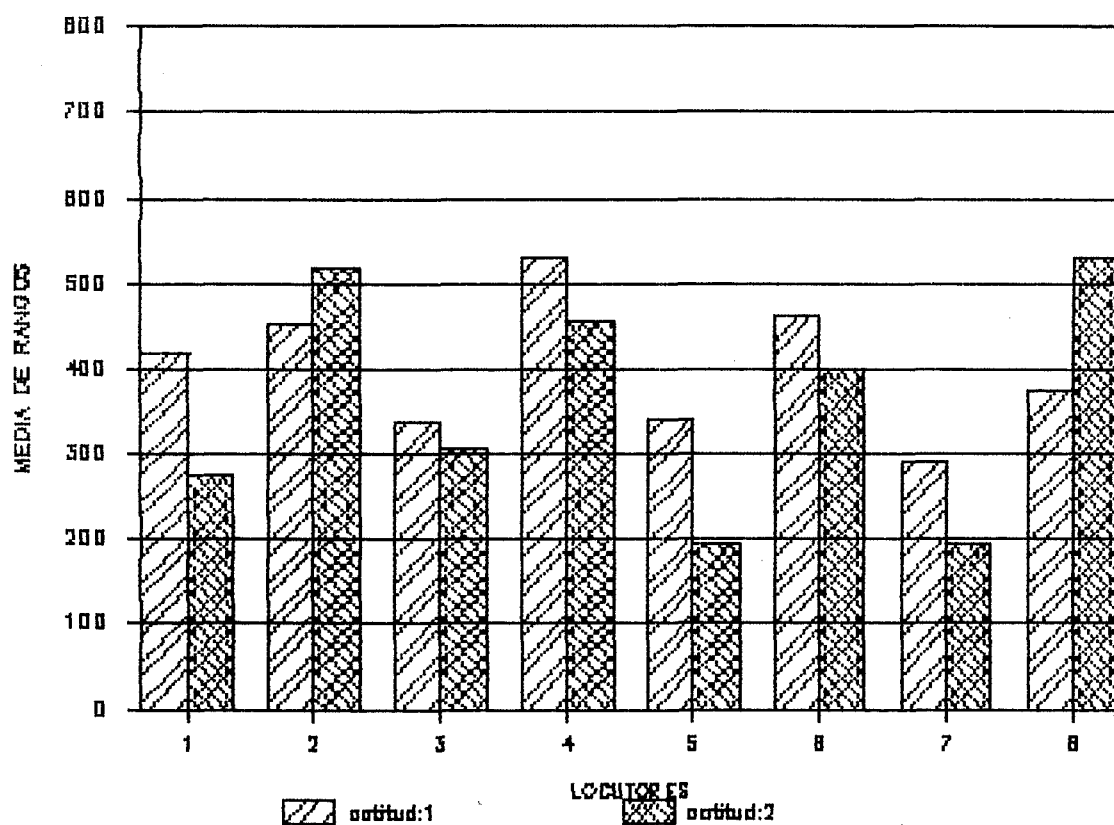
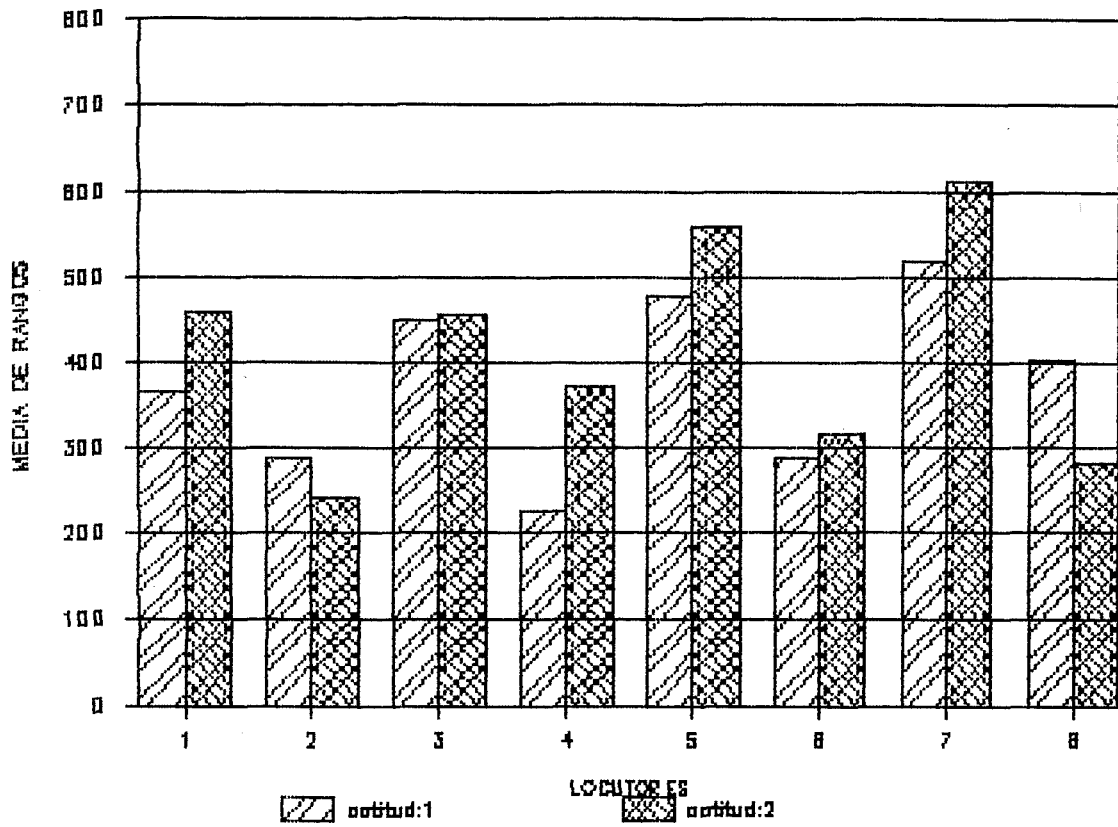


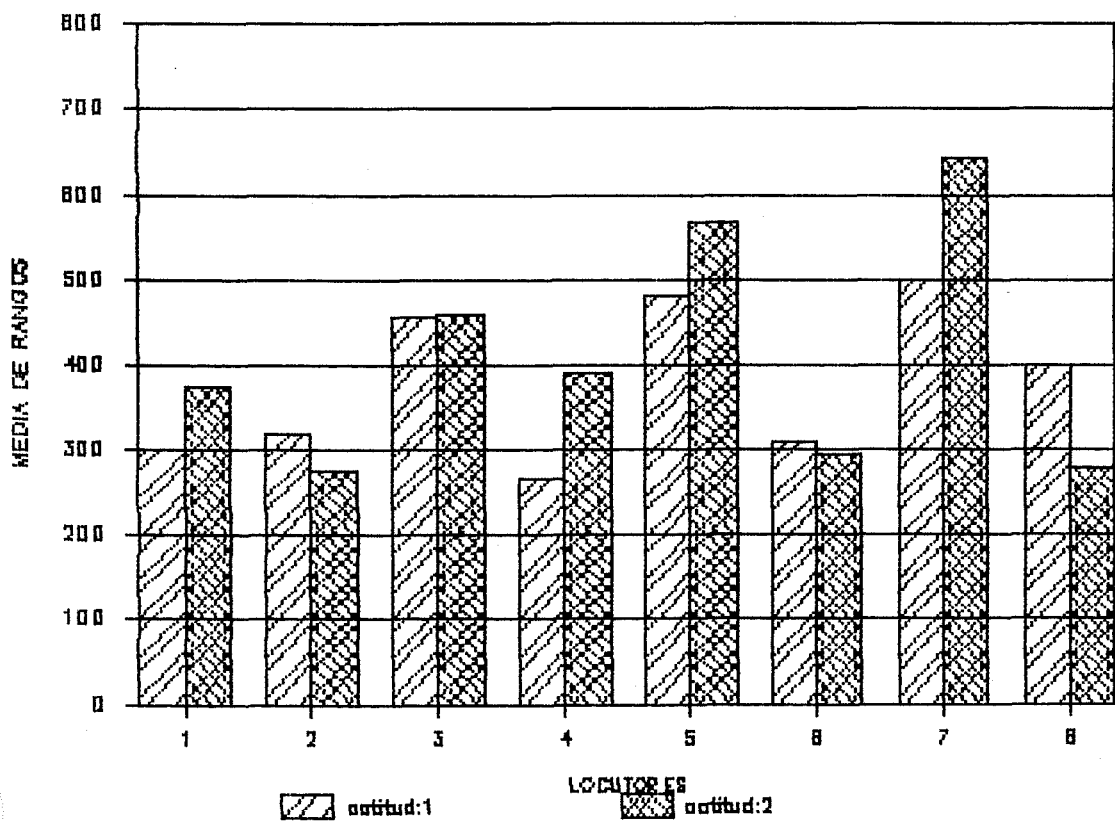
## COLOR DE LA PIEL (oscuro-claro)



## COLOR DEL PELO (rubio—moreno)



## COLOR DE LOS OJOS (claros—oscuros)



Conclusiones: La serie de asociaciones entre variables localizadas mediante el análisis estadístico permiten un considerable grado de previsibilidad sobre la forma en que el receptor se imaginará al locutor de cualquier voz humana emitida a través de un altavoz. O, diciéndolo de otra forma, que las correlaciones encontradas indican que manipulando determinados aspectos formales del sonido de una voz es posible controlar en buena medida la figura que surgirá en la mente del oyente cuando este intente imaginarse al locutor.

6. ANALISIS ELECTROACUSTICO DEL TIMBRE EXPRESIVO DE LA VOZ.

---

### 6.1. NUEVA SELECCION DE VOCES.

Puesto que el objetivo del análisis electroacústico es buscar relaciones entre características concretas del espectro de las voces con los resultados obtenidos por cada voz en los tests de percepción, es importante que las voces que se analicen hayan obtenido previamente en los tests respuestas claras desde el punto de vista estadístico. Lógicamente, sin que se cumpla esta condición, localizar asociaciones entre la percepción sicoacústica de los sujetos experimentales y la forma espectral del sonido resultaría prácticamente imposible. Por esta razón, se han seleccionado para realizar la última parte del experimento las voces que obtuvieron estadísticamente una respuesta mas coherente de los oyentes.

Así, se seleccionaron para ser analizadas las voces de las dos locutoras y los dos locutores que obtuvieron en los tests de percepción unos Indices de Coherencia (Ic) más altos en las respuestas de los sujetos experimentales (ver Pgs.111-112).

El criterio para elegir estas voces se obtuvo sumando los Indices de Coherencia obtenidos por las dos versiones de cada locutor. La cifra obtenida nos permitía saber cuales eran los dos locutores masculinos y los dos femeninos que habían construido sus voces con mayor eficacia y en consecuencia seleccionarlos. Lógicamente las voces

seleccionadas fueron las que daban un "Ic" total más alto,  
es decir:

Locutor	Versión	Ic.	SUMA	SELECCIONADAS
Loc-1	V.11	63,6	124,9	NO
	V.21	61,3		
Loc-2	V.12	45,4	129,5	SI
	V.22	84,1		
Loc-3	V.13	90,9	163,6	SI
	V.23	72,7		
Loc-4	V.14	68,5	147,0	SI
	V.24	78,5		
Loc-5	V.15	81,8	163,6	SI
	V.26	81,8		
Loc-6	V.16	65,9	99,7	NO
	V.26	34,0		
Loc-7	V.17	70,4	145,4	NO
	V.27	75,0		
Loc-8	V.18	41,6	80,2	NO
	V.28	38,6		

De este modo tendríamos la garantía de trabajar con las voces que habían generado en los oyentes una imágenes auditivas más claras.

En resumen, a partir del criterio expuesto un poco más arriba, decidimos que todo el análisis espectral y sus posteriores interpretaciones en función de los test de percepción se realizaría trabajando con las voces de los locutores: 2,3,4 y 5. (Josep Gaya, Mercedes Montalá, Juan A. Bernal y Ma. Rosa Pizà respectivamente).

## 6.2. METODOLOGIA DEL ANALISIS ESPECTRAL.

### 6.2.1. Adquisición de datos.

La adquisición de datos se realizó a partir de las mediciones efectuadas con un analizador de espectro sobre una serie concreta de sonidos en todas las versiones del texto portador seleccionadas.

De cada versión sonora se analizó sistemáticamente una muestra de 32 sonidos vocálicos ya definidos previamente en el apartado que hace referencia al diseño del texto portador (ver Pgs.85 a 89). Del total de los sonidos pre-definidos en

un principio se decidió excluir el estudio a los diptongos y triptongos ya que carecía de sentido analizarlos al no contemplar el método de trabajo la componente temporal y el análisis de las transiciones de una vocal a otra.

Los sonidos estudiados son los que corresponden a la vocal situada entre barras de cada una de las palabras siguientes.- 1: c/a/sa, 2: cas/a/, 3: desp/a/cho, 4: estab/a/, 5: est/a/nte, 6: m/a/no, 7: /a/zules, 8: cansad/a/, 9: acercu/é/, 10: lentam/e/nte, 11: lentament/e/, 12: r/e/ino, 13: /e/ra, 14: tard/e/, 15: p/e/sca, 16: sent/i/a, 17: v/i/sto, 18: fu/i/, 19: cog/i/, 20: /i/ndice, 21: l/i/bros, 22: /o/bras, 23: sorprendi/ó/, 24: olm/o/s, 25: rec/o/rriendo, 26: c/o/mo, 27: aut/ó/mata, 28: titul/o/, 29: tit/u/lo, 30: bamb/ú/, 31: /u/lula, 32: az/u/les.

Todas las mediciones se realizaron con un analizador digital que trabaja utilizando el algoritmo de la Transformada Rápida de Fourier, conocido también con las iniciales: FFT (Fast Fourier Transform). Concretamente, se trabajó con un analizador FFT Brüel & Kjaer tipo 2033. Y la señal acústica se introdujo en el analizador desde un magnetófono Revox tipo B-77.

#### 6.2.2. Método de medición.



Las mediciones se efectuaron trabajando con un fondo de escala de 10.000 Hz, con el que se obtiene una resolución máxima de 25 Hz. El calibrado del analizador se realizó siempre utilizando la misma señal pura de 1 Khz a 90 Db de intensidad. Ya la primera parte del experimento se había calibrado también la cadena de grabación con esta misma señal.

Seleccionado el fondo de escala y calibrado el analizador, se procedía a la toma de muestras registrando en el analizador fragmentos temporales del discurso sonoro de 0,4 segundos que contuviesen los sonidos vocálicos que nos interesaba estudiar. A partir de este momento, una vez tomada una muestra que contuviese la vocal que nos interesaba en cada ocasión, se procedía a su análisis frecuencial haciendo otra nueva selección de un fragmento temporal de la zona más estacionaria del sonido que nos interesaba, mediante una ventana "hanning" de 40 Milisegundos.

Una vez frente al espectro se procedía a medir, siguiendo el método clásico, la frecuencia en Hz a la que estaban situados la frecuencia fundamental y todos y cada uno de los formantes que aparecían en el espectro de 0 a 10.000 Hz. Dada la escasa resolución que nos permitía el fondo de escala de 10.000 Hz, la frecuencia fundamental se calculó midiendo la frecuencia del décimo armónico y dividiéndola por 10. Las mediciones de los formantes se

efectuaron situando el cursor en el centro de gravedad y anotando su frecuencia de situación.

Simultáneamente, se anotaba también la intensidad en Db de la frecuencia fundamental, y la intensidad y la anchura de banda de cada uno de los formantes. Para calcular el ancho de banda de los formantes se tomó como punto de referencia una altura situada a (-10) Db de la intensidad máxima del formante, a esta altura se medía la frecuencia del armónico más alto del formante y a la cifra obtenida se le restaba la frecuencia del armónico más bajo de ese mismo formante.

Siguiendo este método cada formante quedaba caracterizado por tres parámetros distintos: Frecuencia (Hz), Intensidad (Db), y Anchura de Banda (Hz). Puesto que en cada sonido, además de la frecuencia fundamental, aparecen unos 8 formantes, cada uno de los sonidos vocálicos analizado queda a su vez caracterizado, aproximadamente, por unos 26 parámetros.

### 6.2.3. Discusión del método.

El problema esencial con el que se enfrenta cualquier investigación que pretenda abordar la voz a partir del

análisis electroacústico es la enorme cantidad y complejidad de los datos obtenidos.

Ya el análisis de la evolución en el tiempo de un simple oscilograma generado por los impulsos sonoros de un discurso oral con el fin de estudiar la distribución de las pausas, los ritmos orales, la duración de sonidos concretos, las unidades fónicas, etc. plantea serias dificultades de procesamiento de la información obtenida. La variabilidad acústica del texto oral: transiciones entre sonidos, cesuras depresivas en lugar de pausas, enlaces y separaciones aparentemente indiscriminados de los sonidos del texto, alteración de las cantidades sonoras en función de la situación emocional o el estilo del locutor, etc, etc, generan una cantidad de información nada fácil de manejar. Si a todo esto se le añade el análisis frecuencial de un espectro que también evoluciona en el tiempo, la complejidad del procesamiento de los datos se multiplica monstruosamente.

Aparece entonces el problema de disponer de muchos más datos de los que que somos capaces de tratar. El análisis electroacústico mediante analizadores que trabajan con el algoritmo FFT (transformada Rápida de Fourier), genera mucha más información sobre los sonidos de la voz de la que en estos momentos somos capaces de interpretar. La respuesta a este problema, lógicamente, ha sido el trabajo por aproximación.

Desde las aproximaciones más elementales que parten del estudio de los formantes, fragmentando los sonidos del discurso fonema a fonema, y que tiene su origen en los primeros sistemas de análisis espectrográfico aparecidos en los años 40 (MARTINEZ,1983) (técnica que sigue dando actualmente buenos rendimientos); hasta otras mucho más completas y complejas como la predicción lineal (LPC), o los sistemas que integran la variabilidad temporal, como los basados en la "función de alineamiento temporal" (CASACUBERTA,1987), que aparecen con los modernos métodos de tratamiento numérico de señales; absolutamente todas las investigaciones se limitan a diseñar distintos modelos de aproximación que procesan solo una parte de la información numérica disponible con mayor o menor acierto y eficacia.

En esta investigación proponemos una aproximación a la información acústica del discurso oral que pretende exclusivamente estudiar el timbre individual; es decir, que busca localizar aquellos aspectos del sonido de la voz que diferencian acústicamente a un individuo de otro. Puesto que partimos de la hipótesis de que la percepción acústica del timbre personal depende fundamentalmente de los sonidos vocálicos, la adquisición de datos se ha realizado sólo a partir de una muestra de sonidos vocálicos. Y puesto que el estudio de los formantes es una técnica extremadamente cómoda y simple que sigue dando excelentes rendimientos en fonética experimental, es la que hemos elegido para desarrollar esta parte de la investigación,

aunque, lógicamente, con algunas modificaciones en función de nuestras propias necesidades.

Así, la aproximación al espectro real de los sonidos que nos interesaba analizar se realizó siguiendo en la mayoría de los aspectos una metodología clásica en fonética experimental, es decir, partiendo del concepto de "formante acústico", y midiendo sus frecuencias centrales y sus intensidades. No obstante, se han introducido en el método algunas innovaciones respecto a la metodología fonética habitual.

El fondo de escala, que en los estudios fonéticos acostumbra a ser de 5.000 Hz, se ha ampliado en nuestro estudio a 10.000. Esta modificación se apoya en la hipótesis de que las frecuencias altas del espectro aportan al receptor información acústica sobre el timbre personal del sonido de la voz. Esta hipótesis ha sido ya expuesta y mantenida por otros investigadores: J. Martí (1984), por ejemplo, afirma que el tercer y cuarto formantes son puntos muy significativos en la identificación de las personas; en esta misma línea Quilis (1982) dice que los formantes más altos son producto del timbre individual de cada hablante. El resultado de doblar el espacio espectral del estudio es la aparición más o menos regular de otros 4 ó 5 formantes por encima de los tres que suelen estudiarse en fonética experimental, con el consecuente incremento de parámetros que nos aproximan algo más al espectro real de la voz.

También es innovador el método de medición de los anchos de banda. Mientras en la metodología fonética esta medición se realiza a (-3) Db, en esta investigación los anchos se han medido a (-10) Db. La introducción de esta modificación se justifica en la voluntad de conseguir más información para caracterizar cada formante, concretamente se persigue reflejar de forma más completa la energía acústica que concentra cada formante.

### 6.3. PRIMERAS APROXIMACIONES A LOS DATOS Y CONSTRUCCION DE HIPOTESIS.

#### 6.3.1. Distribución de los formantes en la parte alta del espectro.

A medida que íbamos analizando sonidos y configurando el fichero de datos pudimos observar que los formantes no se comportaban de la misma forma en la parte alta que en la parte baja del espectro. Mientras que F1 y F2, es decir el primer y el segundo formantes, y en ocasiones también F3, se sitúan sistemáticamente en frecuencias muy similares en función del sonido analizado, a partir de F3 esto ya no se cumple y la situación de los formantes deja de ser coherente en función de su orden. Así, F5 tanto puede aparecer situado a 3.600 como a 6.175 Hz y F6 puede oscilar

perfectamente entre los 4.450 y los 9550 Hz. Observe el lector, por ejemplo, la siguiente tabla en la que aparecen las frecuencias de los formantes de las "A" y la "E" del locutor 12 (Josep Gaya):

## DISTRIBUCION DE LOS FORMANTES DE LA VOZ-12 (Josep Baya)

Datos ordenados respetando el orden de los formantes.

	F0	F1	F2	F3	F4	F5	F6	F7	F8	F9
C/A/SA	75	450	1500	2400	3250	3600	4450	5100	6025	9000
CAS/A/	75	725	1450	2325	3100					
DESP/A/CHO	82	500	1300	2225	3475	4650	5225	6400	9700	
ESTAB/A/	81	425	1250	2250	3225	4500	5075	6450	8775	
EST/A/NTE	78	750	1400	2325	3650	4175	6150	9125		
M/A/NO	140	925	1450	2250	3925	5700				
/A/ZULES	75	750	1275	2600	3475	4175	6225			
DESV. ESTANDAR:		154,52	91,61	120,16	261,76	642,53	675,09	1464,09	1560,76	0,00

ACERQU/E/	105	225	2200	3600	5500	6675	7800	8175		
LENTAM/E/NTE	82	175	1875	2350	3400	4500	5650	7025	7500	9625
LENTAMENT/E/	125	350	1950	2500	3800	6175	9550	9775		
R/E/IND	77	375	2200	2600	3100	3850	5525	6650	8000	
/E/RA	107	275	2050	2575	3700	4500	5775	6350	8550	
P/E/SCA	81	350	2300	3375	4150	5425	6525	8275		
DESV. ESTANDAR:		73,12	150,29	473,83	769,42	998,20	1451,25	1172,04	428,82	0,00

Datos reagrupados.

	F0	F1	F2	1	2	3	4	5	6	7	8
C/A/SA	75	450	1500	2400	3250	3600	4450	5100	6025	9000	
CAS/A/	75	725	1450	2325	3100						
DESP/A/CHO	82	500	1300	2225		3475	4650	5225	6400		9700
ESTAB/A/	81	425	1250	2250	3225		4500	5075	6450	8775	
EST/A/NTE	78	750	1400	2325		3650	4175		6150		
M/A/NO	140	925	1450	2250		3925		5700			
/A/ZULES	75	750	1275	2600		3475	4175		6225		
DESV. ESTANDAR:		154,52	91,61	120,16	65,62	165,08	187,48	251,87	157,32	112,50	0,00
ACERQU/E/	105	225	2200		3600		5500	6675	7800	8175	
LENTAM/E/NTE	82	175	1875	2350	3400	4500	5650	7025	7500		9625
LENTAMENT/E/	125	350	1950	2500	3800			6175		9550	9775
R/E/IND	77	375	2200	2600	3100	3850	5525	6650		8000	
/E/RA	107	275	2050	2575	3700	4500	5775	6350		8550	
P/E/SCA	81	350	2300		3375	4150	5425	6525		8275	
DESV. ESTANDAR:		73,12	150,29	97,43	232,92	271,57	123,49	267,97	150,00	549,68	75,00



Revisando globalmente la distribución de los datos en la tabla donde se respeta el orden de aparición de los formantes en el espectro, se puede observar como la dispersión crece rápidamente a partir de F3, especialmente para los datos de las /E/, aumentando las desviaciones estandar hasta el punto de hacer pensar en una falta absoluta de relación entre los datos. Frente a esta tabla, parece evidente que la altura en Hz de los formantes altos no se distribuye en función del tipo de sonido vocálico que estamos analizando, como ocurre con los formantes bajos F1 y F2; y que la forma adecuada de analizar estadísticamente estos datos no es tampoco la agrupación clásica que los ordena según su secuencia de aparición de menor a mayor altura en Hz.

A medida que avanzábamos en el análisis y disponíamos de una mayor cantidad de sonidos estudiados pudimos observar como los formantes altos parecían concentrarse en torno a frecuencias concretas en función de la voz con la que estábamos trabajando. Esto nos hizo pensar en ensayar un reagrupamiento de los datos de cada sonido analizado en función de la altura en Hz, en lugar de en función del orden de aparición. El lector puede observar en la segunda tabla de la página anterior como al reagrupar los datos de cada sonido en función de su altura conseguimos que la dispersión disminuya de una forma evidente, pasando ahora las desviaciones estandar a ser perfectamente aceptables.