(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 371 455**

(51) Int. CI.:
***G10L 19/12*** (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **03751533 .5**
(96) Fecha de presentación: **14.10.2003**
(97) Número de publicación de la solicitud: **1554717**
(97) Fecha de publicación de la solicitud: **20.07.2005**

(54) Título: **PRE-PROCESAMIENTO DE DATOS DIGITALES DE AUDIO PARA CODECS DE AUDIO DE MÓVIL.**

(30) Prioridad:
**14.10.2002 KR 2002062507**

(45) Fecha de publicación de la mención BOPI:
**02.01.2012**

(45) Fecha de la publicación del folleto de la patente:
**02.01.2012**

(73) Titular/es:
**REALNETWORKS ASIA PACIFIC CO., LTD.**
**K1 REIT BUILDING 463 CHUNGJEONG-RO-3-GA**
**SEODAEMUN-GU SEOUL, KR**

(72) Inventor/es:
**NAM, Young Han;**
**PARK, Seop Hyeong;**
**HA, Tae Kyoon y**
**JEON, Yun Ho**

(74) Agente: **Pérez Barquín, Eliana**

## DESCRIPCIÓN

Pre-procesamiento de datos digitales de audio para codecs de audio de móvil

**Campo técnico**

La presente invención está dirigida a un método para el pre-procesamiento de datos de audio, con el fin de mejorar la calidad de la música descodificada en los terminales receptores, tales como los teléfonos móviles; y más en particular, a un método para el pre-procesamiento de datos de audio con el fin de mitigar una degradación de la señal musical, que puede originarse cuando los datos de audio se codifican/descodifican en un sistema de comunicaciones inalámbricas que utilice codificadores-decodificadores (o codecs) de habla optimizados solamente para señales de voz humana.

**Técnica anterior**

El ancho de banda del canal de un sistema de comunicaciones inalámbricas es mucho más estrecho que el de un sistema convencional de comunicaciones telefónicas de 64 kbps, y por tanto los datos de audio en un sistema de comunicaciones inalámbricas se comprimen antes de ser transmitidos. Los métodos para comprimir datos de audio en un sistema de comunicaciones inalámbricas incluyen el QCELP (Predicción Lineal Provocada por Código QualComm) de IS-95, el EVRC (Codificación de Velocidad Variable Reforzada), el VSELP (Predicción Lineal Provocada por Vector Suma) de GSM (Sistema Global para las Comunicaciones Móviles), el PRE-LTP (LPC Provocada por Impulsos Normales con Predicción a Largo Plazo), y el ACELP (Predicción Lineal Provocada con Código Algebraico). Todos estos métodos listados están basados en la LPC (Codificación Lineal Predecible). Los métodos de compresión de audio basados en la LPC utilizan un modelo optimizado para las voces humanas y por tanto son eficientes para comprimir la voz a una velocidad baja o media de codificación. En un método de codificación utilizado en un sistema inalámbrico, para utilizar eficientemente el uso del ancho de banda y disminuir el consumo de potencia, los datos de audio se comprimen y transmiten solamente cuando se detecta la voz del que habla utilizando lo que se llama función de VAD (Detección de la Actividad de Voz).

Recientemente, se han hecho disponibles varios servicios para proporcionar música a usuarios de teléfonos. Uno de los cuales es denominado “Servicio de coloración” que facilita al usuario designar una melodía de su elección de manera que las personas que hacen una llamada al abonado escuchen música en lugar de un tono de llamada tradicional, mientras que el abonado no contesta al teléfono. Como este servicio se hizo muy popular primero en Corea, donde se originó, y después en otros países, la transmisión de datos musicales a un teléfono celular ha ido aumentando. Sin embargo, como se ha explicado anteriormente, el método de compresión de audio basado en la LPC es adecuado para la voz humana que tiene componentes de frecuencia limitados. Cuando se procesa música o señales con componentes de frecuencia en la mayoría de la gama de frecuencias audibles (20 ~ 20.000 Hz) en un códec convencional basado en LPC y se transmiten a través de un teléfono celular, ocurre una distorsión de la señal, lo cual origina una pausa en la música o hace un sonido que tiene solamente una parte de los componentes de frecuencia originales.

Hay varias razones por las que se degrada la calidad del sonido de los datos de audio después de haber comprimido los datos de audio utilizando codecs basados en LPC, especialmente los codecs EVRC. La degradación de la calidad del sonido tiene lugar de la siguiente manera.

(i) Pérdida completa de los componentes de frecuencia en un ancho de banda de alta frecuencia.

(ii) Pérdida parcial de componentes de frecuencia en un ancho de banda de baja frecuencia.

(iii) Pausa intermitente de la música.

La primera causa de degradación no puede ser evitada en tanto y cuanto los componentes de alta frecuencia se eliminan utilizando un filtro de paso bajo de 4 kHz (o 3,4 kHz), cuando se comprimen datos de audio utilizando un códec de audio de ancho de banda estrecho.

El segundo fenómeno es debido a la característica intrínseca de los métodos de compresión de audio basados en la LPC. De acuerdo con los métodos de compresión basados en LPC, se obtiene un tono y una frecuencia formante en una señal de entrada y después, a partir de un libro de código, se deduce una señal de excitación para minimizar la diferencia entre la señal de entrada y la señal compuesta calculada por el tono y la frecuencia formante de la señal de entrada. Es difícil extraer un tono desde una señal musical polifónica, mientras que es fácil en el caso de la voz humana. Además, el componente formante de la música es muy diferente del de la voz de una persona. Consecuentemente, se espera que la señal de error de predicción para los datos de música fuera mucho mayor que los de la señal del habla humana, y por tanto se pierden muchos componentes de frecuencia incluidos en los datos de audio originales. Los dos problemas anteriores, es decir, la pérdida de componentes de alta y baja frecuencia, son debidos a la característica inherente al códec de audio optimizado para señales de voz, y son inevitables en cierta medida.

Las pausas en la señal de audio son originadas por la velocidad variable de la codificación utilizada por la EVRC. Un codificador de EVRC procesa los datos de audio con tres velocidades (que son de 1,1/2 y 1/8). ·Entre estas velocidades, la velocidad de 1/8 significa que el codificador EVRC determina que la señal de entrada es un ruido y no una señal de voz. Debido a que los sonidos de un instrumento de percusión, tal como un tambor, incluyen componentes del espectro que tienden a ser percibidos como ruidos por los codecs de audio, la música que incluye este tipo de sonidos hace una pausa frecuentemente. Además, los codecs de audio consideran los sonidos que tienen amplitudes bajas como ruidos, lo cual degrada también la calidad del sonido.

El documento WO 02/065457 divulga un sistema de codificación del habla con un clasificador de música. Se dispone un codificador para recibir una señal de entrada y proporciona una cadena de bits basada en la codificación del habla de una parte de la señal de entrada. El codificador proporciona una clasificación de la entrada como una de ruido, habla y música. El clasificador de música analiza o determina las propiedades de la señal de entrada. El clasificador de música compara las propiedades de la señal con umbrales, para determinar la clasificación de la señal de entrada.

El documento US 5.742.734 divulga un método y un dispositivo para determinar la velocidad de codificación del habla en un codificador de voz de velocidad variable.

**Divulgación de la invención**

La presente invención proporciona un método para el pre-procesamiento de la señal de audio a transmitir por un sistema inalámbrico, con el fin de mejorar la calidad del sonido de los datos de audio recibidos en un terminal receptor de un abonado. La presente invención proporciona un método para mitigar el deterioro de la calidad del sonido musical que tiene lugar cuando la señal musical se procesa con códigos optimizados para la voz humana, tal como los codecs EVRC. Otro objeto de la presente invención es proporcionar un método y un sistema para el pre-procesamiento de datos de audio, de una manera que no interfiera con el sistema de comunicaciones inalámbricas existente. Consecuentemente, el método de pre-procesamiento de la presente invención es útil en cuanto que puede utilizarse sin modificar un sistema existente. La presente invención puede ser aplicada también de una manera similar a otros codecs optimizados para la voz humana distintos al EVRC.

Con el fin de conseguir el objeto anterior, la presente invención proporciona un método y un sistema para el pre-procesamiento de datos de audio a procesar por un códec con velocidad de codificación variable, de acuerdo con las reivindicaciones independientes 1 y 3, respectivamente.

**Breve descripción de los dibujos**

El objeto y características anteriores de la presente invención quedarán más claros a partir de la siguiente descripción de modos de realización preferidos, ofrecidos conjuntamente con los dibujos que se acompañan.

La figura 1 es un diagrama de bloques de un codificador EVRC.

La figura 2A es un gráfico que muestra una señal de trama residual de una señal que tiene un componente de frecuencia dominante.

La figura 2B es un gráfico que muestra una señal de trama residual de una señal que tiene una diversidad de frecuencias.

La figura 3A es un gráfico que muestra la autocorrelación de la residual para una señal que tiene un componente de frecuencia dominante.

La figura 3B es un gráfico que muestra la autocorrelación de la residual para una señal que tiene una diversidad de frecuencias.

La figura 4 es un diagrama de flujo para realizar el pre-procesamiento del AGC (Control Automático de Ganancia), de acuerdo con la presente invención.

La figura 5 es un diagrama de flujo para realizar el pre-procesamiento de AGC selectivo de tramas, de acuerdo con la presente invención.

La figura 6 es un diagrama de bloques para realizar el AGC de acuerdo con la presente invención.

La figura 7 es un gráfico que ilustra una señal de audio muestreada y su nivel de señal.

La figura 8 es un gráfico que explica el cálculo del nivel de la señal en dirección de avance, de acuerdo con la presente invención.

La figura 9 es un gráfico para explicar el cálculo del nivel de la señal en dirección regresiva, de acuerdo con la presente invención.

Las figuras 10A - 10D son gráficos que muestran los resultados del pre-procesamiento del AGC.

**Modos de llevar a cabo la invención**

Como forma de resolver el problema de las pausas intermitentes, la presente invención proporciona un método de pre-procesamiento de datos de audio antes de ser sometidos al códec de audio. Cierto tipo de sonidos (tales como el de un instrumento de percusión) incluyen componentes del espectro que tienden a ser percibidos como ruido por los codecs de audio optimizados para la voz humana (tales como los códigos para un sistema inalámbrico), y los codecs de audio consideran como ruido las partes de la música que tienen amplitudes bajas. Este fenómeno se muestra comúnmente en todos los sistemas que emplean la DTX (transmisión discontinua) basada en la VAD (Detección de la Actividad de Voz), tales como el GSM (Sistema Global para las comunicaciones Móviles). En el caso de la EVRC, si se determina que los datos son ruido, los datos se codifican con una velocidad de 1/8 entre las tres velocidades predeterminadas de 1/8, 1/2 y 1. Los datos musicales se deciden como ruido por el sistema de codificación, los datos transmitidos no pueden ser oídos básicamente en el extremo receptor, deteriorando así seriamente la calidad del sonido.

Este problema puede ser resuelto mediante el pre-procesamiento de datos de audio, de manera que las velocidades de codificación del códec de EVRC pueden ser decididas como 1 (y no como 1/8) para tramas de datos musicales. De acuerdo con la presente invención, la velocidad de codificación de las señales musicales puede ser aumentada por medio del pre-procesamiento y, por tanto, las pausas de la música en el terminal receptor originadas por la EVRC se reducen. Aunque la presente invención se explica con respecto al códec de EVRC, una persona experta en la técnica sería capaz de aplicar la presente invención a otros sistemas de compresión que utilicen la velocidad de codificación variable, especialmente un códec optimizado para la voz humana (tal como un códec de audio para la transmisión inalámbrica).

Con referencia a la figura 1, se explicará el algoritmo RDA (Algoritmo de Decisión de la Velocidad) de la EVRC. La EVRC será explicada como ejemplo de un sistema de compresión que utiliza una velocidad de codificación variable para comprimir los datos que han de transmitirse a través de una red inalámbrica donde puede aplicarse la presente invención. Es importante comprender el algoritmo de decisión de la velocidad del códec convencional utilizado en un sistema existente, porque la presente invención está basada en una idea de que, en un códec convencional, algunos datos de música pueden ser codificados con una velocidad de datos que es demasiado baja para los datos musicales (aunque puede ser adecuada para los datos de voz), e incrementando la velocidad de datos de la música, se puede mejorar la calidad de la música tras la codificación, la transmisión y la descodificación.

La figura 1 es un diagrama de bloques de alto nivel de un codificador EVRC. En la figura 1, una entrada puede ser una señal de audio de 8k, 16 bits con PCM (Modulación de Código por Impulsos), y una salida codificada pueden ser unos datos digitales cuyo tamaño puede ser de 171 bits (cuando la velocidad de codificación es 1), 80 bits ( cuando la velocidad de codificación es 1/2,16 bits (cuando la velocidad de codificación es 1/8), o 0 bits (en blanco) por trama, de acuerdo con la velocidad de codificación decidida por el RDA. El audio de 8k, 16 bits PCM está acoplado a un codificador EVRC en unidades de trama, conde cada trama tiene 160 muestras (correspondientes a 20 ms). La señal de entrada s[n] (es decir, una señal de trama de entrada de orden n) está acoplada a un bloque 110 de supresión de ruido, que comprueba la señal s[n] de la trama de entrada. En el caso de que la señal de la trama de entrada se considere ruido en el bloque 160 de supresión de ruido, multiplica una ganancia inferior a 1 con la señal, y con ello suprime la señal de trama de entrada. Y después, s'[n], (es decir, la señal que ha pasado a través del bloque 110) se acopla con un bloque 120 de RDA, que selecciona una velocidad entre un conjunto predefinido de velocidades de codificación (1, 1/2, 1/8 y ninguna en el modo de realización explicado en este caso. Un bloque 130 de codificación extrae los parámetros apropiados a partir de la señal, de acuerdo con la velocidad de codificación seleccionada por el bloque 120 de PDA, y un bloque 140 de empaquetamiento de bits empaqueta los parámetros extraídos para adaptarse a un formato de salida predeterminado.

Como se ilustra en la tabla siguiente, la salida codificada puede tener 171, 80, 16 o 0 bits por trama, dependiendo de la velocidad de codificación seleccionada por el RDA.

[Tabla 1]

| Tipo de trama | Bits por trama |
|---|---|
| Trama con velocidad de codificación 1 | 171 |
| Trama con velocidad de codificación 1/2 | 80 |
| Trama con velocidad de codificación 1/8 | 16 |
| En blanco | 0 |

El bloque 120 de RDA divide s'[n] en dos anchos de banda (f(1) de 0,3 ~ 2,0 kHz y f(2) de 2,0 ~ 4,0 kHz) utilizando un filtro de paso banda, y selecciona la velocidad de codificación para cada ancho de banda, comparando un valor de energía de cada ancho de banda con un umbral de decisión de la velocidad decidido por una Estimación de Ruido de Fondo ("BNE"). Se utilizan las ecuaciones siguientes para calcular los dos umbrales para f(1) y f(2).

$T_1 = k_1 (SNR_{f(i)} (m-1)) B_{f(i)}(m-1)$ Ec. (1a)

$T_2 = k_2 (SNR_{f(i)} (m-1)) B_{f(i)}(m-1)$ Ec. (1b)

Donde $k_1$ y $k_2$ son factores de escala umbral, que son funciones de la SNR (Relación de Señal a Ruido) y aumentan cuando lo hace la SNR. Además, $B_{f(i)}(m-1)$ es el BNE (estimación del ruido de fondo) para la banda en la trama de orden (m-1). Como se ha descrito en las ecuaciones anteriores, el umbral de decisión de la velocidad se decide multiplicando el coeficiente de escala por la BNE, y por tanto es proporcional a la BNE.

Por otra parte, la energía de la banda puede ser elegida por los coeficientes de autocorrelación de orden 0 a 16 de los datos de audio pertenecientes a cada ancho de bandas de frecuencia.

$$BE_{f(i)} = R_w(0)\, R_{f(i)}(0) + 2.0 \sum_{k=1}^{L_h - 1} R_w(k)\, R_{f(i)}(k) \qquad \text{Ec. (2)}$$

Donde $BE_{f(i)}$ es un valor de energía para el ancho de banda de frecuencias de orden i (i =1, 2), $R_w(k)$ es una función de los coeficientes de autocorrelación de los datos de audio de entrada, y $R_{f(i)}(k)$ es un coeficiente de autocorrelación de una respuesta de impulsos en un filtro de paso de banda. $L_h$ es una constante de valor 17.

Ahora se explicará la actualización de un ruido estimado ($B_{f(i)}$ (m-1)). El ruido estimado ($B_{f(i)}$ (m)) la para la banda de frecuencias de orden i (o f(i)) de la trama de orden m se decide por el ruido estimado ($B_{f(i)}$ (m-1)) de f(i) para la trama de orden (m-1), la energía de la banda alisada ($E^{SM}_{f(i)}(m)$) de f(i) para la trama de orden m, y una relación de señal a ruido ($SNR_{f(i)}(m-1)$ de f(i) de la trama de orden (m-1), que se representa en pseudo-código.

```
if (β < 0.30 para 8 o más tramas consecutivas)
   B_f(i)(m)=min{E^sm_f(i)(m), 80954304, max{1.03B_f(i)(m-1), B_f(i)(m-1)+1}}
else{
   if (SNR_f(i)(m-1) > 3)
      B_f(i)(m)=min{E^SM_f(i)(m), 80954304, max{1.0054B_f(i)(m-1), B_f(i)(m-1)+1}}
   else
      B_f(i)(m)=min{E^SM_f(i)(m), 80954304, B_f(i)(m-1)}
   }
   if (B_f(i)(m) < lownoise(i))
   B_f(i)(m) = lownoise(i)
}
```

Como se ha descrito anteriormente, si el valor de β, que es una ganancia de predicción a largo plazo (se explicará más adelante cómo decidir β) es inferior a 0,3 durante más de 8 tramas, se selecciona como BNE el valor más bajo entre (i) la energía de la banda alisada, (ii) 1,03 veces la BNE de la trama anterior, y (iii) un valor máximo predeterminado de una BNE (80954304 en el caso anterior). En otro caso, (si el valor de β no es inferior a 0,3 en cualquiera de las 8 tramas consecutivas), si la SNR de la trama anterior es mayor que 3, se selecciona como BNE para esta trama el valor más bajo entre (i) la energía de la banda alisada, (ii) 1,00547 multiplicado por la BNE de la trama anterior, y (iii) un valor máximo predeterminado de la BNE. SI la SNR de la trama anterior no es mayor que 3, se selecciona como BNE para esta trama el valor más bajo entre (i) la energía de la banda alisada, (ii) la BNE de la trama anterior, y el valor máximo predeterminado de la BNE.

Por tanto, en el caso de una señal de audio, la BNE tiende a aumentar a medida que pasa el tiempo, por ejemplo en 1,03 veces o en 1,00547 veces de trama a trama, y disminuye solamente cuando la BNE se hace mayor que la energía de la banda alisada. Consecuentemente, si la energía de la banda alisada se mantiene dentro de una gama relativamente pequeña, la BNE aumenta a medida que pasa el tiempo, y por ello el valor del umbral de decisión de la velocidad aumenta (véase la Ec. (1)). Como resultado, se hace más probable que se codifique una trama con una

velocidad de 1/8. En otras palabras, si se reproduce la señal musical durante un tiempo largo, las pausas tenderán a ocurrir más frecuentemente.

La ganancia (β) de la predicción a largo plazo se define por la autocorrelación de las residuales, como sigue:

$$\beta = \max\{o, \min\{1, \frac{R_{max}}{R_{\varepsilon}(0)}\}\}$$ Ec. (3)

Donde ε es una señal residual de predicción, $R_{máx}$ es un valor máximo de los coeficientes de autocorrelación de la señal residual de predicción, y $R_{\varepsilon}(0)$ es un coeficiente de orden 0 de una función de autocorrelación de la señal residual de predicción.

De acuerdo con la ecuación anterior, en el caso de una señal monofónica o una señal de voz en la que existe un tono dominante, el valor de β sería mayor, pero en el caso de música que incluya varios tonos, el valor de β sería menor.

La señal residual (ε) de predicción se define como sigue:

$$\varepsilon[n] = s'[n] - \sum_{i=1}^{10} a_i[k]\, s'[n-i]$$ Ec. (4)

donde s'[n] es una señal de audio pre-procesada por el bloque 110 de supresión de ruido, y $a_i[k]$ es un coeficiente LPC interpolado del segmento de orden k de una trama en curso.

Es decir, la señal residual de predicción es una diferencia entre una señal reconstruida por los coeficientes LCP y una señal original.

La señal residual de la trama parece normal en el caso de que exista un componente de frecuencia dominante en la trama (véase la figura 2A), mientras que es anormal en el caso en que existan varios componentes de frecuencia en la trama (véase la figura 2B). Consecuentemente, en el caso anterior, un valor máximo de pico normalizado de los coeficientes de correlación (es decir, la ganancia β de predicción a largo plazo), sería un valor mayor (por ejemplo, β = 0,6792, véase la figura 3A), mientras que en el último caso, se convierte en un valor menor (por ejemplo, β = 0,2616, véase la figura 3B). En estas figuras 3A y 3B, los coeficientes de autocorrelación están normalizados por R(0). En las figuras 2A y 2B, el eje x representa números y el eje y representa la amplitud de la señal residual, donde los números del gráfico son valores normalizados que dependen del requisito del sistema (por ejemplo, cuántos bits se utilizan para representar el valor), lo cual se aplica a otros gráficos en esta solicitud (por ejemplo en las figuras 7 - 10).

Se explicará ahora cómo decidir la velocidad de codificación. Para cada una de las dos bandas de frecuencia, si la energía de la banda es mayor que los dos valores umbrales, la velocidad de codificación es 1, si la energía de la banda está entre los dos valores umbrales, la velocidad de codificación es 1/2, y si la energía de la banda es inferior a ambos valores umbrales, la velocidad de codificación es 1/8. Después de decidir las velocidades de codificación para las dos bandas de frecuencia, la más alta de las velocidades de codificación decididas para las bandas de frecuencia se selecciona como la velocidad de codificación para esa trama. En un sistema real, la codificación a una velocidad de 1/8 puede significar que la señal relevante se decide como ruido y se transmiten muy pocos datos; la codificación a una velocidad de 1 puede significar que la señal se decide que es voz humana válida; y la codificación a velocidad de 1/2 sucede en un corto intervalo durante la transición entre 1/8 y 1.

Hasta ahora, se ha explicado cómo se decide la velocidad de codificación en un códec EVRC, que es un ejemplo de un sistema de codificación de velocidad variable, donde se puede aplicar la presente invención. A partir de lo anterior, se puede comprender que la velocidad de codificación de una trama se puede maximizar a 1 tanto como sea posible (i) incrementando la energía de la banda y/o (ii) disminuyendo el valor umbral para la decisión de la velocidad de codificación.

La presente invención utiliza un método de AGC (Control Automático de Ganancia) para aumentar la energía de la banda. El AGC es un método de ajuste de la ganancia de la señal en curso mediante la predicción de señales durante un cierto intervalo (Intervalo de ATAQUE). Por ejemplo, si se reproduce música en altavoces con diferentes gamas dinámicas, no se puede procesar apropiadamente sin el AGC (sin AGC, algunos altavoces funcionarán en la región de saturación). Por tanto, es necesario realizar el pre-proceso del AGC basado en la característica del dispositivo generador del sonido, tal como un altavoz, un auricular o un teléfono celular.

En el caso de un teléfono celular, aunque sería ideal medir la gama dinámica del teléfono celular y realizar el AGC

con el fin de asegurar la mejor calidad del sonido, es imposible diseñar un AGC optimizado para todos los teléfonos celulares, porque la característica de un teléfono celular podría variar dependiendo del fabricante y también del modelo en particular. Por tanto, es necesario diseñar un AGC que sea generalmente aplicable a todos los teléfonos celulares.

La figura 4 es un diagrama de flujo de alto nivel para realizar el pre-procesamiento del AGC de acuerdo con un modo de realización de la presente invención. Al principio, se obtienen los datos de audio en el paso 410, y después se clasifican los datos de audio basándose en la característica de los datos de audio del paso 420. Los datos de audio serían procesados de diferentes maneras dependiendo de la clasificación porque, para ciertos tipos de datos de audio, es preferible reforzar la energía de todas las tramas, mientras que, en otros casos, funciona mejor reforzar solamente la energía de la banda de todas las tramas que son codificadas con una velocidad de tramas baja en el codificador de velocidad de codificación variable (tal como el EVRC). La parte derecha 440 del diagrama de flujo muestra el reforzamiento de energía de todas las tramas. En el caso de la música clásica o datos de audio monofónicos que tengan un solo tono, es preferible que se realice la parte derecha 440 del diagrama de flujo. La parte izquierda 430 del diagrama de flujo muestra el reforzamiento de la energía de la banda de tales tramas que son codificadas con una velocidad baja de tramas. En el caso de datos de audio polifónicos, tales como la música rock, es preferible que se realice la parte izquierda 430 del diagrama de flujo.

La figura 5 es un diagrama de flujo del AGC selectivo de trama para el pre-procesamiento de tramas que serían codificadas con una velocidad baja sin pre-procesamiento. El AGC se realiza de diferentes maneras dependiendo de la energía de las tramas de las señales musicales. El intervalo en el cual la energía de las tramas de los datos de audio (antes de la codificación EVRC) es baja (es decir, menor que 1000) se define como intervalo de “SILENCIO” donde no se realiza ningún pre-procesamiento. Para las tramas que no están en el intervalo de “SILENCIO”, se realiza la codificación EVRC para detectar la velocidad de codificación de cada trama. Para tales intervalos, donde las tramas con velocidad de codificación de 1/8 ocurren frecuentemente, (lo que significa que tales intervalos se consideran ruido por el codificador EVRC), la energía de la banda de las tramas se aumenta localmente. Cuando se refuerza la energía de ciertas tramas, sería necesaria la interpolación con otras tramas (a este respecto, lo que se denomina “interpolación de envolvente” se explicará más adelante) para impedir la discontinuidad de la amplitud del sonido entre las tramas reforzadas y las tramas vecinas no reforzadas.

La figura 6 es un diagrama de bloques para el AGC de acuerdo con un modo de realización de la presente invención. En este modo de realización, el AGC es un proceso para ajustar el nivel de la señal de la muestra en curso basándose en una ganancia de control decidida a partir de un conjunto de valores de muestra en una ventana futura. Al principio, se calcula un “nivel de señal en dirección de avance” $l_f[n]$ y un “nivel de señal en dirección de retroceso” $l_b[n]$, utilizando la señal de audio muestreada s[n] de una manera que se explicará más adelante, y a partir de ellos, se calcula un “nivel de señal final” l[n]. Después de calcular l[n], se calcula la ganancia del proceso por muestra (G[n]) utilizando l[n] y después se obtiene la salida y[n] multiplicando G[n] por s[n].

En lo que sigue, se describirán con más detalle las funciones de los bloques de la figura 6.

La figura 7 muestra un ejemplo de nivel de señal (l[n]) calculado a partir de la señal de audio muestreada (s[n]). La envolvente del nivel de la señal l[n] varía dependiendo de cómo se procesan las señales utilizando la supresión exponencial en la dirección de avance (“ATAQUE”) y en la supresión exponencial en la dirección de retroceso (“LIBERACIÓN”). En la figura 7, $L_{máx}$ y $L_{mín}$ se refieren a los valores máximo y mínimo de la señal de salida, después del pre-procesamiento de AGC.

Se obtiene un nivel de señal en el instante n calculando los niveles de señal en dirección de avance (para efectuar la LIBERACIÓN) y calculando los niveles de señal en dirección de retroceso (para realizar el ATAQUE). La constante de tiempo de una “función exponencial” que caracteriza la supresión exponencial será denominada “tiempo de LIBERACIÓN” en la dirección de avance, y “tiempo de ATAQUE” en la dirección de retroceso. El tiempo de ATAQUE es el tiempo que tarda una nueva señal de salida en alcanzar una amplitud de salida apropiada. Por ejemplo, si la amplitud de una señal de entrada disminuye en 30 dB abruptamente, el tiempo de ATAQUE es el tiempo que tarda una señal de salida en disminuir de manera consecuente (en 30 dB). El tiempo de LIBERACIÖN es el tiempo que se tarda en alcanzar un nivel de amplitud apropiado al final de un nivel de salida existente. Esto es, el tiempo de ATAQUE es el periodo para que el inicio de un impulso alcance una amplitud de salida deseada, mientras que el tiempo de LIBERACIÓN es el periodo para que el final de un impulso alcance una amplitud de salida deseada.

En lo que sigue, se describirá cómo calcular un nivel de señal en dirección de avance y un nivel de señal en dirección de retroceso, con referencia a las figuras 8 y 9.

Con referencia a la figura 8, el nivel de señal en dirección de avance se calcula por medio de los pasos siguientes:

En el primer paso, se inicializa (se ponen a 0) un valor de pico en curso y un índice de pico en curso, y el nivel de señal en dirección de avance ($l_f[n]$) se inicializa como |s[n]|, que es el valor absoluto de s[n].

En el segundo paso, el valor de pico en curso y el índice de pico en curso se actualizan. Si |s[n]| es mayor que el

valor de pico en curso (p[n]), se actualiza p[n] al valor |s[n]|, y el índice de pico en curso ($i_p$[n]) se actualiza a n (como se ilustra en el siguiente pseudo-código).

```
if (|s[n]| > p[n]) {
   p[n] = |s[n]|
   ip[n] = n
}
```

En el tercer paso, se calcula el valor de pico en curso suprimido. El valor de pico en curso suprimido $p_d$[n] se decide reduciendo exponencialmente el valor de p[n] de acuerdo con el paso del tiempo, como sigue.

$P_d[n] = p[n] * \exp(-TD/RT)$ Ec. (5)

$TD = n - i_p[n]$

Donde RT representa el tiempo de LIBERACIÓN.

En el cuarto paso, valores mayores de $p_d$[n] y |s[n]| se deciden como el nivel de señal en la dirección de avance, como sigue.

$l_f[n] = \text{máx}(p_d[n], |s[n]|)$ Ec. (6)

A continuación, los pasos segundo a cuarto anteriores se repiten para obtener un nivel de señal en dirección de avance ($l_f$[n]) a medida que n aumenta de uno en uno.

Con referencia a la figura 9, se calcula el nivel de señal en la dirección de retroceso por medio de los pasos siguientes.

En el primer paso, se inicializa en 0 un valor de pico en curso, se inicializa en AT un índice de pico en curso, y se inicializa como |s[n]| un valor absoluto de s[n].

En el segundo paso, se actualiza el valor de pico en curso y el índice de pico en curso. Se detecta un valor máximo de s[n] en la ventana de tiempo desde n hasta n + AT y se actualiza el valor de pico en curso como valor máximo detectado. Además, se actualiza $i_p$[n] como índice de tiempo para el valor máximo.

$p[n] = \text{máx}(\{|s[]|\})$ Ec. (7)

Ip[n] = (un índice de s[], donde |s[]| tiene su valor máximo)

donde el índice de s[] puede tener valores desde n a n+AT.

En el tercer paso, se calcula un valor de pico en curso suprimido como sigue.

$P_d[n] = p[n] * \exp(-TD/AT)$ Ec. (8)

$TD = i_p[n] - n$

donde AT representa el tiempo de ATAQUE.

En el cuarto paso, se decide un valor mayor de $p_d$[n] y |s[n]|, como nivel de la señal en dirección de retroceso.

$l_b[n] = \text{máx}(p_d[n], |s[n]|)$ Ec. (9)

A continuación, se repiten los pasos segundo a cuarto anteriores para obtener un nivel de señal en dirección de retroceso ($l_b$[n]) cuando n aumenta de uno en uno.

El nivel final de la señal (l[n]) se define como el valor máximo del nivel de la señal en dirección de avance y el nivel de la señal en dirección de retroceso para cada índice de tiempo.

$l[n] = \text{máx}(l_f[n], l_b[n])$ para $t = 0, \ldots, t_{máx}$ Ec. (10)

Donde $t_{máx}$ es el índice de tiempos máximo.

La relación tiempo de ATAQUE/tiempo de LIBERACIÓN está relacionada con la característica/calidad del sonido. Consecuentemente, cuando se calculan los niveles de la señal, es necesario fijar apropiadamente el tiempo de ATAQUE y el tiempo de LIBERACIÓN para obtener un sonido optimizado con la característica del medio. Si la suma del tiempo de ATAQUE y el tiempo de LIBERACIÓN es demasiado pequeña, (es decir, si la suma es inferior a 20 ms), se puede escuchar una distorsión en forma de vibración, con una frecuencia de 1000/(tiempo de ATAQUE + tiempo de LIBERACIÓN) en el teléfono celular del usuario. Por ejemplo, si el tiempo de ATAQUE y el tiempo de LIBERACIÓN son de 5 ms cada uno, se puede escuchar una distorsión de vibración con una frecuencia de 100 Hz. Por tanto, es necesario fijar la suma del tiempo de ATAQUE y el tiempo de LIBERACIÓN con una duración mayor que 30 ms para evitar la distorsión de vibración.

Por ejemplo, si el ATAQUE es lento y la LIBERACIÓN es rápida, se obtendría un sonido con una gama dinámica más ancha. Cuando el tiempo de LIBERACIÓN es largo, se suprime el componente de alta frecuencia de la señal de salida, dando como resultado una señal de sonido apagada. Sin embargo, si el tiempo de LIBERACIÓN se hace muy rápido (el significado de "rápido" en este contexto puede variar dependiendo de la característica de la música), en la señal de salida procesada por el AGC sigue al componente de baja frecuencia de la forma de onda de entrada. En este caso, el componente fundamental de la señal se suprime, o puede incluso ser sustituido por una cierta distorsión armónica (el componente fundamental significa el componente de frecuencia más importante que una persona puede oír, que es lo mismo que un tono). A medida que los tiempos de ATAQUE y de LIBERACIÓN se hacen más largos, se impiden bien las pausas, pero el sonido se hace más apagado (pérdida de la alta frecuencia). Consecuentemente, hay un compromiso entre la calidad del sonido y el número de pausas.

Para enfatizar el efecto de un instrumento de percusión, tal como un tambor, el tiempo de ATAQUE debe ser alargado. Sin embargo, en el caso de la voz de una persona, acortar el tiempo de ATAQUE ayudaría a impedir que la ganancia de la parte inicial disminuya innecesariamente. Es importante decidir apropiadamente el tiempo de ATAQUE y el tiempo de LIBERACIÓN para asegurar la calidad del sonido en el proceso del AGC, y se deciden considerando la característica de la música.

El método de pre-procesamiento de la presente invención no implica cálculos muy complicados y puede ser realizado con un retardo muy corto (con el orden del tiempo de ATAQUE y de LIBERACIÓN), y por tanto cuando se retransmite un programa musical, es posible el pre-procesamiento casi en tiempo real.

En cuanto a qué tramas (o intervalos) deben ser procesados utilizando el AGC de acuerdo con la presente invención, es preferible procesar intervalos con amplitud baja y alta (en comparación con un cierto estándar). Cuando se codifican y transmiten datos de audio con una amplia gama dinámica en un sistema de comunicaciones inalámbricas y se reproducen en un teléfono celular, la calidad del sonido se degrada, porque el sonido con amplitudes bajas tiende a no ser oído. Así, para tales tramas con amplitud baja, la amplitud debe aumentarse para una señal de mejor calidad. Y, en el caso de intervalos (tramas) con amplitudes altas, la amplitud debe reducirse para evitar la saturación de los sonidos reproducidos. Para conseguir ambos objetivos, en un modo de realización de la presente invención, se fijan dos valores límite ($L_{mín}$ y $L_{máx}$), y después se procesan los intervalos en los cuales los niveles de la señal son inferiores a $L_{mín}$ o mayores que $L_{máx}$.

Como se ha explicado anteriormente, para evitar el cambio repentino de la amplitud entre los intervalos procesados (por el AGC) y los no procesados, es necesario ajustar la ganancia de control apropiadamente para impedir un cambio abrupto de la amplitud. Además, después del AGC, el nivel máximo no puede exceder del valor límite máximo ($L_{máx}$) y, por tanto, sin suavización del valor de la ganancia, la envolvente de las señales musicales puede ser fijada en el valor límite máximo. Si la envolvente se fija en un valor límite máximo, la calidad del sonido de los intervalos procesados sería diferente de la de intervalos no procesados.

Considerando lo anterior, la ganancia del proceso para cada señal de muestra (G[n]) se decide con la ecuación siguiente.

$$G[n] = c * (L/l[n] + (1-c) \qquad \text{Ec. (11)}$$

Donde c es un coeficiente de ganancia, que tiene un valor entre 0 y 1. Y L se fija en $L_{mín}$ o $L_{máx}$, dependiendo de la característica de la señal en intervalos a procesar.

La señal procesada (s'[n]) se decide con una multiplicación de la señal antes del AGC (s[n]) por la ganancia del proceso.

$$s'[n] = G[n] * s[n] \qquad \text{Ec. (12)}$$

A partir de las ecuaciones anteriores, (Ec. 11 y Ec. 12), se puede saber que a medida que c se aproxima a 1, la envolvente de la salida sería fijada con el valor límite, y a medida que c se aproxima a 0, la envolvente de la señal resultante tras el AGC (utilizando la ganancia en la ecuación anterior), se haría similar a la envolvente de entrada.

Utilizando el método explicado anteriormente, la velocidad de codificación de las señales musicales puede ser reforzada, y por ello el problema de las pausas en la música originados por la EVRC se puede mejorar suficientemente.

Se explicarán ahora los resultados experimentales concernientes al método anterior explicado. En estos experimentos se utilizan señales musicales monofónicas muestreadas de 8 kHz, 16 bits con calidad de CD.

Las figuras 10A - 10D muestran la comparación entre las señales codificadas en el caso de utilizar el pre-procesamiento de AGC de la presente invención y en el caso de no utilizar el pre-procesamiento de AGC. En las figuras 10A - 10D, el eje horizontal es el eje de tiempos, y el eje vertical representa la amplitud de la señal. La figura 10A muestra la señal original, la figura 10B muestra la señal pre-procesada por AGC, la figura 10C muestra la señal codificada por la EVRC a partir de las señales originales, y la figura 10D muestra la señal codificada por EVRC a partir de las señales pre-procesadas por el AGC. En la señal con gama dinámica ancha, como se ilustra en la figura 10A, tienden a ocurrir más pausas, especialmente durante el periodo de baja amplitud que sería considerado ruido. En la figura 10C, se puede observar que la señal con baja amplitud no se oiría. La señal original se pre-procesa con el AGC utilizando los parámetros de la Tabla 2, y la señal pre-procesada está ilustrada en la figura 10B. Tras la codificación/descodificación de EVRC, la señal pre-procesada por AGC se convierte en la de la figura 10D. Como se ilustra en la figura 10D, el pre-procesamiento de AGC refuerza la parte de la señal que tiene amplitud baja, de manera que tras la codificación/descodificación de EVRC, la señal no puede ser pausada. Como se ilustra en la Tabla 3, a través del pre-procesamiento de AGC, el número de tramas codificadas con una velocidad de codificación de 1/8 disminuye desde 356 a 139.

[Tabla 2]

| | |
|---|---|
| Número de muestras de ATAQUE | 160 |
| Número de muestras de LIBERACIÓN | 2000 |
| Valor límite mínimo | 5000 |
| Valor límite máximo | 30000 |
| Coeficiente de suavización de la ganancia | 0,5 |

[Tabla 3]

| | Señales originales | Señales pre-procesadas por AGC |
|---|---|---|
| Número de tramas con una velocidad de codificación de 1/8 | 356 | 139 |

Se ha efectuado una prueba MOS (puntuación media de opinión) a un grupo de prueba de 11 personas con edades de 20 y 30 años, para la comparación entre la música original y la música pre-procesada por el algoritmo sugerido de pre-procesamiento de AGC. Para la prueba, se utilizaron teléfonos celulares de Samsung Anycall®. Las señales musicales no procesadas y pre-procesadas han sido codificadas y proporcionadas a un teléfono celular en una secuencia aleatoria, y evaluadas por el grupo de prueba utilizando un esquema de puntuación de cinco grados como sigue:

(1) malo (2) pobre (3) regular (4) bueno (5) excelente

Se utilizaron tres canciones para la prueba, y la Tabla 4 muestra el resultado del experimento. De acuerdo con el resultado de la prueba, el promedio de puntos para las canciones aumentó desde 3,000 hasta 3,273, desde 1,727 hasta 2,455 y desde 2,091 hasta 2,727.

[Tabla 4]

| Títulos de canciones (Compositor) | Género de canciones | Promedio de puntos para canciones originales | Puntuación media para canciones pre-procesadas |
|---|---|---|---|
| Girl's prayer (Badarzcevska) | Piano solo | 3,000 | 3,273 |
| Sonata Patética Op. 13 (Beethoven) | Piano solo | 1,727 | 2,455 |
| Quinta sinfonía (Destino) (Beethoven) | Sinfonía | 2,091 | 2,727 |

En un modo de realización de la invención, se puede dar servicio al teléfono convencional y al teléfono inalámbrico con un solo sistema para proporcionar señales musicales. En ese caso, se detecta el identificador del usuario en el sistema para procesar la señal musical. En un sistema telefónico convencional, se utiliza una señal de voz no

comprimida con un ancho de banda de 8 kHz, y por tanto, si se transmite una música muestreada de 8 kHz/8 bits/Ley-A, se puede oír música de alta calidad sin distorsión de la señal. En un modo de realización de la invención, un sistema para proporcionar una señal musical a un terminal de usuario determina si se ha originado una petición de música desde una persona que llama desde un teléfono convencional o un teléfono inalámbrico, utilizando un identificador del que llama. En el caso anterior, el sistema transmite la señal de música original, y en el último caso, el sistema transmite la música pre-procesada con AGC.

Sería evidente para una persona de la técnica que el método de pre-procesamiento de la presente invención puede ser implementado utilizando software o bien un hardware exclusivo. Además, en un modo de realización de la invención, se utiliza el sistema VoiceXLM para proporcionar música a los abonados, donde el contenido de audio puede ser cambiado frecuentemente. En tal sistema, el pre-procesamiento por AGC de la presente invención puede ser realizado en base a la demanda. Para realizar esto, se puede definir una etiqueta no estándar, tal como <audio src = “xx.wav” tipo = “música/clásica/”>, para determinar si debe realizarse el pre-procesamiento o los tipos de pre-procesamiento a realizar.

**Aplicación industrial**

La aplicación de la presente invención incluye cualquier servicio inalámbrico que proporcione música u otro sonido que no sea voz humana a través de la red inalámbrica (esto es, utilizando un códec para un sistema inalámbrico). Además, la presente invención puede ser aplicada también a otros sistemas de comunicaciones donde un códec utilizado para comprimir los datos de audio está optimizado para la voz humana y no para la música y otro sonido. Los servicios específicos donde se puede aplicar la presente invención incluyen, entro otros, el “servicio de coloración” y el “ARS (sistema de Respuesta de Audio)”.

El método de pre-procesamiento de la presente invención puede ser aplicado a cualquier dato de audio antes de ser sometido a un códec de un sistema inalámbrico (o a cualquier otro códec optimizado para la voz humana y no para la música). Después de que los datos de audio hayan sido pre-procesados de acuerdo con el método de pre-procesamiento de la presente invención, los datos pre-procesados se pueden procesar y transmitir en un códec inalámbrico normal. Aparte de añadir el componente necesario para realizar el método de pre-procesamiento de la presente invención, no es necesaria ninguna otra modificación al sistema inalámbrico. Por tanto, el método de pre-procesamiento de la presente invención puede ser fácilmente adoptado por un sistema inalámbrico existente.

Aunque la presente invención se ha explicado con respecto a un códec de EVRC, en otro modo de realización de la presente invención se puede aplicar de una manera similar a otros codecs que tengan velocidad de codificación variable.

La presente invención se ha descrito con referencia a los modos de realización preferidos y a los dibujos, pero la descripción no pretende limitar la presente invención a la forma aquí descrita. Debe entenderse que una persona experta en la técnica es capaz de utilizar una diversidad de modificaciones y otros modos de realización iguales a la presente invención. Por tanto, solamente las reivindicaciones anexas están destinadas a limitar la presente invención.

**REIVINDICACIONES**

1. Un método para el pre-procesamiento de datos de audio que contienen datos musicales a procesar por un códec de Codificación de Velocidad Variable Reforzada, para la transmisión en un sistema de comunicaciones inalámbricas, estando dicho códec optimizado para la voz humana y funcionando a tres velocidades de codificación, comprendiendo el método el paso de, para al menos un intervalo de datos que ha de codificarse por el códec a la velocidad de codificación más baja y que no está definido como intervalo de SILENCIO, ajustar las amplitudes de los datos de audio dentro de dicho al menos un intervalo de datos, de forma que los datos de audio dentro del al menos un intervalo de datos, se codifican a la velocidad de codificación máxima y, cuando los datos de audio se descodifican en el terminal receptor, se puede reducir la pausa intermitente de la música.

2. Un método según la reivindicación 1, en el que el paso de ajuste comprende:

- calcular niveles de señal de los datos de audio;

- decidir los coeficientes de ganancia suavizada basándose en los niveles de la señal; y

- generar datos de audio pre-procesados multiplicando los coeficientes de ganancia suavizada de los datos de audio dentro del intervalo decidido.

3. Un dispositivo para el pre-procesamiento de datos de audio que contienen datos musicales para ser codificados por un códec de Codificación de Velocidad Variable Reforzada para su transmisión por un sistema de comunicaciones inalámbricas, estando dicho códec optimizado para la voz humana y funcionando a tres velocidades de codificación, comprendiendo el dispositivo, para el al menos un intervalo de datos que ha de codificarse por el códec a la velocidad de codificación más baja y que no esté definido como intervalo de SILENCIO, medios para ajustar las amplitudes de los datos de audio dentro de dicho al menos un intervalo de datos, de forma que los datos de audio dentro del al menos un intervalo de datos se codifica a la velocidad máxima de codificación y, cuando los datos de audio se descodifican en el terminal receptor, se puede reducir la pausa intermitente de la música.

Fig. 1

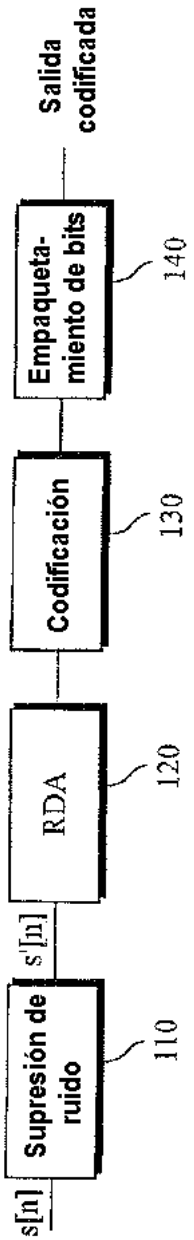

s[n]
Supresión de ruido
110
s'[n]
RDA
120
Codificación
130
Empaqueta-miento de bits
140
Salida codificada

Fig. 2A

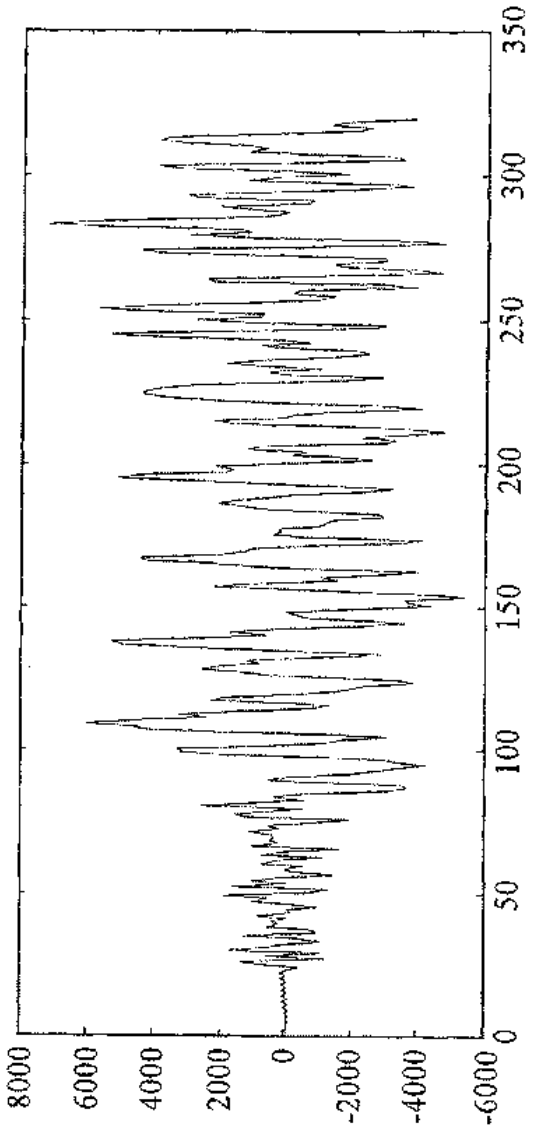

8000
6000
4000
2000
0
-2000
-4000
-6000
0
50
100
150
200
250
300
350

Fig. 2B

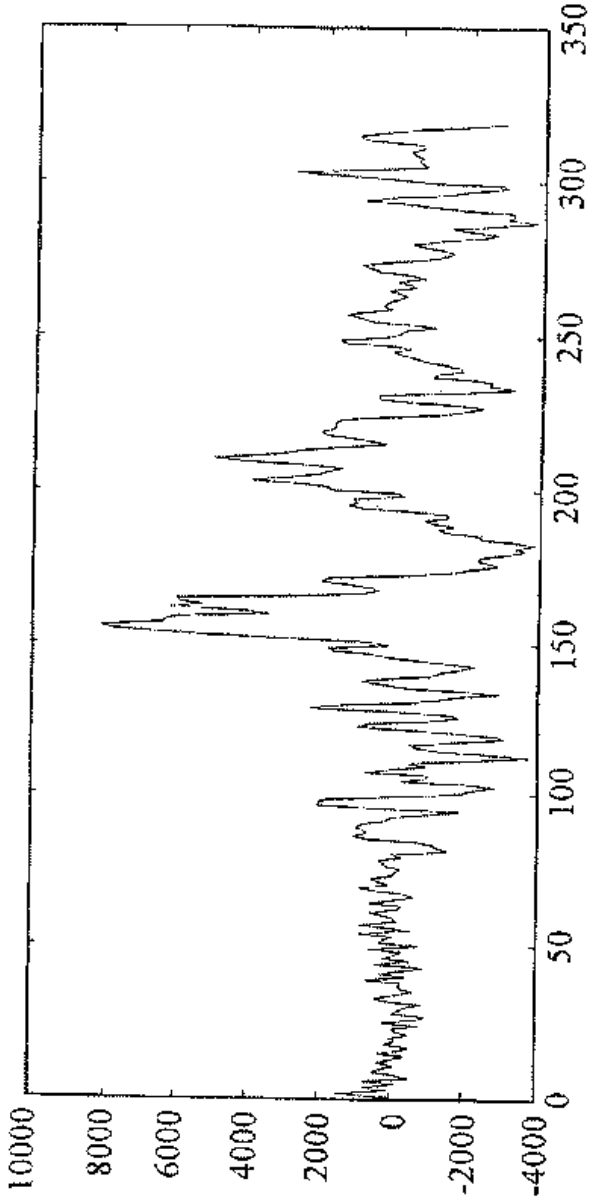

10000
8000
6000
4000
2000
0
-2000
-4000
0
50
100
150
200
250
300
350

**Fig. 3A**

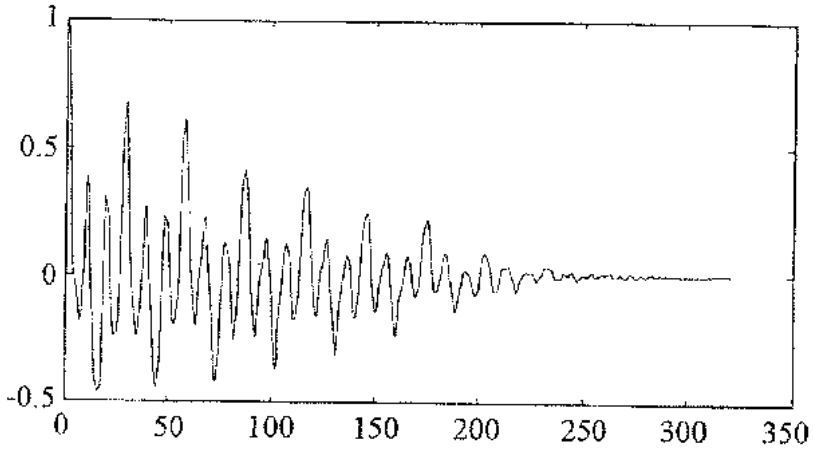
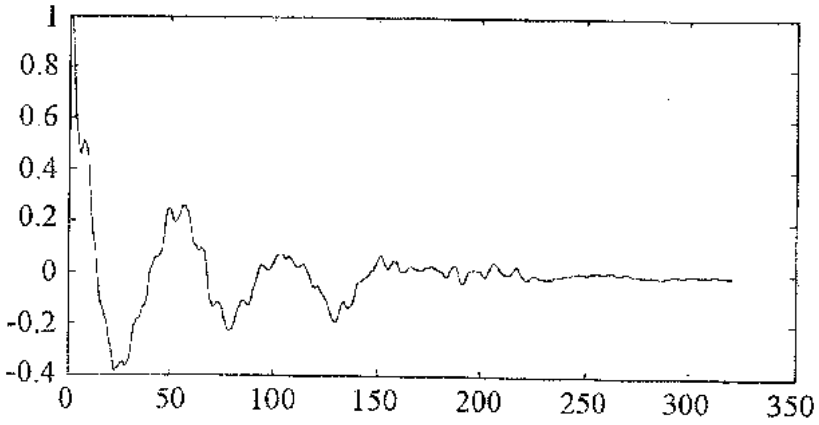
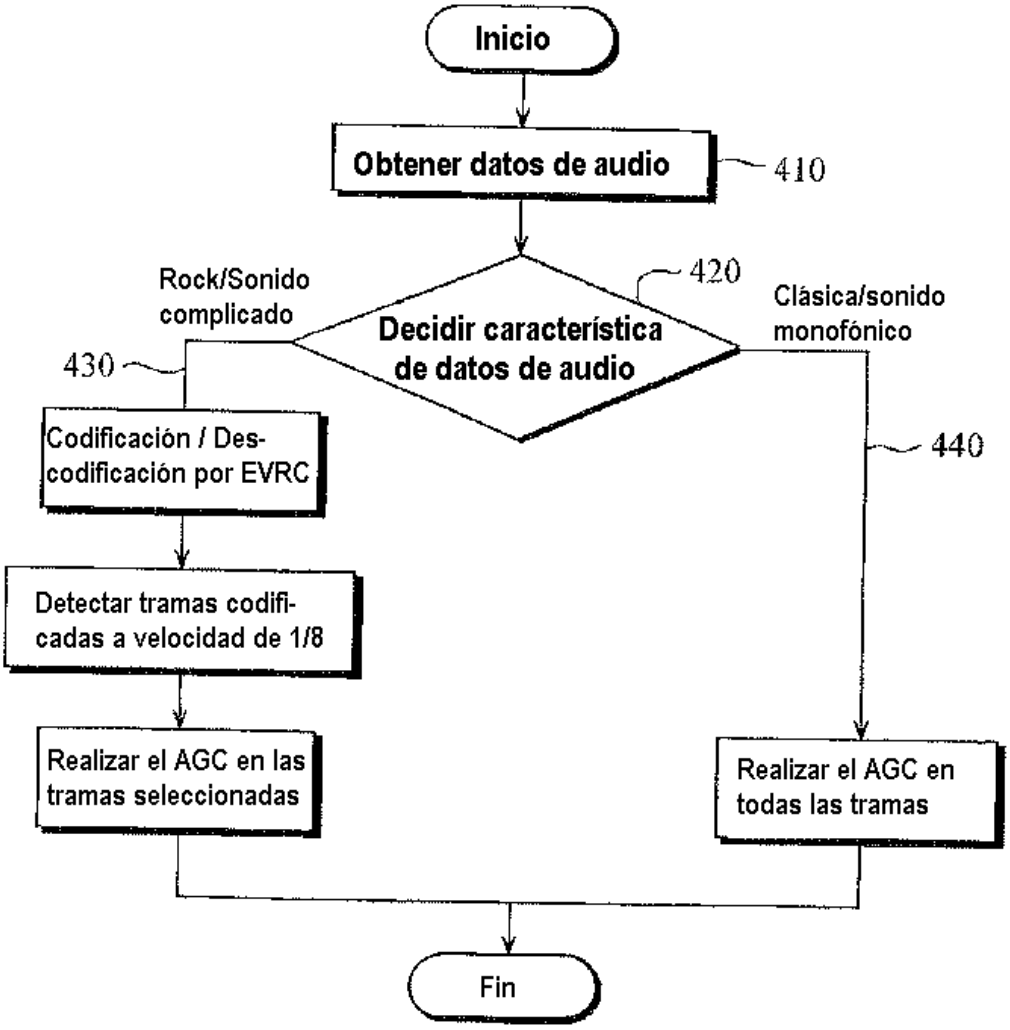

1
0.5
0
-0.5
0
50
100
150
200
250
300
350

**Fig. 3B**


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
50
100
150
200
250
300
350

# Fig. 4


Inicio
Obtener datos de audio
410
Decidir característica de datos de audio
420
Rock/Sonido complicado
Clásica/sonido monofónico
430
440
Codificación / Des-codificación por EVRC
Detectar tramas codificadas a velocidad de 1/8
Realizar el AGC en las tramas seleccionadas
Realizar el AGC en todas las tramas
Fin

# Fig. 5

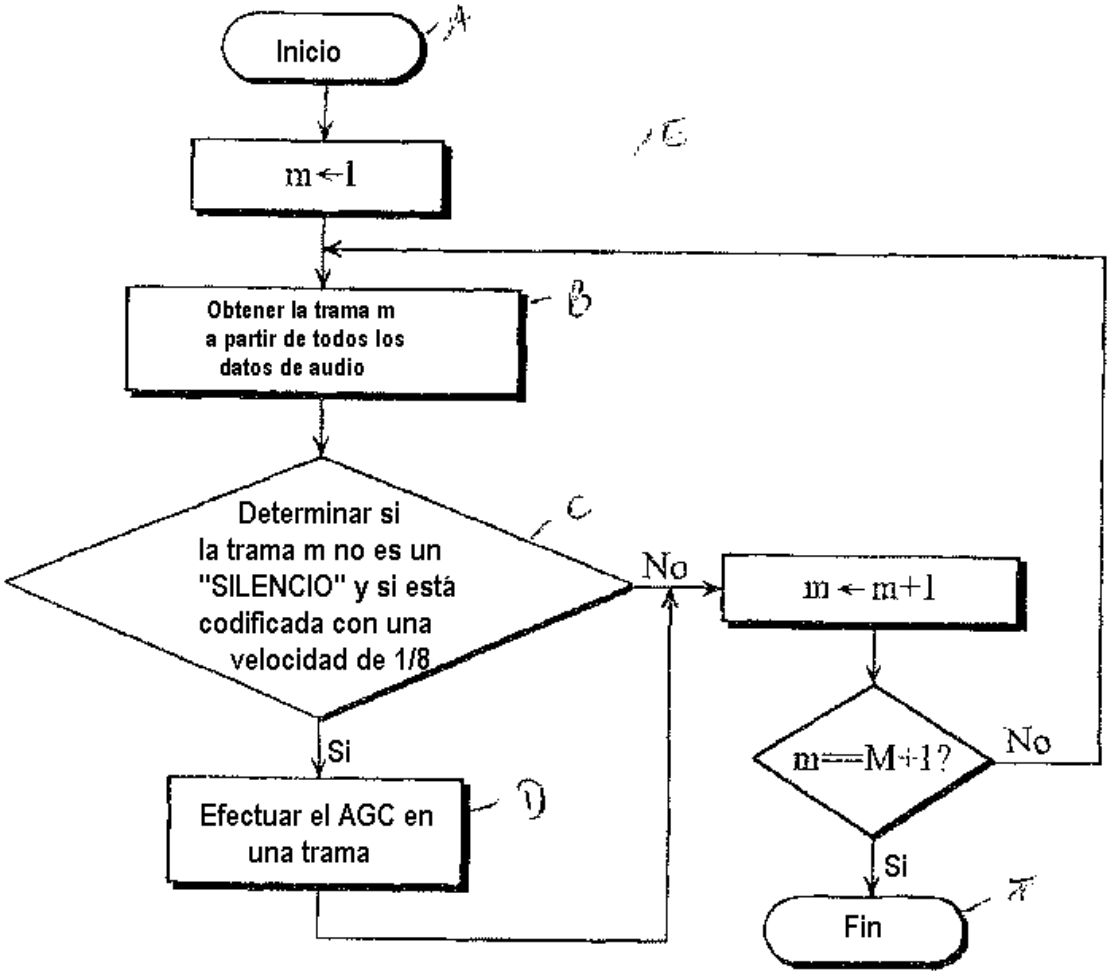
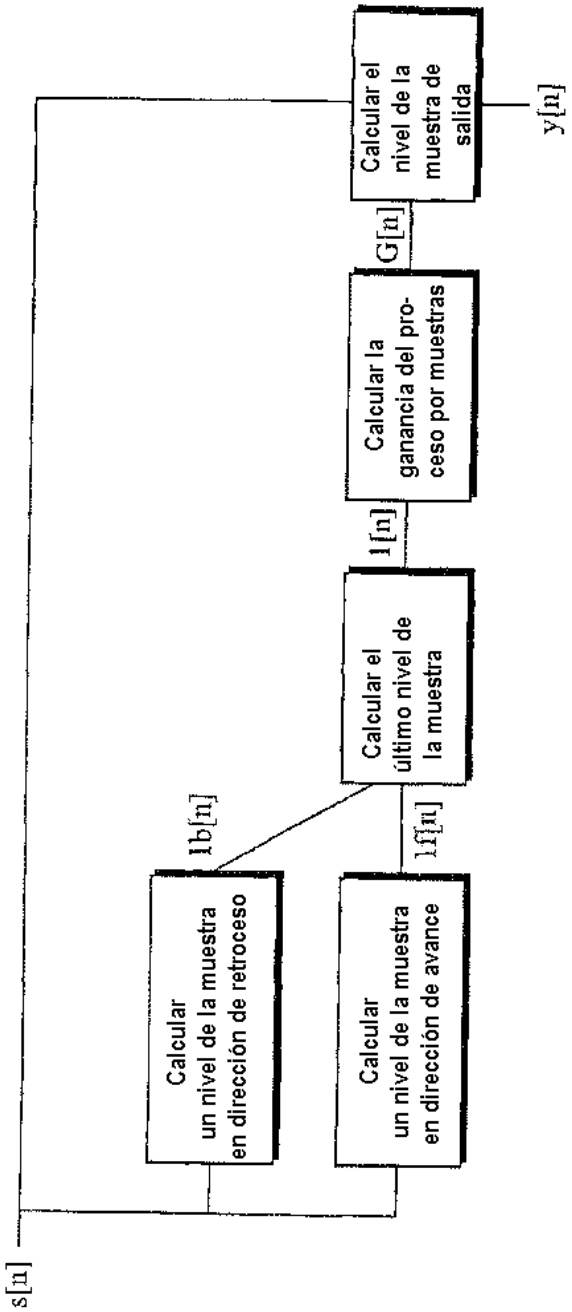
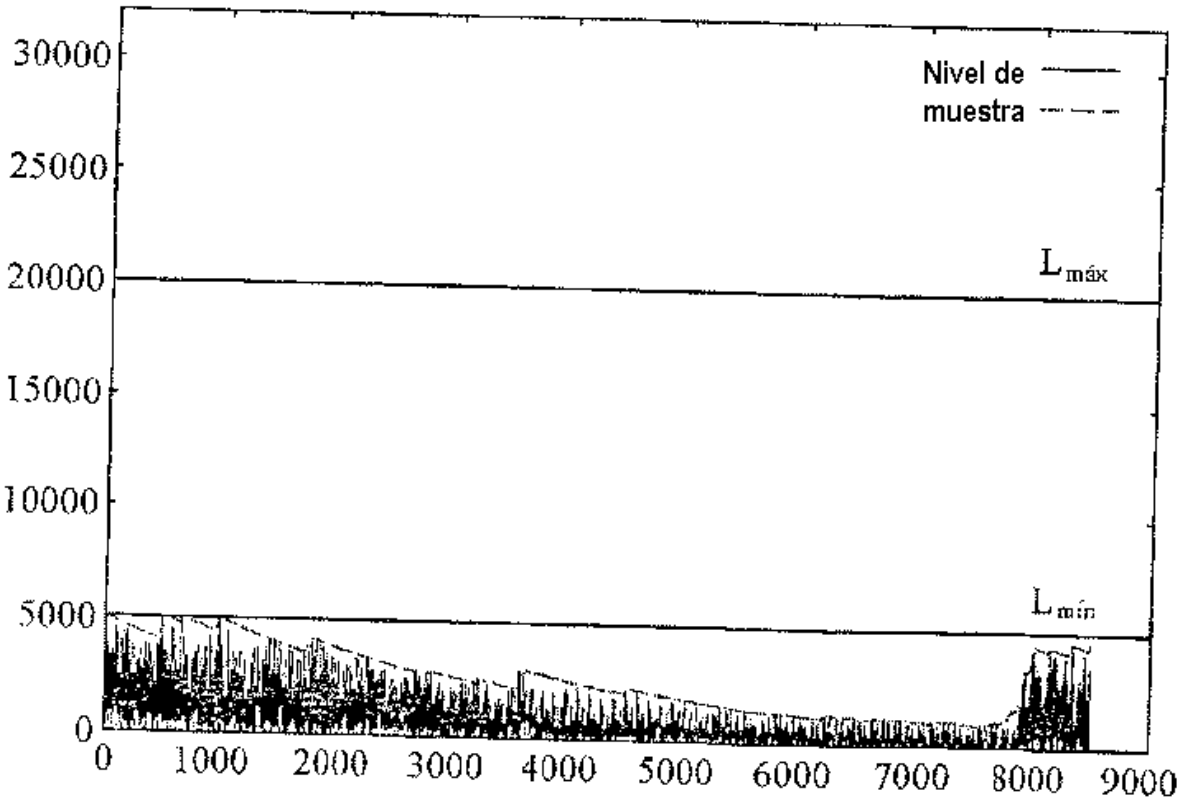

Inicio
A
m←1
C
Obtener la trama m a partir de todos los datos de audio
B
Determinar si la trama m no es un "SILENCIO" y si está codificada con una velocidad de 1/8
C
No
m ← m+1
m==M+1?
No
Si
Efectuar el AGC en una trama
D
Si
Fin
F

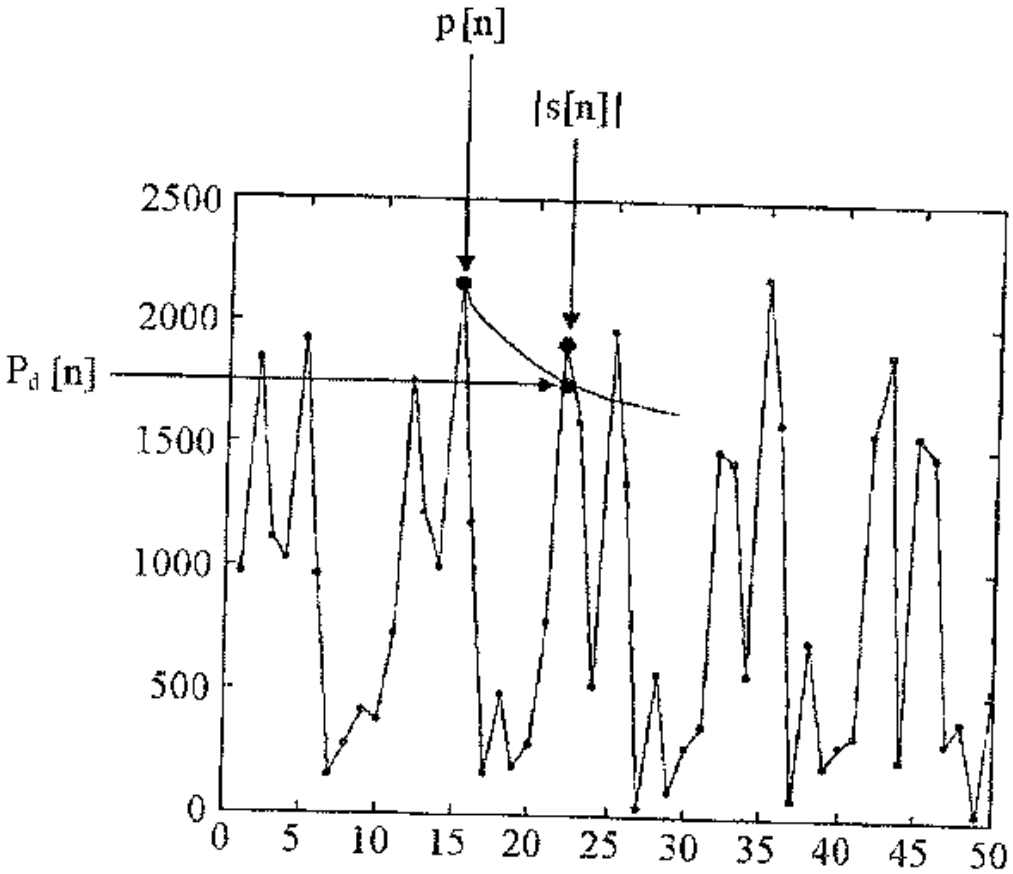
Fig. 6


s[n]
Calcular un nivel de la muestra en dirección de retroceso
Calcular un nivel de la muestra en dirección de avance
lb[n]
lf[n]
Calcular el último nivel de la muestra
l[n]
Calcular la ganancia del proceso por muestras
G[n]
Calcular el nivel de la muestra de salida
y[n]

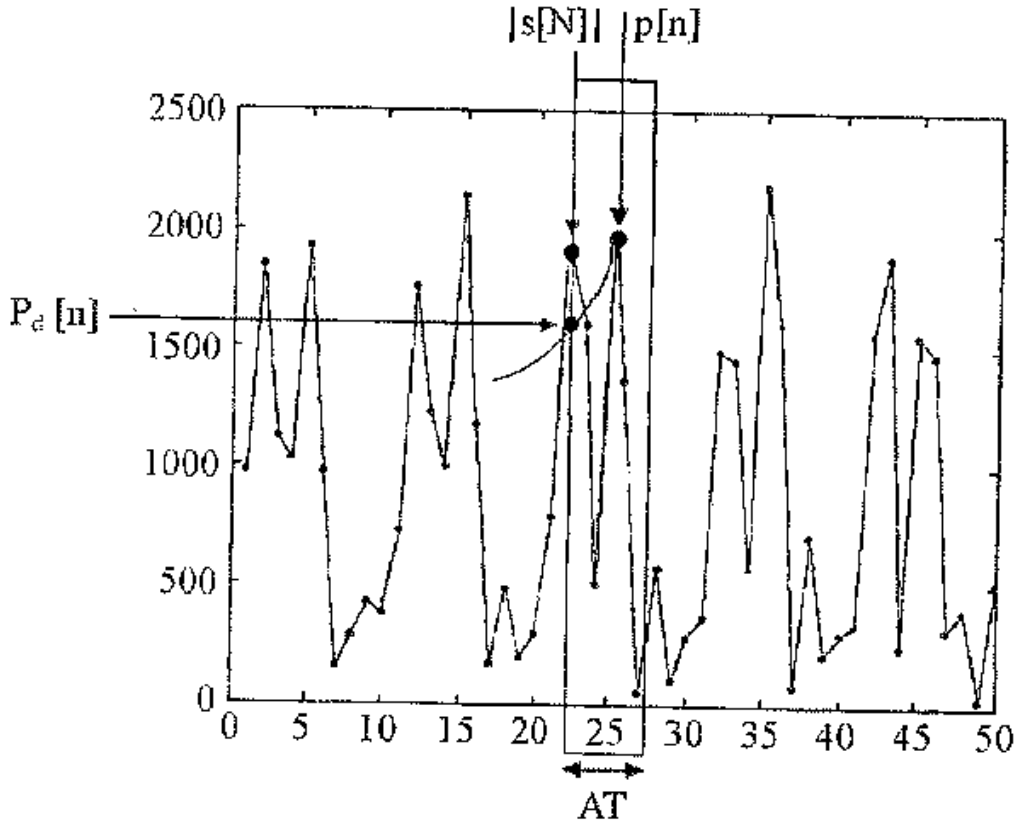
# Fig. 7


30000
25000
20000
15000
10000
5000
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Nivel de
muestra
L máx
L mín

# Fig. 8



# Fig. 9



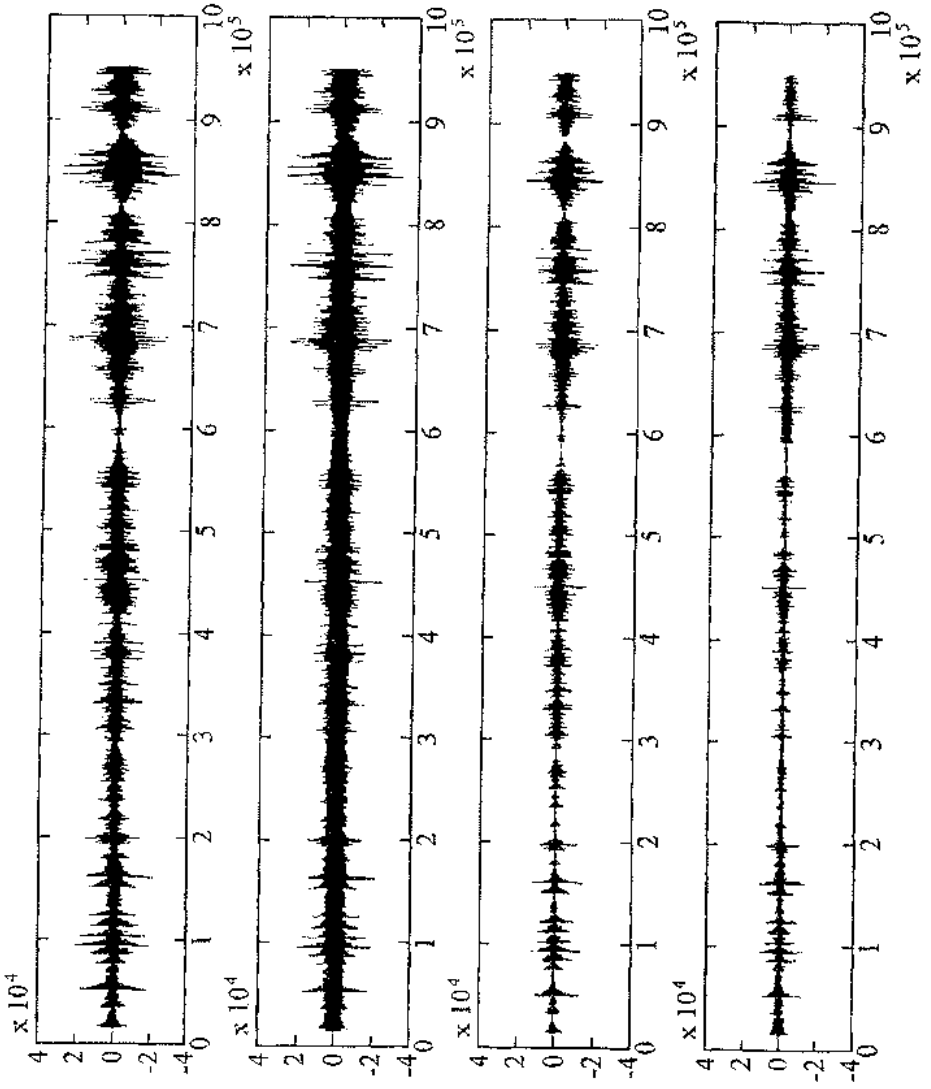



Fig.10A

Fig.10B

Fig.10C

Fig.10D