias de resonancia. Esas componentes o zonas de componentes de mayor energía en el espectro resultante, son los formantes. Por lo tanto, la modulación del espectro de la fuente glotal y las resonancias del tracto vocal afectarán la ubicación de los formantes. Los espectros de los sonidos vocálicos están bien diferenciados entre sí y caracterizan cada vocal.

Una detallada descripción de la teoría Fuente-Filtro de la producción vocálica puede verse en Fant, 1960 (1) y en Kent, 1992 (2).

Los anchos de banda están definidos (3) como una función de la pérdida de energía debida al calor de conducción, a la viscosidad, al movimiento de las paredes de las cavidades resonantes, a la radiación del sonido desde los labios y a la parte real de la impedancia glótica. Acústicamente, las vocales son los sonidos más simples: están representadas por patrones estacionarios y la configuración articulatoria es cuasi estable.

Las descripciones usuales son las que representan los tres primeros formantes F1, F2 y F3 ; el plano F1-F2, es el que suele utilizarse para describir las características analíticas más relevantes.

El sistema vocálico del Español está constituido por cinco sonidos bien diferenciados: /i,e,a,o,u/ de los cuales /i,u/ son vocales cerradas y /e,o,a/ son vocales abiertas. Tomando en cuenta el modo de articulación, las vocales /i/ y /u/ son altas; las /e/ y /o/ son medias y la /a/ es baja. Según el lugar de articulación, las vocales /u/ y /o/ son anteriores, la /a/ es central y las vocales /i/ y /e/ son posteriores.

El espectro acústico de las 5 vocales del Español hablado en Argentina ha sido

analizado por M Guirao y AM Borzone de Manrique en el año 1975 (4). Se determinaron los tres primeros formantes mediante el análisis espectral de vocales aisladas y en contexto consonántico.

Estudios de los patrones formánticos de la lengua inglesa fueron realizados por Peterson y Barney (5), Watrous (6), Hillenbrand J et al (7), entre otros.

En este trabajo, se determinaron los cuatro primeros formantes (F1, F2, F3 y F4), los anchos de banda (B1, B2, B3 y B4) y la frecuencia fundamental (F0), en sujetos adultos normoyentes. Se realizaron estudios para determinar la variabilidad de los resultados obtenidos entre sujetos como así también la obtención de valores formánticos que se superponen entre vocales adyacentes.

Finalmente se realizó una comparación con los resultados obtenidos en nuestro idioma por los autores mencionados (4).

## **ANALISIS ACUSTICO**

### **METODO**

#### **Sujetos**

Participaron en este estudio 45 mujeres y 45 hombres entre 18 y 35 años. Todos los sujetos examinados son normoyentes sin patologías vocales o desórdenes del lenguaje, nativos argentinos, criados en provincias cercanas a la capital del país. Esta consideración se basó en que la fonación de hablantes de algunas provincias argentinas se caracteriza por una prolongación en la emisión de las vocales con marcada variación de la frecuencia fundamental, lo que podría incidir en la determinación de los valores formánticos, aun cuando los registros que se analizaron correspondieran a vocales aisladas.

#### **Registros**

Se realizaron dos registros de las 5 vocales aisladas de 2 segundos de duración por vocal, para cada uno de los sujetos participantes. Se ofreció a cada hablante, material escrito de modo de realizar la fonación en el orden /a, o, u, e, i/ en una primera serie y el orden /i, e, u, o, a/ en una segunda. Esto se hizo como un modo de equalizar la energía de la fonación, evitando registros con energía insuficiente para su medición.

Los sujetos estaban sentados frente a un micrófono, a aproximadamente 25 cm del mismo. Se les indicó emitir los fonemas de modo natural, en un nivel medio, tomando aire cuando fuere necesario. Se realizó una serie de fonaciones a modo de ensayo para que cada sujeto pudiera comprender lo que se requería en relación a intensidad y tiempos de fonación. Antes que el sujeto se retirara del laboratorio se verificó que los registros obtenidos fueran adecuados para su posterior análisis.

Los registros se hicieron utilizando un micrófono Shure SM58 dinámico unidireccional con respuesta en frecuencia en 50-15000 Hz e impedancia de 150 ohms. Se utilizó una tarjeta de alta calidad de audio digital, con sobremuestreo de 128 x y posterior filtrado, resultando en una frecuencia de muestreo final de 11025 Hz. La relación señal/ruido del sistema es mayor de 97 dB.

Las mediciones se realizaron en lo que se juzgó como un segmento estable de la fonación, descartando los primeros y últimos 200 ms. Dentro de ese segmento estable, se eligió un núcleo vocálico cuyos límites se determinaron manualmente, aplicando "zoom" hasta que pudieran visualizarse aproximadamente 10-15 ciclos de fonación. Este segmento se dividió en subsegmentos de 512 puntos dentro de los cuales se realizó el análisis de los valores formánticos y ancho de banda utilizando LPC (linear predictive coding) de 15 polos con ventana de Hamming (7). Los picos espectrales se estimaron mediante interpolación parabólica. Los valores numéricos de

los formantes promedio en el segmento y sus respectivos anchos de banda (F1-B1, F2-B2, F3-B3, F4-B4) se leyeron en una ventana editada por el sistema. Para la lectura de los registros, se utilizaron simultáneamente los picos espectrales LPC y el espectrograma en escala de grises. En algunos casos se leyeron los picos espectrales LPC locales (de cada subsegmento), lo que para algunos registros se completó con la lectura utilizando análisis de Fourier también local. El cuarto formante F4, se midió en el espectrograma y en los picos espectrales LPC cuando fue posible visualizarlo. Para la vocal /u/ sólo pudo medirse el tercer formante, en el 65% de los registros. En todos los casos se utilizó un pre-énfasis de 0.50.

El analizador de voz con el que se efectuaron las mediciones permite la variación de los parámetros de análisis en caso de ser necesario: mayor número de polos para el análisis LPC o variaciones en los anchos de ventana de análisis. En varios casos, los registros fueron reanalizados utilizando esta ventaja del equipo. Pero no fue posible realizar la lectura del ancho de banda del cuarto formante de ninguna vocal de las voces femeninas salvo casos aislados que no se consignan en este trabajo. Los valores de F0 se extrajeron por el método de autocorrelación mediante un trazador de "pitch" del segmento estable (7).

## Resultados

En la Tabla I se muestran los valores promedios de la frecuencia glótica F0 y los formantes F1, F2, F3 y F4 para los dos grupos de sujetos analizados. Se consignan también los promedios de los anchos de bandas de los tres primeros formantes B1, B2, B3 para voces femeninas y B1, B2, B3 y B4 para voces masculinas.

Los diagramas de los espacios vocálicos pueden verse en las figuras que siguen. La Figura 1 a) y b) representa los valores individuales obtenidos para F1 y F2 en voces masculinas y femeninas respectivamente y las elipses que ajustan a los valores hallados en el mapa de formantes. Para graficar las elipses se supuso que los vectores de características (vector aleatorio X cuyas coordenadas son las formantes) poseen una función de densidad de probabilidad (fdp) normal o gaussiana multidimensional de la forma:

$$f(X) = \frac{1}{\sqrt{(2\pi)^n |C|}} e^{\left\{-\frac{1}{2}(X-\mu)C^{-1}(X-\mu)^t\right\}}$$

donde:  $n$  es la dimensión de la secuencia de vector aleatorio X (en nuestro caso 2)

$$\mu = E\{X\}$$

la media o promedio.

$$C_{ij} = E\{(x_i - \mu_i)(x_j - \mu_j)\} = E\{x_i x_j\} - E\{x_i\}E\{x_j\}$$

la matriz de covarianza.

La elipse dibujada corresponde a aquella para la cual la integral doble sobre la región elíptica interior centrada en la media (correspondiente a un corte transversal de la fdp) arroja una probabilidad total del 90%. Esto significa que para este modelo el 90% de los casos caerían dentro de la elipse.

Las gráficas de las fdp para todas las clases (vocales) se obtuvieron mediante mezclas de las gaussianas originales suponiendo igual peso para todas mediante la siguiente expresión:

$$f_{total}(X) = \sum_{k=1}^K p_k \cdot f_k(X)$$

donde  $K = 5$  y  $p_k = 1/K$

En la Figura 2 a) y b) se consignan los valores individuales y las correspondientes elipses para el plano F2-F3 para voces masculinas y femeninas respectivamente. La Figura 3 a) y b) permite visualizar los valores individuales y elipses que ajustan los datos para F1-F3 para ambos grupos.

En la Figura 4 se consignan los valores medios obtenidos para F1-F2 y sólo para este plano, se superponen las elipses dentro de las cuales se encuentran el 90% de los datos registrados para voces femeninas y masculinas.

En las figuras que siguen, se visualizan los valores promedios para el plano F2-F3 (Fig 5) y plano F1-F3 (Fig 6) y se omiten las elipses por razones de claridad del gráfico. El gráfico correspondiente a las funciones de densidad de probabilidad basadas en mezclas gaussianas obtenidas de los datos para las vocales en el plano F1-F2 para voces masculinas, puede verse en la Figura 7 para una vista superior en a) y una vista lateral 3D en b).

La comparación entre los resultados obtenidos en este estudio con los hallados por Guirao y Manrique (4) se hallan en la Figura 8 a) y b) para F1 y F2 respectivamente, voces masculinas y en la figura 9 a) y b) para F1 y F2 correspondiente a voces femeninas.

Para el tercer formante, la comparación se realizó sobre valores promedios entre voces femeninas y masculinas por no hallarse en la bibliografía datos diferenciados por sexo. La Figura 10 muestra los datos de ambos estudios para F3.

### **Análisis de los resultados**

Analizando la distribución de datos para F1 y F2 se observa para las voces masculinas, Figura 1 a) , que las áreas de las elipses quedan bien definidas para cada vocal; en el plano de datos de voces femeninas de la Fig 1 b), puede verse en cambio, áreas que se superponen para las vocales /i,e/ y para las vocales /u,o/. Los valores promedios para ambos grupos muestran la ya conocida distribución triangular de las vocales donde es posible observar una importante región de superposición entre los dos grupos estudiados (Fig 4).

La distribución de datos individuales en el plano F2-F3, nuevamente deja a los valores obtenidos para la vocal /a/ sin superponerse con los de las demás vocales tanto para voces femeninas como masculinas como se ve en la Fig. 2 a) y b). Las vocales /u/ y /o/ muestran ocupar espacios comunes especialmente para voces masculinas. La elipse que ajusta a los valores de las vocales /i/ y /e/ es una misma para voces masculinas. Los valores promedios que se ven graficados en la Fig. 5 muestran zonas diferenciadas para ambos sexos demarcando figuras similares.

Datos individuales para el plano F1-F3 muestran zonas comunes para ambos sexos, correspondientes a las vocales /o,u,e,i/ para voces masculinas (Fig 3 a) y para las vocales /o,u,e/ para voces femeninas (Fig 3 b). Las áreas que corresponden a la vocal /a/, se encuentran diferenciadas del resto de las vocales para voces de ambos sexos. Los valores promedios para ambos grupos que se representan en la Fig 6, configuran dos regiones bien diferenciadas, para las cuales no existen espacios comunes. En la Figura 7 es posible observar las funciones de densidad de probabilidad basadas en las mezclas gaussianas obtenidas de los datos para las vocales en el plano F1-F2 de voces masculinas como se comentó anteriormente. La figura 7 a) presenta una vista superior y la 7 b) una vista lateral 3D. El intensidad del tono gris indica la

probabilidad de encontrar una vocal en esa zona del plano F1-F2.

## Discusión

Comparando los resultados obtenidos por este estudio y los hallados por Guirao y Manrique (Referencias 4 y 9) para primer formante, los que pueden observarse en la Fig 8 a) y b) es posible ver que los valores difieren entre 1,5% y 17%. Los valores de F1 obtenidos en este estudio, son ligeramente menores que los obtenidos por G&M para ambos grupos. La mayor diferencia que se observa para F1, es para la vocal /i/: 6% para voces femeninas y del 17% para voces masculinas. Respecto a F2, cuya comparación se ve en la Figura 9 a) y b), también es la vocal /i/ la que presenta mayor diferencia entre estudios: es de sólo 2% para voces femeninas y 8% para masculinas.

Es interesante notar que pese a la diferencia en los mecanismos de análisis empleados, al número de sujetos participantes y al hecho que median casi 25 años entre estudios (lo que puede incidir en la fonación aun cuando se trate de vocales aisladas), ambos estudios arrojan resultados similares en los dos primeros formantes. Para la comparación del tercer formante que se presenta en la Figura 10, se tomaron los valores promedio entre voces femeninas y masculinas porque, como se mencionó, no se hallaron en la bibliografía datos separados por sexo.

Las diferencias pueden deberse a la falta de precisión en la toma de datos del estudio G & M (9) y a la diferencia en las metodologías de análisis empleadas en ambos estudios. El método usado por Manrique no permite obtener una lectura precisa de formantes de alta frecuencia, pero si se considera el rango de variabilidad de los valores, puede verse que los hallados en este estudio caen dentro del mismo. Ni los anchos de banda ni el cuarto formante han podido ser comparados con los resultados hallados por Manrique.

El objetivo de este estudio piloto ha sido el de obtener una normativa de los valores formánticos de las vocales de hablantes argentinos del Río de la Plata. Los valores normales se utilizan en estudios de diferentes patologías vocales, como elemento clínico objetivo de la medida de desviación de esos valores.

La utilización de las mezclas de densidad de probabilidad calculadas en este trabajo, permitirían obtener una medida cuantitativa más adecuada que la que podría obtenerse utilizando desviaciones de valores medios, para la comparación de resultados de voces patológicas con normales.

Los resultados obtenidos también se han utilizado como patrones de la evolución en la adquisición del lenguaje, en estudios longitudinales de producción vocálica, realizados en pacientes con prótesis coclear implantable.

Se están analizando registros de voces infantiles para completar el estudio y obtener así una normativa que permita evaluar desvíos de voces de niños con prótesis cocleares implantables que se encuentren en un programa de adquisición de lenguaje.

## Agradecimientos

Los autores quieren agradecer a todas las personas que brindaron sus voces claras para la realización de este estudio. Este trabajo es parte del proyecto de Reconocimiento Automático del habla que se desarrolla en conjunto entre el Laboratorio de Cibernética de la Facultad de Bioingeniería de la Universidad Nacional de Entre Ríos y el Departamento de Implante Coclear de la Fundación Arauz.

## Referencias

1. Fant G. *Acoustic Theory of Speech Production*. The Hague: Mouton 1960
2. Kent RD, Read C. *The Acoustic Analysis of Speech*. Singular Publishing Group. 1992.
3. Klatt DH. Software for cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America* 1980; Vol 67, No3: 971-995.

4. Guirao M y Borzone de Manrique AM. Identification of Argentine Spanish Vowels. *Journal of Psycholinguistic Research* 1975; Vol 4, No.1: 17-25.
5. Peterson GE y Barney HL. Control methods used in a study of the vowels. *Journal Acoustical Society of America* 1952; 24: 175-184.
6. Watrous RL. Current status of the Peterson-Barney vowel formant data. *Journal of the Acoustical Society of America* 1991; Vol 89 : 2459-2460.
7. Hillenbrand J, Getty LA, Clark J y Wheeler K. Acoustic characteristics of American English vowels. *Journal Acoustical Soc. Am.* 1995; Vol 97(5), Pt 1: 3099-3111.
8. Markel JD y Gray AH. *Linear Prediction of Speech*. Springer-Verlag. Berlin Heidelberg New York 1976.
9. Borzone de Manrique AM. *Manual de Fonética Acústica*. Hachette 1980.

Fig. 1 - Elipses dentro de las cuales se encuentran el 90% de los datos individuales medidos para las vocales del español en el plano F1-F2

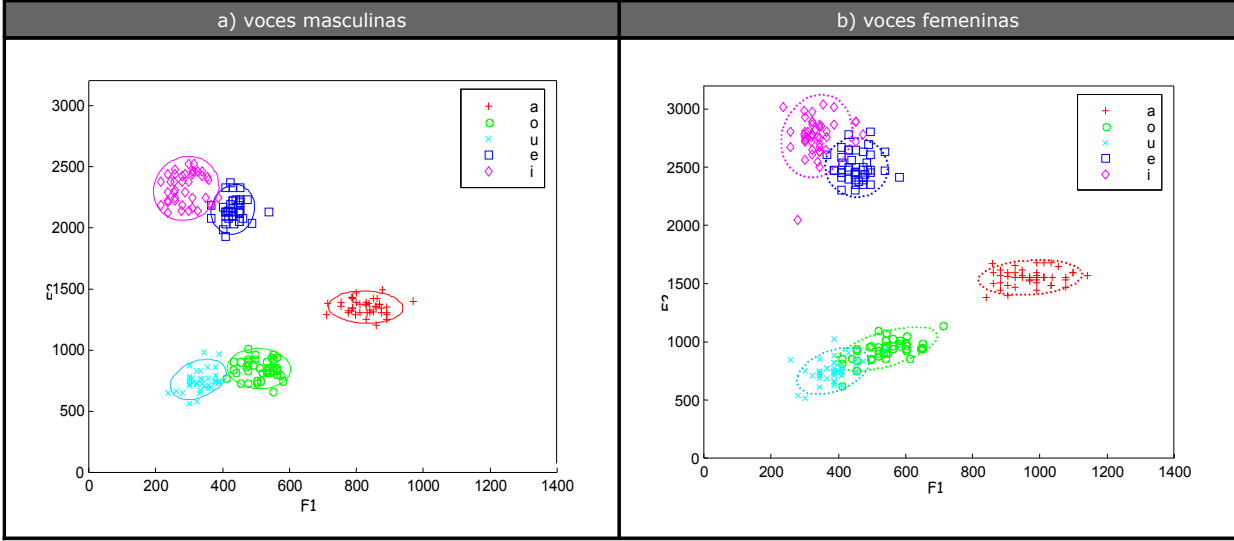


Fig. 2 - Elipses dentro de las cuales se encuentran el 90% de los datos individuales medidos para las vocales del español en el plano F2-F3

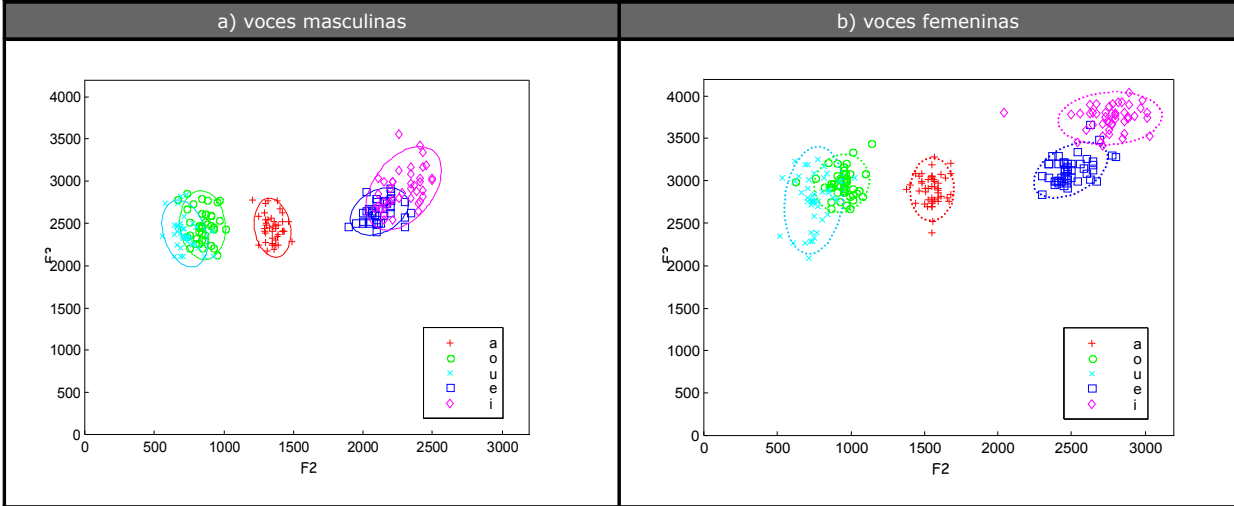


Fig. 3 - Elipses dentro de las cuales se encuentran el 90% de los datos individuales medidos para las vocales del español en el plano F1-F3

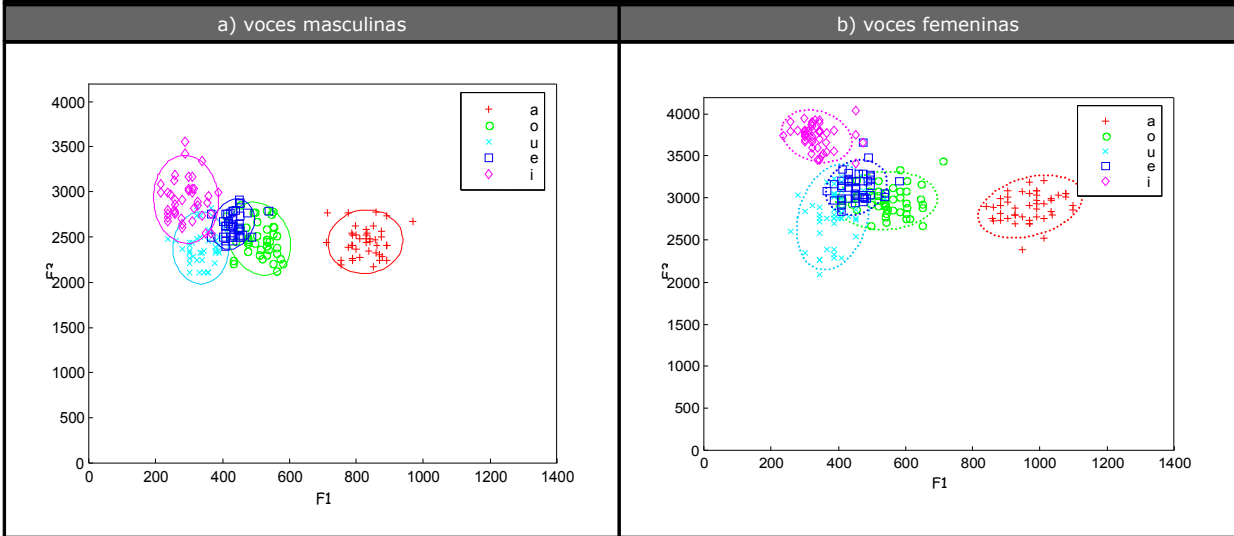


Fig. 4 - Diagrama F1-F2 de los valores promedio y las correspondientes elipses, voces masculinas y femeninas. Puede observarse la configuración triangular para ambos grupos y las áreas de superposición.

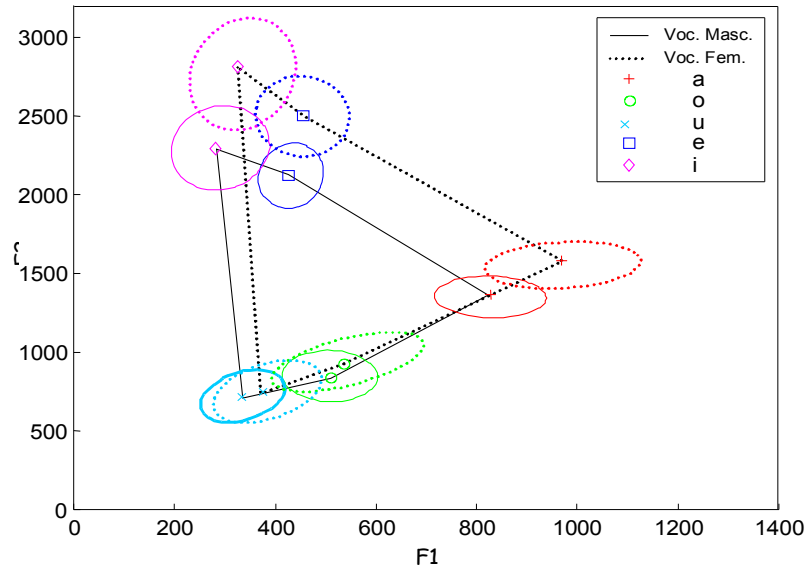


Fig. 5 - Diagrama F2-F3 de los valores medios obtenidos de los datos para las vocales del español de voces masculinas y femeninas.

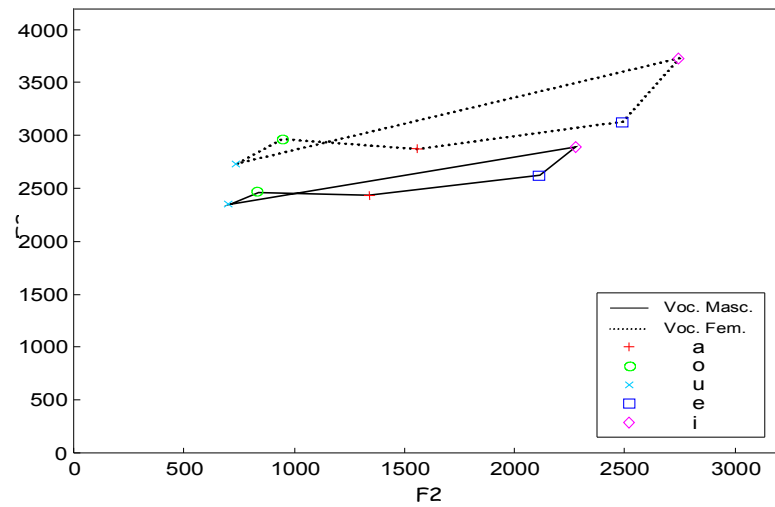




Fig. 6 - Diagrama F1-F3 de los valores medios de los datos medidos para las vocales del español de voces masculinas y femeninas.

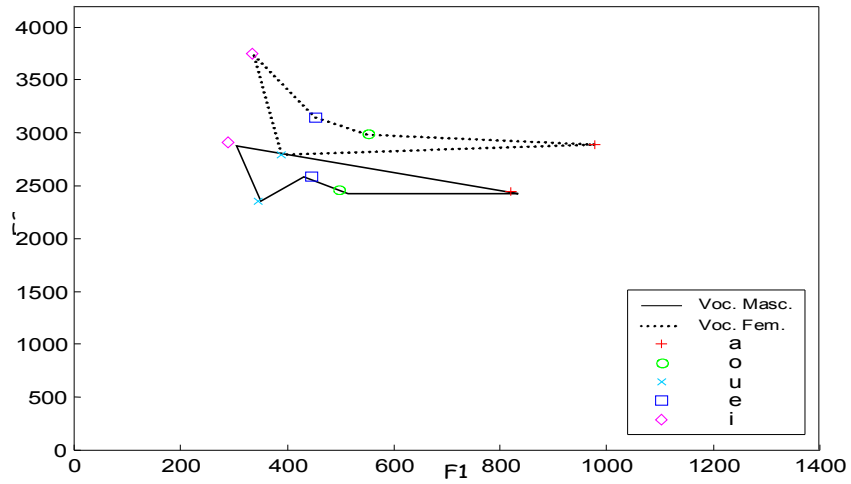


Fig 7 - Funciones de densidad de probabilidad basadas en Mezclas Gaussianas obtenidas de los datos para las vocales en el plano F1-F2 de voces masculinas. (El tono gris indica la probabilidad de encontrar una vocal en esa zona del plano F1-F2)

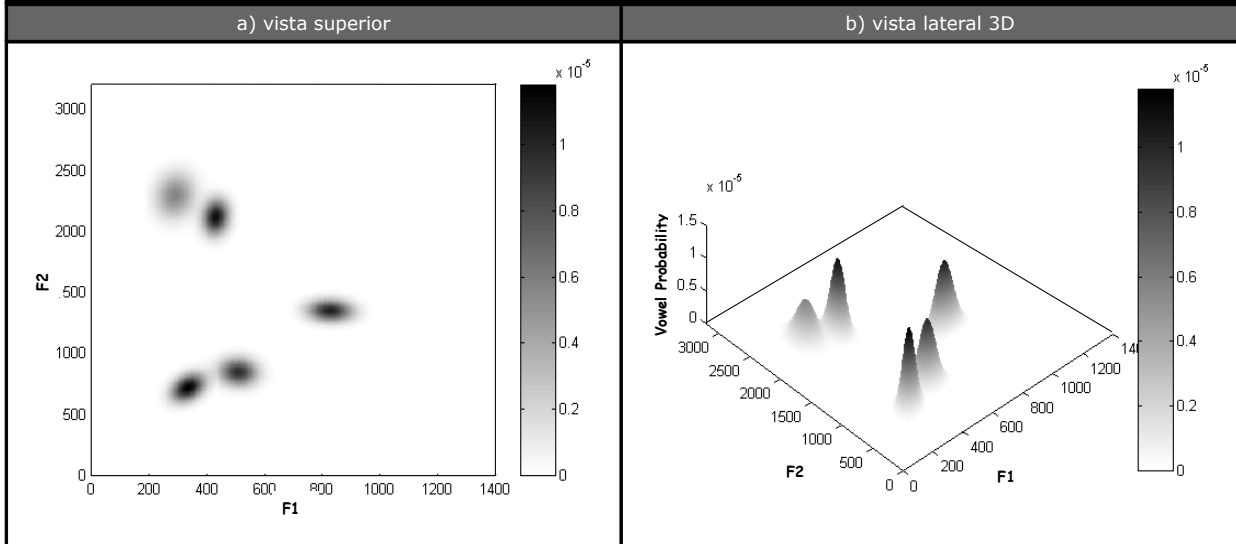


Fig 8 a) y b) - Comparación de los valores obtenidos en este estudio con los de G & M (8) para F1 y F2 voces masculinas respectivamente.

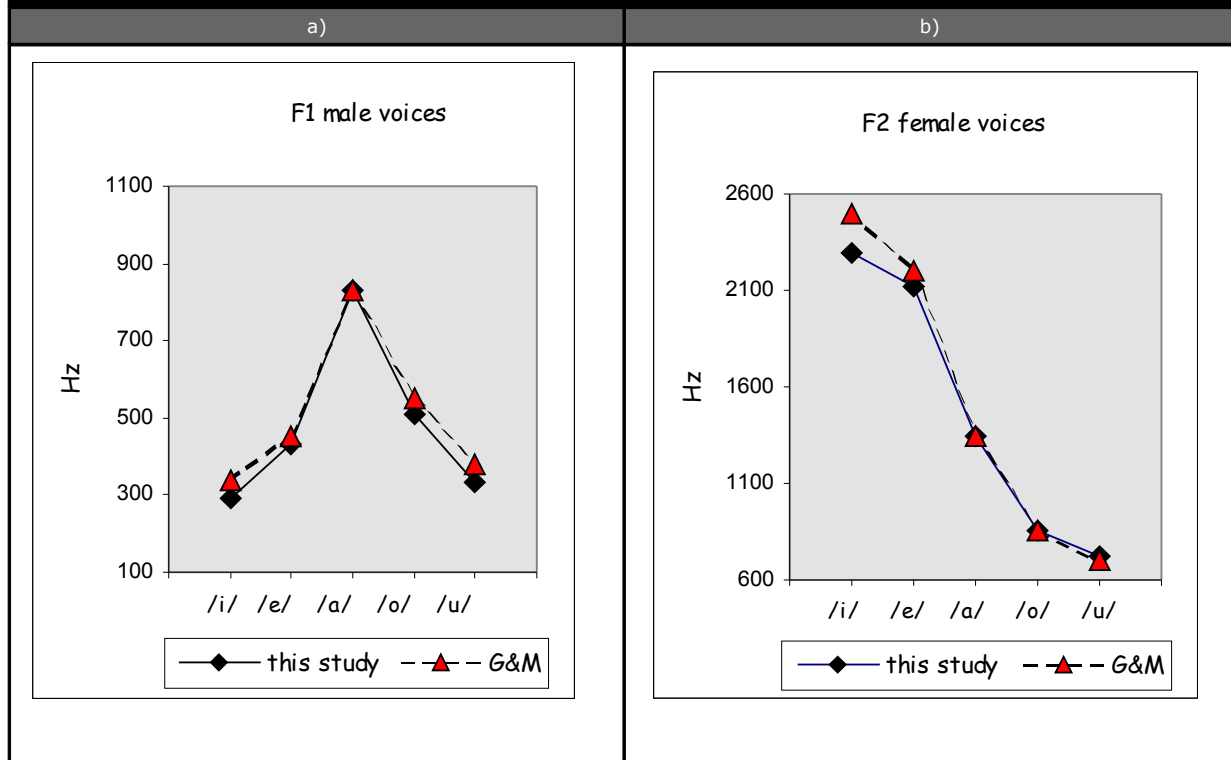


Figura 9 a) y b). Comparación de los valores obtenidos en este estudio con los de G & M (8) para F1 y F2 voces femeninas respectivamente.

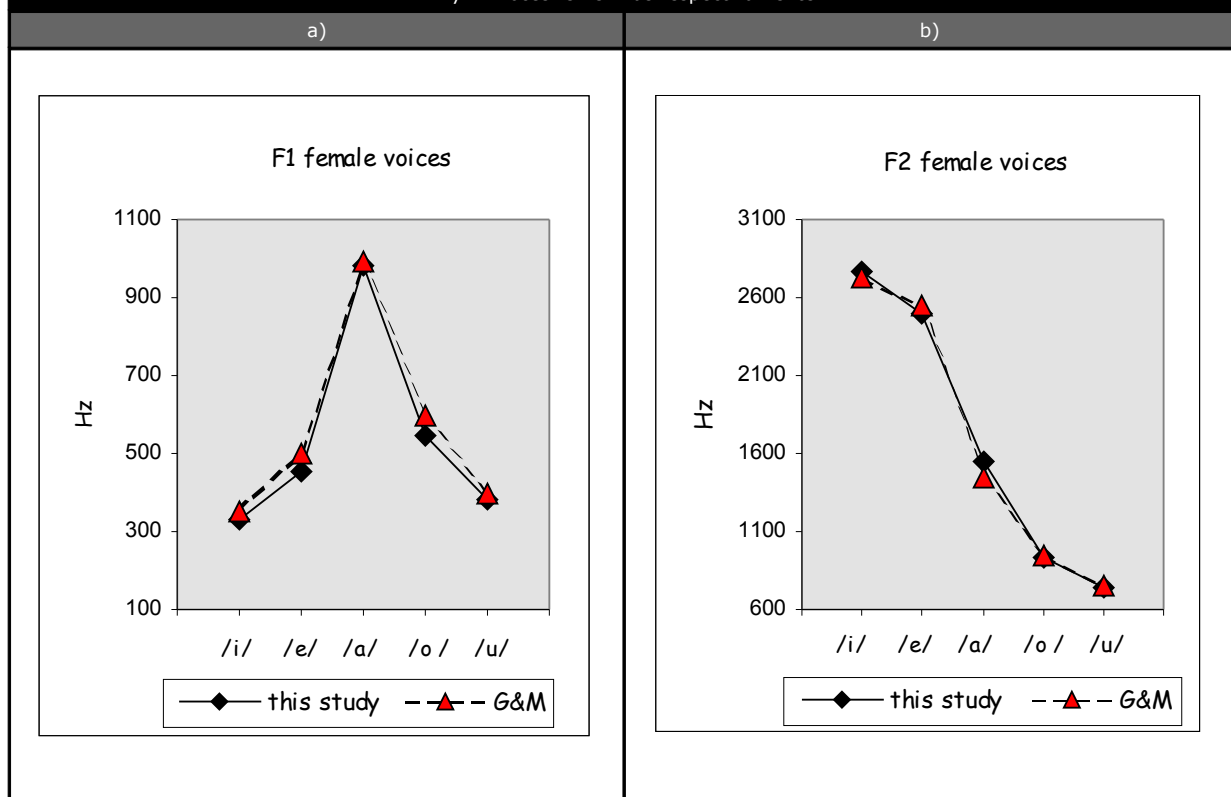
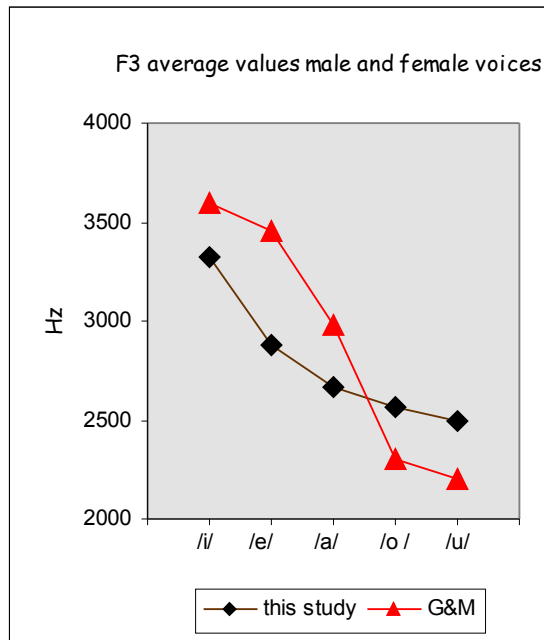


Fig 10) Comparación de los valores F3 obtenidos en este estudio, con los de G & M (8) para el promedio de voces femeninas y masculinas.



**TABLA I.** Valores promedios de la frecuencia glótica F0 y de los formantes F1, F2, F3 y F4 para los dos grupos de sujetos analizados. Promedios de los anchos de banda B1, B2 y B3 para voces femeninas y B1, B2, B3 y B4 para voces masculinas.

	sexo	i	e	a	o	u
F0	Fem	207	205	205	204	204
	Mas	130	125	127	124	124
F1	Fem	330	330	330	546	382
	Mas	290	430	830	510	335
B1	Fem	70	80	110	97	74
	Mas	63	75	105	83	80
F2	Fem	2765	2500	1553	934	740
	Mas	2295	2120	1350	860	720
B2	Fem	130	156	160	130	150
	Mas	103	106	106	105	112
F3	Fem	3740	3130	2890	2966	2760
	Mas	2915	2628	2450	2480	2380
B3	Fem	178	190	210	185	210
	Mas	174	140	142	156	208
F4	Fem	4366	4150	3930	3854	3380
	Mas	3645	3610	3665	3485	3355
B4	Fem					
	Mas	124	180	197	170	150?