

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: 2003.10.14	(73) Titular(es): REALNETWORKS ASIA PACIFIC CO., LTD. K1 REIT BUILDING 463 CHUNGJEONG-RO-3-GA SEODAEMUN-GU SEOUL KR
(30) Prioridade(s): 2002.10.14 KR 2002062507	
(43) Data de publicação do pedido: 2005.07.20	
(45) Data e BPI da concessão: 2011.08.24 225/2011	(72) Inventor(es): YOUNG HAN NAM KR SEOP HYEONG PARK KR TAE KYOON JAMWON HANSHIN HA KR YUN HO JEON KR
	(74) Mandatário: ANTÓNIO JOÃO COIMBRA DA CUNHA FERREIRA RUA DAS FLORES, Nº 74, 4º AND 1249-235 LISBOA PT

(54) Epígrafe: **PRÉ-PROCESSAMENTO DE DADOS DIGITAIS ÁUDIO PARA CODIFICADORES/DESCODIFICADORES DE ÁUDIO MÓVEIS**

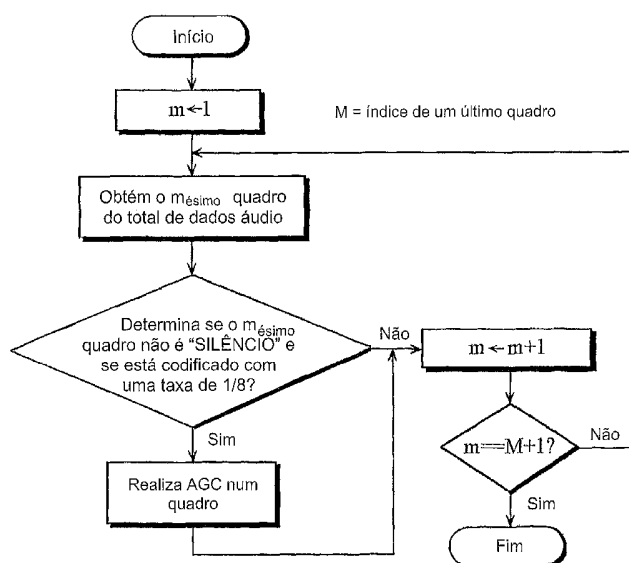
(57) Resumo:

RECENTEMENTE, COM A UTILIZAÇÃO GENERALIZADA DE TELEFONES CELULARES, CADA VEZ MAIS UTILIZADORES OUVEM MÚSICA COM OS SEUS TELEFONES CELULARES E, DESTE MODO, A QUALIDADE DO SOM DA MÚSICA DISPONIBILIZADA ATRAVÉS DOS TELEFONES CELULARES SE TORNA MAIS CRÍTICA. UMA VEZ QUE OS SINAIS DE MÚSICA SÃO CODIFICADOS POR UM MÉTODO DE CODIFICAÇÃO DE VOZ OPTIMIZADO PARA SINAIS DE VOZ HUMANA TAL COMO EVRC (CODIFICAÇÃO DE TAXA VARIÁVEL MELHORADA) NUM SISTEMA DE COMUNICAÇÃO CELULAR, OS SINAIS DE MÚSICA SÃO, MUITAS VEZES, DISTORCIDOS POR ESTES MÉTODOS DE CODIFICAÇÃO E OS OUVINTES REGISTAM PAUSAS NA MÚSICA PROVOCADAS POR ESTES MÉTODOS DE CODIFICAÇÃO OPTIMIZADOS PARA VOZ. PARA MELHORAR A QUALIDADE DO SOM DA MÚSICA, É DISPONIBILIZADO UM MÉTODO PARA PRÉ-PROCESSAR DADOS ÁUDIO A FIM DE EVITAR O PROBLEMA DE PAUSA EM SINAIS DE MÚSICA NUM TELEFONE CELULAR. EM PARTICULAR, É REALIZADO PRÉPROCESSAMENTO AGC (CONTROLO AUTOMÁTICO DE GANHO) AOS DADOS ÁUDIO QUE TENHAM BAIXO INTERVALO DINÂMICO. ATRAVÉS DESTE MÉTODO, O NÚMERO DE PAUSAS NO SINAL DE MÚSICA É REDUZIDO E A QUALIDADE DE SOM DA MÚSICA É MELHORADA.

RESUMO

"Pré-processamento de dados digitais áudio para codificadores/descodificadores de áudio móveis"

Recentemente, com a utilização generalizada de telefones celulares, cada vez mais utilizadores ouvem música com os seus telefones celulares e, deste modo, a qualidade do som da música disponibilizada através dos telefones celulares se torna mais crítica. Uma vez que os sinais de música são codificados por um método de codificação de voz optimizado para sinais de voz humana tal como EVRC (codificação de taxa variável melhorada) num sistema de comunicação celular, os sinais de música são, muitas vezes, distorcidos por estes métodos de codificação e os ouvintes registam pausas na música provocadas por estes métodos de codificação optimizados para voz. Para melhorar a qualidade do som da música, é disponibilizado um método para pré-processar dados áudio a fim de evitar o problema de pausa em sinais de música num telefone celular. Em particular, é realizado pré-processamento AGC (controlo automático de ganho) aos dados áudio que tenham baixo intervalo dinâmico. Através deste método, o número de pausas no sinal de música é reduzido e a qualidade de som da música é melhorada.



DESCRIÇÃO

"Pré-processamento de dados digitais áudio para codificadores/descodificadores de áudio móveis"

CAMPO TÉCNICO

O presente invento refere-se a um método para pré-processar dados áudio a fim de melhorar a qualidade da música descodificada em terminais de recepção tais como telefones móveis e, mais em particular, a um método para pré-processar dados áudio a fim de mitigar uma degradação do sinal de música que pode ser provocada quando os dados áudio são codificados/descodificados num sistema de comunicação sem fios que utiliza codificadores/descodificadores de voz otimizados apenas para sinais de voz humana.

TÉCNICA ANTECEDENTE

A largura de banda de canal de um sistema de comunicação sem fios é muito mais estreita do que a de um sistema de comunicação de telefone convencional de 64 kbps e, deste modo, dados áudio num sistema de comunicação sem fios são comprimidos antes de serem transmitidos. Métodos para compressão de dados áudio num sistema de comunicação sem fios incluem QCELP (predição linear excitada por código QualComm) de IS-95, EVRC (codificação de taxa variável melhorada), VSELP (predição linear excitada por vector soma) de GSM (sistema global para comunicações móveis), PRE-LTP (LPC excitado com impulso regular com predição de horizonte estendido) e ACELP (predição linear excitada por código algébrico). Todos estes métodos listados são baseados em LPC (codificação preditiva linear). Os métodos de compressão áudio baseados em LPC utilizam um modelo otimizado para voz humana e, deste modo, são eficientes para comprimir voz a uma taxa de codificação média ou baixa. Num método de codificação utilizado num sistema sem fios, para utilizar de forma eficiente a largura de banda limitada e diminuir o consumo de energia, os dados áudio são comprimidos e transmitidos apenas quando é detectada voz humana através da utilização da que é designada a função de VAD (detecção de actividade de voz).

Recentemente, ficaram acessíveis vários serviços para disponibilizar música para utilizações em telefones sem fios. Um dos quais é o que é designado "Coloring service" que habilita um assinante a designar uma melodia à sua escolha de modo que quem efectuar chamadas para o assinante poderá ouvir música em vez de um som de toque tradicional enquanto o assinante não atende o telefone. Uma vez que este serviço se tornou muito popular primeiro na Coreia de onde é originário e depois noutros países, a transmissão de dados de música para um telefone celular tem vindo a aumentar. No entanto, como explicado acima, o método de compressão de áudio baseado em LPC é adequado para voz humana que tem componentes de frequência limitados. Quando música ou sinais que têm componentes de frequência na maior parte do intervalo de frequência audível (20 a 20.000 Hz) são processados num codificador/descodificador baseado em LPC convencional e transmitidos através de um telefone celular, ocorre distorção de sinal, o que provoca uma pausa na música ou faz com que o som tenha apenas parte dos componentes de frequência originais.

Existem vários motivos para que a qualidade de som de dados áudio se degrade depois de dados áudio serem comprimidos com a utilização de codificadores/descodificadores áudio baseados em LPC, em especial codificadores/descodificadores EVRC. A degradação da qualidade de som ocorre na seguinte forma.

- i) Perda total de componentes de frequência numa largura de banda de alta-frequência;
- ii) Perda parcial de componentes de frequência numa largura de banda de baixa frequência;
- iii) Pausa intermitente de música

A primeira causa da degradação não pode ser evitada uma vez que os componentes de alta-frequência são removidos com a utilização de um filtro passa-baixo de 4 kHz (ou 3,4 kHz) quando os dados áudio são comprimidos com a utilização de um codificador/descodificador de áudio de largura de banda estreita.

O segundo fenómeno é devido às características intrínsecas dos métodos de compressão áudio baseados em LPC. De acordo com os métodos de compressão baseados em LPC, são obtidos um pitch (frequência fundamental da fonte excitadora) e uma frequência formante de um sinal de entrada e depois um sinal de excitação para minimizar a diferença entre o sinal de entrada e o sinal composto calculado pelo pitch e pela frequência formante do sinal de entrada, são obtidos a partir de um índice de códigos. É difícil extrair um pitch de um sinal de música polifónico, uma vez que é um para cada voz humana. Além disso, o componente formante de música é muito diferente daquele da voz de uma pessoa. Em consequência, é expectável que o sinal de erro de predição para dados de música possa ser muito maior do que no caso de sinal de voz humana e, deste modo, muitos componentes de frequência incluídos nos dados áudio originais são perdidos. Os dois problemas acima, isto é, perda de componentes de alta e baixa frequência são devidos a características inerentes de codificação/descodificação áudio otimizada para sinais de voz e inevitáveis até um certo grau.

As pausas no sinal áudio são provocadas pela taxa de codificação variável utilizada por EVRC. Um codificador EVRC processa os dados áudio com três taxas (designadamente, 1, 1/2 e 1/8). De entre estas taxas, a taxa 1/8 significa que o codificador EVRC determina que o sinal de entrada é um ruído e não um sinal de voz. Devido a que sons de um instrumento de percussão, tal como um tambor, incluem componentes de espectro que tendem a ser percebidos como ruídos pelos codificadores/descodificadores áudio, a música que inclui este tipo de sons é frequentemente pausada. Do mesmo modo, codificadores/descodificadores áudio consideram sons com baixas amplitudes como ruídos, o que também degrada a qualidade de som.

A WO 02/065457 apresenta um sistema de codificação de voz com um classificador de música. Um codificador está previsto para receber um sinal de entrada e disponibilizar uma sequência de bits com base numa codificação de voz de uma porção do sinal de entrada. O codificador disponibiliza uma classificação da entrada como ruído, voz e música. O classificador de música analisa ou determina propriedades de

sinal do sinal de entrada. O classificador de música compara as propriedades de sinal com limiares para determinar a classificação do sinal de entrada.

A US 5 742 734 apresenta um método e um aparelho para determinar a taxa de codificação de voz num codificador de voz de taxa variável.

APRESENTAÇÃO DO INVENTO

O presente invento disponibiliza um método para pré-processamento de sinais áudio que se destinam a serem transmitidos através de sistemas sem fios a fim de melhorar a qualidade de som de dados áudio recebidos num terminal de recepção de um assinante. O presente invento disponibiliza um método para mitigar a deterioração da qualidade de som de música que ocorre quando o sinal de música é processado por codificadores/descodificadores otimizados para voz humana, tais como codificadores/descodificadores EVRC. Outro objecto do presente invento é proporcionar um método e um sistema para pré-processar dados áudio de uma forma que não interfere com o sistema de comunicação sem fios existente. Em consequência, o método de pré-processamento do presente invento é útil por o mesmo poder ser utilizado sem modificação de um sistema existente. O presente invento também pode ser aplicado de uma forma semelhante a outros codificadores/descodificadores otimizados para voz humana diferentes de EVRC.

A fim de alcançar o objecto acima, o presente invento disponibiliza um método e um sistema para pré-processar dados áudio que se destinam a serem processados por um codificador/descodificador com uma taxa de codificação variável de acordo com as reivindicações independentes 1 e 3, respectivamente.

BREVE DESCRIÇÃO DOS DESENHOS

O objecto e as funcionalidades acima do presente invento ficarão mais evidentes a partir da seguinte descrição das concretizações preferidas vistas em conjunto com os desenhos em anexo.

A Fig. 1 é um diagrama de blocos de um codificador EVRC.

A Fig. 2A é um gráfico que mostra um sinal residual de quadro para um sinal que tem um componente de frequência dominante.

A Fig. 2B é um gráfico que mostra um sinal residual de quadro para um sinal que tem uma variedade de frequências.

A Fig. 3A é um gráfico que mostra a autocorrelação de resíduos para um sinal que tem um componente de frequência dominante.

A Fig. 3B é um gráfico que mostra a autocorrelação de resíduos para um sinal que tem uma variedade de frequências.

A Fig. 4 é um fluxograma para realizar pré-processamento AGC (controlo automático de ganho) de acordo com o presente invento.

A Fig. 5 é um fluxograma para realizar pré-processamento AGC de quadro selectivo de acordo com o presente invento.

A Fig. 6 é um diagrama de blocos para realizar AGC de acordo com o presente invento.

A Fig. 7 é um gráfico que mostra um sinal áudio amostrado e o seu nível de sinal.

A Fig. 8 é um gráfico para explicar o cálculo de um nível de sinal no sentido de avanço de acordo com o presente invento.

A Fig. 9 é um gráfico para explicar o cálculo de um nível de sinal no sentido do retorno de acordo com o presente invento.

As Figs. 10A a 10D são gráficos que mostram resultados de pré-processamento AGC.

MODOS DE CONCRETIZAÇÃO DO INVENTO

Como uma forma de resolver o problema de pausas intermitentes, o presente invento disponibiliza um método para pré-processar dados áudio antes dos mesmos serem sujeitos a codificação áudio. Certo tipo de sons (tais como os de um instrumento de percussão) incluem componentes de espectro que tendem a ser percebidos como ruído pelos codificadores/descodificadores de áudio otimizados para voz humana (tais como codificadores/descodificadores para sistemas sem fios) e os codificadores/descodificadores áudio consideram as porções de música que têm baixas amplitudes como ruído. Este fenómeno é evidente habitualmente em todos os sistemas que utilizam DTX (transmissão descontínua) baseada em VAD (detecção de actividade de voz) tal como GSM (sistema global para comunicações móveis). No caso de EVRC, se os dados forem determinados como ruído, os dados são codificados com uma taxa de 1/8 de entre as três taxas pré-determinadas de 1/8, 1/2 e 1. Se for decidido pelo sistema de codificação que os dados de música são ruído, os dados transmitidos basicamente não podem ser ouvidos no terminal receptor, o que deste modo deteriora de forma grave a qualidade do som.

Este problema pode ser resolvido por pré-processamento de dados áudio de modo que as taxas de codificação do codificador/descodificador EVRC podem ser definidas como 1 (e não 1/8) para quadros de dados de música. De acordo com o presente invento, a taxa de codificação de sinais de música pode ser aumentada através de pré-processamento, e, por conseguinte, as pausas de música no terminal de recepção provocadas por EVRC são reduzidas. Apesar do presente invento ser explicado em relação a um codificador/descodificador EVRC, alguém especializado na técnica poderia aplicar o presente invento a outro sistema de compressão com a utilização de taxa de codificação variável, em especial um codificador/descodificador otimizado para voz humana (tal como um codificador/descodificador de áudio para transmissão sem fios).

Com referência à Fig. 1, será explicado RDA (algoritmo de decisão de taxa) de EVRC. EVRC será explicado como um

exemplo de um sistema de compressão que utiliza uma taxa de codificação variável para comprimir dados que se destinam a serem transmitidos através de rede sem fios onde o presente invento pode ser aplicado. A compreensão do algoritmo de decisão da taxa do codificador/descodificador convencional utilizado num sistema existente é importante devido ao presente invento ser baseado numa ideia de que, num codificador/descodificador convencional, alguns dados de música podem ser codificados a uma taxa de dados que é demasiado baixa para dados de música (embora talvez adequada para dados de voz) e que através do aumento da taxa de dados para os dados de música, a qualidade da música pode ser aumentada depois da codificação, transmissão e descodificação.

A Fig. 1 é um diagrama de blocos de alto nível de um codificador EVRC. Na Fig. 1, uma entrada pode ser um sinal áudio PCM (modulação por código de impulsos) de 16 bit, 8k, e uma saída codificada pode ser dados digitais cuja dimensão pode ser de 171 bits (quando a taxa de codificação é 1), 80 bits (quando a taxa de codificação é 1/2), 16 bits (quando a taxa de codificação é 1/8) ou 0 bits (em branco) por quadro de acordo com a taxa de codificação decidida pelo RDA. O áudio PCM de 16 bit, 8k, está ligado ao codificador EVRC em unidades de quadros onde cada quadro tem 160 amostras (o que corresponde a 20 ms). O sinal de entrada $s[n]$ (isto é, um $n^{\text{ésimo}}$ sinal de quadro de entrada) está ligado a um bloco de supressão de ruído 110, que verifica o sinal de quadro de entrada $s[n]$. No caso do sinal de quadro de entrada ser considerado ruído no bloco de supressão de ruído 160, o mesmo multiplica um ganho inferior a 1 com o sinal e, deste modo, suprime o sinal de quadro de entrada. E então, $s'[n]$ (isto é, um sinal que passou através do bloco 110) é ligado a um bloco RDA 120, que selecciona uma das taxas do conjunto pré-definido de taxas de codificação (1, 1/2, 1/8 e branco na concretização explicada aqui). Um bloco de codificação 130 extrai parâmetros adequados do sinal de acordo com a taxa de codificação seleccionada pelo bloco PDA 120, e um bloco de empacotamento de bit 140 junta os parâmetros extraídos para conformar um formato de saída pré-determinado.

Como mostrado na tabela seguinte, a saída codificada pode ter 171, 80, 16 ou 0 bits por quadro em função da taxa de codificação seleccionada pelo RDA.

[Tabela 1]

Tipo de quadro	Bits por quadro
Quadro com taxa de codificação 1	171
Quadro com taxa de codificação 1/2	80
Quadro com taxa de codificação 1/8	16
Branco	0

O bloco RDA 120 divide $s'[n]$ em duas larguras de banda ($f(1)$ de 0,3 a 2,0 kHz e $f(2)$ de 2,0 a 4,0 kHz) através da utilização de um filtro passa banda e selecciona a taxa de codificação para cada largura de banda ao comparar um valor de energia de cada largura de banda com um limiar de decisão de taxa decidido por uma estimativa de ruído de fundo ("BNE"). As seguintes equações são utilizadas para calcular os dois limiares para $f(1)$ e $f(2)$.

$$T_1 = k_1(\text{SNR}_{f(i)}(m-1)) B_{f(i)}(m-1) \quad \text{Eq. (1a)}$$

$$T_2 = k_2(\text{SNR}_{f(i)}(m-1)) B_{f(i)}(m-1) \quad \text{Eq. (1b)}$$

Em que k_1 e k_2 são factores de escala de limiares, que são funções da RSR (relação sinal ruído) e aumentam com o aumento da RSR. Além disso, $B_{f(i)}(m-1)$ é BNE (estimativa de ruído de fundo) para a banda $f(i)$ no quadro $(m-1)$ ésimo. Como descrito nas equações acima, o limiar de decisão de taxa é decidido pela multiplicação do coeficiente de escala por BNE e, deste modo, proporcional a BNE.

Por outro lado, a energia de banda pode ser decidida por coeficientes de correlação $0^{\text{ésimo}}$ a $16^{\text{ésimo}}$ de dados áudio para cada largura de banda de frequência.

$$BE_{f(i)} = R_w(0) R_{f(i)}(0) + 2.0 \sum_{k=1}^{L_i-1} R_w(k) R_{f(i)}(k) \quad \text{Eq. (2)}$$

Em que $BE_{f(i)}$ é um valor de energia para a largura de banda de frequência i ésima ($i=1,2$), $R_w(k)$ é uma função de coeficientes de autocorrelação de dados áudio de entrada e $R_{f(i)}(k)$ é um

coeficiente de autocorrelação de uma resposta impulsiva num filtro passa banda. L_h é uma constante de 17.

Então, a actualização de um ruído estimado ($B_{f(i)}(m-1)$) será explicada. O ruído estimado ($B_{f(i)}(m)$) para a i ésima banda de frequência (ou $f(i)$) do m ésimo quadro é decidido pelo ruído estimado ($B_{f(i)}(m-1)$) para $f(i)$ do $(m-1)$ ésimo quadro, pela energia de banda alisada ($E_{f(i)}^{SM}(m)$) para $f(i)$ do m ésimo quadro e uma relação sinal ruído ($RSR_{f(i)}(m-1)$) para $f(i)$ de $(m-1)$ ésimo quadro, que está representada no pseudo código.

```

if ( $\beta < 0.30$  para 8 ou mais quadros consecutivos)
   $B_{f(i)}(m) = \min\{E_{f(i)}^{sm}(m), 80954304, \max\{1.03B_{f(i)}(m-1), B_{f(i)}(m-1)+1\}\}$ 
else{
  if ( $SNR_{f(i)}(m-1) > 3$ )
     $B_{f(i)}(m) = \min\{E_{f(i)}^{SM}(m), 80954304, \max\{1.0054B_{f(i)}(m-1), B_{f(i)}(m-1)+1\}\}$ 
  else
     $B_{f(i)}(m) = \min\{E_{f(i)}^{SM}(m), 80954304, B_{f(i)}(m-1)\}$ 
  }
  if ( $B_{f(i)}(m)$  baixo ruído (i))
     $B_{f(i)}(m) =$  baixo ruído (i)
}

```

Como descrito acima, se o valor de β , um ganho preditivo de horizonte estendido (como definir β será explicado mais tarde) for inferior a 0,3 para mais do que 8 quadros, o valor mais pequeno de entre (i) a energia de banda alisada, (ii) 1,03 vezes da BNE do quadro anterior e (iii) um valor máximo pré-determinado de uma BNE (80954304 no acima) é seleccionado como a BNE. Caso contrário (se o valor de β não for inferior a 0,3 em qualquer dos 8 quadros consecutivos), se a RSR do quadro anterior for maior do que 3, o valor mais pequeno de entre (i) a energia de banda alisada, (ii) 1,00547 multiplicado pela BNE do quadro anterior e (iii) um valor máximo pré-determinado de uma BNE é seleccionado como a BNE para este quadro. Se a RSR do quadro anterior não for maior do que 3, o valor mais pequeno de entre (i) a energia de banda alisada, (ii) a BNE do quadro anterior e o valor máximo pré-determinado de BNE é seleccionado como a BNE para este quadro.

Por conseguinte, no caso de um sinal áudio, a BNE tende a aumentar à medida que o tempo passa, por exemplo, por 1,03 vezes ou por 1,00547 vezes de quadro para quadro e diminui apenas quando a BNE se torna maior do que a energia de banda alisada. Em consequência, se a energia de banda alisada é mantida dentro de um intervalo relativamente pequeno, a BNE aumenta à medida que o tempo passa e, por esse motivo, o valor do limiar de decisão de taxa aumenta (ver Eq. (1)). Em consequência, torna-se mais provável que um quadro seja codificado com uma taxa de 1/8. Por outras palavras, se o sinal de música for exibido durante um longo período de tempo, as pausas tendem a ocorrer com maior frequência.

O ganho de predição de horizonte estendido (β) é definido pela autocorrelação de resíduos como se segue.

$$\beta = \max\{0, \min\{1, \frac{R_{\max}}{R_{\varepsilon}(0)}\}\} \quad \text{Eq. (3)}$$

Em que ε é um sinal residual de predição, R_{\max} é um valor máximo dos coeficientes de autocorrelação do sinal residual de predição e $R_{\varepsilon}(0)$ é um 0ésimo coeficiente de uma função de autocorrelação do sinal residual de predição.

De acordo com a equação acima, no caso de um sinal monofónico ou de um sinal de voz onde existe um pitch dominante, o valor de β seria maior, mas no caso de música que inclui vários pitches, o valor de β seria menor.

O sinal residual de predição (ε) é definido como se segue:

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

em que $s'[n]$ é um sinal áudio pré-processado pelo bloco de supressão de ruído 110 e $a_i[k]$ é um coeficiente LPC interpolado do k ésimo segmento de um quadro do momento.

Isto é, o sinal residual de predição é uma diferença entre um sinal reconstruído pelos coeficientes LPC e um sinal original.

O sinal residual de quadro parece regular no caso em que existe um componente de frequência dominante no quadro (ver Fig. 2A), enquanto que o mesmo é irregular no caso em que existem vários componentes de frequência no quadro (ver Fig. 2B). Em consequência, no primeiro caso, um valor de pico máximo regulado de coeficientes de autocorrelação (isto é ganho de predição de horizonte estendido β) seria um valor maior (tal como $\beta = 0,6792$, ver Fig. 3A), enquanto que no último caso, seria um valor menor (tal como $\beta = 0,2616$, ver Fig. 3B). Nestas Figs. 3A e 3B, os coeficientes de autocorrelação são normalizados por $R(0)$. Nas Figs. 2A e 2B, o eixo x representa números amostrados e o eixo y representa a amplitude de sinal residual onde os números no gráfico são valores normalizados que dependem da necessidade do sistema (por exemplo, quantos bits são utilizados para representar o valor), o que se aplica a outros gráficos neste pedido (tais como Figs. 7 a 10).

Será agora explicado como definir a taxa de codificação. Para cada uma das duas bandas de frequência, se a energia de banda for maior do que os dois valores de limiar, a taxa de codificação é 1, se a energia de banda estiver entre os dois valores de limiar, a taxa de codificação é 1/2 e se a energia de banda for inferior aos dois valores de limiar, a taxa de codificação é 1/8. Depois das taxas de codificação estarem decididas para duas bandas de frequência, a maior das duas taxas de codificação decididas para as bandas de frequência é seleccionada como uma taxa de codificação para aquele quadro. Num sistema real, codificar a uma taxa de 1/8 pode significar que o sinal relevante é definido como ruído e são transmitidos muito poucos dados; codificar a uma taxa de 1 pode significar que o sinal é definido como voz humana válida; e codificar a uma taxa de 1/2 acontece para um curto intervalo durante a transmissão entre 1/8 e 1.

Até agora, foi explicado como é definida a taxa de codificação num codificador/descodificar EVRC, que é um exemplo de um sistema de codificação de taxa variável onde o presente invento pode ser aplicado. Do antecedente, pode ser compreendido que a taxa de codificação de um quadro pode ser maximizada a 1 tanto quanto possível por (i) incremento da

energia de banda e/ou (ii) diminuição do valor de limiar para a definição da taxa de codificação.

O presente invento utiliza um método AGC (controlo automático de ganho) para aumentar a energia de banda. AGC é um método para ajustar o ganho de sinais actuais por sinais de predição durante um certo intervalo (intervalo ATTACK). Por exemplo, se for exibida música em altifalantes com diferentes intervalos dinâmicos, a mesma pode ser processada de forma adequada sem AGC (sem AGC, alguns altifalantes funcionam na região de saturação). Por conseguinte, é necessário realizar pré-processamento AGC com base na característica do dispositivo que gera o som, tal com um altifalante, um auscultador ou um telefone celular.

No caso de um telefone celular, embora seja ideal para medir o intervalo de dinâmica do telefone celular e realizar AGC a fim de garantir melhor qualidade de som, é impossível conceber AGC optimizado para todos os telefones celulares devido às características de um telefone celular variarem em função do fabricante e também de cada modelo em particular. Por conseguinte, é necessário conceber um AGC de forma geral aplicável a todos os telefones celulares.

A Fig. 4 é um fluxograma de alto nível para realizar pré-processamento AGC de acordo com uma concretização do presente invento. No início, dados áudio são obtidos no passo 410 e depois os dados áudio são classificados com base na característica dos dados áudio no passo 420. Os dados áudio são processados em diferentes formas em função da classificação devido, para certo tipo de dados áudio, a ser preferível melhorar a energia de todos os quadros, enquanto noutros casos, funciona melhor melhorar apenas a energia de banda de quadros que são codificados com uma taxa de quadro baixa no codificador de taxa variável (tal como EVRC). A parte direita 440 do fluxograma mostra a melhoria de energia de todos os quadros. No caso de música clássica ou de dados de áudio monofónico com um pitch, é preferível que a parte direita 440 do fluxograma seja realizada. A parte esquerda 430 do fluxograma mostra a melhoria da energia de banda destes quadros que são codificados com uma taxa de quadro baixa. No caso de dados de áudio polifónico, tal como música

rock, é preferível que a parte esquerda 430 do fluxograma seja realizada.

A Fig. 5 é um fluxograma para AGC de quadro selectivo para pré-processar quadros que poderiam ser codificados com taxa baixa sem o pré-processamento. AGC é realizado em diferentes formas em função da energia de quadros de sinais de música. O intervalo em que a energia de quadros dos dados áudio (antes da codificação EVRC) é baixa (isto é, inferior a 1.000) é definido como um intervalo de "SILÊNCIO" onde não é realizado processamento. Para os quadros que não estão no intervalo de "SILÊNCIO", é realizada codificação EVRC para detectar a taxa de codificação para cada quadro. Para estes intervalos onde os quadros com taxa de codificação de 1/8 ocorrem com frequência (o que significa que estes intervalos são considerados um ruído pelo codificador EVRC), a energia de banda dos quadros é localmente aumentada. Quando se melhora a energia para certos quadros, poderia ser necessária a interpolação com outros quadros (a este respeito, o que é referido como "interpolação de envelope" será explicado mais tarde) para evitar descontinuidade de amplitude de som entre os quadros melhorados e os quadros vizinhos não melhorados.

A Fig. 6 é um diagrama de blocos para AGC de acordo com uma concretização do presente invento. Nesta concretização, AGC é um processo para ajustar o nível de sinal da amostra corrente com num controlo de ganho decidido de um conjunto de valores de amostra numa janela de pesquisa. No início, um "nível de sinal no sentido de avanço" $I_f[n]$ e um "nível de sinal no sentido de retorno" $l_b[n]$ são calculados com a utilização de sinal áudio amostrado $s[n]$ de uma forma explicada mais tarde, e a partir dos quais é calculado um "nível de sinal final" $l[n]$. Depois de $l[n]$ ser calculado, é calculado o ganho de processamento por amostra ($G[n]$) com a utilização de $l[n]$ e depois é obtida a saída $y[n]$ por multiplicação de $G[n]$ e $s[n]$.

No seguinte, as funções dos blocos na Fig. 6 serão descritas com maior detalhe.

A Fig. 7 mostra um nível de sinal exemplificativo ($l[n]$) calculado a partir do sinal áudio amostrado ($s[n]$). O

envelope do nível de sinal $l[n]$ varia em função do modo como são processados os sinais através da utilização da supressão exponencial no sentido de avanço ("ATTACK") e da supressão exponencial no sentido de retorno ("RELEASE"). Na Fig. 7, L_{\max} e L_{\min} referem-se aos valores máximo e mínimo do sinal de saída depois do pré-processamento AGC.

Um nível de sinal no instante n é obtido através do cálculo de níveis de sinal no sentido de avanço (para realizar RELEASE) e do cálculo de níveis de sinal no sentido de retorno (para realizar ATTACK). A constante de tempo de uma "função exponencial" que caracteriza a supressão exponencial será referida como "tempo de RELEASE" no sentido de avanço e como "tempo de ATTACK" no sentido de retorno. O tempo de ATTACK é um tempo necessário para um novo sinal de saída atingir uma amplitude de saída adequada. Por exemplo, se uma amplitude de um sinal de entrada diminui abruptamente de 30 dB, o tempo de ATTACK é um tempo para um sinal de saída diminuir de forma correspondente (por 30 dB). O tempo de RELEASE é um tempo para atingir um nível de amplitude adequado no fim de um nível de saída existente. Isto é, o tempo de ATTACK é um período para um início de um impulso atingir uma amplitude de saída desejada visto que o tempo de RELEASE é um período para uma extremidade de um impulso atingir uma amplitude de saída desejada.

No seguinte, será descrito como calcular um nível de sinal no sentido de avanço e um nível de sinal no sentido de retorno com referência às Figs. 8 e 9.

Com referência à Fig. 8, um nível de sinal no sentido de avanço é calculado pelos passos seguintes.

No primeiro passo, um valor de pico actual e um índice de pico actual são inicializados (colocados a 0) e um nível de sinal no sentido de avanço ($l_f[n]$) é inicializado como $|s[n]|$, um valor absoluto de $s[n]$.

No segundo passo, o valor de pico actual e o índice de pico actual são actualizados. Se $|s[n]|$ for maior do que o valor de pico actual ($p[n]$), $p[n]$ é actualizado para $|s[n]|$ e o índice de pico actual ($i_p[n]$) é actualizado para n (como

mostrado no pseudo código seguinte) if
 $|s[n]| > p[n]$ { $p[n] = |s[n]|$; $i_p[n] = n$ }

No terceiro passo, é calculado um valor de pico actual suprimido. O valor de pico actual suprimido $p_d[n]$ é decidido por redução exponencial do valor de $p[n]$ de acordo com a passagem de tempo como se segue.

$$p_d[n] = p[n] * \exp(-TD/RT) \quad \text{Eq. (5)}$$

$$TD = n - i_p[n]$$

Em que RT representa o tempo de RELEASE.

No quarto passo, um valor maior do que $p_d[n]$ e $|s[n]|$ é definido como um nível de sinal no sentido de avanço, como se segue.

$$l_f[n] = \max(p_d[n], |s[n]|) \quad \text{Eq. (6)}$$

A seguir, os segundo a quarto passos acima são repetidos para se obter um nível de sinal no sentido de avanço ($l_f[n]$) à medida que n aumenta por um de cada vez.

Com referência à Fig. 9, um nível de sinal no sentido do retorno é calculado pelos passos seguintes.

No primeiro passo, um valor de pico actual é inicializado a 0, um índice de pico actual é inicializado a AT e um nível de sinal no sentido de retorno ($l_b[n]$) é inicializado a $|s[n]|$, um valor absoluto de $s[n]$.

No segundo passo, o valor de pico actual e o índice de pico actual são actualizados. Um valor máximo de $s[n]$ na janela de tempo de n a $n + AT$ é detectado e o valor de pico actual $p(n)$ é actualizado como o valor máximo detectado. Também $i_p[n]$ é actualizado como o índice de tempo para o valor máximo.

$$p[n] = \max(\{|s[]|\}) \quad \text{Eq. (7)}$$

$i_p[n]$ = (um índice de $s[]$, onde $|s[]|$ tem o seu valor máximo)

Em que o índice de $s[]$ pode ter valores de n a $n + AT$.

No terceiro passo, um valor de pico actual suprimido é calculado como se segue.

$$\begin{aligned} p_d[n] &= p[n] * \exp(-TD/AT) && \text{Eq.(8)} \\ TD &= i_p[n] - n \end{aligned}$$

Em que AT representa o tempo de ATTACK.

No quarto passo, um valor maior do que $p_d[n]$ e $|s[n]|$ é definido como um nível de sinal no sentido de retorno.

$$l_b[n] = \max(p_d[n], |s[n]|) \quad \text{Eq. (9)}$$

A seguir, os segundo a quarto passos acima são repetidos para se obter um nível de sinal no sentido de retorno ($l_b[n]$) à medida que n aumenta por um de cada vez.

1 O nível de sinal final ($l[n]$) é definido como um valor máximo do nível de sinal no sentido de avanço e do nível de sinal no sentido de retorno para cada índice de tempo.

$$l[n] = \max(l_f[n], l_b[n]) \text{ for } t = 0, \dots, t_{\max} \quad \text{Eq. (10)}$$

Em que t_{\max} é um índice de tempo máximo.

Tempo de ATTACK/tempo de RELEASE está relacionado com a qualidade/característica do som. Em consequência, quando se calculam os níveis de sinal, é necessário definir tempo de ATTACK e tempo de RELEASE de forma adequada para assim se obter som optimizado para a característica de um meio. Se a soma do tempo de ATTACK e do tempo de RELEASE for demasiado pequena (isto é se a soma for inferior a 20 ms), pode ser ouvida uma distorção na forma de vibração com uma frequência de $1000/(\text{tempo de ATTACK} + \text{tempo de RELEASE})$ por um utilizador de telefone celular. Por exemplo, se o tempo de ATTACK e o tempo de RELEASE forem 5 ms cada, pode ser ouvida uma distorção de vibração com uma frequência de 100 Hz. Por conseguinte, é necessário definir a soma do tempo de ATTACK e do tempo de RELEASE maior do que 30 ms para evitar distorção de vibração.

Por exemplo, se ATTACK for baixo e RELEASE for rápido, poderia ser obtido som com intervalo dinâmico mais amplo. Quando o tempo de RELEASE é longo, o componente de alta-frequência do sinal de saída é suprimido o que resulta num som de sinal lento. No entanto, se o tempo de RELEASE ficar muito rápido (significando "rápido" a este respeito poder variar em função da característica da música), no sinal de saída processado por AGC segue-se o componente de baixa frequência da forma de onda de entrada. Neste caso, o componente fundamental do sinal é suprimido ou pode até ser substituído por uma certa distorção harmónica (o componente fundamental significa o componente de frequência mais importante que uma pessoa pode ouvir, o que é o mesmo que um pitch). À medida que os tempos de ATTACK e de RELEASE se tornam maiores, as pausas são bem evitadas mas o som torna-se lento (perda de frequência alta). Em consequência, existe um compromisso entre a qualidade do som e o número de pausas.

Para enfatizar o efeito de um instrumento de percussão, tal como um tambor, o tempo de ATTACK deverá ser alongado. No entanto, no caso da voz de uma pessoa, o tempo de ATTACK encurtado deveria ajudar a evitar uma diminuição desnecessária do ganho na porção inicial. É importante decidir o tempo de ATTACK e o tempo de RELEASE de forma adequada para garantir a qualidade de som no processamento AGC, e os mesmos são decididos tendo em consideração a característica da música.

O método de pré-processamento do presente invento não envolve cálculos muito complicados e pode ser realizado com um atraso muito curto (na ordem dos tempos de ATTACK e de RELEASE) e, deste modo durante a emissão de um programa de música, é possível quase pré-processar em tempo real.

Quanto a que quadros (ou intervalos) processar com a utilização de AGC de acordo com o presente invento, é preferível processar intervalos tanto com alta como com baixa amplitude (comparado com um certo padrão). Quando dados áudio com um intervalo dinâmico amplo são codificados e transmitidos num sistema de comunicação sem fios e exibidos num telefone celular, a qualidade de som torna-se degradada devido ao som com baixas amplitudes tender a não ser ouvido.

Deste modo, para quadros deste tipo com baixa amplitude, a amplitude deverá ser aumentada para melhor qualidade de sinal. E, no caso de intervalo (quadros) com amplitudes elevadas, a amplitude deverá ser reduzida para se evitar a saturação dos sons exibidos. Para se alcançar os dois objectivos, numa concretização do presente invento, são definidos dois limites (L_{\min} e L_{\max}) e então os intervalos, em que os níveis de sinal são inferiores a L_{\min} ou superiores a L_{\max} , são processados.

Como explicado acima, para evitar a alteração súbita na amplitude entre os intervalos processados (por AGC) e não processados, é necessário ajustar o controlo de ganho de forma adequada para evitar variação brusca na amplitude. Também, depois do AGC, o nível máximo não pode exceder o valor limite máximo (L_{\max}) e, por conseguinte, sem alisamento do valor de ganho, o envelope de sinais de música pode ser fixado no valor limite máximo. Se o envelope for fixado no valor limite máximo, a qualidade de sinal de intervalos processados poderá ser diferente daquela dos intervalos não processados.

Considerando o acima, o processamento de ganho por cada um dos sinais amostrados ($G[n]$) é decidido pela equação seguinte.

$$G[n] = c * (L / I[n]) + (1-c) \quad \text{Eq. (11)}$$

Em que c é um coeficiente de ganho, que tem um valor entre 0 e 1. E, L é definido como L_{\min} ou L_{\max} em função da característica do sinal nos intervalos a processar.

O sinal processado ($s'[n]$) é decidido por uma multiplicação do sinal antes de AGC ($s[n]$) e do processamento de ganho.

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

Das equações acima (Eq. 11 e Eq. 12) sabe-se que à medida que c se torna próximo de 1, o envelope de saída poderá ser fixo ao valor limite e à medida que c se torna próximo de 0, o envelope do sinal resultante depois de AGC

(com a utilização do ganho na equação acima) deverá ficar semelhante ao envelope de entrada.

Através da utilização do método explicado acima, a taxa de codificação de sinais de música pode ser melhorada e, desse modo, o problema de pausa de música originado por EVRC pode ser suficientemente melhorado.

Resultados experimentais relativos ao método explicado acima serão explicados. Na experiência foram utilizados sinais de música monofônicos amostrados a 16 bit, 8 kHz, com qualidade de CD.

As Figs. 10A a 10D mostram a comparação entre os sinais codificados no caso de utilização de pré-processamento AGC do presente invento e no caso da não utilização de pré-processamento AGC. Nas Figs. 10A a 10D, o eixo horizontal é um eixo de tempo e o eixo vertical representa uma amplitude de sinal. A Fig. 10A mostra o sinal original, a Fig. 10B mostra o sinal pré-processado AGC, a Fig. 10C mostra o sinal codificado EVRC a partir dos sinais originais e a Fig. 10D mostra um sinal codificado EVRC a partir dos sinais pré-processados AGC. No sinal que tem intervalo dinâmico amplo como mostrado na Fig. 10A, tendem a ocorrer mais pausas, em especial durante o período de baixa amplitude que poderá ser considerado ruído. Na Fig. 10C, pode-se verificar que o sinal com baixas amplitudes poderá não ser ouvido. O sinal original é pré-processado por AGC com a utilização dos parâmetros da Tabela 2, e o sinal pré-processado é mostrado na Fig. 10B. Depois da codificação/descodificação EVRC, o sinal pré-processado por AGC torna-se um na Fig. 10D. Como mostrado na Fig. 10D, o pré-processamento AGC melhora a porção de sinal que tem baixa amplitude de modo que depois da codificação/descodificação EVRC o sinal pode não ser pausado. Como mostrado na Tabela 3, através de pré-processamento AGC, o número dos quadros codificados com uma taxa de codificação de 1/8 diminui de 356 para 139.

[Tabela 2]

Número de amostras ATTACK	160
Número de amostras RELEASE	2000
Valor limite mínimo	5000
Valor limite máximo	30000
Coefficiente de alisamento de ganho	0,5

Tabela 3

	Sinais originais	Sinais pré-processados AGC
Número de quadros com uma taxa de codificação de 1/8	356	139

O teste MOS (resultado médio da opinião) aplicado a um grupo de teste de 11 pessoas entre os 20 e 30 anos foi realizado para a comparação entre a música original e a música pré-processada pelo algoritmo de pré-processamento de AGC sugerido. Foram utilizados para o teste telefones celulares Samsung Anycall™. Sinais de música não processados e pré-processados foram codificados e disponibilizados para um telefone celular numa sequência aleatória e avaliados pelo grupo de testes através da utilização de um esquema de pontuação com cinco níveis como se segue:

(1) mau (2) fraco (3) razoável (4) bom (5) excelente

Três canções foram utilizadas para o teste e a Tabela 4 mostra o resultado da experiência. De acordo com o resultado do teste, através do pré-processamento de AGC, a pontuação média para as canções sobe de 3,000 para 3,273, de 1,727 para 2,455 e de 2,091 para 2,727.

Tabela 4

Título das canções (Compositor)	Género de canções	Pontuação média para as canções originais	Pontuação média para as canções pré-processadas
Girl's Prayer (Badarczevska)	Piano solo	3,000	3,273
Sonata Patética Op13 (Beethoven)	Piano solo	1,727	2,455
Quinta sinfonia (Fate) (Beethoven)	Sinfonia	2,091	2,727

Numa concretização do invento, o telefone convencional e o telefone sem fios podem conter um sistema para disponibilizar um sinal de música. Neste caso, um ID de quem chama é detectado no sistema para processamento do sinal de música. Num sistema telefónico convencional, é utilizado um sinal de voz não comprimido com 8 kHz de largura de banda e, deste modo, se for transmitida música amostrada de 8 kHz/8bit/a-law (método europeu de quantificação/compressão), pode ser ouvida música de alta qualidade sem distorção de sinal. Numa concretização do invento, um sistema para disponibilizar sinal de música para o terminal de utilizador determina se um pedido para música foi originado por um assinante de um telefone convencional ou de um telefone sem fios, através da utilização de um ID de quem chama. No primeiro caso, o sistema transmite o sinal de música original e no último caso, o sistema transmite música pré-processada por AGC.

Será evidente para quem for especializado na técnica que o método de pré-processamento do presente invento pode ser implementado através da utilização quer de suporte lógico quer de um suporte físico dedicado. Também, numa concretização do invento é utilizado o sistema VoiceXML para disponibilizar música aos assinantes, onde os conteúdos de áudio podem ser alterados com frequência. Num sistema deste tipo, o pré-processamento AGC do presente invento pode ser realizado a pedido. Para realizar isto, uma etiqueta sem norma, tal como <áudio src="xx.wav" tipo="music/classical/">, pode ser definida para determinar se é para realizar pré-processamento ou que tipo de pré-processamento realizar.

APLICABILIDADE INDUSTRIAL

O pedido do presente invento inclui qualquer serviço sem fios que disponibilize música ou outro som de voz não humana através de uma rede sem fios (isto é, através da utilização de um codificador/descodificador para um sistema sem fios). Além disso, o presente invento também pode ser aplicado a outro sistema de comunicação onde um codificador/descodificador utilizado para comprimir os dados áudio esteja otimizado para voz humana e não para música ou outro som. Serviços específicos onde o presente invento pode

ser aplicado incluem, entre outros, "coloring service" e "ARS (sistema de resposta áudio)".

O método de pré-processamento do presente invento pode ser aplicado a quaisquer dados áudio antes dos mesmos serem sujeitos a um codificador/descodificador de um sistema sem fios (ou qualquer outro codificador/descodificador otimizado para voz humana e diferente de música). Depois dos dados áudio serem pré-processados de acordo com o método de pré-processamento do presente invento, os dados pré-processados podem ser processados e transmitidos num codificador/descodificador sem fios normal. Para além de adicionar o componente necessário para realizar o método de pré-processamento do presente invento, não é necessária qualquer outra modificação ao sistema sem fios. Por conseguinte, o método de pré-processamento do presente invento pode ser facilmente adoptado por um sistema sem fios existente.

Apesar do presente invento ser explicado em relação ao codificador/descodificador EVRC, noutra concretização do presente invento, o mesmo pode ser aplicado de forma semelhante a outros codificadores/descodificadores com taxa de codificação variável.

O presente invento é descrito com referência às concretizações preferidas e aos desenhos, mas a descrição não se destina a limitar o presente invento à forma apresentada aqui. Deverá ser entendido que alguém especializado na técnica pode utilizar uma variedade de modificações e outras concretizações idênticas ao presente invento. Por conseguinte, apenas as reivindicações anexas se destinam a limitar o presente invento.

Lisboa, 2011-11-08

REIVINDICAÇÕES

1 - Método para pré-processamento de dados áudio que contenham dados de música para serem processados por um codificador/descodificador de codificação de taxa variável melhorado para transmissão num sistema de comunicação sem fios, estando o dito codificador/descodificador otimizado para voz humana e a funcionar em três taxas de codificação, compreendendo o método o passo de, para pelo menos num intervalo de dados que se destine a ser codificado pelo codificador/descodificador à taxa de codificação mais baixa e que não esteja definido como um intervalo de silêncio, ajustar amplitudes de dados áudio dentro do dito, pelo menos um, intervalo de dados tal que os dados áudio dentro do, pelo menos um, intervalo de dados sejam codificados à taxa de codificação máxima e, quando os dados áudio forem descodificados num terminal de recepção, uma pausa intermitente de música possa ser reduzida.

2 - Método de acordo com a reivindicação 1, em que o passo de ajuste compreende:

- o cálculo de níveis de sinal dos dados áudio;
- a decisão de coeficientes de ganho alisado; e
- a geração de dados áudio pré-processados pela multiplicação dos coeficientes de ganho alisado com os dados áudio no interior do intervalo decidido.

3 - Aparelho para pré-processamento de dados áudio que contenham dados de música para serem codificados por um codificador/descodificador de codificação de taxa variável melhorado para transmissão num sistema de comunicação sem fios, estando o dito codificador/descodificador otimizado para voz humana e a funcionar em três taxas de codificação, compreendendo o aparelho, para pelo menos num intervalo de dados que se destine a ser codificado pelo codificador/descodificador à taxa de codificação mais baixa e que não esteja definido como um intervalo de silêncio, meios para ajustar amplitudes de dados áudio dentro do dito pelo menos um intervalo de dados, tal que os dados áudio dentro

do, pelo menos um, intervalo de dados sejam codificados à taxa de codificação máxima e, quando os dados áudio forem descodificados num terminal de recepção, uma pausa intermitente de música possa ser reduzida.

Lisboa, 2011-11-08

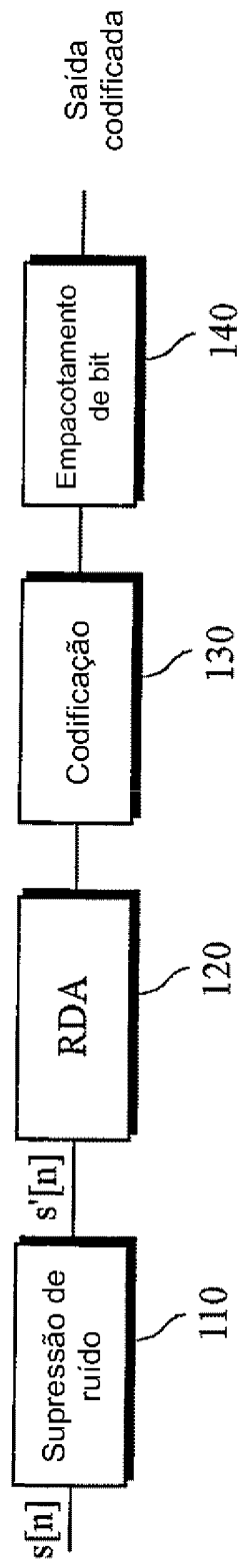
Fig. 1

Fig. 2A

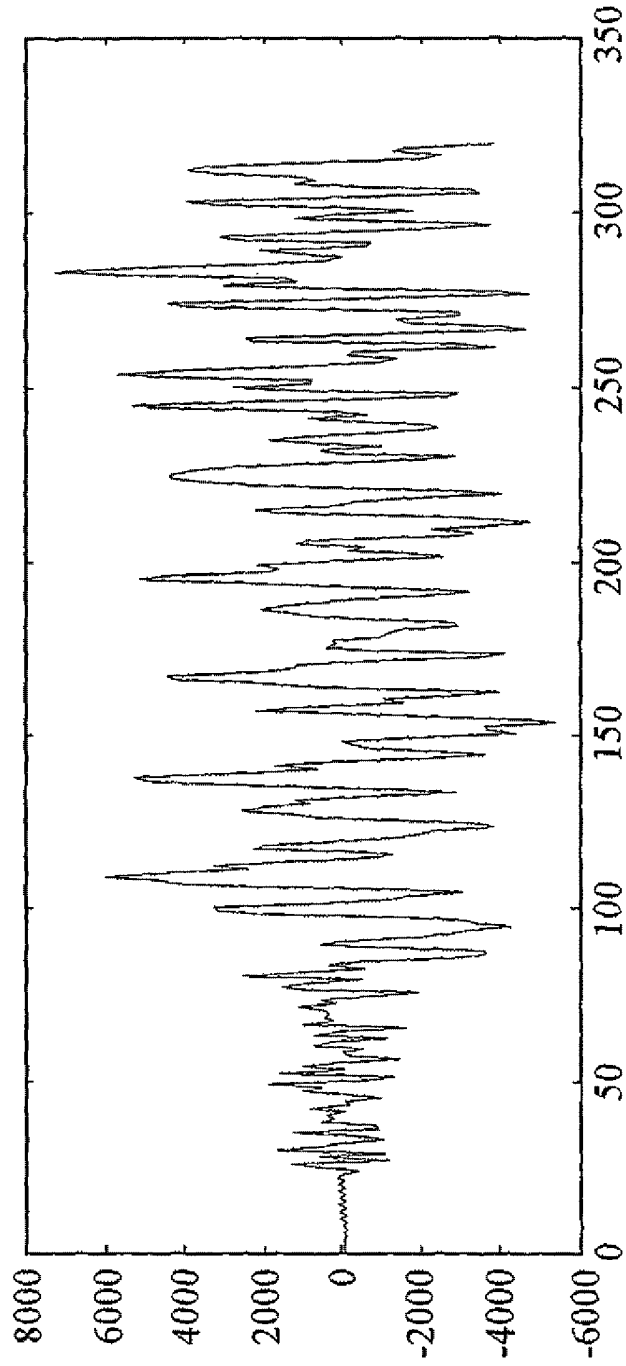


Fig. 2B

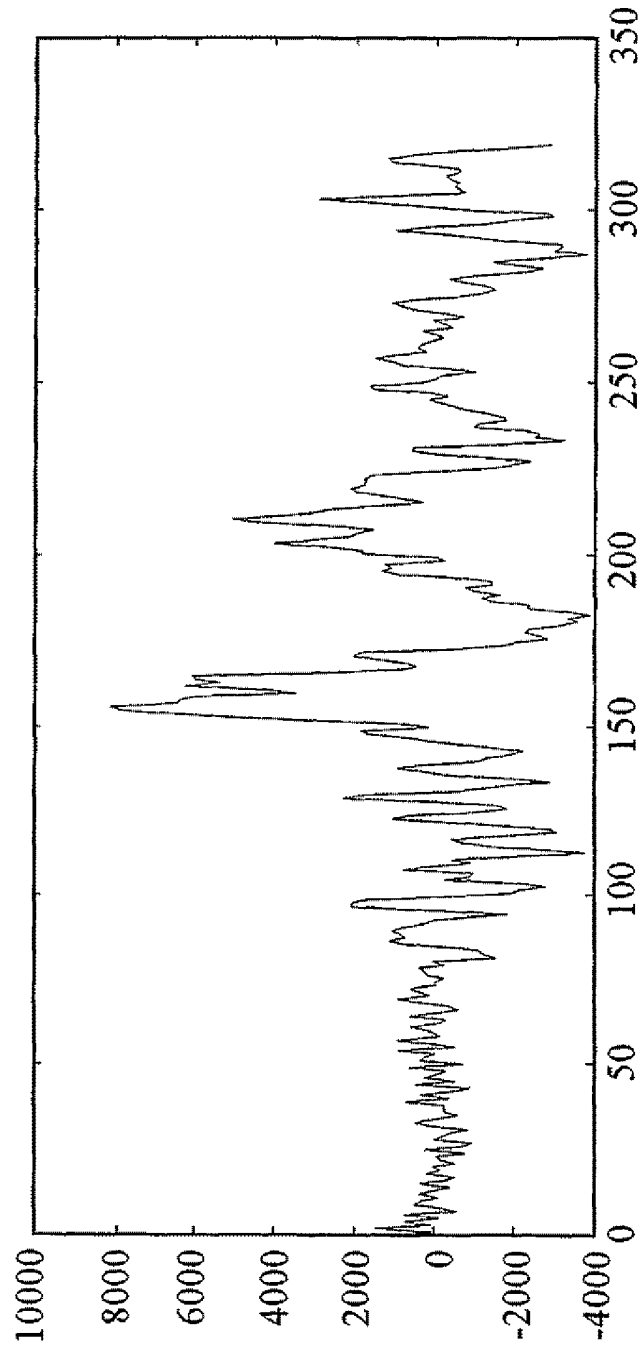


Fig. 3A

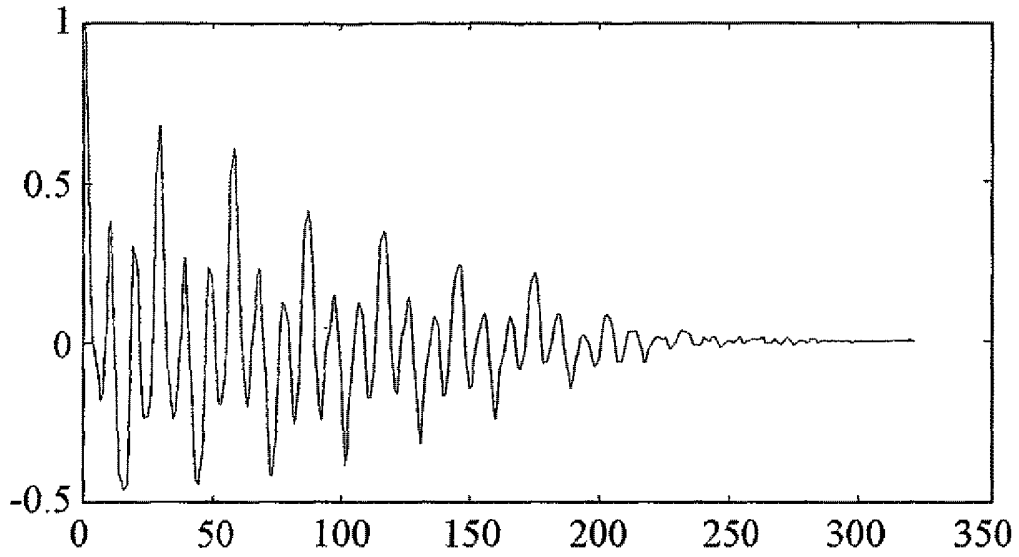


Fig. 3B

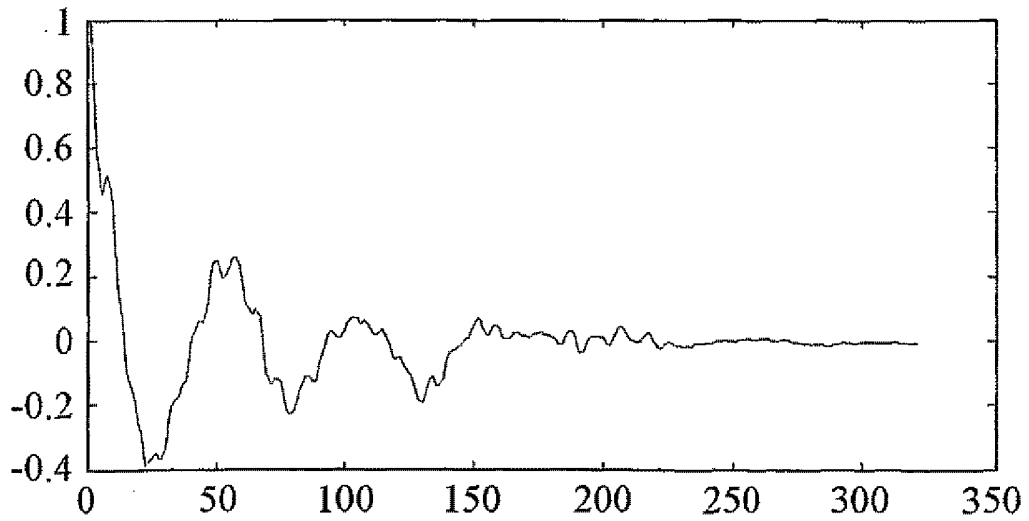


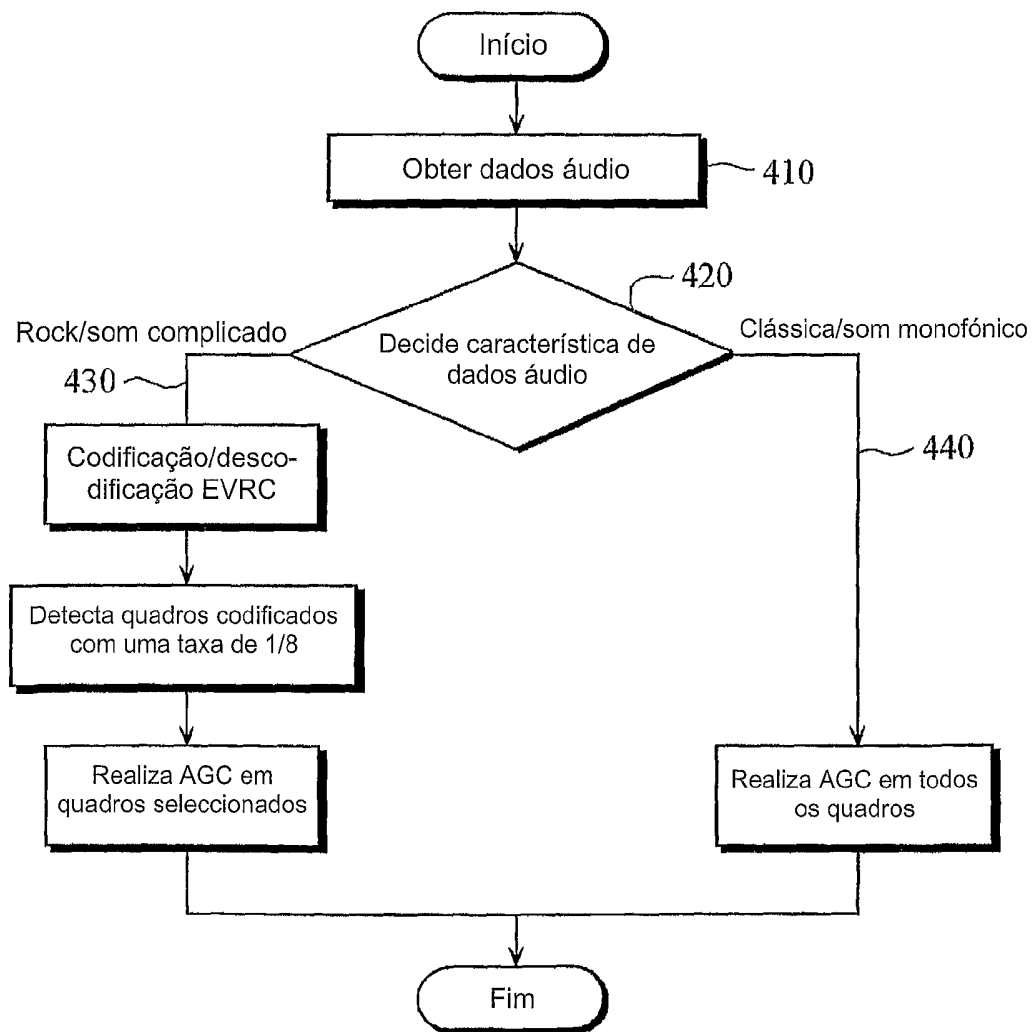
Fig. 4

Fig. 5

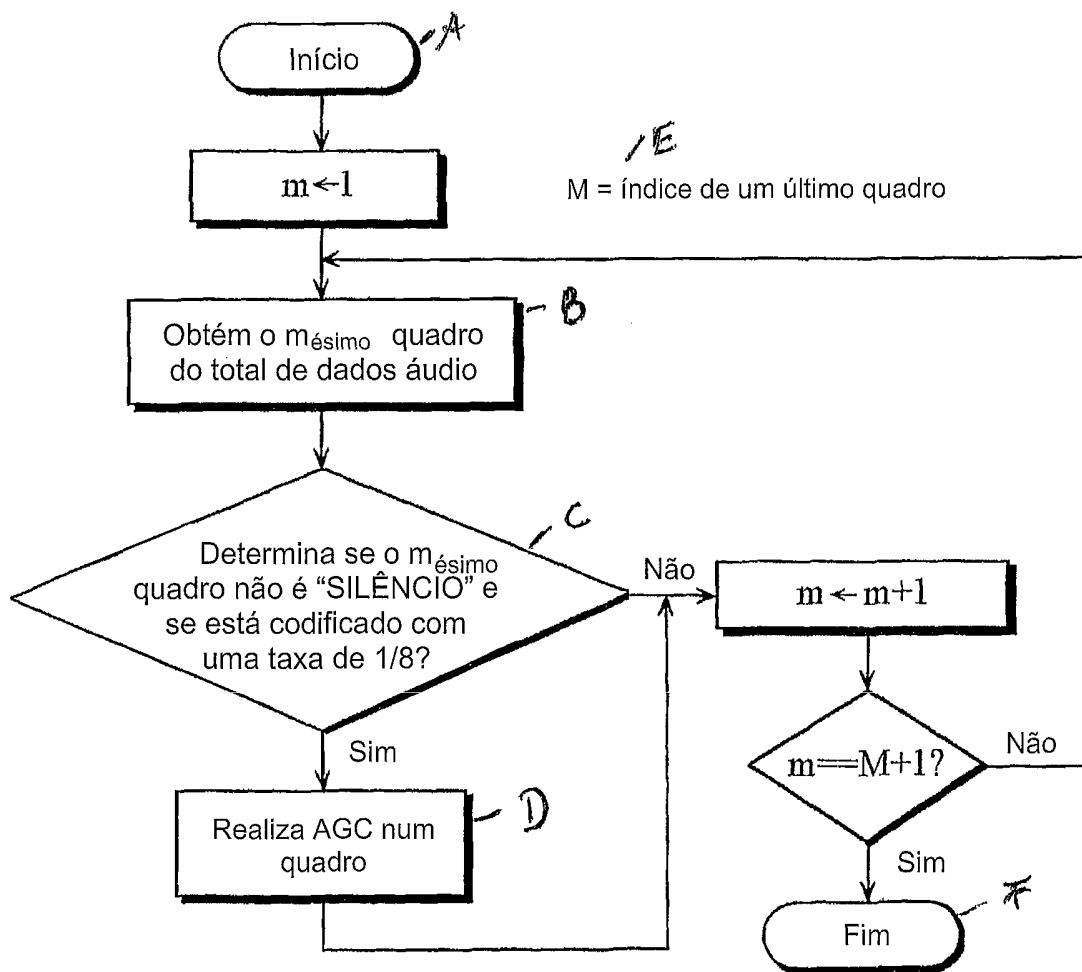


Fig. 6

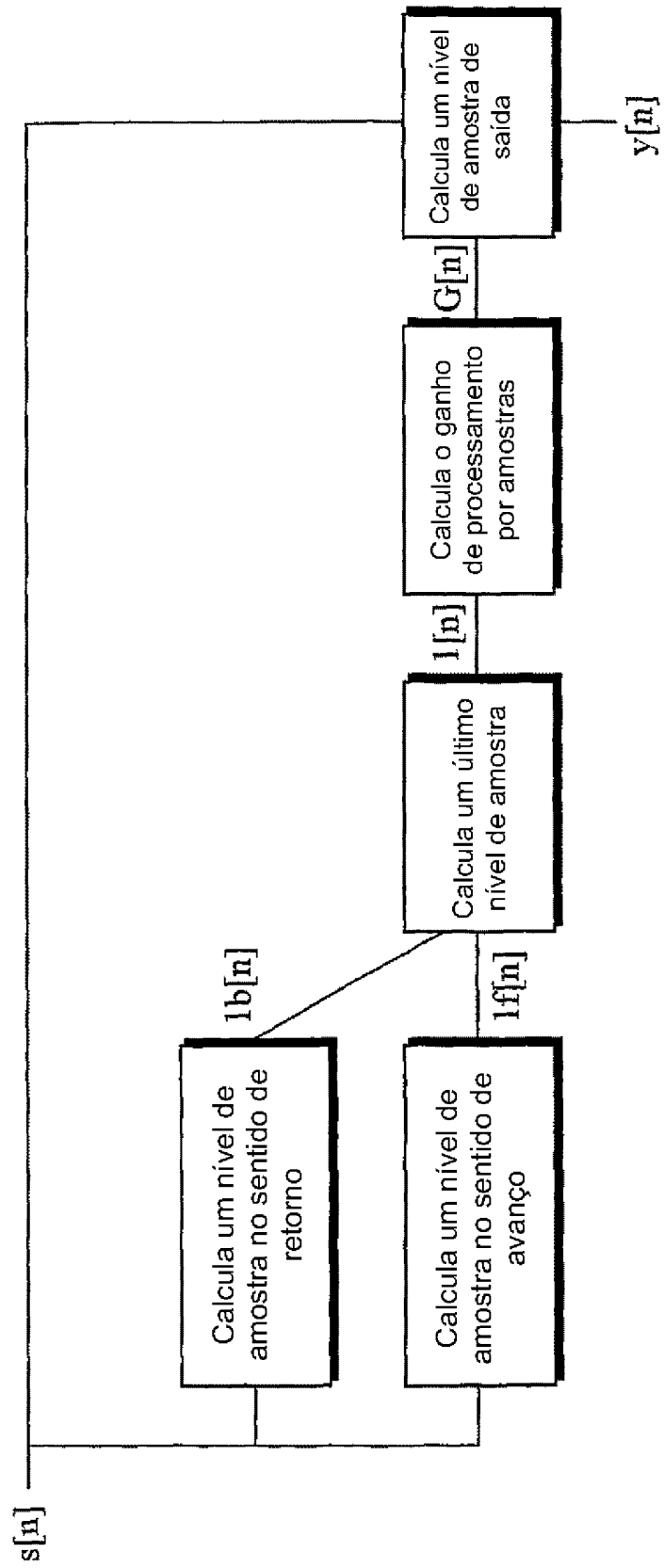


Fig. 7

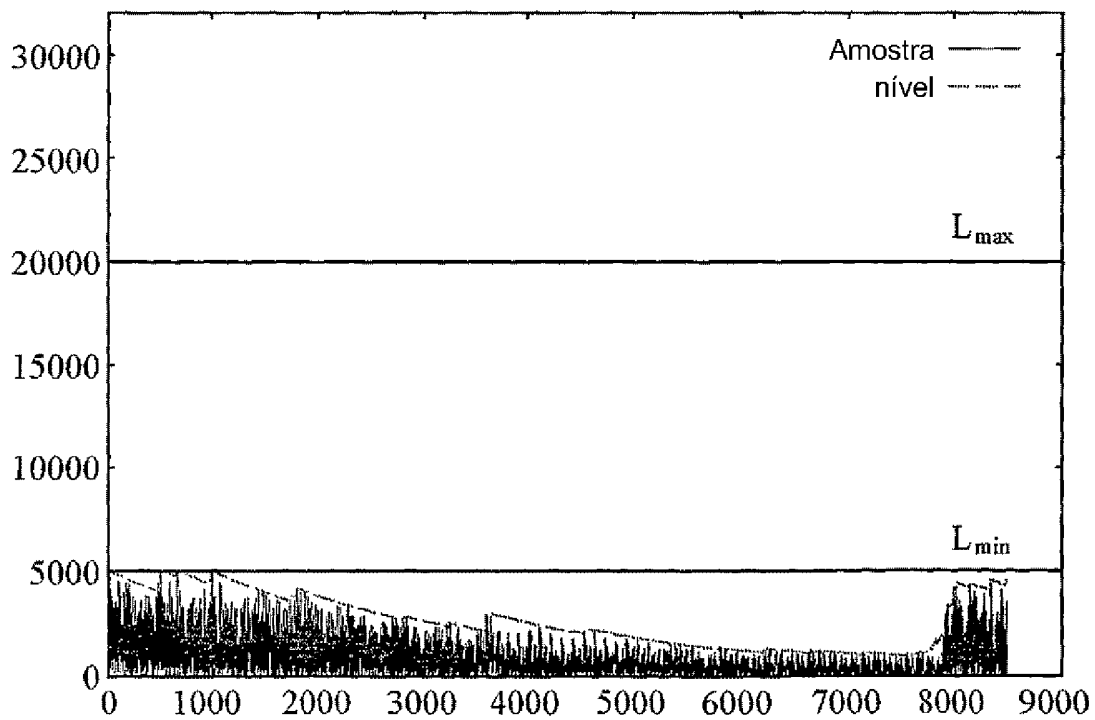


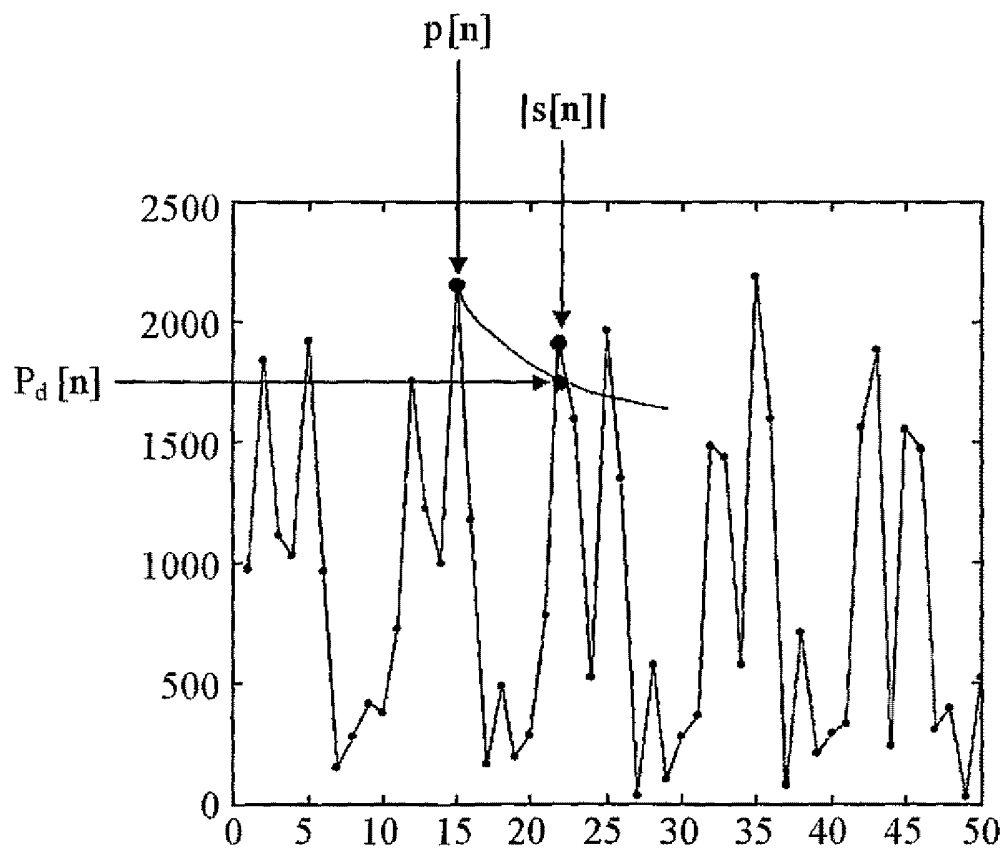
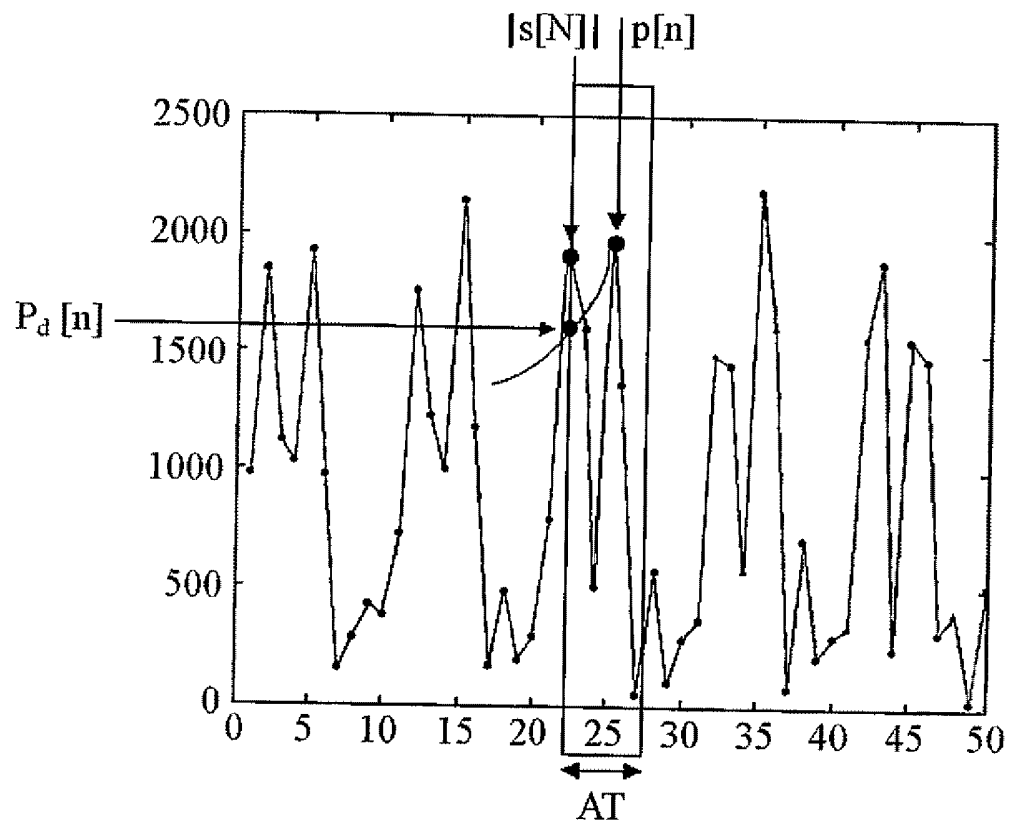
Fig. 8

Fig. 9

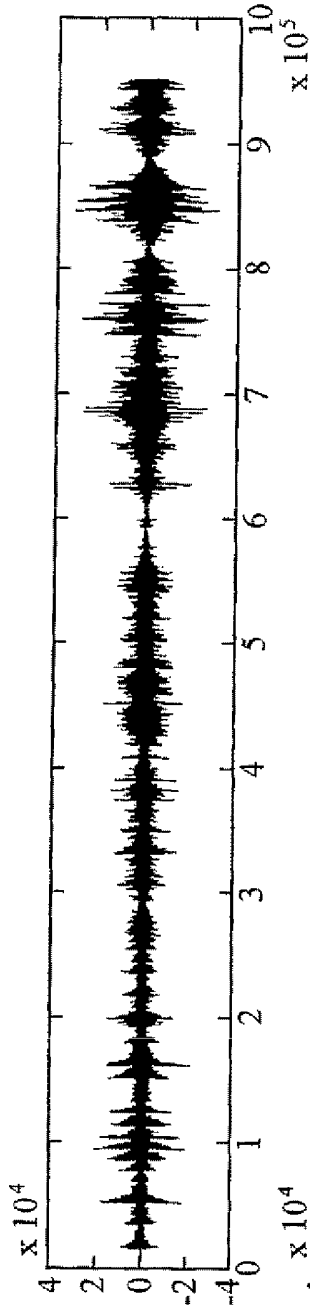


Fig.10A

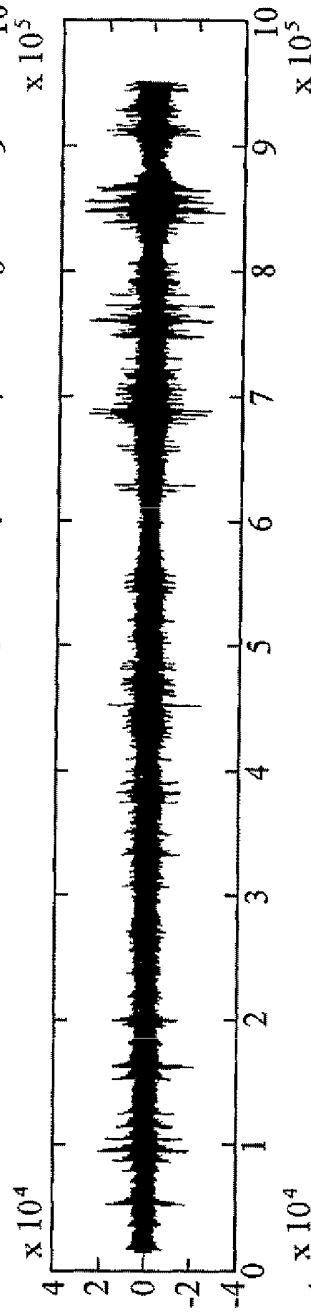


Fig.10B

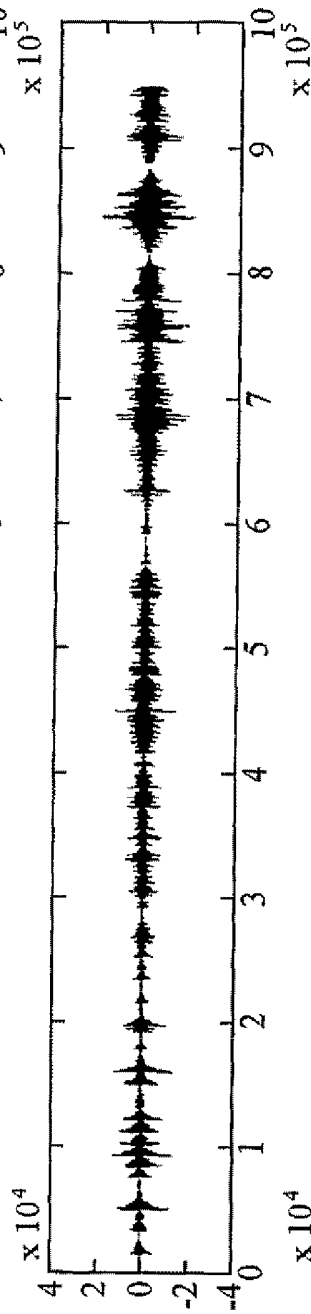


Fig.10C

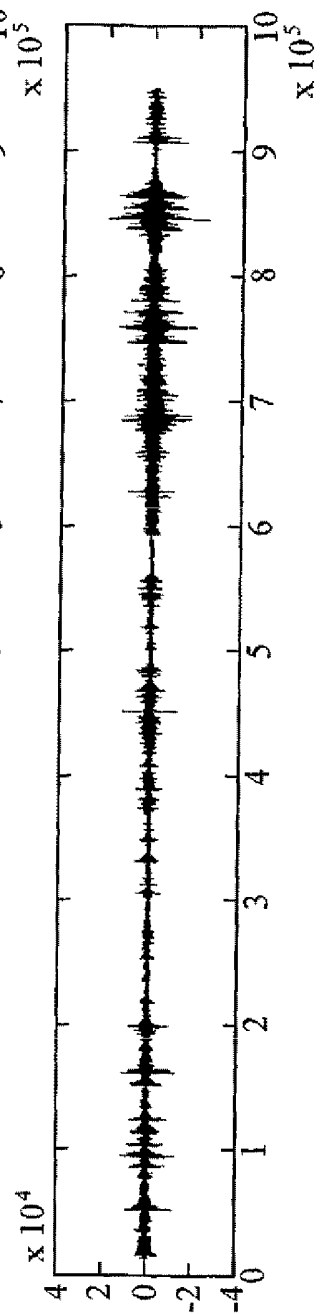


Fig.10D