



MANUAL
de FONÉTICA
ESPAÑOLA

Articulaciones y sonidos del español

2.^a EDICIÓN

EUGENIO MARTÍNEZ CELDRÁN

ANA M. FERNÁNDEZ PLANAS

Ariel
LETRAS

ÍNDICE

<i>Prólogo</i>	11
CAPÍTULO 1. Introducción	13
1. Presentación	13
2. Medios instrumentales utilizados	15
2.1. Interpretación de los gráficos que contienen información articulatoria	15
2.1.1. Electropalatogramas	15
2.1.2. Electroglotogramas	21
2.2. Interpretación de los gráficos que contienen información acústica	22
2.2.1. Espectrogramas	22
2.2.2. Espectros	24
2.2.3. La curva de intensidad	25
2.2.4. Los pulsos glotales	26
2.2.5. Las curvas de F0	27
2.3. Interpretación de los gráficos que contienen información psicoacústica	28
CAPÍTULO 2. Las obstruyentes	31
1. Oclusivas	31
1.1. Las oclusivas bilabiales, dentoalveolares y velares [p, b, t, d, k, g]	31
1.2. La africada [tʃ]	47
2. Aproximantes espirantes	53
2.1. Las aproximantes bilabial, interdental y velar [β, ð, ʋ]	53
2.2. La aproximante y la articulación doble palatales: [j - j̟]	58
3. Sonoridad/sordez vs. tensión/relajación	63
3.1. Concepto de sonoridad y tensión en las consonantes	64
3.2. Sonoridad y tensión en las consonantes obstruyentes del castellano	69
3.3. Evidencias contra la sonoridad	71
4. Puntos de articulación de las oclusivas	86
4.1. Las barras de explosión y el VOT	88
4.2. Las transiciones y el «locus»	89
4.3. Las plantillas espectrales de Blumstein y Stevens	92
4.4. Las ecuaciones de locus	95
5. Fricativas	98
5.1. Naturaleza de las fricativas	98
5.2. La sibilancia	105
5.3. El punto de articulación	107

5.4. Variables	110
5.5. La tensión en las fricativas	119
CAPÍTULO 3. Las sonantes	123
1. Nasaes.	123
2. Laterales.	134
3. Róticas.	143
3.1. La cuestión terminológica: vibrante simple y vibrante múltiple	143
3.2. La posición articulatoria	144
3.2.1. Descripciones clásicas.	145
3.2.2. El efecto de Bernoulli	147
3.2.3. El teorema de Bernoulli en fonética.	148
3.2.4. El efecto de Bernoulli en el mecanismo de la vibrante múltiple.	149
3.2.5. Descripción a través de la electropalatografía.	151
3.3. La posición acústica	156
4. Semivocales.	161
4.1. Diferencias entre hiato y diptongo	161
4.2. Diferencias entre la semivocal palatal prenuclear y la espirante palatal	166
CAPÍTULO 4. Las vocales	171
1. Descripción articulatoria de las vocales del español.	171
2. Valores acústicos canónicos de las vocales españolas	173
2.1. Análisis de la voz masculina	174
2.2. Análisis de la voz femenina	177
2.3. Diferencias entre voz masculina y femenina	179
2.4. Discusión	180
3. Vocales abiertas y cerradas.	183
4. Relaciones entre vocales acentuadas y no acentuadas.	188
CAPÍTULO 5. La prosodia	193
1. Aspectos generales sobre entonación.	193
1.1. Pausas	195
1.2. Declinación	195
1.3. Rango	196
1.4. Movimientos de subidas y bajadas	196
1.5. Duración e intensidad	196
1.6. Núcleo o tonema	197
2. La entonación del español	199
2.1. Enunciativas neutras en frases simples	199
2.1.1. Los tipos de palabras y la entonación.	199
2.1.2. Pico en la postónica	203
2.2. Interrogativas absolutas en frases simples	204
2.2.1. Características y diferencias en las enunciativas neutras	204
2.2.2. Estructuras entonativas en el pretonema	206
2.3. Comparación con los perfiles de las enunciativas e interrogativas de Navarro Tomás.	211
2.4. La expresión del foco y otros aspectos conexos	214
2.4.1. Generalidades.	214
2.4.2. Foco ancho en las declarativas.	215

2.4.3.	Foco estrecho contrastivo	216
2.4.4.	Frase entonativa y frase intermedia	217
2.4.5.	Reajuste y declinación.	220
2.5.	Otros tipos de entonaciones	221
2.5.1.	Interrogativas relativas	221
2.5.2.	Entonaciones volitivas	221
2.5.3.	Entonaciones emocionales	223
	<i>Referencias bibliográficas</i>	225
	<i>Cuadro de sonidos del español.</i>	237
	<i>Índice de figuras</i>	239
	<i>Índice de tablas.</i>	247
	<i>Índice alfabético de términos.</i>	249

CAPÍTULO 1

INTRODUCCIÓN

1. Presentación

La fonética estudia los sonidos que utiliza el ser humano en el habla. Este estudio puede adoptar tres perspectivas diferentes, pero complementarias. La primera de ellas, tradicionalmente conocida como fonética articulatoria, estudia los sonidos desde el punto de vista de su producción por parte del emisor del mensaje. Si se estudian las modificaciones establecidas en el medio elástico en que se transmiten los sonidos articulados por el emisor, se tratará de la fonética acústica, la segunda perspectiva, encargada de estudiar la parte física del mensaje que viaja entre el emisor y el receptor. Por último, si se adopta la óptica del receptor del mensaje, se hablará de fonética perceptiva, la tercera. A pesar de este panorama tripartito, nosotros vamos a adoptar en este libro una visión de conjunto procurando mezclar las tres distintas perspectivas para que el fenómeno estudiado no se fragmente, de modo que, en general, veremos qué consecuencias acústicas tienen unos hechos articulatorios concretos y, además, qué efectos tienen esos hechos en la percepción por parte del receptor.

Como reza el subtítulo, el objetivo fundamental consiste en estudiar las articulaciones y los sonidos del español, básicamente del español peninsular, aunque no se ha dejado de aludir a las variantes más extendidas desde el punto de vista dialectal. En los últimos cien años se ha trabajado mucho sobre la fonética del español; en esta obra nosotros hemos tratado de recoger lo más relevante de lo que se ha dicho, pero sobre todo hemos intentado exponer con cierto detalle todo aquello que se ha investigado en nuestro Laboratorio con nuestro patrocinio durante los últimos años y también algunas investigaciones realizadas para esta obra expresamente, sobre todo en el terreno de la fonética articulatoria. El estilo de habla que fundamentalmente tiene en cuenta el libro es el formal y, aunque estemos en una comunidad bilingüe, siempre hemos tenido en cuenta en nuestros trabajos que los hablantes que a lo largo de treinta años han prestado y prestan su voz tuvieran el español como lengua materna y primera y que, por tanto, lo hablasen de forma natural y espontánea.

La investigación es innovadora en su planteamiento y, en muchas ocasiones, contradice las afirmaciones del pasado. Pedimos, pues, al lector que se acerque a estas páginas con la mente predispuesta a ver bastantes cambios respecto de lo que la tradición ha establecido y que los manuales al uso, y los profesores que se fían de ellos ciegamente, han ido repitiendo un poco miméticamente. Aunque buena parte de lo que se dice ya se ha publicado en forma de artículos de revista o comunicaciones a congresos (especialmente los datos acústicos), ahora queremos recogerlo y ordenarlo para ofrecer una visión de conjunto, visión que probablemente se desvanece en la fragmentación de los diversos artículos. Además, un manual permite ofrecer explicaciones más amplias de los fenómenos y prescindir algo más de la engorrosa estadística, obligatoria y esclarecedora en los trabajos de investigación, pero prescindible en los panoramas generales que ofrece un manual que permite un cierto grado de divulgación científica.

El libro aparece dividido en cuatro grandes apartados. Los tres primeros examinan los llamados segmentos: obstruyentes, sonantes y vocales. Los segmentos obstruyentes (que pueden ser sordos o sonoros) son aquellos que oponen algún obstáculo a la salida del aire, aunque sea mínimo. Se encuadran dentro de esta categoría las oclusivas (se incluyen en ellas las africadas), fricativas y aproximantes de tipo espirante. Las oclusivas son las que oponen el mayor obstáculo, puesto que impiden totalmente la salida del aire; las fricativas oponen una gran resistencia a la salida del aire por su estrechamiento en el punto de articulación, pero no la impiden por completo; y las aproximantes espirantes o no poseen el mismo grado de estrechamiento que las fricativas o no ofrecen gran resistencia a la salida del aire. Por otra parte, su alternancia en la lengua con las oclusivas o su posible realización fricativa en pronunciaciones enfáticas ofrece evidencias claras para su inclusión en la categoría de las obstruyentes.

Sonante es el nombre de la categoría de los sonidos que aun presentando algún obstáculo en el tracto vocal dejan paso libre a la salida del aire y, por regla general, se realizan con vibración de las cuerdas vocales, esto es, son sonidos sonoros. En las nasales, el aire queda retenido en algún punto de la boca, pero el aire sale de forma totalmente libre por la nariz. Las laterales obstruyen un lado de la boca o su centro, pero por los lados abiertos el aire fluye franco. Los sonidos róticos se desarrollan de forma vibrante, aproximante o fricativa, todos ellos con un cierto grado de obstáculo, pero con un elemento de tipo vocálico que permite el escape del aire ampliamente. Las semivocales poseen un grado mayor de abertura que las aproximantes espirantes, pero menor que las vocales más cerradas, por eso son clasificadas también dentro de esta categoría. Laterales, róticas y semivocales se suelen integrar dentro de los sonidos aproximantes. Obstruyentes y sonantes han sido clasificadas tradicionalmente dentro del grupo de las consonantes con la única salvedad de las semivocales que, si eran prenucleares, se dejaban dentro de las consonantes y, si eran posnucleares, se encuadraban en el conjunto de las vocales. La organización de estos segmentos en obstruyentes y sonantes los subdivide de forma más adecuada y organizada que su inclusión en la categoría global de consonantes.

Las vocales son los segmentos más abiertos. En su producción no hay ningún tipo de obstáculo importante a la salida del aire, por eso pueden aparecer de forma aislada sin el apoyo de ningún otro segmento; por el contrario, se constituyen en el sostén de las otras clases de sonidos, principalmente de las obstruyentes y, en menor medida, de las sonantes. Esto es sobre todo verdad en español; por eso mismo, el único segmento que puede aparecer en el núcleo de la sílaba y constituir una sola sílaba por sí misma es la vocal; ningún otro segmento posee esta capacidad en esta lengua.

El último capítulo del libro trata de los suprasegmentos, entre los cuales destaca la melodía que sustenta la entonación. De alguna forma, el acento y sus características fonéticas se estudian como soporte de la melodía, pues se constituyen en su armazón en la estructura prosódica de la frase. Lo verdaderamente fonético es la melodía, pues la entonación es la interpretación lingüística de la melodía; es decir, el hecho fonológico; no obstante, el uso tradicional confunde estos dos niveles y nosotros también nos permitiremos muchas veces el uso del término entonación por melodía, cuando no tengamos como objetivo prioritario hacer hincapié en la diferencia entre esos niveles.

2. Medios instrumentales utilizados

En cualquier estudio de fonética se hace imprescindible acudir a gráficos y esquemas proporcionados por medios instrumentales que apoyen las explicaciones que se dan y que demuestren las afirmaciones que se lanzan, especialmente cuando resultan de estudios experimentales que no siempre están de acuerdo con la visión tradicional comúnmente conocida de los hechos fónicos. En este libro aparecen abundantes gráficos de diferente naturaleza para cuya interpretación proporcionamos en este capítulo unas claves.

2.1. INTERPRETACIÓN DE LOS GRÁFICOS QUE CONTIENEN INFORMACIÓN ARTICULATORIA

2.1.1. *Electropalatogramas*

Los gráficos articulatorios que incorpora el texto están obtenidos fundamentalmente mediante electropalatografía o palatografía indirecta dinámica. El electropalatógrafo utilizado para obtener los gráficos que contiene este libro es el modelo EPG3 de Reading.

La electropalatografía es una técnica instrumental que proporciona información de los contactos que efectúa la lengua, un articulador móvil fundamental en la articulación de los sonidos, con la superficie palatina desde los incisivos superiores hasta el velo del paladar en una secuencia dinámica de habla. La elec-

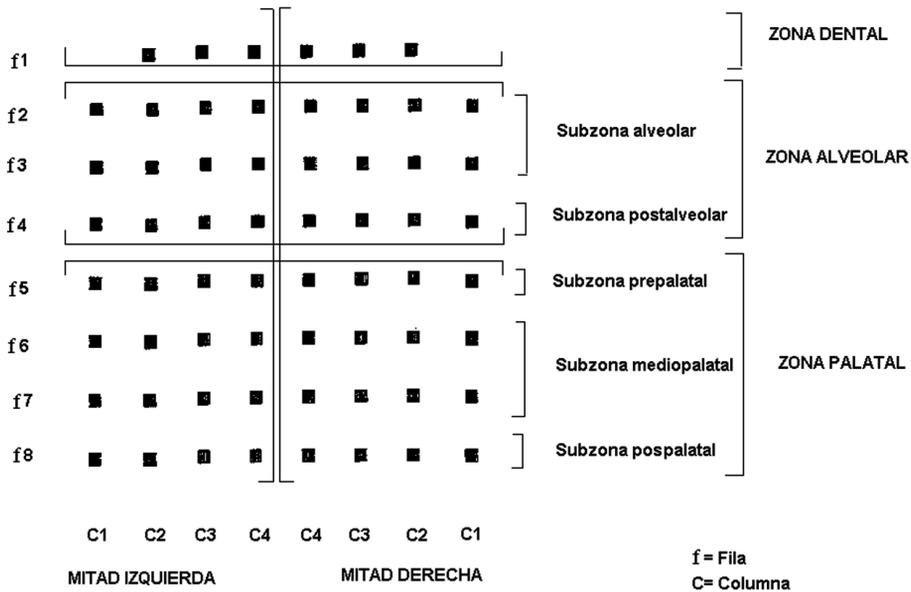


FIG. 1.1. *Clasificación de los electrodos del paladar artificial.*

tropalatografía ha superado a la antigua palatografía estática, por tanto, en fiabilidad y precisión de contacto linguopalatal y entre sus logros están la incorporación del parámetro temporal y, en las últimas versiones, la inclusión de información acústica fundamentalmente a través de oscilogramas.

Para obtener los datos de contacto, los informantes, a partir de los cuales se han conseguido las muestras de habla utilizadas, llevaban en la boca un paladar artificial provisto de sesenta y dos electrodos clasificados en filas y columnas y distribuidos por la superficie palatina atendiendo a esquemas anatómicos prefijados, con el objetivo de asegurar la comparación entre diferentes informantes analizados, cada uno de los cuales posee su propio paladar artificial, que se adapta a la dimensión morfológica de su cavidad bucal. Aunque electropalatográficamente existe la posibilidad de trabajar con paladares universales aptos para múltiples informantes, todos los estudios realizados en el Laboratorio de Fonética de la UB (de los que extraemos los gráficos que aparecen en este libro) utilizan paladares personalizados para cada uno de los hablantes. De esta forma, los datos adquieren una mayor fiabilidad ya que los paladares artificiales consiguen un ajuste mucho más adecuado en la cavidad bucal de su poseedor. La clasificación de los electrodos seguida, una de las clasificaciones posibles, es la que aparece en la figura 1.1 (adaptada de Fernández Planas, 2001).

En la figura 1.1, los electrodos aparecen dispuestos en filas y columnas equidistantes puesto que se trata de una visión esquemática; sin embargo, la separación real de las filas sobre el paladar artificial no es así. En realidad, la distancia que hay entre las cuatro primeras filas (correspondientes a las zonas dental y alveolar)

es aproximadamente la mitad de la distancia que hay entre las filas de la zona palatal (f5 a f8). La propia introspección coincide con el hecho de otorgar a la zona propiamente palatal un área mucho mayor que la que cubre los alveolos. La disposición real de las columnas tampoco se realiza en líneas rectas ya que el paladar artificial sigue la configuración anatómica, es decir, cuanto más posterior es la fila más gana en anchura, de tal manera que la figura real que forman los electrodos, si observamos el paladar artificial longitudinalmente, es más bien trapezoidal.

La electropalatografía es una técnica que presenta algunas limitaciones a causa del espacio que realmente cubre el paladar artificial, a saber: no informa de la sonoridad/sordez, ni del comportamiento de los articuladores labiales, ni de lo que sucede en el velo del paladar propiamente dicho. Estas limitaciones son compensadas con la información acústica oscilográfica que incorpora el EPG3 de las secuencias grabadas y con la posibilidad de grabar conjuntamente información electropalatográfica y electroglotográfica. Por otra parte, la información relativa a la dentalidad se deduce de la primera fila de electrodos (que se sitúa justo en el punto de unión entre los incisivos superiores y las encías) aunque es cierto que el paladar artificial no recubre los dientes. Algunas articulaciones velares aparecen parcialmente; cuando se aprecian mejor es cuando se muestran en contacto con vocales anteriores. Finalmente, el electropalatógrafo tampoco informa de la región lingual que efectúa el contacto o la constricción; sin embargo, se asume que la región lingual implicada es la que queda inmediatamente debajo (o muy próxima) de la zona palatina activada, como se observa en la figura 1.2 adaptada de Fernández Planas (2001).

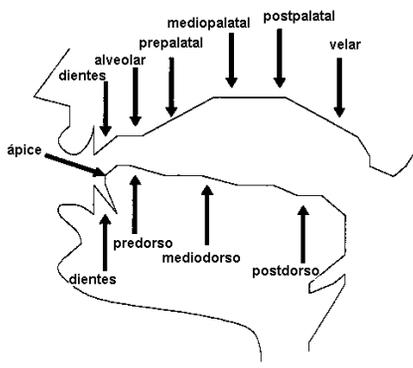


FIG. 1.2. *Regiones linguales y zonas palatinas.*¹

1. Como se ve en esta figura, se sigue la tradición hispánica en la clasificación de las regiones linguales. El seguimiento de la tradición anglosajona incluiría la consideración de la «lámina», una pequeña zona de unos 15 mm inmediatamente siguiendo al ápice; es decir, situada en la parte anterior de lo que aquí cubre la región lingual predorsal. En este libro solemos utilizar la tradición hispánica en este punto aunque a veces podemos emplear el término laminal cuando nos referimos a alguna cuestión de detalle que implica a dicha región de la lengua.

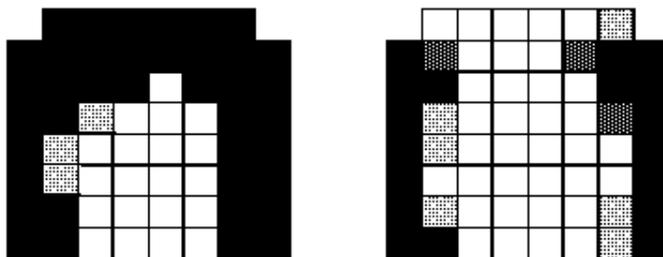


FIG. 1.3. *Electropalatogramas correspondientes a [t] (a la izquierda) y a [s] (a la derecha)*

Respecto a los gráficos que proporciona, la electropalatografía nos permite por un lado estudiar una configuración determinada, ya sea en un momento puntual de su desarrollo temporal o de forma promediada a lo largo de dicho desarrollo; y, por otro lado, el hecho de incorporar el factor temporal (aspecto crucial que distingue la electropalatografía de la palatografía estática clásica y la supera ampliamente) nos posibilita el estudio del paso de una configuración a otra en la cadena fónica y su evolución en el tiempo (Fernández Planas, 1999). En cualquier caso, constituye una herramienta fundamental en el estudio del fenómeno de la coarticulación. En estos dos sentidos, considerando o no la dimensión temporal, encontraremos a lo largo del trabajo dos tipos de gráficos como los que aparecen en las figuras 1.3 y 1.4.

Los gráficos de la figura 1.3 pueden corresponder o bien a la media de un momento puntual del desarrollo del segmento entre varias repeticiones de la secuencia de la que está extraída la consonante (normalmente el punto de máximo contacto de electrodos que se suele producir hacia la mitad de su evolución temporal, en su estado estacionario característico), o bien a la media en una emisión (o en varias) de los contactos realizados durante todo su desarrollo temporal. En cualquier caso, los gráficos resultantes incorporan diferencias de grado de contacto o de permanencia en el tiempo de activación de un electrodo determinado con dos sombreados extremos: blanco, que indica que en ningún momento el electrodo indicado ha sufrido activación; y negro, que señala que el electrodo en cuestión ha estado activo en todo momento y/o en todas las emisiones promediadas. Los diferentes niveles de grises señalan diferencias de porcentaje de activación entre los dos extremos: el sombreado más claro indica un nivel de activación hasta un 50 % y el más oscuro a partir de un 51 %.

En la figura 1.4 aparece el desarrollo temporal de una secuencia de electropalatogramas de la palabra *casita* y su oscilograma acústico. Cada uno de los seis segmentos que contiene la señal aparece asociado en la sucesión articulatoria de electropalatogramas a un electropalatograma representativo de cada segmento. Así se pueden ver claramente varios aspectos: la diferencia de punto de articulación entre una oclusiva velar (A) y una oclusiva dentoalveolar (E) ambas sordas, según se desprende del oscilograma; la diferencia de modo de articulación de (A)

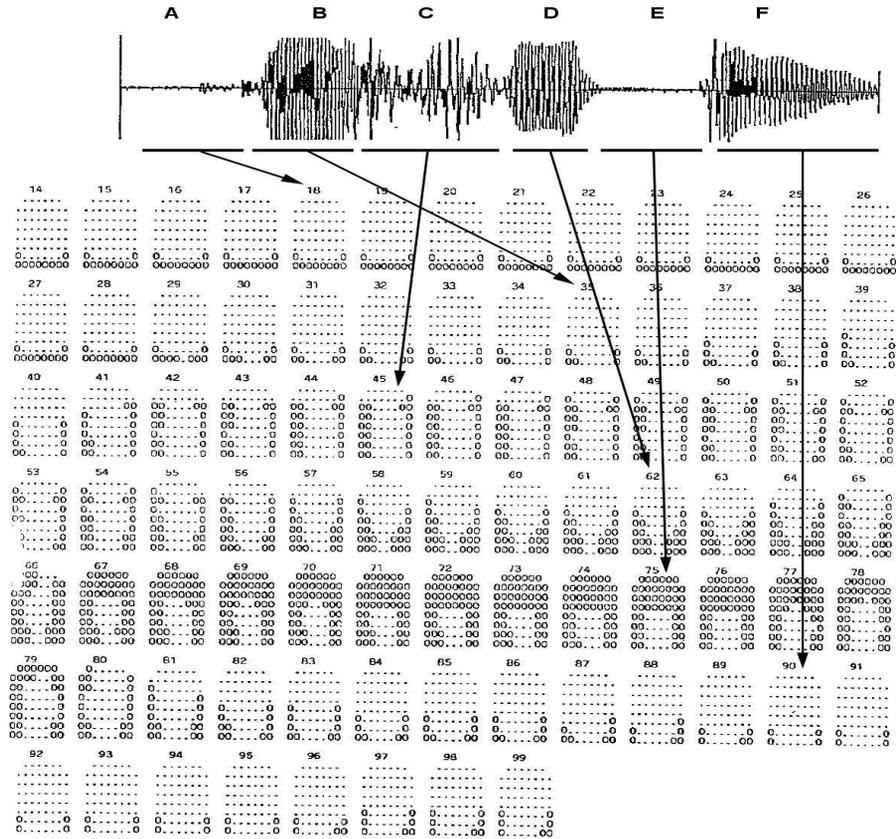


FIG. 1.4. *Desarrollo temporal de una secuencia de electropalatogramas y su oscilograma correspondiente.*²

y (E) respecto de la fricativa alveolar (C); la mayor anterioridad y altura de la vocal [i] (D) respecto a la vocal baja central (B) y (F); y, entre estas últimas, la mayor duración de la vocal final de emisión puesto que obtenemos un electropalatograma cada 10 ms. Un análisis más minucioso nos puede informar de otros aspectos como, por ejemplo, la tensión deducida a partir del mantenimiento en el tiempo del máximo contacto de una determinada configuración linguopalatal.

Los estudios que utilizan la electropalatografía como herramienta instrumental necesitan poder reducir de algún modo los gráficos obtenidos a un índice numérico que cumpla dos requisitos: en primer lugar, constituir un índice representativo de la configuración que representa el electropalatograma en cuestión; y, en segundo lugar, ser susceptible de sometimiento a un estudio estadístico que será el que determinará la validez de los datos obtenidos respecto a la hipótesis ex-

2. Agradecemos la inestimable colaboración de Toni Fernández Planas en la preparación de buena parte de los gráficos que contiene este libro.

<i>ÍNDICES</i>	
CAa = Contactos Anteriores (a = f1 – f4)	CAP = Contactos Anteriores (p = f5 – f8)
CPa = Contactos Posteriores (a = f4 – f1)	CPp = Contactos Posteriores (p = f8 – f5)
CCa = Contactos Centrales (a = f1 – f4)	CCp = Contactos Centrales (p = f5 – f8)

FIG. 1.5. *Significado de los índices desdoblados.*

perimental inicial y, de alguna forma, ordenará la enorme variabilidad del habla, como exponen Martínez Celdrán (1991), Llisterri (1991) o Hardcastle, Gibbon y Nicolaidis (1991) entre otros. En el terreno de la electropalatografía han sido muchos los autores que han propuesto variados índices, siempre con el objetivo de explicar mejor los aspectos más relevantes del contacto linguopalatal efectuado (Hardcastle, Gibbon y Nicolaidis, 1991; Fontdevila, Pallarès y Recasens, 1994). Los índices a los que se hará alguna referencia en este libro (véase la figura 1.5) serán los índices CA (índice de anterioridad), CP (índice de posterioridad) y CC (índice de centralidad) (desdoblados en cada mitad del paladar, CAa, CCa, CPa —en la mitad anterior (a)—, CAP, CCp y CPp —en la mitad posterior (p)—) propuestos por Fontdevila, Pallarès y Recasens (1994) y ya utilizados en una aplicación exhaustiva a las secuencias VCV del español por Fernández Planas (2001).

El índice que da idea de la anterioridad de los contactos (CA) nos informa específicamente, en primer lugar, del número de fila que presenta contacto(s) más anterior(es); en segundo lugar, del número de electrodos activados en esa fila; y, en tercer lugar, del nivel de contactos tras de sí en la dimensión longitudinal. De modo inverso, el índice que nos orienta sobre la posterioridad de los contactos (CP) nos indica el número de fila que presenta contacto(s) más posterior(es), el número de electrodos contactados en ella y el número de contactos delante de sí en la dimensión longitudinal. El tercer índice (CC) trata la centralidad de los contactos en el eje transversal sin atender a si su distribución es más anterior o más posterior.

Su cálculo está basado en razonamientos matemáticos considerando que la aportación de un solo electrodo activado (en una fila o en una columna) es más importante que el conjunto de electrodos que funcionan, respectivamente, después (para CA), antes (para CP) o más hacia los bordes palatinos (para CC) que él. Por lo tanto, según las fórmulas que calculan estos índices, su valor aumenta más en función del incremento de electrodos activados en una fila o columna dada que en función de variaciones en el esquema linguopalatal conjunto.

Los valores obtenidos en la aplicación de las fórmulas oscilan entre 0 y 1. Un valor más cercano a 1 indica mayor anterioridad (en el caso de CA), mayor posterioridad (CP) o mayor centralidad (CC) de activación de electrodos (véase la figura 1.6).

La ventaja de estos índices sobre otros, a nuestro modo de ver (coincidimos en este aspecto con Nicolaidis, 1999), se fundamenta en la observación bidimensional (longitudinal y transversal) de las distintas configuraciones, con lo cual, además de informar del grado de contacto (como muchos índices), informan tam-

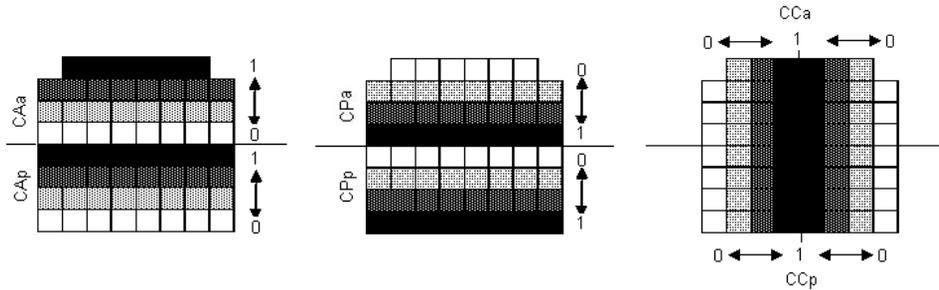


FIG. 1.6. Especificación esquemática de los valores de los índices. La escala de grises señala, en cada caso, la localización de los electrodos que más pesan en el cálculo del índice correspondiente.

bién de la distribución de los contactos y permiten diferenciar variaciones sutiles de punto de articulación, por ejemplo. Otra ventaja reside en la posibilidad de no utilizarlos todos a la vez necesariamente, como hacen por ejemplo Nicolaidis (1999) y Recasens y Pallarès (2001).

2.1.2. *Electroglotogramas*

En una ocasión hemos incorporado otro tipo de información articulatoria a través de un electroglotograma. Aparece unida a otra información articulatoria, electropalatogramas, y a información acústica, oscilograma y espectrograma. Este tipo de gráfico informa del movimiento de las cuerdas vocales, de las sucesivas aberturas y cierres y, por lo tanto, coincide con los sonidos sonoros. La figura 1.7 constituye un ejemplo de este tipo de gráficos. Como se ve, permite observar una onda muy parecida a la típica en forma de sierra del tono laríngeo, cuando la ventana temporal establecida es adecuada. En el ejemplo propuesto, correspondiente a la vocal [a] sostenida, aparece en la parte inferior del gráfico, mientras que la

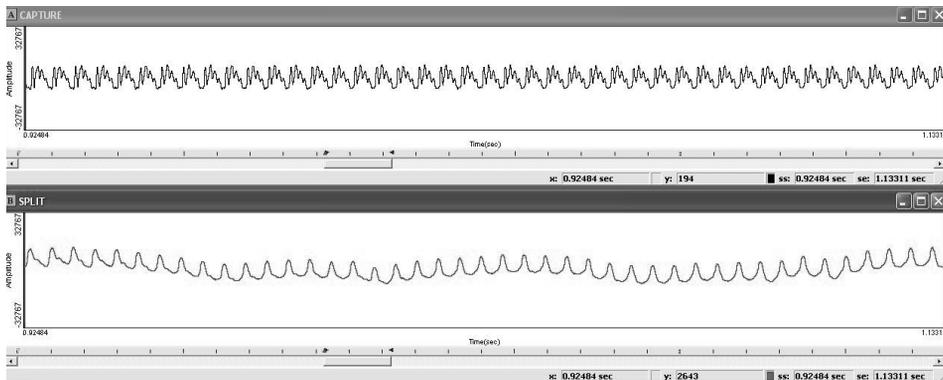


FIG. 1.7. Oscilograma superior y electroglotograma inferior.

parte superior está ocupada por un oscilograma; es evidente la diferencia entre la forma de onda compleja acústica (arriba) y la forma del tono laríngeo (abajo).

2.2. INTERPRETACIÓN DE LOS GRÁFICOS QUE CONTIENEN INFORMACIÓN ACÚSTICA

Como existe una amplia bibliografía de introducción al análisis y a la interpretación de los gráficos utilizados en fonética acústica (véanse Martínez Celdrán, 1996, 1998; Fernández Planas, 2005; Román 2011, etc.), no vamos a extendernos demasiado en este apartado, en todo caso invitamos a todos aquellos que encuentren dificultades en la lectura de dichos gráficos a que acudan a cualquiera de los libros citados.

2.2.1. *Espectrogramas*

Los espectrogramas utilizados en este libro son de banda ancha (300 Hz) siempre, puesto que son los idóneos para mostrar la cualidad de los sonidos: formantes vocálicos, ruidos fricativos, silencios oclusivos, pulsos glotales de los sonidos armónicos, explosiones, etc.

En la figura 1.8 presentamos dos espectrogramas de banda ancha que, como cualquier gráfico de este tipo, representan en la ordenada las frecuencias (Hz) y, en la abscisa, el tiempo (segundos). En la parte superior, el espectrograma está sin analizar; en la parte inferior, mostramos el mismo espectrograma segmentado y transcrito con algunos detalles incorporados. Hemos reseguído en blanco la trayectoria de los dos primeros formantes de las vocales, con lo cual se pueden seguir perfectamente las transiciones debidas a las influencias consonánticas y, en la penúltima vocal, hemos indicado cuál es el F1 (formante primero) y el F2 (formante segundo); hemos enmarcado la explosión de la oclusiva y el VOT y hemos señalado su fase de silencio, además hemos indicado la ausencia de barra de sonoridad en este sonido a través del rasgo ‘sordo’. Hemos señalado la presencia de pulsos glotales en los sonidos armónicos para que se compare con el ruido de las fricativas que no tiene dichos pulsos en forma de estrías. La intensidad es un parámetro que no se puede medir de forma precisa solamente a partir del espectrograma, se observa por el mayor o menor negror de los pulsos glotales o por la de cualquiera de los eventos señalados: formantes, explosión, ruido, etc.

Es conveniente tener muy en cuenta los datos de la ordenada, pues la frecuencia representada en el espectrograma suele ser diferente según lo que se pretenda mostrar; en general, la gama frecuencial mostrada en los espectrogramas puede oscilar entre 4.000 y 8.000 Hz. El de la figura 1.8, va de 0 a 4.250 Hz como muestra la ordenada, puesto que cada marca de la izquierda representa 250 Hz, aunque los números se escriban de 500 en 500. Por ejemplo, el F1 de la primera vocal [a] posee 719 Hz (obsérvese que el centro de la curva del F1 se sitúa sólo un poquito por debajo de la marca de los 750 Hz). En la abscisa, los valores temporales tam-

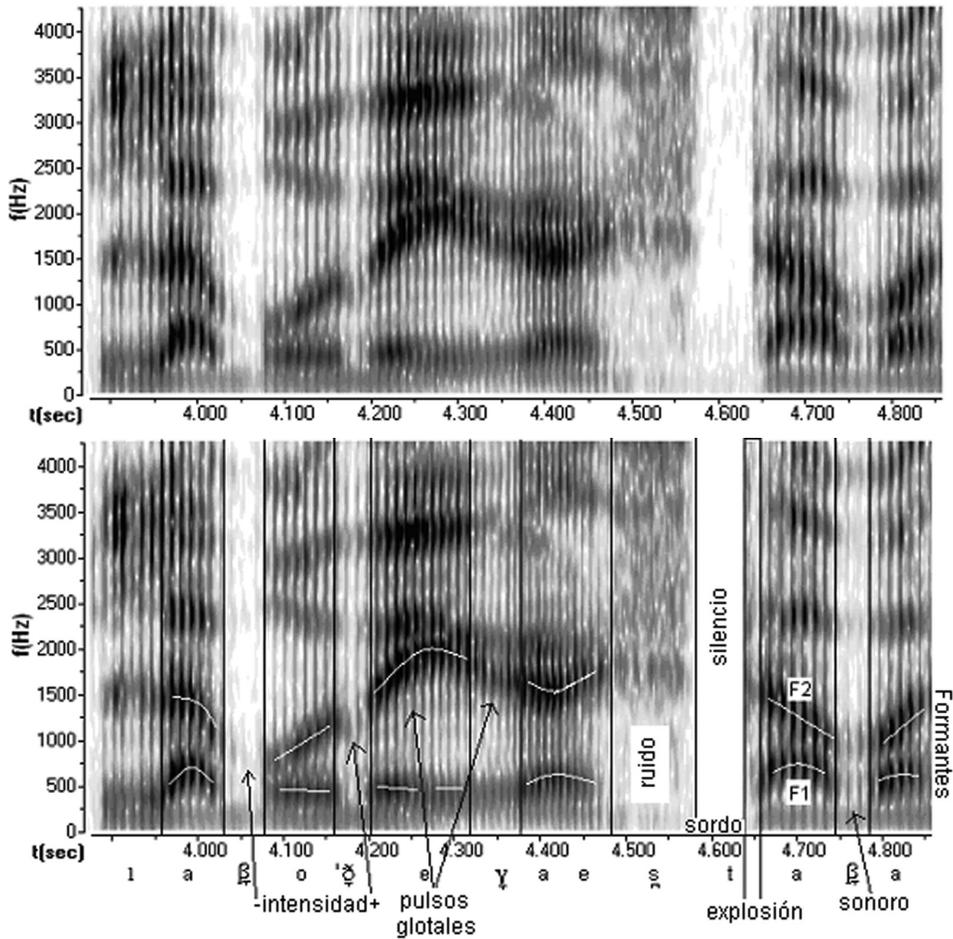


FIG. 1.8. *Espectrograma sin especificaciones arriba y con algunas especificaciones abajo.*

bién son totalmente variables. Se graba en el ordenador durante unos segundos, mediante un programa adecuado, pero el tiempo total de la memoria es muy superior; por tanto, los datos concretos del espectrograma corresponden al punto de la memoria donde se ha grabado la frase, lo cual no tiene ninguna trascendencia en el análisis. Lo que sí es relevante, obviamente, es la diferencia entre esos valores absolutos que corresponden a la memoria donde se ha grabado la frase porque indican la duración de los segmentos. Obsérvese que en el espectrograma de la figura 1.8, cada marca de la abscisa posee 50 milésimas de segundo (ms). El espectrograma comienza un poco antes de los 3,9 segundos, aunque el primer dato que vemos sea 4 segundos. Como el segundo constituye una unidad demasiado grande para describir los eventos del habla, habitualmente lo transformamos en milésimas, así pues, diremos, por ejemplo, que la [e] de *bodega dura* 115 ms (desde la marca de 4.200 hasta la marca de 4.315 en la abscisa del espectrograma).

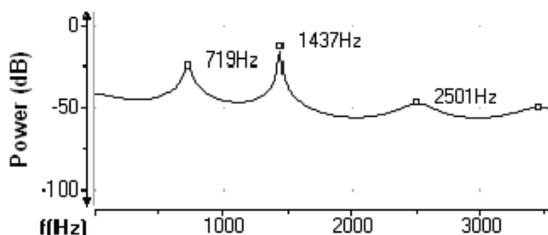


FIG. 1.9. *Espectro LPC de la vocal [a] en el artículo de la bodega.*

2.2.2. Espectros

El espectro LPC se consigue mediante una técnica de predicción lineal en la que aparecen destacados los picos de máxima intensidad que están formados por un conjunto de armónicos enfatizados por un resonador bucal. Por este motivo, podemos estar seguros de que cada pico representa un formante en caso de los sonidos armónicos o una zona frecuencial amplificada en caso de sonidos no armónicos.

En la figura 1.9 aparece el espectro de una vocal [a]. El mismo analizador proporciona los valores frecuenciales de cada formante. En la ordenada, este analizador representa la intensidad referida a un valor 0, que representa el máximo valor admitido sin distorsión por el instrumento. Por tanto, los valores son en general negativos y cuanto más próximos estén del cero mayor intensidad representan. Por ejemplo, el F1 mide -20 dB y tiene menor intensidad que el F2 que posee -12 dB; es decir, posee 8 dB más el F2 que el F1. En la abscisa, se representan las frecuencias, concretamente hay 3.500 Hz aproximadamente y, en este caso, el analizador sitúa el F1 en 719 Hz y el F2 en 1.437 Hz. El analizador puede proporcionar, según la voluntad del investigador, los tres parámetros de cada formante: frecuencia central, intensidad y ancho de banda.

En la figura 1.10, presentamos dos espectros superpuestos para que se puedan comparar dos eventos fonéticos. En este caso concreto, la realización de los formantes vocálicos correspondientes a una [a], la misma del espectro anterior,

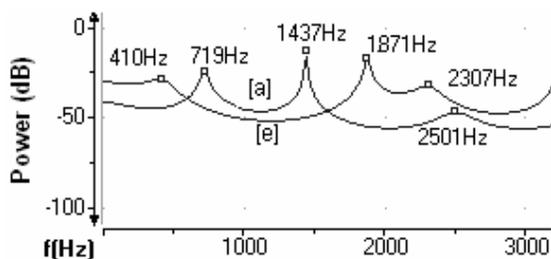


FIG. 1.10. *Superposición de los espectros de [a] y de [e].*

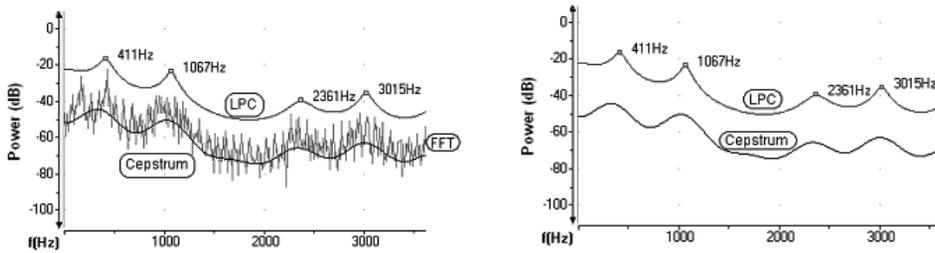


FIG. 1.11. Espectros LPC y cepstrum de la vocal [o]: con FFT (a la izquierda) y sin ella (a la derecha).

y una [e]. Pertenecen a la misma persona y se trata de la primera [a] y la [e] del sintagma *la bodega*. Estos gráficos se consiguen con un programa de tratamiento de gráficos, pues el analizador utilizado no permite superponer espectros. No obstante, los hemos utilizado en múltiples ocasiones en el libro para mostrar las diferencias entre diversas variantes del mismo sonido. La imagen de la figura 1.10 nos proporciona una visión magnífica de la situación respectiva de los dos primeros formantes de dichas vocales: [a] posee estos formantes situados entre los formantes de [e].

En el estudio de las nasales hemos utilizado en más de una ocasión el gráfico de la derecha de la figura 1.11, en el que además de la curva LPC aparece otra denominada *cepstrum*. En la imagen de la izquierda se incluye también la serie de armónicos que constituyen la vocal, obtenidos mediante la técnica denominada FFT (*Fast Fourier Transform*). De hecho, las dos curvas se ajustan a los bordes de dichos armónicos; por la parte superior la LPC y, por la inferior, la cepstrum. En las nasales, esta última curva muestra perfectamente los denominados antifor-mantes constituidos por caídas considerables de la intensidad de los armónicos, como veremos más adelante.

2.2.3. La curva de intensidad

Otro de los gráficos que en algún momento utilizamos muestra la curva de intensidad global o energía de un sonido determinado; esa intensidad es, como decimos, global, mientras que en cualquier espectro se muestra la intensidad relativa de cada formante o evento fonético que se represente. En la figura 1.12 mostramos un oscilograma de una porción de la vocal [a] y, debajo, la curva de intensidad correspondiente. Por supuesto, la curva no es plana, lo cual indica que hay unos puntos más intensos que otros a lo largo de la duración de la vocal.

En la parte inferior de la figura 1.12, se observa que este analizador acústico mide la intensidad desde el valor 0, presentando valores negativos, como sucede en los espectros. En la ordenada aparecen los valores de intensidad en decibelios (dB) y en la abscisa, los valores temporales relativos en segundos.

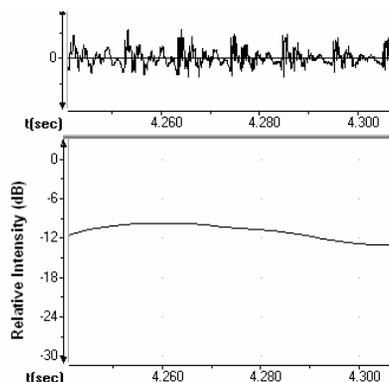


FIG. 1.12. *Oscilograma y curva de intensidad de la vocal [a].*

2.2.4. *Los pulsos glotales*

En alguna ocasión ha sido muy útil presentar algún espectrograma y su oscilograma con las líneas que marcan los pulsos glotales para mostrar de forma más evidente la existencia de sonoridad (figura 1.13).

Sobre el oscilograma de la figura 1.13 aparecen unas líneas verticales que marcan cada onda de los sonidos armónicos. Normalmente se las denomina pulsos glotales porque señalan los movimientos cuasi periódicos de las cuerdas vocales. Coinciden totalmente con las estrías que se ven en el espectrograma: la parte negra de la estría responde al momento en que las cuerdas vocales están abiertas y pasa

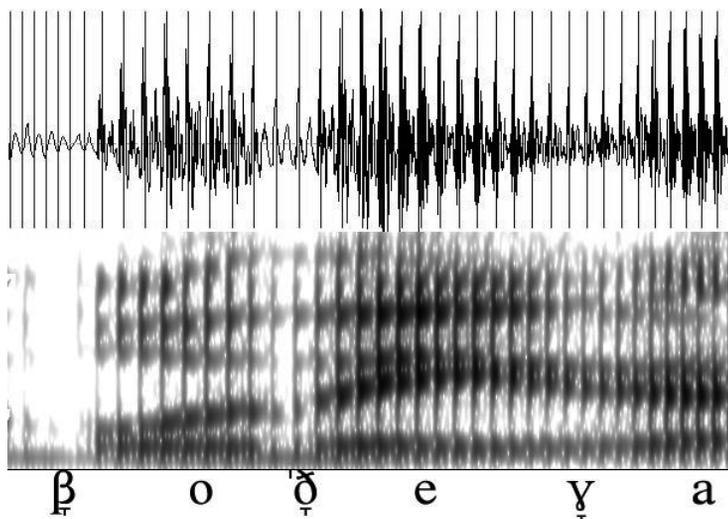


FIG. 1.13. *Los pulsos glotales sobre un oscilograma y su correspondencia en un espectrograma.*

por la laringe la cantidad máxima de energía; por el contrario, la parte blanca se produce por el cierre de las cuerdas vocales que impide momentáneamente la salida de energía. Cada estría constituye un ciclo de la onda sonora compuesta, producido por la abertura y el cierre de las cuerdas vocales. A este fenómeno se le denomina pulso glotal, y un analizador acústico del habla detecta y coloca de forma automática las líneas verticales que corresponden a cada inicio del ciclo. De ahí que sea una de las mejores formas de detectar automáticamente la presencia de sonoridad. En el espectrograma de la figura 1.13 todos los sonidos son sonoros, por esta razón todos ellos poseen las líneas verticales que indican la sonoridad, incluido el sonido inicial.

2.2.5. Las curvas de F_0

En el capítulo de prosodia, el gráfico más utilizado es el que muestra las oscilaciones del F_0 o tono fundamental. Además, se suele incorporar el oscilograma en la parte superior con las divisiones por sílabas de la frase que se estudia y la escritura ortográfica de dichas sílabas situadas a la izquierda de la línea de separación. Esto permite alinear visualmente cada movimiento tonal con la sílaba que lo produce, mostrando así todo el detalle de la trayectoria de dicho movimiento.

La figura 1.14, por ejemplo, muestra perfectamente que en la pronunciación de la frase *la bodega estaba cerrada* hay dos picos destacados que corresponden al primer acento y al último de la frase, más otro menos prominente que pertenece al verbo. Los picos aparecen alrededor de la sílaba tónica, aunque en los dos primeros el pico se desplaza a la posición postónica, mientras que el último se eleva en la misma sílaba tónica.

En la ordenada aparecen las frecuencias, en este caso muy graves por tratarse

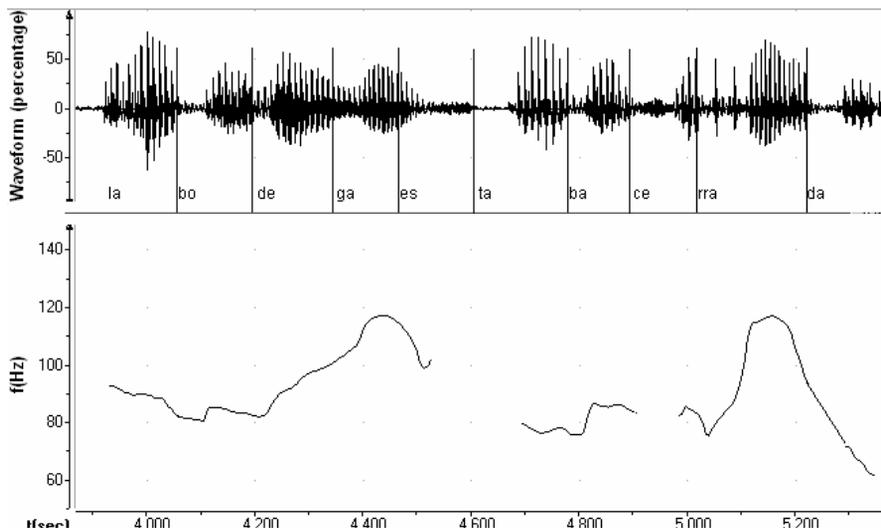


FIG. 1.14. Oscilograma y curva de F_0 .

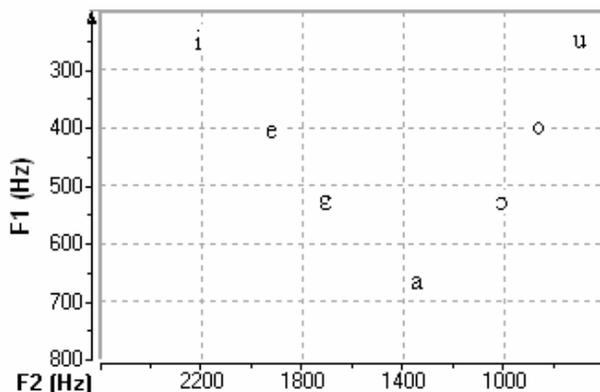


FIG. 1.15. Carta de formantes con espacios equidistantes entre los valores de referencia.

de voz masculina: la ordenada va desde 50 Hz hasta 150 Hz, aunque la voz del hablante discurre entre los 60 y los 120 Hz: un intervalo de una octava. En la abscisa se representa el parámetro temporal como en muchos de estos gráficos.

2.3. INTERPRETACIÓN DE LOS GRÁFICOS QUE CONTIENEN INFORMACIÓN PSICOACÚSTICA

Sólo hemos utilizado gráficos de este tipo para representar las cartas de formantes de las vocales. Se trata de gráficos que sirven para representar espacialmente los valores de los primeros formantes vocálicos. Se coloca en la ordenada el F1 ordenado de arriba abajo y en la abscisa el F2 ordenado de derecha a izquierda.

Si se representan esos valores de forma lineal, obtenemos una carta de tipo acústico simplemente, como se aprecia en la figura 1.15.

En la figura 1.15 se observa que las vocales anteriores se extienden en un espacio más amplio que las posteriores, puesto que los valores de F2 mantienen mayores distancias entre ellas.

Sin embargo, existen gráficos, como el de la figura 1.16, en los que las distancias entre los valores de referencia de los ejes, tanto en la ordenada como en la abscisa, no son iguales, siguen una progresión logarítmica que afecta a los dos formantes, pero sobre todo destaca en el segundo: la distancia entre 500 y 1.000 es mucho mayor que la existente entre 2.000 y 2.500. Esto se aprecia igual, pero en menor medida, en el F1: la distancia entre 200 y 300 es mayor que la existente entre 600 y 700. Éste es el tipo de gráfico que el lector encontrará en el apartado de las vocales en este libro y se puede observar que ahora las distancias entre las vocales anteriores y posteriores son más homogéneas.³ Esa progresión logarítmica

3. Estas cartas se han obtenido mediante un programa inspirado por Ladefoged y utilizado en el Laboratorio de Fonética de UCLA (Ladefoged, 2003).

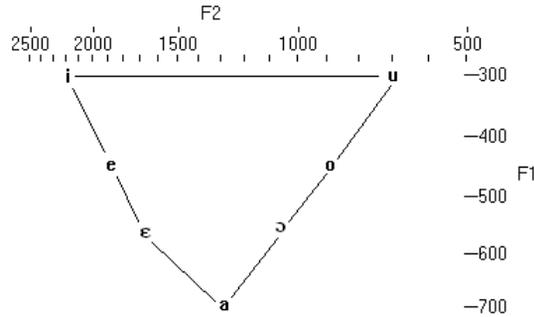


FIG. 1.16. Carta de formantes con ejes basados en la unidad bark.

está basada en la unidad bark. El bark (en honor al físico alemán Georg Heinrich Barkhausen) es una unidad de frecuencia perceptiva. La escala bark relaciona la frecuencia absoluta (en Hz) con las frecuencias medidas perceptivamente. Usando el bark, un sonido en el dominio de la frecuencia puede ser convertido a sonido en el dominio psicoacústico. De modo que esto es lo que pretenden representar las cartas de formantes como la de la figura 1.16: representar el espacio perceptivo en el que se distribuyen las vocales.

Fórmula para convertir $f(\text{Hz})$ en barks (Hayward, 2000: 142):

$$z = (26,81 \times f) / (1960 + f) - 0,53$$

La tabla 1.1 muestra la equivalencia entre frecuencia y barks. La diferencia entre los distintos valores de barks proporciona las distancias entre ellos. Si transformáramos las diferencias en milímetros, tendríamos, por ejemplo, que, para la

TABLA 1.1. Equivalencia entre frecuencia y barks

F2			F1		
$f(\text{Hz})$	bk	distancia	$f(\text{Hz})$	bk	distancia
1	250 = 2,50	0	1	100 = 0,77	0
2	500 = 4,91	2,41	2	200 = 1,97	1,18
3	750 = 6,88	1,97	3	300 = 3,02	1,07
4	1.000 = 8,52	1,63	4	400 = 4,01	0,98
5	1.250 = 9,91	1,38	5	500 = 4,91	0,90
6	1.500 = 11,09	1,18	6	600 = 5,75	0,83
7	1.750 = 12,11	1,02	7	700 = 6,52	0,77
8	2.000 = 13,01	0,89	8	800 = 7,24	0,71
9	2.250 = 13,79	0,78	9	900 = 7,90	0,66
10	2.500 = 14,49	0,69			
11	2.750 = 15,12	0,62			
12	3.000 = 15,68	0,56			

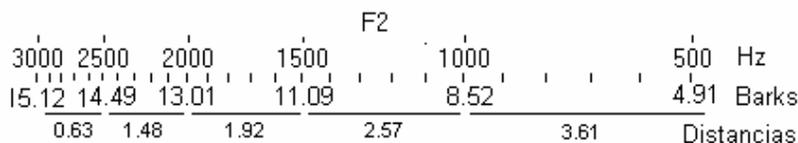


FIG. 1.17. *Escala de frecuencias comparada con sus equivalencias en barks.*

representación de F2, la marca de 500 Hz habría que colocarla a 21,1 milímetros de los 250 Hz; la de 750 Hz a 19,7 mm; la de 1.000 Hz a 16,3 mm, la de 1.250 a 13,8 mm, etc. En F1, los 300 Hz los colocaríamos a 10,7 mm de los 200 Hz; los 400 Hz, a 9,8 mm; los 500 Hz, a 8,3 mm, etc. Así se puede construir una escala relacionando barks y frecuencias, como la presentada en la carta de formantes que reproducimos en la figura 1.17 mediante la escala del F2. Esta figura muestra los Hz, la equivalencia en barks y las distancias existentes entre los distintos tramos de barks que, como se ve, no son lineales. De hecho, las distancias a las que se colocan los Hz de 500 en 500 responden a esas diferencias entre los barks.

La figura 1.18 muestra cómo se perciben los dos primeros formantes de una schwa situados a 500 y 1.500 Hz, lo cual equivale en la escala bark aproximadamente a 5 y 11 respectivamente, calculándolos con la fórmula anterior. En esa figura no están en relación la escala bark y la escala de frecuencia. La figura muestra simplemente la relevancia que tienen para el oído las frecuencias más graves, por eso mismo los dos primeros formantes se producen en las frecuencias más graves. El F2 no suele superar en las voces normales los 3.000 Hz.

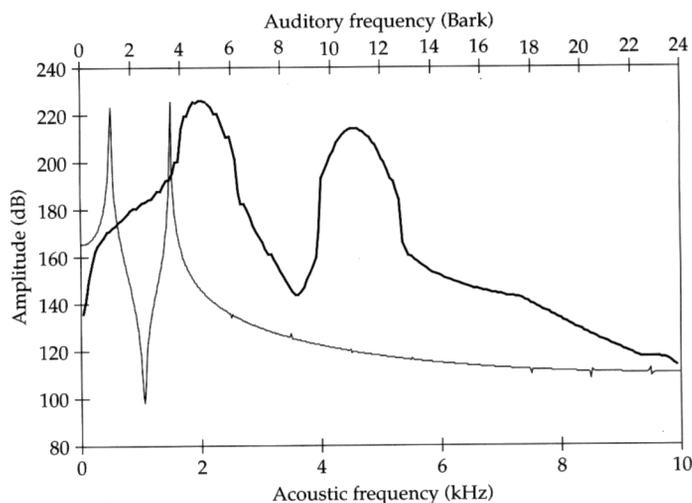


FIG. 1.18. *Formantes acústicos vs perceptivos (apud Johnson, 2003: 53).*