

República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial.

(21) **PI 0621563-7 A2**



(22) Data de Depósito: 23/10/2006
(43) Data da Publicação: 10/07/2012
(RPI 2166)

(51) *Int.Cl.:*
G10L 11/06
G10L 15/20
G10L 21/02
H04B 1/00

(54) **Título:** MÉTODO E SISTEMA PARA REDUZIR EFEITOS DE ARTEFATOS QUE PRODUZEM RUÍDO EM UM CODEC DE VOZ

(30) **Prioridade Unionista:** 20/03/2006 UA 11/385,553

(73) **Titular(es):** Mindspeed Technologies, Inc.

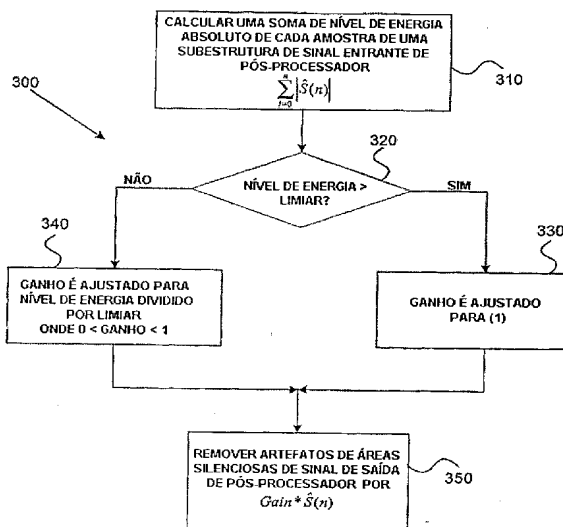
(72) **Inventor(es):** Gao, Yang, Shlomot, Eyal

(74) **Procurador(es):** Guerra Adv.

(86) **Pedido Internacional:** PCT US2006041434 de 23/10/2006

(87) **Publicação Internacional:** WO 2007/111645de 04/10/2007

(57) **Resumo:** MÉTODO E SISTEMA PARA REDUZIR EFEITOS DE ARTEFATOS QUE PRODUZEM RUÍDO EM UM CODEC DE VOZ. É fornecido um método de reduzir efeitos de artefatos que produzem ruído em um sinal de fala. O método compreende obter uma pluralidade (310) de amostras entrantes de uma subestrutura de fala, somar um nível de energia (310) para cada um dentre a pluralidade de amostras entrantes para gerar um nível de entrada total; comparar o nível de entrada total (320) com um limiar predeterminado; ajustar um valor de ganho (340) como uma função do nível de entrada total, onde o valor de ganho é entre zero (0) e um (1), e onde a função resulta em um valor de ganho mais baixo quando o nível de entrada total é indicativo de uma área silenciosa que quando o nível de entrada total é indicativo de uma área não silenciosa; e multiplicar a pluralidade (350) de amostras entrantes da subestrutura de fala pelo valor de ganho.



MÉTODO E SISTEMA PARA REDUZIR EFEITOS DE ARTEFATOS QUE PRODUZEM RUÍDO EM UM CODEC DE VOZ

FUNDAMENTOS DA INVENÇÃO

1. CAMPO DA INVENÇÃO

5 A presente invenção se refere geralmente à codificação de fala. Mais particularmente, a presente invenção se refere a redução de efeitos de artefatos que produzem ruído em um codec de voz.

2. TÉCNICA ANTERIOR

10 A redução da fala pode ser usada para reduzir o número de bits que representa o sinal de fala que reduz desse modo a largura da banda necessária para transmissão. Entretanto, a redução da fala pode resultar em degradação da qualidade da fala descomprimida. Geralmente, uma taxa de bit mais elevada conduzirá em alta qualidade, quando uma taxa de bit mais baixa resultará em má qualidade. Entretanto, as técnicas de redução de fala modernas, tais como técnicas
15 de codificação, podem produzir a fala descomprimida de qualidade relativamente alta em taxas de bit relativamente baixas. Geralmente, as técnicas de codificação modernas tentam representar as características importantes perceptualmente do sinal de fala, sem preservar a forma de onda real da fala. Os sistemas de redução de fala, chamados geralmente codec, incluem um codificador e um decodificador e
20 podem ser usados para reduzir a taxa de bit de sinais de fala digitais. Os algoritmos numerosos foram desenvolvidos para os codecs de fala que reduzem o número de bits exigido para codificar digitalmente a fala original ao tentar manter a alta qualidade de fala reconstruída.

FIG. 1 ilustra sistema de decodificação de fala convencional 100, que inclui o
25 decodificador de excitação 110, filtro de síntese 120 e pós-processador 130. Como

mostrado, o sistema de decodificação 100 recebe o fluxo de bits de fala codificado 102 sobre um meio de comunicação (não mostrado) de um codificador, onde o sistema de decodificação 100 possa ser parte de um dispositivo de comunicação móvel, uma estação base ou outro dispositivo de comunicação sem fio ou com fio que é capaz de receber o fluxo de bit de fala codificado 102. O sistema de decodificação 100 opera para decodificar o fluxo de bit de fala codificado 102 e gerar o sinal de fala 132 na forma de um sinal digital. O sinal de fala 132 pode então ser convertido a um sinal análogo por um conversor numérico-analógico (não mostrado). A saída análoga do conversor numérico-analógico pode ser recebida por um receptor (não mostrado) que possa ser uma orelha humana, um registrador de banda magnética, ou qualquer outro dispositivo capaz de receber um sinal análogo. Alternativamente, um dispositivo de gravação digital, um dispositivo de reconhecimento de fala, ou qualquer outro dispositivo capaz de receber um sinal digital pode receber sinal de fala 132.

O decodificador de excitação 110 decodifica o fluxo de bit de fala codificado 102 de acordo com o algoritmo de codificação e taxa de bit de fluxo de bit de fala codificado 102, e gera a excitação decodificada 112. O filtro de síntese 120 pode ser um filtro de predição a curto prazo que gera fala sintetizada 122 baseada na excitação decodificada 112. O pós-processador 130 pode incluir a filtração, o realce do sinal, a redução de ruído, a amplificação, a correção da inclinação e outras técnicas similares capazes de melhorar a qualidade perceptual da fala sintetizada 122. O pós-processador 130 pode diminuir o ruído audível sem visivelmente degradar a fala sintetizada 122. Diminuir o ruído audível pode ser realizado enfatizando a estrutura formadora da fala sintetizada 122 ou suprimindo o ruído nas regiões de frequência que não são perceptualmente relevantes para a fala

sintetizada 122.

Em codificadores de fala de taxas variáveis, as partes perceptualmente importantes da fala (por exemplo, fala expressa, plosivos, ou inícios expressos) são codificadas com um número mais elevado de bits, e as partes menos importantes da
5 fala (por exemplo, partes não expressas ou palavras entre silêncio) são codificadas com um número mais baixo de bits. A supressão do ruído melhora a qualidade do sinal reconstruído de voz e ajuda os codificadores de fala de taxas variáveis distinguirem as partes de vozes das peças de ruído. A supressão do ruído igualmente ajuda baixos codificadores de fala bit-taxa a produzir uma saída de alta
10 qualidade melhorando a qualidade de fala perceptual. Tipicamente, as técnicas da supressão do ruído removem o ruído por métodos de subtração espectrais no domínio de frequência. Um detector da atividade da voz (VAD) determina no tempo-domínio se uma estrutura do sinal inclui fala ou ruído. As estruturas do ruído são analisadas no domínio de frequências para determinar características do sinal de
15 ruído. Destas características, os espectros das estruturas do ruído são subtraídos dos espectros das estruturas de fala, fornecendo um sinal de fala limpo nas estruturas de fala.

Também, a atenuação do ruído tempo-domínio pode ser aplicada para melhorar a qualidade de um sinal de fala. Por exemplo, em um sistema de
20 codificação de fala com a atenuação do ruído tempo-domínio, descrito no Pedido E.U.A. nº de série 09/782.791, depositado em 13 de Fevereiro de 2001, que é incorporado por este meio pela referência em sua totalidade, os ganhos da codificação de fala de predição linear são ajustados por um fator de ganho para suprimir o ruído de fundo. Como descrito aqui, o sistema de codificação de fala usa
25 supressão de ruído domínio-frequência junto com a atenuação de voz tempo-

domínio para reduzir mais o ruído de fundo. Depois que um sinal análogo é convertido em um sinal digitado, um pré-processor suprime o ruído no sinal digitado usando uma supressão de ruído domínio-freqüência VAD. Quando o VAD identificar uma estrutura associada com somente o ruído (nenhum fala), uma estrutura de

5 janela que inclui a estrutura identificada de aproximadamente 10 ms que é transformada no domínio de freqüência. As magnitudes espectrais do sinal de fala ruidoso são modificadas para reduzir o nível de ruído de acordo com um SNR estimado, e os valores espectrais modificados são combinados com as fases espectrais não modificadas. O espectro modificado é transformado então de volta ao

10 tempo-domínio. Um esquema de análise por síntese escolhe a melhor representação para diversos parâmetros tais como um ganho de livro de código fixo ajustado, um índice fixo de livro de código, um parâmetro de retardação, e parâmetro ajustado de ganho do preditor à longo prazo. Os ganhos podem ser ajustados por um fator de ganho antes da quantização. O fator de ganho G_f pode suprimir o ruído de fundo no

15 domínio de tempo ao manter o sinal de fala, onde G_f é definido por $G_f = 1 - C \cdot \text{NSR}$, onde NSR é a relação ruído para sinal baseado na estrutura, e C é uma constante na escala de 0 a 1 e controla o grau de redução de ruído. NSR tem um valor de aproximadamente 1 quando somente o ruído de fundo está detectado na estrutura, e quando a fala está detectada na estrutura, NSR é a raiz quadrada da energia de

20 ruído de fundo dividida pela energia de sinal na estrutura.

Entretanto, as aproximações existentes não endereçam corretamente a diminuição de efeitos de artefatos que produzem ruídos nas áreas silenciosas de um sinal de fala, e especialmente quando um VAD não está disponível. Em áreas silenciosas, os valores da amostra são completamente pequenos e seus erros de

25 quantização são relativamente muito grandes. Os efeitos destes erros são

agravados mais pela execução de ponto fixo de um algoritmo de codificação de fala com sua exatidão limitada por meio dos valores de número inteiro que conduzem aos erros de quantização relativos das áreas silenciosas para se tornar muito maior, que faz com que uma energia de saída mais elevada comparada à energia de entrada de silêncio original, e de um ruído mais audível.

Conformemente, há uma necessidade intensa na técnica para os métodos e os sistemas de codificação de fala de baixa complexidade que podem eficazmente reduzir efeitos de artefatos em um sinal de fala, e especialmente em áreas silenciosas do sinal de fala e/ou sem utilizar um VAD.

10 SUMÁRIO DA INVENÇÃO

A presente invenção é dirigida a um método de reduzir o efeito de artefatos que produzem ruído em áreas silenciosas de um sinal de fala para uso por um sistema de decodificação de fala. Em um aspecto, o método compreende a obtenção de uma pluralidade de amostras entrantes de uma subestrutura de fala; somando um valor absoluto de um nível de energia para cada um dentre a pluralidade de amostras entrantes para gerar um nível de entrada total (*gain_in*); atenuando o nível de entrada total para gerar um nível atenuado (*Level_in_sm*); determinando que a subestrutura de fala está em uma área silenciosa baseada no nível de entrada total, no nível atenuado e em um parâmetro de inclinação espectral; definindo um ganho usando $k1 * (Level_in_sm / 1024) + (1 - k1)$, onde *K1* é uma função do parâmetro de inclinação espectral; e modificando um nível de energia da subestrutura de fala usando o ganho.

Em um aspecto mais adicional, a atenuação é executada usando ($Level_in_sm = 0,75 * Level_in_sm + 0,25 * gain_in$), e a determinação é executada usando ($Level_in_sm < 1024$) && ($gain_in < 2 * Level_in_sm$) && (*parcor0*

<512./32768).

Em um outro aspecto, o método ainda compreende atribuição *Level_in_sm* to *again_in* ($gain_{in} = Level_{in_sm}$) if $Level_{in_sm} < gain_{in}$. Contudo, em um outro aspecto, o método igualmente compreende a soma de um valor absoluto de um nível de energia para cada um dentre a pluralidade de amostras de saída, antes da alteração, para gerar um nível de saída total (*gain_out*); determinando um ganho inicial usando ($gain_{in}/gain_{out}$); e modificando o ganho usando o ganho inicial para gerar um ganho modificado (g_0), onde a alteração compreende a multiplicação *sig_out* para cada um dentre a pluralidade de amostras de saída por um ganho atenuado (g_{sm}), onde o g_{sm} é obtido usando iterações de 0 a n-1 de ($g_{sm} anterior * 0,95 \ g_0 * 0,05$), onde n é o número de amostras, e o g_{sm} anterior é zero (0) antes da primeira iteração.

Em um aspecto separado, é fornecido um método de reduzir efeitos de artefatos que produzem ruído em um sinal de fala. O método compreende obter uma pluralidade representante de amostras entrantes de uma subestrutura de fala; somar um nível de energia para cada um dentre a pluralidade de amostras para gerar um nível de entrada total; comparar o nível de entrada total com um limiar predeterminado; ajustar um valor de ganho como uma função do nível de entrada total, onde o valor do ganho está entre zero (0) e um (1), e onde a função resulta em um valor mais baixo de ganho quando o nível de entrada total for indicativo de uma área silenciosa do que quando o nível de entrada total for indicativo de uma área não silenciosa; multiplicar a pluralidade de representante das amostras de subestrutura de fala pelo valor de ganho.

Em um aspecto relacionado, o ajuste divide o nível de entrada total pelo limiar predeterminado se o nível de entrada total não for maior do que o limiar

predeterminado, e o ajuste ajusta o valor de ganho a um (1) se o nível de entrada total for maior do que o limiar predeterminado.

Em um aspecto mais adicional, o somatório soma um valor absoluto do nível de energia para cada um dentre a pluralidade de amostras para gerar o nível de
5 entrada total.

Em um aspecto, o método é executado por um sistema de decodificação de fala. Contudo, em um outro aspecto, o método é executado por um sistema de decodificação de fala.

Em um aspecto adicional, o método mais adicional compreende determinar se
10 o sinal de fala é um sinal de banda estreita ou um sinal de banda larga; e executar a obtenção, a soma, a comparação, o ajuste e a multiplicação somente se a determinação determina que o sinal de fala é o sinal de banda estreita.

Contudo, em um outro aspecto adicional, o método mais adicional compreende a detectar uma transição do sinal de fala entre um sinal de banda
15 estreita e um sinal de banda larga; e gradualmente mudar o valor de ganho baseado na transição.

Outras características e vantagens da presente invenção tornar-se-ão mais prontamente aparentes àquelas de versados na técnica após ter revisto a seguinte descrição detalhada e desenhos em anexo.

20 BREVE DESCRIÇÃO DOS DESENHOS

As características e as vantagens da presente invenção tornar-se-ão mais prontamente aparentes àquelas de versados na técnica após ter revisto a seguinte descrição detalhada e desenhos em anexo, onde:

FIG. 1 ilustra um diagrama em bloco de um sistema de decodificação
25 convencional para decodificar e pós-processar fluxo de bits de fala codificado;

FIG. 2 ilustra um diagrama em bloco de um pós-processador de fala, de acordo com uma modalidade do presente pedido; e

FIG. 3 ilustra um diagrama de fluxo de um método de pós-processar para uso pelo pós-processador de fala da FIG. 2, de acordo com uma modalidade do presente
5 pedido.

DESCRIÇÃO DETALHADA DA INVENÇÃO

Embora a invenção seja descrita no que diz respeito às modalidades específicas, os princípios da invenção, como definidos pelas reivindicações adicionadas aqui, podem obviamente ser aplicados além das modalidades
10 especificamente descritas da invenção descrita aqui. Além disso, na descrição da presente invenção, certos detalhes foram deixados a fim de não obscurecer os aspectos inventivos da invenção. Os detalhes deixados estão dentro do conhecimento de uma pessoa versada na técnica.

Os desenhos no presente pedido e sua descrição detalhada de
15 acompanhamento são dirigidos meramente às modalidades de exemplo da invenção. Para manter a brevidade, outras modalidades da invenção que usam os princípios da presente invenção não são descritas especificamente no presente pedido e não são ilustradas especificamente pelos desenhos atuais. Deve-se ter que, a menos que notáveis de outra maneira, como ou os elementos
20 correspondentes entre as figuras podem ser indicados por ou numerais de referência correspondentes.

FIG. 2 ilustra um diagrama em bloco do pós-processador de fala 220, de acordo com uma modalidade do presente pedido. Como mostrado, o pós-processador de fala 220 recebe o sinal entrante (*sig_in*) 210 e gera sinal de saída
25 (*sig_out*) 230 após pós-processamento de *sig_in* 210 para reduzir os efeitos audíveis

de artefatos em áreas silenciosas de sig_in 210. Com referência à FIG. 3, que ilustra um diagrama de fluxo de exemplo do método de pós-processamento 300 para uso pelo pós-processador de fala 220, o calculador de nível de energia de subestrutura 222 recebe sig_in 210, na etapa 310, e calcula uma soma do nível de energia absoluto de cada amostra de uma subestrutura de sig_in 210, que podem ser definidos por:

$$L = \sum_{i=0}^n |\hat{S}(n)|$$

Equação 1,

onde L é o nível de energia de subestrutura, S(n) designa sig_in 210, e (n) é o número de amostras.

Em seguida, o comparador de nível de energia de subestrutura 224 recebe o nível de energia de subestrutura (L) do calculador de nível de energia de subestrutura 222, e na etapa 320, o comparador de nível de energia de subestrutura 224 compara o nível de energia de subestrutura (L) com um limiar predeterminado (TH), por exemplo, 1.024, para uma determinação se o nível de energia de subestrutura (L) for indicativo de uma área silenciosa.

A saída do comparador de nível de energia de subestrutura 224 é recebida então pelo modificador de nível de energia de subestrutura 226. Se o modificador de nível de energia de subestrutura 226 determina que o nível de energia de subestrutura (L) é maior do que o limiar predeterminado (TH), na etapa 320, o método de pós-processamento 300 move-se para etapa 330, que é indicativa da área não silenciosa de fala. Na etapa 330, um valor de ganho (g) é ajustado a um (1). De um lado, se o modificador de nível de energia de subestrutura 226 determina que o nível de energia de subestrutura (L) não é maior do que o limiar predeterminado (TH), na etapa 320, o método de pós-processamento 300 move-se

para etapa 340, que é indicativa da área silenciosa de fala. Na etapa 340, o valor de ganho (g) é ajustado de acordo com o resultado do nível de energia de subestrutura (L) dividido pelo limiar predeterminado (TH), onde $0 < g < 1$, como mostrado abaixo:

$$g = \begin{cases} 1 & L > TH \\ L/TH & L \leq TH \end{cases}$$

Equação 2.

5 Em seguida, o método de pós-processamento300 move-se para etapa 350, onde o modificador de nível de energia de subestrutura 226 modifica o nível de energia de subestrutura (L), para reduzir efeitos de artefatos nas áreas silenciosas de sinal de saída de pós-processador (sig_out) 230, por exemplo, multiplicando o nível de energia de subestrutura (L) pelo limiar predeterminado (TH), como mostrado
10 na etapa 350, que é definida por:

$$g * \hat{S}(n)$$

Equação 3.

Em uma modalidade preferida da presente invenção, as modalidades de FIG. 2 e FIG. 3 são executadas em um decodificador de fala; entretanto, em outras modalidades, a presente invenção pode igualmente ser executada por um
15 codificador. Mais, embora a equação 2 mostre que g é uma função de L para áreas silenciosas, g pode igualmente ser uma função L nas áreas não silenciosas ($L > TH$) em outras modalidades. Por exemplo, nas modalidades de FIG. 2 e FIG. 3, g é ajustado a um (1) nas áreas não silenciosas ($L > TH$), de modo que $S(n)$ permaneça não modificado após a operação da Equação 3 em áreas não silenciosas. Em
20 adição, embora a equação 2 mostre que g é definido pela função de L/TH , outras funções de L/TH , outras funções de L podem ser utilizadas por outras modalidades.

Os apêndices A e B mostram uma execução de uma modalidade da presente invenção usando linguagem de programação "C" em ponto fixo e em ponto flutuante,

respectivamente. Em uma modalidade, no lado do decodificador, a energia do sinal é reduzida após detectar o sinal de baixo nível de silêncio. Por exemplo, o nível de sinal anterior ao pós-processamento de fala puder ser definido como:

$$gain_in = \sum_{i=0}^{Lsub-1} |sig_in(i)|$$

Equação 4,

- 5 onde $Lsub$ é o tamanho de subestrutura ou o número de amostras de fala para cada subestrutura, e $sig_in()$ é o sinal antes do desempenho do pós-processamento de fala. Em seguida, o nível atenuado de sig_in é calculado por:

$$Level_in_sm = 0,75 * Level_in_sm + 0,25 * gain_in \quad \text{Equação 5,}$$

- onde o valor inicial do $Level_in_sm$ é zero (0). Em uma modalidade, as equações 4 e 10 5 podem ser executadas pelo calculador do nível de energia de subestrutura 222 do pós-processador 200 na FIG. 2.

Em tal modalidade, a detecção de silêncio é baseada no nível de sinal e parâmetro de inclinação espectral, que é referido igualmente como o primeiro coeficiente de reflexão ($parcor0$)), como mostrado abaixo:

$$15 \quad Sil_Deci = (Level_in_sm < 1024) \&\& (gain_in < 2 * Level_in_sm) \\ \&\& (parcor0 < 512./32768); \quad \text{Equação 6,}$$

onde, $Sil_Deci = 1$ é indicativo da detecção de baixo nível de silêncio. Em uma modalidade, o comparador de nível de energia de subestrutura 224 pode determinar Sil_Deci , de acordo com a equação 6.

- 20 Mais, se a seguinte circunstância for satisfatória, o $gain_in$ (ou o nível de sinal de referência) é substituído com o nível de sinal atenuado, que pode ser executado pelo modificador de nível de energia de subestrutura 226:

$$\text{Se } ((Sil_Deci = 1) \&\& (Level_in_sm < gain_in))$$

$$gain_in = Level_in_sm;$$

Equação 7,

Onde, ($Level_in_sm < gain_in$) é indicativo de inexistência de picos grandes no sinal. Agora, sig_out presumido (i), $i=0, 1, \dots, Lsub-1$, é o sinal pós-filtrado antes do ajuste de ganho discutido abaixo, o nível de sinal pós-filtrado inicial pode ser calculado pelo calculador de nível de energia de subestrutura 222, como segue:

$$gain_out = \sum_{i=0}^{Lsub-1} |sig_out(i)| \quad \text{Equação 8,}$$

onde, o ganho inicial para ajustar a energia de sinal pós-filtrada pode ser determinado pelo modificador de nível de energia de subestrutura 226 como:

$$g0 = \frac{gain_in}{gain_out} \quad \text{Equação 9,}$$

Uma vez que o silêncio de baixo nível é detectado, isto é $Sil_Deci = 1$, modificador de nível de energia de subestrutura 226 pode reduzir ainda o ganho inicial aplicando o seguinte fator de ganho ($gain$), como mostrado abaixo:

$$gain = k1 * (Level_in_sm / 1024) + (1 - k1); \quad \text{Equação 10,}$$

onde $0 \leq gain \leq 1$, e $k1$ ($0 \leq k1 \leq 1$) é uma função de $parcor0$, e:

$$k1 = (512 / 32768) - parcor0;$$

$$\begin{aligned} & \text{if } (k1 > (2047 / 32768)) \\ & \quad k1 = (2047 / 32768); \\ & k1 = k1 / (2047 / 32768); \end{aligned}$$

Finalmente, o modificador de nível de energia de subestrutura 226 aplicará um ajuste de ganho ao sinal pós-filtrado, como segue, onde g_sm é o ganho atenuado:

```

if (Sil_Deci==1) {
    g0 = g0 * gain
    for (i=0; i<Lsub;i++) {
        g_sm = g_sm*0,95 + g0*0,05;
        sig_out(i) = sig_out(i)*g_sm;
    }
}

```

Em outras palavras, sig_out é modificado multiplicando sig_out para cada um dentre a pluralidade de amostras de saída por um ganho atenuado (g_{sm}), onde g_{sm} é obtido usando iterações de 0 a n-1 de (g_{sm} anterior * 0,95 + g_0 *0,05), onde n é o número de amostras ou o tamanho de subestrutura, e g_{sm} anterior é zero (0) antes da primeira iteração.

Em uma modalidade preferida da presente invenção, a redução de ganho de silêncio descrita acima é executada somente para o sinal de fala (0-4 KHz) de banda estreita no decodificador, mas não para o sinal de fala (4-8 KHz) de banda larga.

Entretanto, outras modalidades da presente invenção podem incluir execuções de codificador e/ou de banda larga. Mais, quando uma transição de/para sinal de fala de banda estreita de/para sinal de fala de banda larga é detectada, o ganho pode gradualmente ser mudado ou ajustado um pouco do que uma aplicação abrupta (transição de banda larga à banda estreita) ou não aplicação (transição de banda estreita à banda larga) do ganho reduzindo os efeitos de artefatos nas áreas silenciosas, onde comutar entre a banda estreita e a banda larga é ainda descrito no Pedido de Patente dos E.U.A. Nº de série _____, depositado _____, Nº do procurador 0160148, intitulado "Seamless Switching Between Narrowband and Wideband Speech", o qual é incorporado por este meio pela referência em sua totalidade.

Da descrição acima da invenção manifesta-se que várias técnicas podem ser

usadas executando os conceitos da presente invenção sem partir de seu escopo. Além disso, quando a invenção for descrita com referência específica a determinadas modalidades, uma pessoa versada na técnica reconheceria que as mudanças podem ser feitas na forma e no detalhe sem partir do espírito e do escopo da invenção. Por exemplo, contempla-se que os circuitos divulgados aqui podem ser executados em software, ou vice-versa. As modalidades descritas devem ser consideradas em todos os respeitos como ilustrativas e não restritivas. Deve-se igualmente compreender que a invenção não está limitada às modalidades particulares descritas aqui, mas é capaz de muitos rearranjos, modificações, e substituições sem partir do escopo da invenção.

ANEXO A

```

5  /*-----*/
/*-----*/
/* Fixed-Point Silence Cleaning */
/*-----*/
/*-----*/

Word16Level_in_sm=1024; /* temporally put this variable here */
10 Word16 PostNB=0; /* temporally set to 0; real value : 0<=PostNB<=1 */

/*-----*/
* G729EV_G729_scale_st - control of the subframe gain
* gain[n] = G729EV_G729_AGC_FAC * gain[n-1] + (1 - AGC_FAC) g_in/g_out
15 *-----*/
*/
static void G729EV_G729_scale_st(Word16 *sig_in, /* input : postfilter input signal */
                                Word16 *sig_out, /* in/out: postfilter output signal */
                                Word16 *gain_prec, /* in/out: last value of gain for subframe */
20 #ifdef SILENCE_CLEANING
                                Word16 parcor0,
                                Word16 PostNB,
                                Word32 *Level_in_sm
#endif
25 #endif
)
{
Word32 L_acc, L_temp;
Word16 i;
Word16 scal_in, scal_out;
30 Word16 s_g_in, s_g_out, temp, sh_g0, g0;

```

```

#ifdef SILENCE_CLEANING
Word16 gain;
Word16 Cond;
#endif

35     /* compute input gain */
     L_acc = 0L;
#ifdef WMOPS
     move32();
40 #endif
     for (i = 0; i < G729EV_G729_L_SUBFR; i++)
     {
         L_temp = L_abs(L_deposit_l(sig_in[i]));
         L_acc = L_add(L_acc, L_temp);
45     }

#ifdef SILENCE_CLEANING

     /* Smooth level */

     *Level_in_sm=L_add(L_shr(*Level_in_sm, 1), L_shr(*Level_in_sm, 2));
     *Level_in_sm=L_add(*Level_in_sm, L_shr(L_acc, 2));

     /* Detect silence*/
5     Cond = (*Level_in_sm<1024) && (L_acc<L_shl(*Level_in_sm, 1)) && (parcor0<512);

     /* If silence is detected, replace the original level with smoothed level*/
     if (Cond == 1)
         L_acc = *Level_in_sm;
10 #endif

     #ifdef WMOPS
         test();
     #endif
15     if (L_acc == 0L)
     {
         g0 = 0;
         #ifdef WMOPS
             move16();
20 #endif
         }
         else
         {
25             scal_in = norm_l(L_acc);
             L_acc = L_shl(L_acc, scal_in);
             s_g_in = extract_h(L_acc); /* normalized */

             /* Compute output gain */
             L_acc = 0L;
30 #ifdef WMOPS
                 move32();
             #endif

```



```

-----
    for (i = 0; i < G729EV_G729_L_SUBFR; i++)
    {
35         L_temp = L_abs(L_deposit_l(sig_out[i]));
        L_acc = L_add(L_acc, L_temp);
    }
    #ifndef WMOPS
        test();
40    #endif
        if (L_acc == 0L)
        {
            *gain_prec = 0;
        #ifndef WMOPS
45            move16();
        #endif
            return;
        }
        scal_out = norm_l(L_acc);
50        L_acc = L_shl(L_acc, scal_out);

        s_g_out = extract_h(L_acc); /* normalized */

        sh_g0 = add(scal_in, 1);
        sh_g0 = sub(sh_g0, scal_out); /* scal_in - scal_out + 1 */
5    #ifndef WMOPS
        test();
    #endif
        if (sub(s_g_in, s_g_out) < 0)
        {
10            g0 = div_s(s_g_in, s_g_out); /* s_g_in/s_g_out in Q15 */
        }
        else
        {
            temp = sub(s_g_in, s_g_out); /* sufficient since normalized */
15            g0 = shr(div_s(temp, s_g_out), 1);
            g0 = add(g0, (Word16) 0x4000); /* s_g_in/s_g_out in Q14 */
            sh_g0 = sub(sh_g0, 1);
        }
        /* L_gain_in/L_gain_out in Q14 */
20        /* overflows if L_gain_in > 2 * L_gain_out */
        g0 = shr(g0, sh_g0); /* sh_g0 may be >0, <0, or =0 */

    #ifndef SILENCE_CLEANING

25        if (Cond==1)
        { /* Apply a gain reduction for silence; the gain is defined as
            gain = (Level_in_sm/MAX_SILENCE_LEVEL)*k1 + (1-k1);
            k1 (0<=k1<=1) is a function of PARCOR0 */
            /* k1 in Q15*/
30            temp=sub(512, parcor0);
            if (temp>2047) temp=2047;
            temp=shl(temp, 4);

```

```

35      /* gain = (Level_in_sm/MAX_SILENCE_LEVEL) in Q15 */
      if (*Level_in_sm > 1023) gain = 1023;
      else gain = extract_l(*Level_in_sm);
      gain = shl(gain, 5);

      /* gain = gain*k1 + 1-k1 */
40      gain = mult_r(gain, temp);
      gain = add(gain, sub(32767, temp));

      gain = mult_r(gain, sub(32767, PostNB));
45      gain = add(gain, PostNB);

      g0 = mult_r(g0, gain);
    }

    #endif

50

      g0 = mult_r(g0, G729EV_G729_AGC_FAC1); /* L_gain_in/L_gain_out *
AGC_FAC1 */
    }

5      /* gain(n) = G729EV_G729_AGC_FAC gain(n-1) + G729EV_G729_AGC_FAC1
gain_in/gain_out */
      /* sig_out(n) = gain(n) sig_out(n) */
      gain = *gain_prec;
      for (i = 0; i < G729EV_G729_L_SUBFR; i++)
10      {
          temp = mult_r(G729EV_G729_AGC_FAC, gain);
          gain = add(temp, g0); /* in Q14 */
          L_temp = L_mult(gain, sig_out[i]);
          L_temp = L_shl(L_temp, 1);
15          sig_out[i] = round(L_temp);
      }
      *gain_prec = gain;

    #ifdef WMOPS
20      move16();
      move16();
    #endif
    return;
25
}

```

ANEXO B

```

/*****
/*****
5  /*
Floating-Point Silence Cleaning */
/*****
/*****

```

```

10  REAL Level_in_sm=1024.; /* temporally put this variable here */
    REAL PostNB=0.; /* temporally set to 0; real value : 0<=PostNB<=1 */

    /**-----
15  * Function G729EV_G729_scale_st - control of the subframe gain
    * gain[n] = G729EV_G729_AGC_FAC * gain[n-1] + (1 - G729EV_G729_AGC_FAC)
    g_in/g_out
    *-----
    */
20  void
    G729EV_G729_scale_st (REAL *sig_in,      /**< input : postfilter input signal */
                        REAL *sig_out,      /**< in/out: postfilter output signal */
                        REAL *gain_prec     /**< in/out: last value of gain for subframe */
#ifdef SILENCE_CLEANING
25      ,
                        REAL parcor0,
                        REAL PostNB,
                        REAL *Level_in_sm
#endif
30      )
    {
        int i;
        REAL gain_in, gain_out;
        REAL g0, gain;
35  #ifdef SILENCE_CLEANING
        short Cond;
        REAL k1;
    #endif

40      /* compute input gain */
        gain_in = (REAL) 0.;
        for (i = 0; i < G729EV_G729_L_SUBFR; i++)
        {
45          gain_in += (REAL) fabs (sig_in[i]);
        }

        #ifdef SILENCE_CLEANING

            /* Smooth level */
50          *Level_in_sm = 0.75*(*Level_in_sm) + 0.25*gain_in;

            /* Detect silence*/
            Cond = (*Level_in_sm < 1024.) && (gain_in < *Level_in_sm*2.) &&
            (parcor0<512./32768);
5          /* If silence is detected, replace the original level with smoothed level*/
            if (Cond == 1)
                gain_in = *Level_in_sm;

10  #endif

```

```

15   if (gain_in == (REAL) 0.)
      {
          g0 = (REAL) 0.;
      }
      else
20   {
      /* Compute output gain */
      gain_out = (REAL) 0.;
      for (i = 0; i < G729EV_G729_L_SUBFR; i++)
      {
25         gain_out += (REAL) fabs (sig_out[i]);
      }
      if (gain_out == (REAL) 0.)
      {
          *gain_prec = (REAL) 0.;
30         return;
      }
      g0 = gain_in / gain_out;

#ifdef SILENCE_CLEANING
35   if ( Cond==1 )
      { /* Apply a gain reduction for silence; the gain is defined as
          gain = (Level_in_sm/MAX_SILENCE_LEVEL)*k1 + (1-k1);
          k1 (0=<k1<=1) is a function of PARCOR0 */

40         /*k1*/
          k1=(512./32768) - parcor0;
          if (k1>(2047./32768)) k1=(2047./32768);
          k1 /= (2047./32768);

45         /* gain = (Level_in_sm/MAX_SILENCE_LEVEL)*/
          if (*Level_in_sm>1023) gain = 1.;
          else gain = *Level_in_sm/1024.;

50         gain = gain*k1 + 1-k1;

          gain = gain *(1.-PostNB) + PostNB;

          g0 *= gain;
      }
5   #endif

      g0 *= G729EV_G729_AGC_FAC1;
  }
10

  /* compute gain(n) = G729EV_G729_AGC_FAC gain(n-1) + (1-
  G729EV_G729_AGC_FAC)gain_in/gain_out */
  /* sig_out(n) = gain(n) sig_out(n) */

```

```
gain = *gain_prec;
15 for (i = 0; i < G729EV_G729_L_SUBFR; i++)
    {
        gain *= G729EV_G729_AGC_FAC;
        gain += g0;
        sig_out[i] *= gain;
20    }
    *gain_prec = gain;
    return;
}
25
```

REIVINDICAÇÕES

1. **“Método para reduzir efeitos de artefatos que produzem ruído em áreas silenciosas de um sinal de fala para uso por um sistema de decodificação de fala”**, dito método **caracterizado por** compreender uma pluralidade de amostras entrantes de uma subestrutura de fala; somar um valor absoluto de um nível de energia para cada um dentre a pluralidade de amostras entrantes para gerar um nível de entrada total (*gain_in*); atenuar o nível de entrada total para gerar um nível atenuado (*Level_in_sm*); determinar que a subestrutura de fala está em uma área silenciosa baseada no nível de entrada total, o nível atenuado e um parâmetro de inclinação espectral; definir um ganho usando $kl * (Level_in_sm / 1024) + (1 - kl)$, onde *kl* é uma função do parâmetro de inclinação espectral; e modificar um nível de energia da subestrutura de fala usando o ganho.
2. **“Método para reduzir efeitos de artefatos que produzem ruído em áreas silenciosas de um sinal de fala para uso por um sistema de decodificação de fala”** da reivindicação 1, **caracterizado pela** atenuação ser realizada usando $Level_in_sm = 0,75 * Level_in_sm + 0,25 * gain_in$.
3. **“Método para reduzir efeitos de artefatos que produzem ruído em áreas silenciosas de um sinal de fala para uso por um sistema de decodificação de fala”** da reivindicação 1, **caracterizado pela** determinação ser realizada usando $(Level_in_sm < 1024) \&\& (gain_in < 2 * Level_in_sm) \&\& (parcor0 < 512./32768)$.
4. **“Método para reduzir efeitos de artefatos que produzem ruído em áreas silenciosas de um sinal de fala para uso por um sistema de decodificação de fala”** da reivindicação 1, **caracterizado por** compreender atribuição $Level_in_sm$ to $again_in$ ($gain_in = Level_in_sm$) if $Level_in_sm < gain_in$.
5. **“Método para reduzir efeitos de artefatos que produzem ruído em áreas**

silenciosas de um sinal de fala para uso por um sistema de decodificação de fala”

da reivindicação 4, **caracterizado por** compreender soma de um valor absoluto de um nível de energia para cada um dentre a pluralidade de amostras de saída, antes da alteração, para gerar um nível de saída total (*gain_out*); determinação de um ganho inicial usando (*gain_in/gain_out*); e modificação do ganho usando o ganho inicial para gerar um ganho modificado (*g0*).

6. “Método para reduzir efeitos de artefatos que produzem ruído em áreas silenciosas de um sinal de fala para uso por um sistema de decodificação de fala”

da reivindicação 5, **caracterizado pela** alteração compreender a multiplicação *sig_out* para cada um dentre a pluralidade de amostras de saída por um ganho atenuado (*g_sm*), em que *g_sm* é obtido usando iterações de 0 a n-1 de (*g_sm* anterior * 0,95 + *g0**0,05), onde n é o número de amostras, e *g_sm* anterior é zero (0) antes da primeira iteração.

7. “Sistema de decodificação de fala para reduzir efeitos de artefatos produzindo

ruído em áreas silenciosas de um sinal de fala”, dito sistema de decodificação de fala **caracterizado por** compreender um calculador de nível de energia de subestrutura configurado para obter uma pluralidade de amostras entrantes de uma subestrutura de fala, e configurado para somar um valor absoluto de um nível de energia para cada um dentre a pluralidade de amostras entrantes para gerar um nível de entrada total (*gain_in*), e mais configurado para atenuar o nível de entrada total para gerar um nível atenuado (*Level_in_sm*); um comparador de nível de energia de subestrutura configurado para determinar que a subestrutura de fala está em uma área silenciosa baseada no nível de entrada total, no nível atenuado e um parâmetro de inclinação espectral; e um modificador de nível de energia de subestrutura configurado para definir um ganho utilizando $k1*(Level_in_sm/1024) + (1-k1)$, onde *K1* é uma

função do parâmetro de inclinação espectral, e mais configurado para modificar um nível de energia da subestrutura de fala usando o ganho.

8. “Sistema de decodificação de fala para reduzir efeitos de artefatos produzindo ruído em áreas silenciosas de um sinal de fala” da reivindicação 7, caracterizado pelo calculador de nível de energia de subestrutura atenua o nível de entrada total utilizar $Level_in_sm = 0,75 * Level_in_sm + 0,25 * gain_in$.

9. “Sistema de decodificação de fala para reduzir efeitos de artefatos produzindo ruído em áreas silenciosas de um sinal de fala” da reivindicação 7, caracterizado pelo comparador de nível de energia de subestrutura determina que a subestrutura de fala está na área silenciosa utilizar $(Level_in_sm < 1024) \ \&\& \ (gain_in < 2 * Level_in_sm) \ \&\& \ (parcor0 < 512./32768)$.

10. “Sistema de decodificação de fala para reduzir efeitos de artefatos produzindo ruído em áreas silenciosas de um sinal de fala” da reivindicação 7, caracterizado pelo modificador de nível de energia de subestrutura atribui $Level_in_sm$ to $again_in$ $(gain_in = Level_in_sm)$ if $Level_in_sm < gain_in$.

11. “Sistema de decodificação de fala para reduzir efeitos de artefatos produzindo ruído em áreas silenciosas de um sinal de fala” da reivindicação 10, caracterizado pelo calculador de nível de energia de subestrutura ser configurado para somar um valor absoluto de um nível de energia para cada um dentre a pluralidade de amostras de saída, antes da modificação pelo modificador de nível de energia de subestrutura, para gerar um nível de saída total $(gain_out)$, e o modificador de nível de energia de subestrutura ser configurado para determinar um ganho inicial usando $(gain_in/gain_out)$ e para modificar o ganho usando o ganho inicial para gerar um ganho modificado $(g0)$.

12. “Sistema de decodificação de fala para reduzir efeitos de artefatos produzindo

ruído em áreas silenciosas de um sinal de fala” da reivindicação 11, **caracterizado pelo** modificador de nível de energia de subestrutura modificar o nível de energia de subestrutura de fala multiplicando *sig_out* para cada um dentre a pluralidade de amostras de saída por um ganho atenuado (g_{sm}), onde g_{sm} é obtido usando

5 iterações de 0 a $n-1$ de $(g_{sm} \text{ anterior} * 0,95 + g_0 * 0,05)$, onde n é o número de amostras, e o g_{sm} anterior é zero (0) antes da primeira iteração.

13. “Método de reduzir efeitos de artefatos que produzem ruído em um sinal de fala”, o método **caracterizado por** compreender a obtenção uma pluralidade de representante entrante de amostras de uma subestrutura de fala; somar um nível de

10 energia para cada um dentre a pluralidade de amostras para gerar um nível de entrada total; comparar o nível de entrada total com um limiar predeterminado; ajustar um valor de ganho como uma função do nível de entrada total, onde o valor de ganho está entre zero (0) e um (1), e onde a função resulta em um valor inferior de ganho quando o nível de entrada total for indicativo de uma área silenciosa do que quando o nível de

15 entrada total for indicativo de uma área não silenciosa; multiplicar a pluralidade de representante das amostras da subestrutura de fala pelo valor de ganho.

14. “Método de reduzir efeitos de artefatos que produzem ruído em um sinal de fala”, da reivindicação 13, **caracterizado pelo** ajuste dividir o nível de entrada total pelo limiar predeterminado, se o nível de entrada total não for maior do que o limiar

20 predeterminado.

15. “Método de reduzir efeitos de artefatos que produzem ruído em um sinal de fala” da reivindicação 14, **caracterizado pelo** ajuste ajustar o valor de ganho a um (1) se o nível de entrada total for maior do que o limiar predeterminado.

16. “Método de reduzir efeitos de artefatos que produzem ruído em um sinal de fala” da reivindicação 13, **caracterizado pelo** somatório somar um valor absoluto do

25

nível de energia para cada um dentre a pluralidade de amostras para gerar o nível de entrada total.

5 **17. “Método de reduzir efeitos de artefatos que produzem ruído em um sinal de fala”** da reivindicação 13, **caracterizado pelo** método ser executado por um sistema de decodificação de fala.

18. “Método de reduzir efeitos de artefatos que produzem ruído em um sinal de fala” da reivindicação 13, **caracterizado pelo** método ser executado por um sistema de decodificação de fala.

10 **19. “Método de reduzir efeitos de artefatos que produzem ruído em um sinal de fala”** da reivindicação 13, **caracterizado por** compreender a determinação se o sinal de fala é um sinal de banda estreita ou um sinal de banda larga; e executar a obtenção, a soma, a comparação, o ajuste e a multiplicação somente se a determinação determinar que o sinal de fala é o sinal de banda estreita.

15 **20. “Método de reduzir efeitos de artefatos que produzem ruído em um sinal de fala”** da reivindicação 13, **caracterizado por** compreender a detecção de uma transição do sinal de fala entre um sinal de banda estreita e um sinal de banda larga; e mudar gradualmente o valor do ganho baseado na transição.

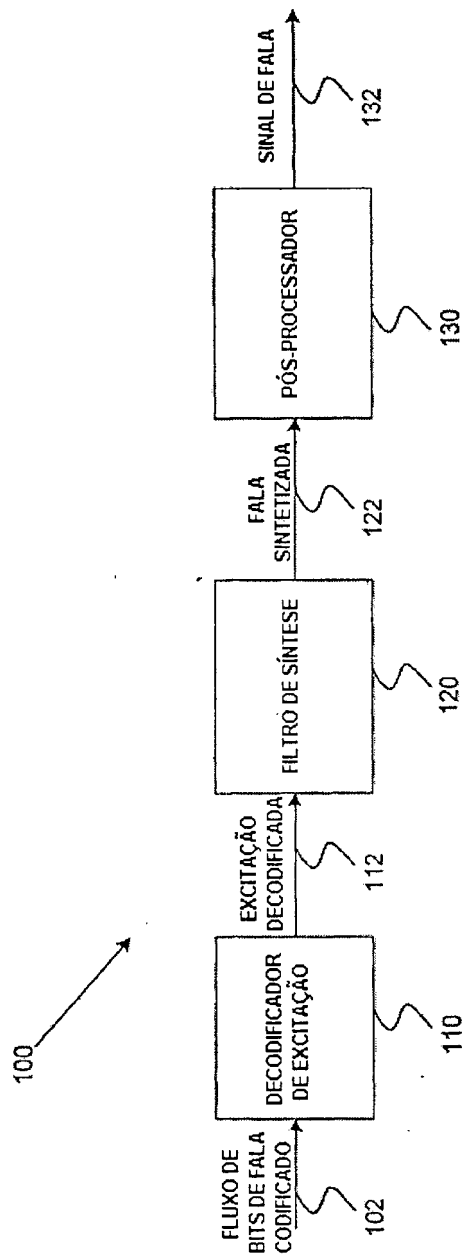


FIG. 1
(TÉCNICA ANTERIOR)

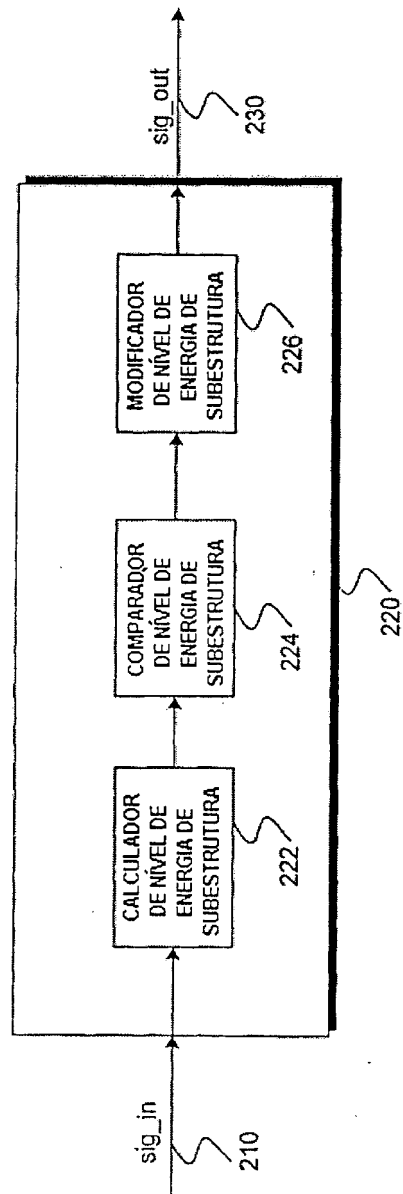


FIG. 2

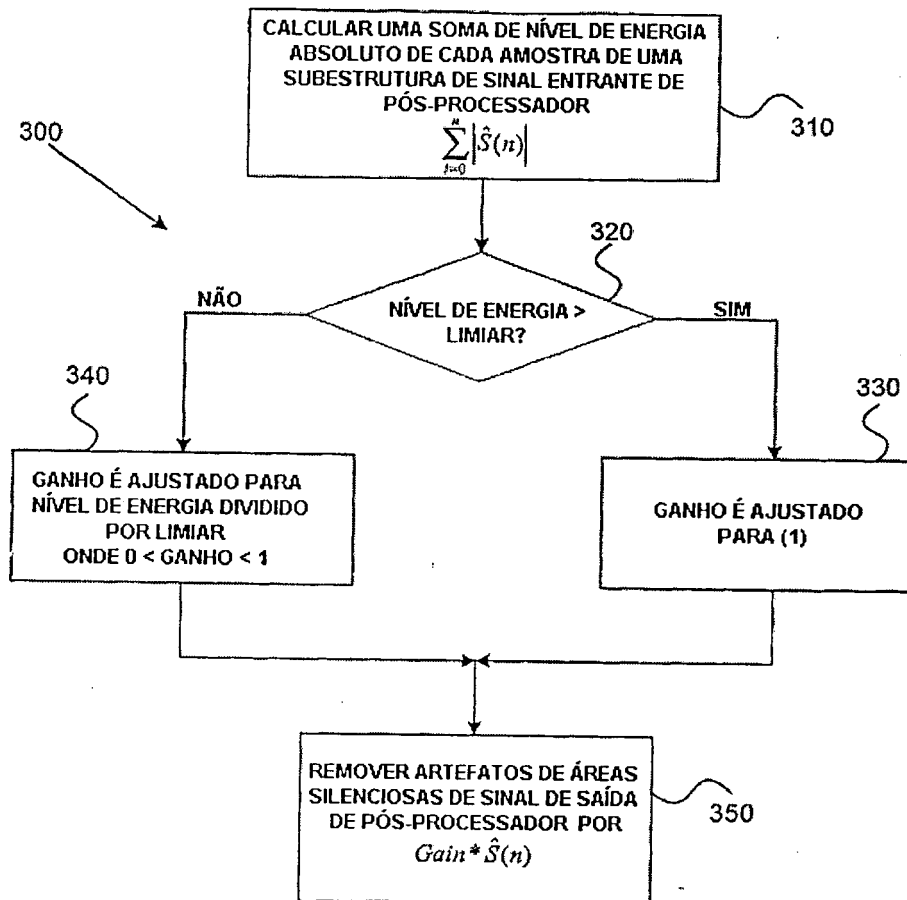


FIG. 3

RESUMO

MÉTODO E SISTEMA PARA REDUZIR EFEITOS DE ARTEFATOS QUE PRODUZEM RUÍDO EM UM CODEC DE VOZ. É fornecido um método de reduzir

efeitos de artefatos que produzem ruído em um sinal de fala. O método compreende

- 5 obter uma pluralidade (310) de amostras entrantes de uma subestrutura de fala, somar um nível de energia (310) para cada um dentre a pluralidade de amostras entrantes para gerar um nível de entrada total; comparar o nível de entrada total (320) com um limiar predeterminado; ajustar um valor de ganho (340) como uma função do nível de entrada total, onde o valor de ganho é entre zero (0) e um (1), e
- 10 onde a função resulta em um valor de ganho mais baixo quando o nível de entrada total é indicativo de uma área silenciosa que quando o nível de entrada total é indicativo de uma área não silenciosa; e multiplicar a pluralidade (350) de amostras entrantes da subestrutura de fala pelo valor de ganho.