Parecía evidente que la situación dentro del espectro de los formantes altos no estaba en absoluto asociada de una forma rígida a los fonemas, es decir, a la posición del resonador vocal en función del sonido lingüístico que estaba produciendo. Pero tampoco quedaba totalmente claro que esta situación dependiese de las características individuales de la voz de cada informante, puesto que, a medida que íbamos acumulando mediciones de las muestras sonoras de cada voz, la gama de alturas en Hz que alcanzaban sus formantes era progresivamente más similar a la de las otras voces; no obstante, el número de ocasiones en que se situaba un formante a determinada altura del espectro sí que parecía depender de cada informante en particular.

Se optó entonces por hacer un nuevo tratamiento de los datos, reagrupándolos en bandas de frecuencia de una anchura determinada, para estudiar si la distribución de datos en las bandas era un instrumento capaz de discriminar unas voces de otras. En la tabla de la Pg. siguiente el lector puede ver una nueva reagrupación de los datos de la voz 12, esta vez prescindiendo de la organización en función de los sonidos en el eje horizontal y ordenados exclusivamente en bandas de 1.000 Hz:

DISTRIBUCION DE LOS FORMANTES DE LA VOZ-12 (Josep Gaya)

Nueva reagrupación de los datos en bandas de 1.000 Hz

| 16% | 23% | 12% | 14% | 15% | 5% | 7% | 7% |
|---|---|---|---|---|---|---|---|
| 2400 | 3250 | 4450 | 5100 | 6025 | 7800 | 8775 | 9000 |
| 2325 | 3100 | 4650 | 5225 | 6400 | 7500 | 8175 | 9700 |
| 2225 | 3225 | 4500 | 5075 | 6450 | 7025 | 8000 | 9625 |
| 2250 | 3600 | 4175 | 5700 | 6150 | | 8550 | 9775 |
| 2325 | 3400 | 4175 | 5500 | 6225 | | 8275 | 9550 |
| 2250 | 3100 | 4500 | 5650 | 6675 | | | |
| 2600 | 3375 | 4150 | 5525 | 6175 | | | |
| 2350 | 3600 | 4500 | 5775 | 6650 | | | |
| 2500 | 3475 | | 5425 | 6350 | | | |
| 2500 | 3475 | | | 6525 | | | |
| 2575 | 3650 | | | | | | |
| | 3800 | | | | | | |
| | 3925 | | | | | | |
| | 3700 | | | | | | |
| | 3850 | | | | | | |

En la parte superior de la tabla puede leerse el porcentaje de ocasiones en que aparecen formantes en esa determinada banda de frecuencias del espectro respecto al total de datos de la tabla. Frente a esta última tabla parece evidente, por ejemplo, que la banda de 3.000 a 4.000 Hz (23%) es mucho más definitoria de la voz 12 que la banda de 7.000 a 8.000 (5%). Si esta distribución de los datos sobre los formantes altos en bandas de frecuencia fuese significativamente distinta entre una voces y otras, esto implicaría que en la parte alta del espectro sonoro existe realmente información sobre el timbre personal y nos daría además una referencia importante para analizar más profundamente el espectro. El problema que aparecía entonces era la elección de los intervalos de las bandas de frecuencia para agrupar los datos, ya que ésta de ningún modo podía ser arbitraria.

6.3.1.1. Aplicación de la teoría de las "bandas críticas".

La única forma aceptable desde un punto de vista teórico de establecer intervalos de frecuencia para agrupar los datos recogidos en el análisis espectral es apoyándonos sólidamente en la teoría de la percepción.

Landercy (1973), recuperando los trabajos sobre percepción acústica de Zwicker (1960), y de forma concreta su diagrama, propone un método de análisis del espectro de

los sonidos del habla que se ajusta eficazmente a la estructura perceptiva del oído humano. Albert Landercy superpone sistemáticamente todas las mediciones realizadas sobre el espectro de un análisis frecuencial (frecuencias e intensidades) sobre el diagrama de Zwicker, y a partir de esta superposición reconstruye totalmente el espectro desembocando en lo que él llama "espectros de sonía". En este diagrama el espectro auditivo esta fragmentado en bandas de frecuencia que ascienden en progresión logarítmica, tal y como percibe la evolución frecuencial el oído humano. Los intervalos que configuran estas bandas están definidos en función de la relación que existe entre la percepción de la frecuencia y de la intensidad sonora subjetiva.

Puesto que haciendo evolucionar la frecuencia de una señal acústica de intensidad mecánica constante el oyente humano percibe, a determinadas frecuencias, cambios subjetivos de la intensidad que mecánicamente no existen; estos puntos de la gama frecuencial auditiva actúan como puntos críticos (umbrales perceptivos), tanto para la percepción de la intensidad como para la de la frecuencia. En cada uno de estos puntos críticos, el diagrama de Zwicker establece un limite de la banda crítica, y el siguiente límite aparece a la altura frecuencial en que la intensidad subjetiva deja de percibirse como constante y el oído humano vuelve a escuchar una variación.

La fragmentación del espectro acústico que se hace en el diagrama de Zwicker se aproxima casi exactamente a una fragmentación en tercios de octava.

Así, siguiendo parcialmente la propuesta de Landercy (en esta investigación solo ajustaremos al diagrama las frecuencias), utilizaremos como intervalos de frecuencia para agrupar nuestros datos las "bandas críticas" establecidas por Zwicker.

Los intervalos de frecuencia de las bandas son los siguientes:

0 a 45 Hz

45 a 90 Hz

90 a 180 Hz

180 a 280 Hz

280 a 355 Hz

355 a 450 Hz

450 a 560 Hz

560 a 710 Hz

710 a 900 Hz

900 a 1120 Hz

1120 a 1400 Hz

1400 a 1800 Hz

1800 a 2240 Hz

2240 a 2800 Hz

2800 a 3550 Hz

3550 a 4500 Hz

4500 a 5600 Hz

5600 a 7100 Hz

7100 a 9000 Hz

9000 a 11200 Hz

Esta primera aproximación a los datos nos llevó a definir la siguiente doble hipótesis de trabajo que debería ser probada o descartada por el análisis estadístico de los datos:

HIPOTESIS 1(a).

A partir del tercero, los formantes del espectro frecuencial de los sonidos vocálicos no se organizan en función de los rasgos acústicos lingüísticos, sino que dependen fundamentalmente de características sonoras individuales del locutor.

HIPOTESIS 1(b).

La distribución en bandas críticas de los datos obtenidos al medir los formantes altos del espectro de los sonidos vocálicos, constituyen un instrumento estadístico capaz de discriminar entre sí voces de distintos locutores o de asociar las voces que pertenezcan al mismo locutor.

Naturalmente, el trabajo con esta hipótesis se apoya en la fuerte variabilidad acústica del habla humana. La hipótesis parte teóricamente de la idea de que el timbre personal no es en absoluto algo estacionario o constante, sino que sufre permanentes variaciones acústicas al estar influenciado por factores como la tensión emocional, el nivel de fatiga física, el matiz expresivo que el locutor intenta imprimir a su voz, etc. No obstante, en tanto que cualquier oyente humano es capaz de reconocer determinados timbres de voz y de discriminar unos timbres de otros simplemente escuchándolos, eso supone que dentro del alto nivel de variabilidad al que nos referimos existen parámetros acústicos que se repiten con regularidad y que nuestro oído reconoce fácilmente; entonces, en consecuencia lógica con lo anterior, es coherente pensar que estos parámetros han de ser estadísticamente detectables.

### 6.3.2. Distibución de los formantes en la parte baja del espectro.

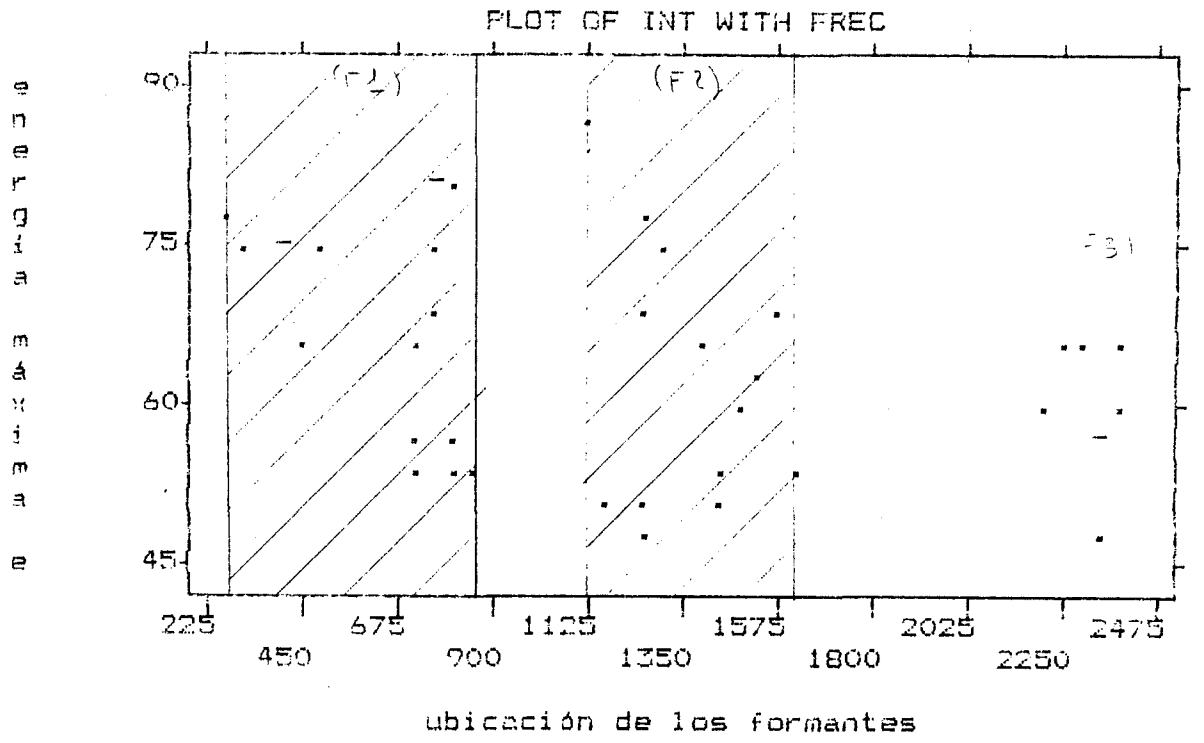
La aproximación inicial a los datos, que en una primera fase se realizó simplemente a partir de la observación global de las cifras y de su reagrupamiento según criterios diferentes, posteriormente se hizo a partir de gráficos globales que interrelacionasen la situación de todos los formantes en el espectro de frecuencia con su intensidad y su ancho de banda. Estos gráficos se realizaron aplicando el

procedimiento "PLOT" del paquete estadístico SPSS. La finalidad de estos gráficos era intentar descubrir, observando la situación de las nubes de puntos que representaban a los formantes, agrupamientos significativos que nos permitieran avanzar otras hipótesis concretas sobra la relación entre el espectro acústico y el timbre personal de los locutores.

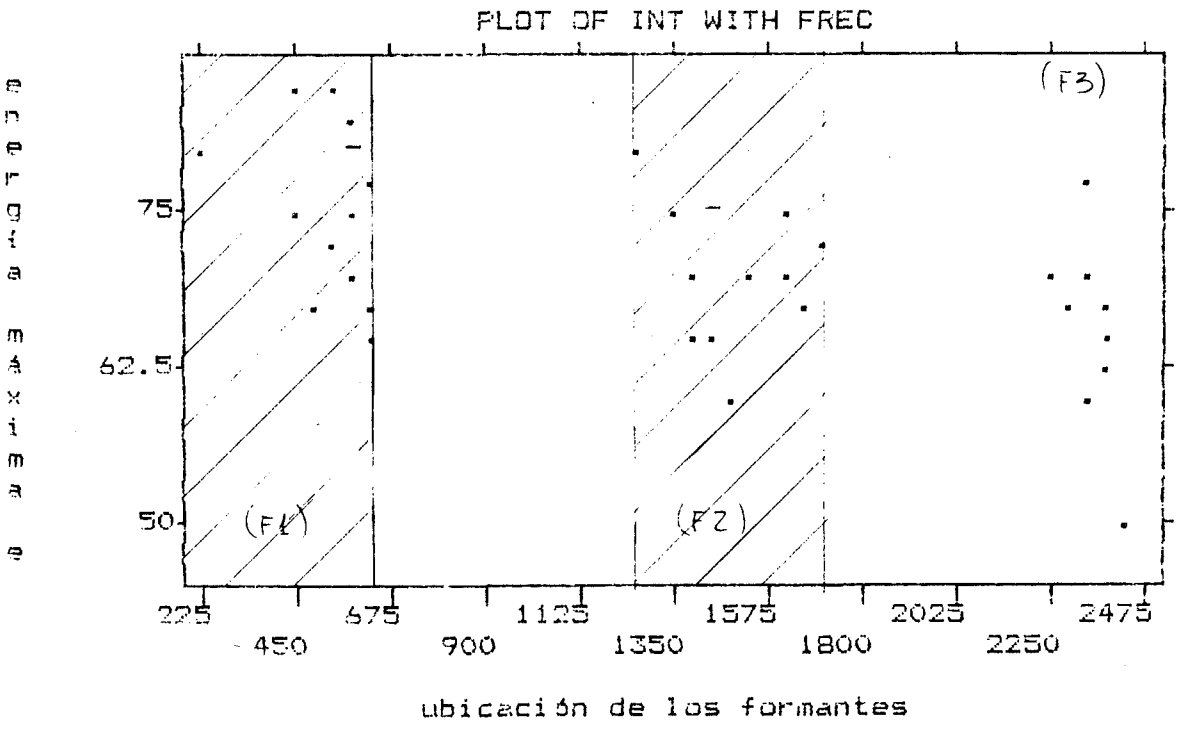
Se realizaron en principio gráficos para los datos de cada locutor que comprendían simultáneamente toda la gama de frecuencias estudiada (0 a 10.000 Hz), posteriormente se estudiaron los datos fragmentando el espectro en cuatro gráficos de 2.500 Hz y, por último, partiendo del conocimiento fonético de que la zona baja del espectro está absolutamente influenciada por la posición del resonador bucal al construir los sonidos lingüísticos, los primeros 2.500 Hz se investigaron separando los datos en gráficos distintos para cada sonido vocálico.

La revisión de estos últimos gráficos vocal a vocal, reflejó interesantes diferencias de un locutor a otro en la distancia que separa la nube de puntos del primer formante de la nube de puntos del segundo. Comparando, por ejemplo, los gráficos que relacionan intensidad y frecuencia de los locutores 2 y 4 el lector puede observar como la distancia que separa las nubes de puntos del primer y el segundo formantes es considerablemente distinta en cada caso (ver los gráficos de la Pg. siguiente).

Page 8 ANALISIS DE LAS "A" EN LA VOZ DEL LOCUTOR 2
DISTRIBUCION DE LOS FORMANTES F1 Y F2 DE LAS 'A'




Page 4 ANALISIS DE LAS "A" EN LA VOZ DEL LOCUTOR 4
DISTRIBUCION DE LOS FORMANTES F1 Y F2 DE LAS 'A'



En estos dos gráficos se representa la intensidad en el eje de ordenadas (energía máxima) y la frecuencia en el de abscisas (ubicación de los formantes), cada uno de los puntos (.) representa a un formante y el guión (-) representa a dos formantes situados exactamente a la misma frecuencia y que tienen la misma intensidad; los puntos situados a la derecha son valores de F3. El lector podrá ver con claridad como las bandas que contienen F1 y F2 en el gráfico del locutor 2 están mucho más cercanas entre sí que las bandas del locutor 4, que aparecen considerablemente distantes.

La mayor proximidad entre los puntos de F1 y los de F2 para el locutor 2, que entre los F1 y F2 del locutor 4, fue posible observarla sistemáticamente al comparar los gráficos de todos los sonidos estudiados del Loc.2 con los del Loc.4. La observación de los gráficos hacía pensar en que la distancia entre F1 y F2, dentro de los límites que marca el reconocimiento auditivo de cada fonema, podía ser también un parámetro capaz de diferenciar unas voces de otras.

Obviamente, la distancia entre F1 y F2 puede ser expresada con un solo dato que extraiga la diferencia entre ambas frecuencias, es decir:

DISTANCIA = (F2-F1)

Este nuevo parámetro, fácil de calcular mediante una simple recodificación de los datos iniciales, nos permitió avanzar otra doble hipótesis de trabajo:

HIPOTESIS 2(a).

La variabilidad de los dos primeros formantes en los sonidos vocálicos, dentro de los márgenes del reconocimiento auditivo del fonema, depende de características sonoras individuales del locutor.

HIPOTESIS 2(b).

La distancia (D) entre los dos primeros formantes de cada sonido vocálico F1 y F2, expresada como la resta de frecuencias D = (F2-F1) es un parámetro capaz de discriminar voces de distintos locutores, o de asociar voces que pertenezcan al mismo locutor.

## 6.4. ANALISIS ESTADISTICO: COMPROBACION DE LAS HIPOTESIS.

### 6.4.1. Hipótesis de la distancia entre formantes.

La investigación fonética y la experimentación en síntesis de voz han establecido sólidamente las estructuras

acústicas de los sonidos vocálicos. No obstante, esas estructuras están definidas con unos márgenes de variación considerablemente amplios; es cierto que los márgenes con los que se trabaja resultan eficaces a nivel de síntesis, pero los datos sobre la situación de los formantes vocálicos son siempre el resultado de establecer valores medios entre diferentes mediciones. Ya Alarcos (1968), cuando publicó sus primeros resultados sobre mediciones de formantes de las vocales castellanas hablaba siempre de valores medios y ,especialmente sobre el segundo formante, sus aportaciones lo que hicieron fue definir centros de diferentes niveles de variación para cada vocal; Martinez Celdrán (1983), por ejemplo, en el 83 seguía afirmando que todavía no existe un consenso respecto a los datos espectrográficos de las vocales del castellano. Quilis (1981), en esta misma línea, dice que la identificación lingüística de las vocales no depende solo de la frecuencia absoluta de los formantes sino de la frecuencia relativa a la estructura total de los formantes del sujeto hablante, y afirma que esta estructura puede variar de una persona a otra.

Evidentemente, nuestra hipótesis sobre la dependencia entre el timbre individual y la distancia entre F1 y F2 no es en absoluto contradictoria con la teoría fonética y suscribe esta última propuesta de Quilis intentando profundizar en ella.

6.4.1.1. Pruebas de comprobación de la hipótesis.

Obviamente, si el análisis estadístico demuestra que la segunda parte de nuestra hipótesis es cierta, automáticamente quedará probada la parte primera.

El primer paso para decidir si la hipótesis de la diferencia de formantes era o no aceptable fue recodificar los datos de modo que pudiéramos disponer de un nuevo parámetro para cada vocal analizada que expresase la distancia entre F1 y F2. Procedimos pues a efectuar la resta: F2-F1 en la cadena de datos de cada fonema, lo que nos permitió disponer de un nuevo dato asociado a cada vocal estudiada del discurso portador, y el conjunto de estos nuevos datos ya posibilitaba el trabajo con otra variable cuantitativa, la variable "(F2-F1)".

Si podíamos probar, respectivamente, que la media de las (F2-F1) de las "A", las "E",las "I", las "O" y las "U" de cada locutor es significativamente distinta de la media de las (F2-F1) de las "A", las "E", las "I", las "O" y las "U" de todos los locutores restantes, nuestra hipótesis quedaría aceptada. Para realizar esta comparación entre (F2-F1) de locutores distintos se aplicaron una serie de pruebas T-Test que debían comprobar, vocal a vocal, que los datos de (F2-F1) del locutor "X" no pertenecían a la misma población que los datos de (F2-F1) del locutor "Y".

Puesto que, previsiblemente, la variabilidad acústica que caracteriza al habla humana haría que los resultados del test no fuesen coherentes al 100%, se decidió aplicar también una prueba alternativa que demostrase el corolario de la hipótesis de trabajo; es decir, se quiso comprobar si se cumplía también que comparando medias vocálicas de (F2-F1) de diferentes grupos de datos de un mismo locutor, los resultados reflejaban cada vez que estos datos sí pertenecían a la misma población.

Los datos estudiados eran de grupos independientes, en consecuencia, para realizar estas pruebas se aplicó el procedimiento T-TEST GROUPS del paquete estadístico "SPSS", comparando en parejas la medias vocálicas de (F2-F1) de cada locutor con los tres restantes, es decir: Loc.2 con Loc.4, Loc.2 con Loc.5, Loc.4 con Loc.3, Loc.4 con Loc.5 y Loc.3 con Loc.5. Y, a continuación, se compararon entre sí, también por el mismo procedimiento, las medias vocálicas (F2-F1) de las dos versiones sonoras del texto que había realizado cada locutor, o sea: V.12 con V.22, V.14 con V.24, V.13 con V.23 y V.15 con V.25.

En la página siguiente se presentan las tablas de resultados de estas pruebas.

---------TABLAS DE PRUEBAS T-TEST GROUPS---------
ESTUDIO DE LAS DISTANCIAS ENTRE EL PRIMER Y EL SEGUNDO FORMANTES
(Comparaciones realizadas a partir de F(2)-F(1) sonido a sonido)

Comparacion de (F2-F1) entre locutores distintos.

| | L2/L4 | L2/L3 | L2/L5 | L4/L3 | L4/L5 | L3/L5 |
|---|---|---|---|---|---|---|
| A | 0 | 0 | 0 | 0 | 0 | 0,001 |
| E | 0,96 | 0 | 0 | 0 | 0 | 0,01 |
| I | 0,917 | 0,001 | 0 | 0 | 0 | 0,085 |
| O | 0 | 0,01 | 0 | 0,446 | 0,032 | 0,022 |
| U | 0,604 | 0,473 | 0,266 | 0,8 | 0,386 | 0,534 |

(Grados de libertad entre 20 y 30, excepto la U que oscila entre 10 y 12

Comparación de (F2-F1)entre distintas versiones del mismo locutor.

| | V12/V22 | V14/V24 | V13/V23 | V15/V25 |
|---|---|---|---|---|
| A | 0,136 | 0,831 | 0,706 | 0,048 |
| E | 0,891 | 0,368 | 0,624 | 0,018 |
| I | 0,819 | 0,521 | 0,648 | 0,379 |
| O | 0,798 | 0,917 | 0,148 | 0,938 |
| U | 0,103 | 0,655 | 0,882 | 0,015 |

(Grados de libertad entre 10 y 17, excepto la U que oscila entre 3 y 6)

Comparacion de (F2-F1) entre distintos subgrupos de una misma version.
(Pruebas alternativas a las de V15/V25)

| | V15(1)/V15(2) | V25(1)/V25(2) |
|---|---|---|
| A | 0,13 | 0,626 |
| E | 0,476 | 0,44 |
| I | - | 0,142 |
| O | 0,121 | 0,21 |
| U | 0,725 | 0,111 |

(Grados de libertad entre 5 y 2)

Como el lector habrá podido observar, la primera tabla, en la parte superior de la hoja, expone los resultados de las comparaciones entre locutores trabajando con todos los datos disponibles de cada locutor, es decir, con los (F2-F1) de las dos versiones del texto acumulados.

En la parte superior de la tabla (eje de abscisas) se expresa qué locutores son los comparados en cada ocasión, y a la izquierda (eje de ordenadas) las letras explican qué fonema es el que se está comparando.

Pero vayamos a los resultados: si, teniendo en cuenta la enorme variabilidad de la información acústica, tomamos como limite aceptable una probabilidad de error de 10%, el lector puede comprobar que el 70% de las pruebas realizadas indican que la distancia entre F1 y F2 comparada vocal a vocal es significativamente distinta entre un locutor y otro, así, la revisión de esta primera tabla hace pensar ya en que nuestra hipótesis es aceptable.

Otra observación interesante a la vista de esta tabla es que las pruebas que no se ajustan a la hipótesis están concentradas casi exclusivamente en la fila de la vocal "U"; si tenemos en cuenta que el fonema /U/ de todos los sonidos vocálicos es el que en los análisis frecuenciales muestra sus F1 y F2 en una zona más baja del espectro (ambos formantes suelen estar ubicados entre los 200 y los 800 Hz), podríamos interpretar la falta de poder discriminante del

parámetro (F2-F1) en este sonido como el reflejo de cierta incapacidad del resonador bucal para matizar la distancia entre formantes a esa gama tan baja de frecuencias.

En la segunda tabla, los datos de cada locutor están subdivididos en dos grupos, dependiendo de la versión del texto de la que estaban extraídos; y lo que se hace es comparar entre sí las distancias (F2-F1) de cada versión, vocal a vocal y locutor a locutor. La lógica de ordenación y de presentación de esta tabla es exactamente igual a la anterior.

A la vista de los resultados de las 20 pruebas estadísticas que componen la segunda tabla hemos de deducir que se cumple también el corolario de nuestra segunda hipótesis, es decir, que comparando las distancias entre F1 y F2 de las vocales de dos versiones sonoras distintas construidas por el mismo locutor no aparecen diferencias significativas. El lector puede ver en la tabla que solamente un 15% de las pruebas "T-Test" realizadas reflejan que las muestras que hemos comparado pertenecen a poblaciones distintas, o lo que es lo mismo, que el 85% de las pruebas estadísticas que presentamos en la tabla indican que a partir del (F2-F1) de los sonidos vocálicos se puede decidir si dos voces pertenecen o no al mismo locutor.

Si el lector ha observado la tabla de comparación de distintas versiones del mismo locutor con cierto