

**FFT COMO HERRAMIENTA DE ANÁLISIS EN  
FONÉTICA**

**JOSÉ MARTÍ ROCA**

**Jefe del Departamento de Acústica de la Escuela de Telecomunicación**

**LA SALLE BONANOVA**

**Barcelona - mayo de 1987.**

## 1.INTRODUCCION

La aparición de los métodos numéricos en el tratamiento de señales ha permitido un gran avance de las posibilidades de análisis y síntesis del sonido y particularmente de la voz humana. El punto de partida estriba en la posibilidad de adquirir valores numéricos de la onda acústica a intervalos discretos, pero suficientemente próximos en el tiempo, que nos permitan analizar y recomponer posteriormente esta señal. Una vez adquiridas estas muestras pueden ser almacenadas en una memoria digital y procesadas debidamente para obtener toda la información que nos transmiten. La cantidad de cualquier magnitud (la presión acústica, por ejemplo) a lo largo del tiempo queda perfectamente identificada por sus componentes frecuenciales. Es decir cualquier fenómeno cuantificado por un determinado parámetro puede interpretarse como una superposición de periodicidades de distintas frecuencias, (incluso de infinitas frecuencias) que están como encubiertas en la complejidad de un devenir temporal. Este es el gran descubrimiento del científico Fourier que ha dado nombre a la herramienta matemática denominada Transformada de Fourier.

En términos matemáticos esta transformada es un operador (F) que aplicado a una función temporal  $g(t)$  la convierte en otra función de la frecuencia  $G(f)$  que nos aporta la misma información que la primera:

$$F [g(t)] = G (f)$$

Esta función  $G(f)$  es la que nos informa de todas las periodicidades de la función  $g(t)$ .

Recíprocamente, existe también el operador antitransformada ( $F^{-1}$ ) que aplicado a una función frecuencial  $G(f)$  permite recomponer su función temporal original  $g(t)$ :

$$F^{-1} [G(f)] = g(t)$$

Podemos interpretar esta transformada como un camino para pasar del dominio del tiempo al dominio de la frecuencia, como si se tratase de dos mundos distintos, pero capaces de aportarnos la misma información. Vistos desde el dominio frecuencial, ciertos fenómenos se explican más fácilmente y tienen un tratamiento matemático más ágil. El camino de regreso al dominio temporal siempre está garantizado por el operador antitransformada (Fig. 1).

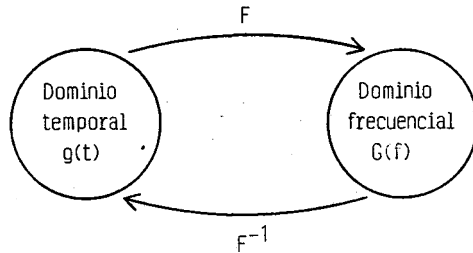


FIG. 1 TRANSFORMADA Y ANTITRANSFORMADA DE FOURIER

El proceso de cálculo para realizar este viaje entre ambos mundos es complejo, pero factible por métodos numéricos con la ayuda de un ordenador. El descubrimiento, relativamente reciente (1965) del algoritmo llamado Transformada Rápida de Fourier (conocido por las iniciales FFT de "Fast Fourier Transform") ha facilitado enormemente la rapidez de este cálculo que hoy en día se realiza perfectamente en tiempo real; de tal forma que podemos visualizar sobre la pantalla la forma de la función frecuencial a medida que se produce la evolución del correspondiente parámetro a estudiar. Evidentemente que la simultaneidad en sentido estricto nunca será posible, ya que la Transformada de Fourier se realiza sobre un determinado número de muestras temporales que siempre serán anteriores al proceso de cálculo sobre las mismas.

EL logaritmo FFT impone la restricción de que el número de estas muestras sea exactamente una potencia de 2: 2, 4, 8, 16, 32, 64, 128, ... La función obtenida vendrá dada también en forma discreta; exactamente con

una cantidad de muestras igual a la mitad de las que tenía la función temporal.

Ofrecemos a continuación una información sobre los equipos de laboratorio que realizan este proceso. Como ejemplo concreto nos referimos particularmente al analizador FFT Brüel & Kjaer tipo 2033. Todos los gráficos con registros espectrales presentados se han obtenido mediante este equipo en los Laboratorios de Acústica de la Escuela Universitaria de Telecomunicaciones La Salle de Bonanova.

## 2. LA ADQUISICION DE DATOS.

### 2.1.El registro temporal.

Generalmente todos los analizadores digitales FFT permiten una visualización de la señal temporal adquirida con una cuantificación precisa del tiempo y de los valores instantáneos. Normalmente se dispone de un cursor para la exploración de la señal, que puede ser visualizada en diferentes escalas temporales o de nivel. Véase por ejemplo la forma de onda de la expresión "el papel" (Fig. 2) y una amplificación temporal de la "a" comprendida entre las dos "p" de "papel" (Fig. 3).

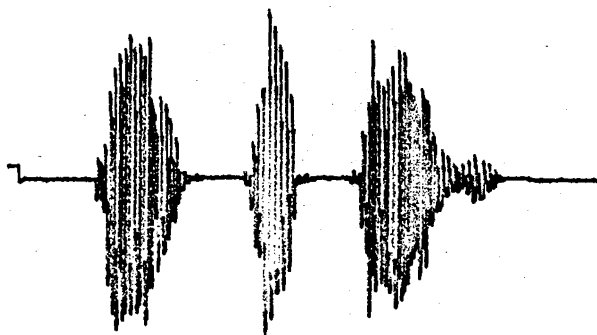


Fig. 2 SEÑAL TEMPORAL CORRESPONDIENTE A "EL PAPEL"

(TIEMPO TOTAL = 800 ms)

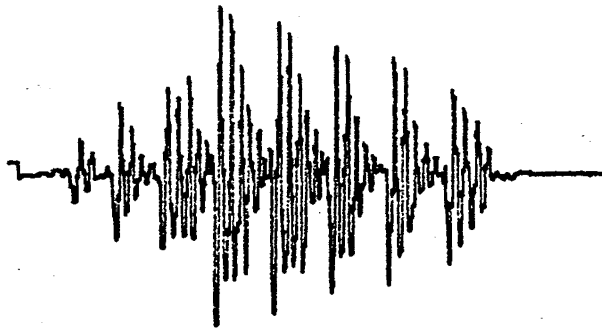


Fig. 3 SEÑAL TEMPORAL AMPLIADA DE LA "A" DE "PAPEL"  
(TIEMPO TOTAL = 80 ms)

Una señal temporal puede ser adquirida en fragmentos y transformada inmediatamente para un análisis en tiempo real. También puede ser adquirida y guardada en memoria para un análisis posterior (análisis de "scan"), al ritmo y en los intervalos que se crean convenientes. Este segundo modo es muy interesante para el estudio detenido de los fenómenos de transición o de coarticulación entre dos sonidos contiguos. Por ejemplo la localización de la explosión de una oclusiva ("burst") y el estudio de su espectro. (Fig. 4 y 5). El analizador Brüel 2033 permite un almacenamiento máximo de 10240 muestras de la señal temporal, lo que corresponde a un tiempo variable según el fondo de escala de frecuencia seleccionado. El tiempo de registro correspondiente a cada fondo de escala puede verse en la Tabla 1.

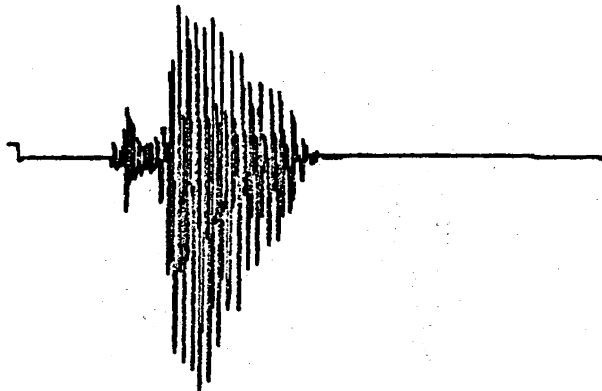


Fig. 4 LOCALIZACION DEL "BURST" CORRESPONDIENTE A [K A]  
(TIEMPO TOTAL = 80 ms)

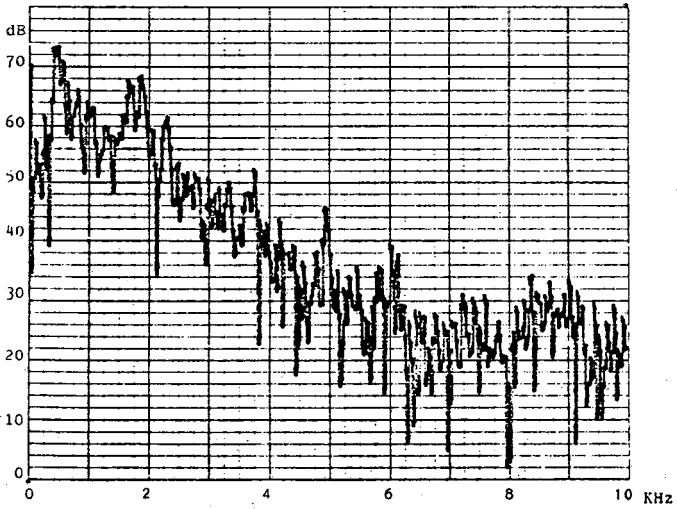


Fig. 5 ESPECTRO DEL RUIDO CORRESPONDIENTE A LA EXPLOSIÓN DE [k].

Fondo de escala en Hz	Tiempo de registro en seg.
10	400
20	200
50	80
100	40
200	20
500	8
1000	4
2000	2
5000	0.8
10000	0.4
20000	0.2

Tabla 1 Relación entre fondo de escala y tiempo de registro

## 2.2. Enventanado

La selección de los fragmentos temporales de señal acústica se realiza mediante un proceso de "enventanado". Estos fragmentos pueden ser analizados tal como han sido adquiridas a lo largo del tiempo (ventana "flat") o mediante una ponderación de las mismas que reduzca la discontinuidad del inicio y del final de la ventana. La elección de este segundo tipo de ventanas ofrece unos resultados más correctos cuando la señal presenta una continuidad temporal. En este caso el hecho de cortar bruscamente el inicio y el final del fragmento seleccionado equivale a considerar otra forma de onda distinta de la que realmente se pretende analizar. EL proceso de ponderación consiste simplemente en multiplicar todas las muestras por una función de valores comprendidos entre 0 (o un valor próximo a 0) en los extremos de la ventana y 1 en el centro de la ventana.

La ponderación utilizada por el 2033 de Brüel & Kjaer se puede escoger entre una ventana "flat" y una ventana "hanning". Esta última corresponde a un factor igual a  $\cos^2 x$  para valores de  $x$  comprendidos entre  $-\pi/2$  y  $+\pi/2$ . Las ventanas se pueden escoger de tal modo que los fragmentos de voz pueden que queden yuxtapuestos o solapados. En el segundo caso se obtiene una mayor resolución temporal en la visión evolutiva del espectro.

La selección de un determinado fragmento de voz puede crear problemas para su localización. Generalmente los analizadores espectrales disponen de un sistema de disparo ("trigger") para captar las muestras temporales a partir del instante que se alcanza un cierto nivel (positivo o negativo) seleccionable manualmente. Hay equipos que presentan facilidades para retener muestras a partir de un tiempo anterior ("pretrigger") o posterior ("after trigger") al disparo. Concretamente el 2033 de Brüel & Kjaer permite una grabación de muestras "after trigger" seleccionable entre 0 y 64000 temporales a intervalos de 100 muestras.

## 2.3. Frecuencia de muestreo.

El sistema de análisis de señales muestreadas temporalmente supone una deformación considerable de la señal, que, de por sí, suele presentar una perfecta continuidad. El problema se va paliando a medida que el muestreo se pueda realizar a intervalos más pequeños. Ahora bien, como que el algoritmo FFT trabaja con un número determinado de muestras (concretamente 1024 para el modelo que comentamos), la comprensión de las mismas en un tiempo relativamente corto no nos permite detectar fenómenos que manifiestan su periodicidad en un tiempo más largo (como, por ejemplo, la periodicidad fundamental de las cuerdas vocales).

El análisis de una señal muestreada a una frecuencia  $f_s$  será correcto y coincidirá con el de la verdadera señal no muestreada, siempre y cuando se eliminen todas las componentes espectrales con frecuencias superiores a  $f_s/2$ . Dicho a la inversa: para analizar correctamente señales muestreadas con componentes espectrales no superiores a una determinada frecuencia ( $f_n$ ) necesitamos utilizar una frecuencia de muestreo que como mínimo sea superior a  $2f_n$ . La solución que se suele adoptar consiste en eliminar previamente los componentes espectrales superiores a la frecuencia del fondo de escala que se quiere analizar, y muestrear posteriormente la señal a un ritmo igual o superior al doble de este fondo de escala. El proceso de filtrado para eliminar la alta frecuencia ha de preceder siempre al proceso de muestreo. Estos filtros se denominan "antialiasing" y suelen estar incorporados ya en el equipo de análisis. El 2033 de Brüel, una vez conectado, selecciona automáticamente el filtro correspondiente, de acuerdo con el fondo de escala de frecuencia elegido; pero permite también su desactivación mediante el pulsador "Filter off" si en algún caso se cree conveniente.

### 3. PROCESO DE ANALISIS.

#### 3.1. Dinámica espectral.

La diferencia entre los niveles espectrales máximos y mínimos detectables depende de la dinámica del convertor analógico/digital que realiza la adquisición numérica de datos. EL analizador Brüel 2033 discretiza las señales de entrada en un margen de 12 bits, lo que supone una relación de  $2^{12} = 4096$  niveles distintos entre el valor máximo y el valor mínimo detectables. Logarítmicamente esta relación equivale a una dinámica de  $20 \log_{10} 4096 = 72.2$  dB.

La visualización en pantalla del espectro puede hacerse de forma lineal, lo cual no permite percibir toda esta dinámica, sino únicamente los valores más altos. Los valores pequeños se confunden con el-cero de la escala. Si se opta por una escala logarítmica se percibe mejor toda esta dinámica, lo cual está más de acuerdo con nuestra percepción auditiva. La escala puede ser de 80 o 40 dB según convenga. Véanse las figuras 6 y 7 que presentan el mismo espectro a escala lineal y a escala logarítmica respectivamente.



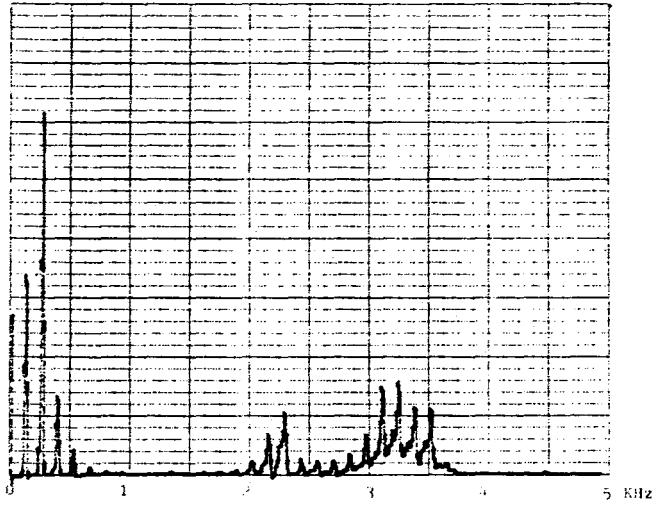


Fig. 6. ESPECTRO A ESCALA LINEAL DE UNA [i]

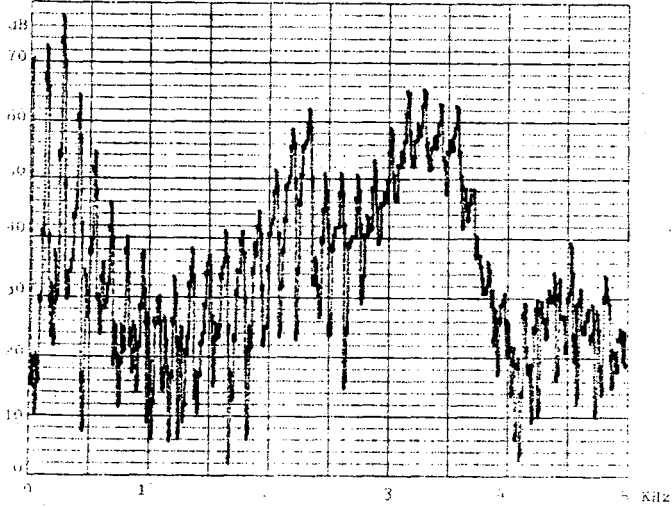


Fig. 7 ESPECTRO A ESCALA LOGARÍTMICA DE LA MISMA [i] DE LA FIGURA ANTERIOR

Los valores espectrales se pueden leer mediante la ayuda de un cursor deslizable por la pantalla con una resolución en el nivel de 0.1 dB.

### 3.2 Resolución espectral.

La resolución entre frecuencias próximas depende siempre del fondo de escala seleccionado. Naturalmente esta resolución es tanto más fina cuanto inferior es la banda de frecuencias analizada. Si se opta por una frecuencia de muestreo  $f_s$  igual al doble de esta banda, la resolución espectral o separación entre dos valores discretos del espectro vendrá dada por  $f_s/2$ .

El equipo aquí comentado trabaja con paquetes temporales de 1024 muestras, por tanto ofrece espectros de 512 líneas, con todo, las líneas mostradas por pantalla son únicamente las 400 primeras. Por tanto, su resolución espectral se puede determinar simplemente por el cociente:

$$\text{Frecuencia del fondo de escala} = \frac{\text{Resolución espectral}}{400}$$

Así, por ejemplo en los espectros de las figuras 6 y 7 la resolución es de 12.5 Hz y en la figura 5, de 25 Hz. Nuestro equipo lleva incorporada una posibilidad de "zoom" que permite una resolución 10 veces mayor para el mismo fondo de escala y en torno a una frecuencia seleccionada. Es como si obtuviéramos una ampliación por 10 de una zona del espectro (una décima parte de la pantalla pasa a ocupar toda la pantalla). Para ello realiza un análisis con todas las 10240 muestras temporales que puede almacenar en su memoria, lo cual limita su utilización a casos de señales que mantienen constante su espectro durante este tiempo. En el caso de la voz solamente sería aplicable a fragmentos muy estacionarios.

### 3.3 Sonidos periódicos y sonidos aperiódicos.

La diferenciación entre los sonidos que presentan una periodicidad temporal y los que simplemente presentan variaciones aleatorias se consigue fácilmente por la forma del espectro. Los primeros presentan barras espectrales equidistantes (armónicos) que pueden cubrir una zona más o menos amplia del espectro, pero que siempre aparecen más claras en la parte

baja del mismo (Fig.8). Los sonidos (el ruido) se caracteriza por un espectro más continuo que puede limitarse a algunas bandas dominantes (Fig.9).

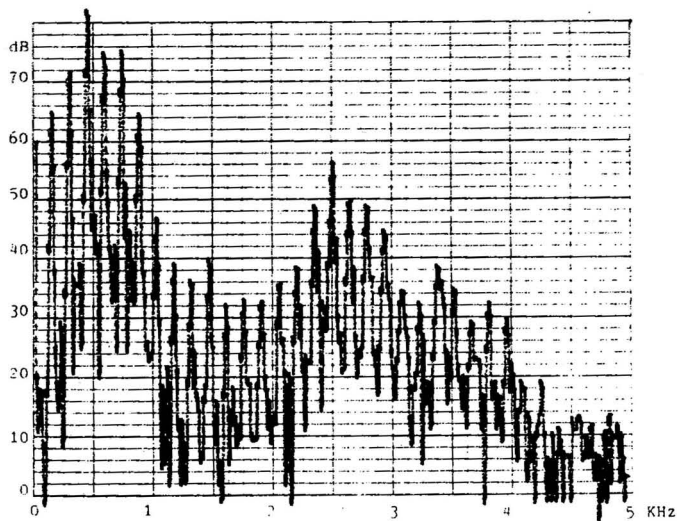


Fig. 8 ESPECTRO DE UNA [o]

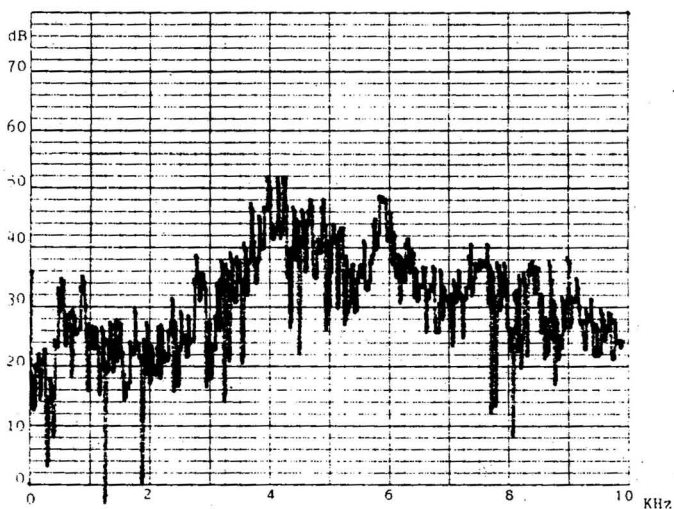


Fig. 9 ESPECTRO DE UNA [s]

Con todo, dentro de un espectro pueden darse a la vez zonas con armónicos y zonas con bandas de ruido. A veces puede surgir la duda sobre si el ruido que aparece en el análisis tiene un origen acústico o es efecto de un enmascaramiento posterior de la señal. Cuando el ruido no presenta características estacionarias bien definidas puede reducirse por un efecto de cancelación promediando muchos espectros sucesivos. Este efecto da lugar a una presentación más nítida de los espectros.

El analizador que comentamos aquí tiene esta posibilidad de promediar hasta 2048 espectros instantáneos, o bien hasta 1153 espectros seleccionados a lo largo de una grabación temporal o incluso de sucesivas grabaciones. Véase, por ejemplo, la diferencia entre un espectro instantáneo de la fricativa sonora [z] (fig.10) y de un promediado de 200 espectros de la misma consonante (Fig.11). Obsérvese como en este último caso quedan más definidos tanto los armónicos que evidencian la parte periódica, como las bandas de ruido situadas en la parte más alta del espectro.

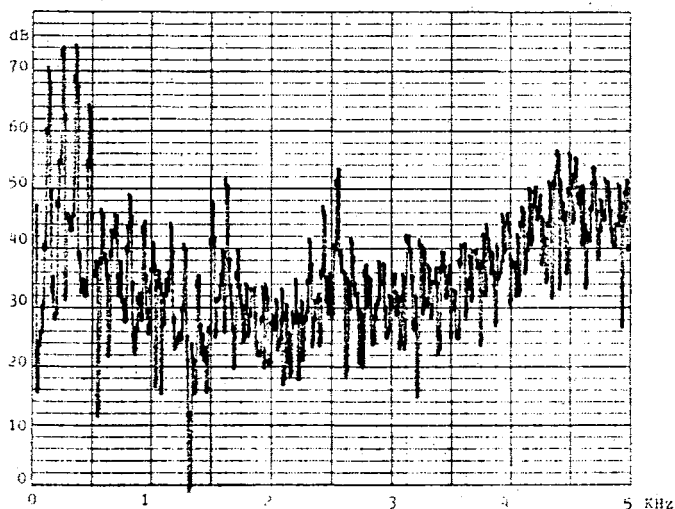


Fig. 10 ESPECTRO INSTANTANEO DE [z]

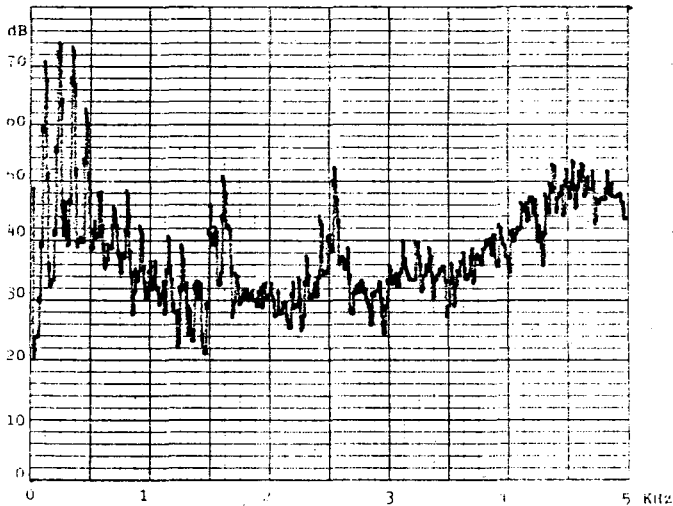


Fig. 11 PROMEDIADO DE 200 ESPECTROS CORRESPONDENCIA [z]

### 3.4. Almacenamiento en memoria.

Los resultados numéricos de la FFT pueden ser guardados fácilmente en memoria para una posterior visualización en pantalla. Concretamente el 2033 puede almacenar un espectro (pulsador "memory") y compararlo con otro espectro posterior mediante una presentación sucesiva de ambos por pantalla (pulsadores "fast" o "slow"). Se puede obtener y visualizar también el cociente de los mismos que corresponde a una resta de sus respectivos valores logarítmicos pulsando la tecla I/M. Las mismas funciones pueden realizarse con los resultados pro-mediados de diferentes espectros.

Cualquier grabación en memoria borra los resultados guardados anteriormente sobre la misma. Si se quiere evitar esta eventualidad existe la posibilidad de proteger el último resultado (pulsador "protect") y evitar posteriores grabaciones.

#### **4.CONEXION A UN ORDENADOR.**

Normalmente los analizadores espectrales digitales tienen prevista la conexión a un ordenador con un doble finalidad:

- a) Posibilidad de un control remoto desde el ordenador.
- b) Posibilidad de trasvase mutuo de datos entre el ordenador y el analizador.

Con esta conexión las posibilidades de utilización de un equipo de análisis espectral se multiplican extraordinariamente, sin otro límite que la imaginación del propio usuario; el cual puede definir cualquier proceso de investigación controlado por un programa informático.

El control por ordenador del FFT de Brüel & Kjaer se realiza a través del "bus" de comunicaciones IEEE Std. 488 que está incorporado ya en el equipo. Para ello se requiere que el ordenador esté dotado con el mismo tipo de conexión externa y la confección de programas adecuados para el proceso de comunicación.

#### **5.APLICACIONES EN EL CAMPO DE LA FONETICA.**

##### **5.1.Estudio detallado de la función temporal.**

Gracias a la precisión de los sistemas digitales, la evolución de la presión sonora a lo largo del tiempo queda perfectamente registrada con una cuantificación precisa. Nuestro analizador permite analizar en pantalla el registro completo de las 10240 muestras. Mediante un cursor se puede seleccionar cualquier punto para una visualización más amplia de una zona del registro. Se puede visualizar un intervalo de 1024 o 390 muestras respectivamente en torno a la posición del cursor, de tal forma que ocupen toda la pantalla. Con cualquiera de estas dos últimas opciones se puede recorrer toda la señal grabada que irá circulando por la pantalla hasta el punto deseado. El tiempo correspondiente a la posición del cursor, en milisegundos, y el número de la muestra correspondiente figuran siempre en la parte superior de la pantalla.

La forma de onda temporal, tal como aparece en pantalla, puede ser volcada siempre a un registrador gráfico X-Y para dejar constancia de la

misma. Lo mismo se puede hacer con cualquier espectro visualizado por pantalla.

### **5.2. Precisión de la frecuencia fundamental.**

Los fragmentos periódicos de la voz presentan un espectro de barras o armónicos claramente marcados. La frecuencia correspondiente al primer armónico es la que determina el tono de voz o frecuencia fundamental ( $F_0$ ). Como que todos los armónicos son múltiplos del primero, se puede obtener una gran precisión en la determinación del tono de voz leyendo la frecuencia correspondiente a un armónico superior y dividiendo por el número de orden del mismo. Así, por ejemplo, se puede leer la frecuencia correspondiente al décimo armónico y dividir por 10; con lo cual el margen de imprecisión de la medida se divide también por 10.

Igualmente se puede determinar con gran precisión la curva de entonación siguiendo la evolución de los armónicos a lo largo de una evolución. En este caso es aconsejable reducir el fondo de escala, ya que los armónicos de más baja frecuencia son siempre los más destacados. Con ello se tiene la ventaja de un mayor tiempo de registro de acuerdo con la Tabla 1.

### **5.3. Localización de formantes.**

Los formantes correspondientes a los fragmentos periódicos de la voz se caracterizan por la elevación de los armónicos de una determinada zona del espectro. Dado el carácter discreto de estos armónicos, muchas veces la frecuencia central del formante cae entre dos armónicos consecutivos. Para su determinación precisa hay que representar la línea envolvente y detectar su punto máximo. En mayor parte de las mediciones este proceso se puede hacer por una estimación visual, sin más. Para una mayor precisión siempre se puede realizar un proceso de "suavizado" del espectro mediante algún algoritmo sencillo de cálculo.

Las resonancias correspondientes a fragmentos fricativos de la voz quedan también marcados dentro de un espectro más o menos continuo, pero altamente fluctuante entre emisiones de voz distintas. De todas formas, las zonas dominantes del espectro, características de cada uno de los sonidos fricativos, pueden localizarse con más facilidad y de una forma más repetitiva.

#### 5.4. Estudio de transiciones.

Las transiciones entre sonidos contiguos se caracterizan por una variación progresiva que se puede cuantificar a partir de espectros sucesivos tomados sobre una señal temporal previamente registrada. Particularmente interesa definir la evolución de la frecuencia central de los formantes.

El analizador Brüel 2033 permite visualizar la evolución del espectro sobre la pantalla al ritmo deseado, detenerla en cualquier momento y continuar la evolución hasta el final del registro. Para dejar una constancia de las transiciones se puede utilizar la conexión a un ordenador que almacene los sucesivos espectros y dar algún tipo de representación posterior de todos ellos. Una posibilidad es realizar una representación espectrográfica plana sobre unos ejes de frecuencia y tiempo, con diferentes grados de ennegrecimiento en cada punto según el nivel espectral del mismo. La figura 12 ofrece un espectrograma concreto obtenido por este método. Otra posibilidad es acudir a una representación tridimensional de la evolución espectral a lo largo del tiempo como la de la figura 13.

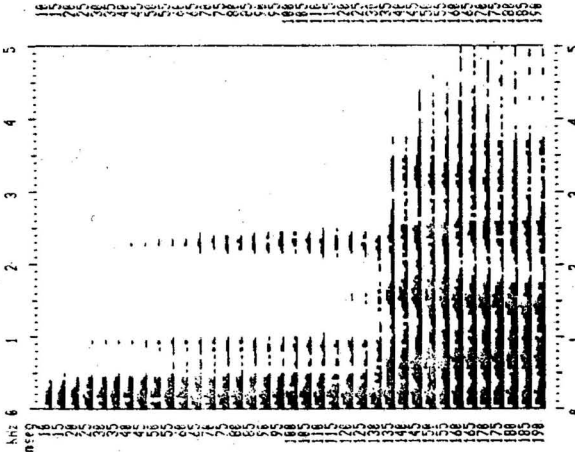


Fig. 12 ESPECTROGRAMA DE LA TRANSICION [na]



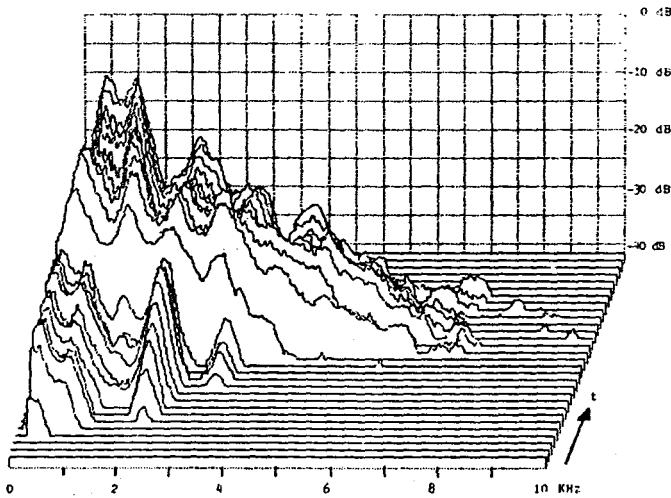


Fig.13 REPRESENTACION TRIDIMENSIONAL DE LA TRANSICION [NA]

### 5.5. Filtrado.

A partir de los valores espectrales (400 valores numéricos para el caso del 2033) se puede contabilizar la energía que corresponde a una zona determinada del espectro, lo que equivale a un filtro de una banda pasante seleccionable según las necesidades. Así, por ejemplo, se puede obtener un espectro por octavas, por bandas de octava, por bandas críticas, etc, a partir del espectro inicial.

Incluso, si interesa, estos valores pueden volver a la pantalla para una visualización del espectro según estas bandas en forma de histograma. Finalmente, como sucede con cualquier espectro, se pueden volcar los resultados sobre un registrador gráfico X-Y.

### 6. CONCLUSION.

Con esta exposición hemos pretendido ofrecer una visión rápida de las posibilidades que ofrece un analizador espectral FFT para estudios fonéticos. Nos hemos referido particularmente a la caracterización de los sonidos, en su doble vertiente temporal y frecuencial. Los diferentes equipos existentes en el mercado ofrecen numerosas dificultades para el análisis de los datos en ambos campos con una cuantificación precisa, garantizada por el tra-

tamiento digital de las ondas sonoras. Las posibilidades de cualquiera de estos se multiplican hasta el infinito a partir de la conexión externa al ordenador. Este permite avanzar tanto en el control automatizado de la toma de datos como en el procesado posterior de los mismos.

Nos hemos centrado particularmente en el analizador Brüel & Kjaer 2033 como un ejemplo concreto, aunque cada vez están apareciendo más equipos que realizan estas mismas funciones, incorporando siempre nuevas prestaciones. Una gran ventaja de este equipo, en cuanto al tratamiento de la voz, es el registro temporal con 10240 muestras que permite la exploración de un tiempo diez veces más largo al que corresponde a una sola ventana de análisis (1024 muestras). La voz es un fenómeno esencialmente fluctuante a lo largo del tiempo, y para caracterizarlo es necesario disponer de unos intervalos mínimos en los que se puedan incluir las transiciones entre sonidos contiguos. Creemos que esto es indispensable para un analizador de Fourier pueda utilizarse satisfactoriamente en estudios fonéticos.

## 7. BIBLIOGRAFIA

- BRÜEL & KJAER: "High resolution signal analyzer type 2033. Instruction manual". (1980).
- DIESTRE, N.: "Introducción a la transformada rápida de Fourier". Mundo electrónico, 80 (1978) p. 56-60
- OPPENHEIM, A. V., Editor. "Signals and systems". Prentice-Hall Internacional. (1983).
- PAPOULIS, A.: "Sistemas digitales y analógicos, transformadas de Fourier, estimación espectral". Marcombo. (1978)
- RABINER, L.R. & GOL, B.: "Theory and application of digital signal processing". Prentice-Hall Internacional. (1975).
- RANDALL, R.B.: "Application of B & K equipment to frequency analysis". Brüel & Kjaer. (1977)

