



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

(11) PI 0114827-3 B1

(22) Data do Depósito: 17/10/2001

(45) Data de Concessão: 11/09/2018



(54) Título: MÉTODO E APARELHO PARA ENCOBRIR OS EFEITOS DOS ERROS DE QUADRO NOS QUADROS A SEREM DECODIFICADOS PELO DECODIFICADOR PARA PROPORCIONAR VOZ SINTETIZADA

(51) Int.Cl.: G10L 13/00; G10L 19/005; G10L 19/04

(30) Prioridade Unionista: 23/10/2000 US 60/242,498

(73) Titular(es): NOKIA TECHNOLOGIES OY

(72) Inventor(es): JARI MÄKINEN; JANI ROTOLA-PUKKILA; HANNU J. MIKKOLA; JANNE VAINIO

(85) Data do Início da Fase Nacional: 22/04/2003

“MÉTODO E APARELHO PARA ENCOBRIR OS EFEITOS DOS ERROS DE QUADRO NOS QUADROS A SEREM DECODIFICADOS PELO DECODIFICADOR PARA PROPORCIONAR VOZ SINTETIZADA”.

CAMPO DA INVENÇÃO

5 A presente invenção refere-se a decodificadores de voz, e mais particularmente a métodos usados para controlar os quadros ruins recebidos pelos decodificadores de voz.

DESCRIÇÃO DA TÉCNICA ANTERIOR

10 Nos sistemas celulares digitais, um fluxo de bits é dito para ser transmitido através de um canal de comunicação conectando uma estação móvel a uma estação base sobre a interface aérea. O fluxo de bits é organizado em quadros, incluindo os quadros de voz. Se ou não um erro ocorre durante a transmissão depende das condições do canal prevaletentes. O quadro de voz que é detectado contendo erros é simplesmente chamado de um quadro ruim. De acordo com a técnica anterior, no caso de um quadro ruim, os
15 parâmetros de voz derivados dos últimos parâmetros corretos (quadros corretos de voz sem erro) são substituídos para os parâmetros de voz do quadro ruim. O objetivo do controle do quadro ruim ao realizar tal substituição é encobrir os parâmetros corrompidos de voz do quadro de voz errôneo sem causar uma degradação observável na qualidade da voz.

20 Os codecs de voz modernos operam através do processamento do sinal de voz em pequenos segmentos, isto é, os quadros acima mencionados. O comprimento de quadro típico de um codec de voz é de 20 ms, o qual corresponde a 160 amostras de voz, assumindo uma frequência de amostragem de 8 KHz. Nos chamados codecs de banda larga, o comprimento do quadro pode novamente ser de 20 ms, mas pode corresponder a
25 320 amostras de voz, assumindo uma frequência de amostragem de 16 KHz. O quadro também pode ser dividido em um número de sub-quadros.

Para cada quadro, o codificador determina uma representação paramétrica do sinal de entrada. Os parâmetros são quantizados e então transmitidos através do canal de comunicação na forma digital. O codificador produz um sinal de voz sintetizado
30 baseado nos parâmetros recebidos (ver Fig. 1).

Um conjunto típico de parâmetros de codificação extraído inclui os parâmetros espectrais (denominados de parâmetros de codificação preditiva linear, ou parâmetros LPC (linear predictive coding)) usados na predição a curto prazo, os parâmetros usados na predição a longo prazo do sinal (os chamados de parâmetros de predição a longo prazo ou parâmetros LTP (long-term prediction parameters)), os vários parâmetros de ganho, e finalmente, os parâmetros de excitação.

O que é chamado codificação preditiva linear é um método amplamente usado e bem sucedido para codificar a voz para transmissão sobre o canal de comunicação; este representa os atributos do modelo de frequência do trato vocal. A parametrização LPC caracteriza o modelo do espectro de um segmento curto de voz. Os parâmetros LPC podem ser representados como LSF (Line Spectral Frequencies - Frequências de Linha Espectral) ou, equivalentemente, como ISPs (Immittance Spectral Pairs - Pares de Imitância Espectral). Os ISPs são obtidos através da decomposição da função $A(z)$ de transferência de filtro inverso para estabelecer duas funções de transferência, uma possuindo simetria par e a outra possuindo simetria ímpar. Os ISPs, também chamados de Frequências de Imitância Espectral (ISFs) são as raízes destes polinomiais no círculo de unidade z . Os Pares de Linha Espectral (também chamado de Frequências de Linha Espectral) podem ser definidos da mesma forma que os Pares de Imitância Espectral; a diferença entre estas representações é o algoritmo de conversão, o qual transforma os coeficientes de filtro LP em outra representação de parâmetro LPC (LSP ou ISP).

Algumas vezes, a condição do canal de comunicação, através do qual são transmitidos os parâmetros de voz codificados é pobre, causando erros no fluxo de bit, isto é, ocasionando erros no quadro (e então ocasionando quadros ruins). Existem dois tipos de erros de quadro: os quadros perdidos e os quadros corrompidos. Em um quadro corrompido, apenas alguns dos parâmetros que descrevem um segmento de voz em particular (tipicamente de 20 ms de duração) são corrompidos. No tipo de quadro perdido do erro de quadro, um quadro é totalmente corrompido ou não recebido de forma alguma.

No sistema de transmissão baseado em pacote para comunicar voz (um sistema no qual um quadro é usualmente carregado como um pacote único), que em

algumas das vezes é proporcionado por uma conexão Internet ordinária, é possível que um pacote de dados (ou quadro) nunca alcance o receptor pretendido ou que o pacote de dados (ou quadro) chegue tão atrasado que não possa ser usado, devido à natureza em tempo real da voz falada. Tal quadro é chamado de quadro perdido. O quadro corrompido em tal situação é um quadro entrante (usualmente dentro de um pacote único) no receptor, mas que contém alguns parâmetros que estão com erro, como indicado, por exemplo, pela verificação de redundância cíclica (CRC – cyclic redundancy check). Está é usualmente uma situação em uma conexão comutada por circuito, tal como uma conexão ao sistema global para comunicação móvel (GSM), onde a taxa de erro de bit (BER) em um quadro corrompido é tipicamente abaixo de 5%.

Desse modo, pode ser visto que a resposta corretiva ótima para uma incidência de um quadro ruim é diferente para os dois casos de quadros ruins (o quadro corrompido e o quadro perdido). Existem diferentes respostas, no caso dos quadros corrompidos, e existe informação não confiável sobre os parâmetros, e no caso de quadros perdidos, nenhuma informação está disponível.

De acordo com a técnica anterior, quando um erro é detectado em um quadro de voz recebido, uma substituição e um procedimento de silêncio são iniciados; os parâmetros de voz do quadro ruim são substituídos por valores modificados ou atenuados do quadro bom anterior, embora alguns dos parâmetros menos importantes do quadro errôneo sejam usados, por exemplo, os parâmetros de predição linear excitados por código (CELPs – code excited linear prediction parameters), ou mais simplesmente os parâmetros de excitação.

Em alguns métodos de acordo com a técnica anterior, uma memória é usada (no receptor) denominada de histórico do parâmetro, onde os últimos parâmetros de voz recebidos sem erro são armazenados. Quando um quadro é recebido sem erro, o histórico do parâmetro é atualizado e os parâmetros de voz carregados pelo quadro são usados para decodificação. Quando um quadro ruim é detectado, através da verificação CRC ou de algum outro método de detecção de erro, um indicador de quadro ruim (BFI – bad frame indicator) é estabelecido para verdadeiro e o encobrimento do parâmetro (substituição e silêncio dos quadros ruins correspondentes) é então iniciado; os métodos da técnica

anterior para encobrimento do parâmetro usam o histórico do parâmetro para encobrir os quadros corrompidos. Como mencionado acima, quando um quadro recebido é classificado como um quadro ruim (BFI estabelecido como verdadeiro), alguns dos parâmetros de voz podem ser usados do quadro ruim; por exemplo, no exemplo de solução para a substituição do quadro corrompido de um codec de voz GSM AMR (múltiplas-taxas adaptativas – adaptive multi-rate) fornecido pelo ETSI (Instituto Europeu de Normas de Telecomunicações) na especificação 06.91, onde o vetor de excitação do canal é sempre usado. Quando um quadro de voz está perdido (incluindo a situação aonde um quadro chega tarde demais para ser usado, tal como, por exemplo, em alguns sistemas de transmissão baseados em IP), obviamente nenhum dos parâmetros do quadro perdido estão disponíveis para ser usado.

Em alguns dos sistemas da técnica anterior, os últimos parâmetros bons espectrais recebidos são substituídos pelos parâmetros espectrais do quadro ruim, após ser ligeiramente desviado em direção a um meio constante pré-determinado. De acordo com a especificação ETSI 06.91 GSM, o encobrimento é realizado no formato LSF (Frequências de Linha Espectral), e é fornecido pelo seguinte algoritmo:

Para $i = 0$ a $N-1$:

LSF_q1(i)

$$= \alpha * \text{última_LSF_q}(i) + (1 - \alpha) * \text{meio_LSF}(i); \quad (\text{eq. 1.0})$$

1.0)

$$\text{LSF_q2}(i) = \text{LSF_q1}(i);$$

Onde $\alpha = 0,95$ e N é a ordem do filtro linear preditivo (LP- predictive linear) sendo usado. A quantidade LSF_q1 é o vetor LSF quantizado do segundo sub-quadro e a quantidade LSF_q2 é o vetor LSF quantizado do quarto sub-quadro. Os vetores LSF do primeiro e do terceiro sub-quadros são interpolados destes dois vetores. (O vetor LSF para o primeiro sub-quadro no quadro n é interpolado do vetor LSF do quarto sub-quadro no quadro $n-1$, isto é, do quadro anterior). A quantidade de última_LSF_q é a quantidade LSF_q2 do quadro anterior. A quantidade do meio_LSF é um vetor, cujo os componentes são constantes pré-determinadas; os componentes não dependem da seqüência de voz decodificada. A quantidade meio_LSF com componentes constantes gera um espectro de

voz constante.

Tais sistemas da técnica anterior sempre desviam os coeficientes do espectro em direção às quantidades constantes, aqui indicadas como meio_LSF(i). As quantidades da constante são construídas para calcular a média sobre um período de tempo longo e sobre vários emissores de voz sucessivos. Tais sistemas, portanto oferecem apenas uma solução de compromisso, não uma solução que é ótima para qualquer alto falante ou situação particular; a transação do compromisso é entre deixar os artefatos perturbadores na voz sintetizada, e tornar a voz mais natural ao soar (por exemplo, a qualidade de voz sintetizada).

O que é necessário é uma substituição do parâmetro espectral melhorado no caso do quadro de voz corrompido, possivelmente uma substituição baseada tanto em uma análise do histórico do parâmetro de voz quanto do quadro errôneo. Uma substituição adequada aos quadros errôneos de voz possui um efeito significativo na qualidade de voz sintetizada produzida do fluxo de bits.

RESUMO DA INVENÇÃO

Conseqüentemente, a presente invenção proporciona um método e um aparelho correspondente para encobrir os efeitos de erros do quadro nos quadros a serem decodificados pelo decodificador para proporcionar uma voz sintetizada, os quadros sendo proporcionados sobre o canal de comunicação pelo codificador, onde cada quadro proporciona os parâmetros usados pelo decodificador na sintetização da voz, o método inclui as etapas de: determinar se um quadro é um quadro ruim; e proporcionar a substituição dos parâmetros do quadro ruim baseada em um meio pelo menos parcialmente adaptativo dos parâmetros espectrais de um número pré-determinado dos quadros bons recentemente recebidos.

Em um aspecto adicional da invenção, o método também inclui a etapa de determinar se o quadro ruim carrega voz não-estacionária ou estacionária, e, em adição, a etapa de proporcionar a substituição do quadro ruim é realizada de uma maneira que depende se o quadro ruim carrega voz não-estacionária ou estacionária. Ainda em um aspecto adicional da invenção, no caso de um quadro ruim carregando voz estacionária, a etapa de proporcionar a substituição do quadro ruim é realizada usando um meio de

parâmetros de um número pré-determinado dos quadros bons recentemente recebidos. Ainda em outro aspecto adicional da invenção, no caso de um quadro ruim carregando uma voz não-estacionária, a etapa de proporcionar a substituição do quadro ruim é realizada usando, quando muito, uma parte pré-determinada dos parâmetros de um número pré-determinado dos quadros bons recebidos recentemente.

Em outro aspecto da invenção, o método também inclui a etapa de determinar se o quadro ruim encontra um critério pré-determinado, e neste caso, usando o quadro ruim ao invés de substituir o quadro ruim. Ainda em um aspecto adicional da invenção com semelhante etapa, o critério pré-determinado envolve preparar um ou mais quadros de comparação: uma comparação interquadro, uma comparação intraquadro, uma comparação de dois pontos, e uma comparação de ponto único.

De outra perspectiva, da invenção um método para encobrir os efeitos de erros de quadros nos quadros a serem decodificados pelo decodificador para proporcionar voz sintetizada, onde os quadros são proporcionados sobre o canal de comunicação pelo decodificador, cada quadro proporcionando os parâmetros usados pelo decodificador na sintetização de voz, o método inclui as etapas de: determinar se um quadro é um quadro ruim; e proporcionar a substituição dos parâmetros do quadro ruim, a substituição na qual as últimas frequências de imitância espectral (ISF) são desviadas em direção a um meio parcialmente adaptativo fornecido por:

$$ISF_q(i) = \alpha * \text{última_}ISF_q(i) + (1 - \alpha) * ISF_{\text{meio}}(i), \text{ para } i = 0,16,$$

onde

$$\alpha = 0,9,$$

$ISF_q(i)$ é o componente $i^{\text{ésimo}}$ do vetor ISF para um quadro atual,

$\text{última_}ISF_q(i)$ é o componente $i^{\text{ésimo}}$ do vetor ISF para um quadro anterior,

$ISF_{\text{meio}}(i)$ é o componente $i^{\text{ésimo}}$ do vetor que é uma combinação do meio adaptativo e os vetores ISF constantes do meio pré-determinado, e é calculado usando a fórmula:

$$ISF_{\text{meio}}(i) = \beta * ISF_{\text{meio_const}}(i) + (1 - \beta) * ISF_{\text{meio_adaptativo}}(i), \text{ para } i = 0,16,$$

onde $\beta = 0,75$, onde $ISF_{\text{meio_adaptativo}}(i) = \frac{1}{3} \sum_{i=0}^2 \text{última_}ISF_q(i)$ e é adaptado sempre

que $BFI=0$, onde BFI é um indicador de quadro ruim, e onde $ISF_{meio_const}(i)$ é o componente $i^{ésimo}$ de um vetor formado de uma média a longo prazo de vetores ISF .

BREVE DESCRIÇÃO DOS DESENHOS

Os objetos acima e outros, os aspectos e as vantagens da invenção tornar-se-ão aparentes ao considerar a descrição detalhada subsequente apresentada em conexão com os desenhos apensos, nos quais:

A Fig. 1 é um diagrama em blocos dos componentes do sistema de acordo com a técnica anterior para transmitir ou armazenar sinal de voz ou de áudio;

A Fig.2 é um gráfico ilustrando os coeficientes LSF [0 ... 4 kHz] dos quadros adjacentes no caso de voz estacionária, o eixo Y sendo a frequência e o eixo X sendo os quadros;

A Fig.3 é um gráfico ilustrando os coeficientes LSF [0...4kHz] dos quadros adjacentes no caso da voz não-estacionária, o eixo Y sendo a frequência e o eixo X sendo os quadros;

A Fig.4 é um gráfico ilustrando o erro de desvio espectral absoluto no método da técnica anterior;

A Fig.5 é um gráfico ilustrando o erro de desvio espectral absoluto na presente invenção (mostrando que a presente invenção fornece uma substituição melhor dos parâmetros espectrais do que o método da técnica anterior), onde a barra mais elevada no gráfico (indica o resíduo mais provável) é de aproximadamente zero;

A Fig. 6 é um fluxograma esquemático ilustrando como os bits são classificados de acordo com a técnica anterior quando um quadro ruim é detectado;

A Fig. 7 é o fluxograma do método completo da invenção; e

A Fig. 8 é um conjunto de dois gráficos ilustrando os aspectos do critério usado para determinar se ou não é aceitável a LSF do quadro indicado que possui erros.

DESCRIÇÃO DETALHADA DA INVENÇÃO

De acordo com a invenção, quando um quadro ruim é selecionado pelo decodificador após a transmissão do sinal de voz através do canal de comunicação (Fig. 1), os parâmetros espectrais corrompidos do sinal de voz são encobertos (ao substituir outros parâmetros por estes) baseados na análise dos parâmetros espectrais recentemente

comunicados através do canal de comunicação. É importante encobrir os parâmetros espectrais corrompidos do quadro ruim não apenas porque os parâmetros espectrais corrompidos podem causar artefatos (sons audíveis que não são obviamente voz), mas também porque a qualidade subjetiva dos quadros subsequentes de voz livre de erro diminui (pelo menos quando a quantização preditiva linear for usada).

A análise de acordo com a invenção também faz uso da natureza localizada do impacto espectral dos parâmetros espectrais, tais como as frequências de linha de espectral (LSFs). O impacto espectral LSFs é dito para estar localizado nesta se um parâmetro LSF é adversamente alterado por uma quantização e pelo processo de codificação, o espectro LP irá mudar apenas próximo da frequência representada pelo parâmetro LSF, deixando o resto do espectro inalterado.

A invenção em geral, para o quadro perdido ou o quadro corrompido.

De acordo com a invenção, um analisador determina o encobrimento do parâmetro espectral no caso do quadro ruim baseado no histórico de parâmetros de voz anteriormente recebido. O analisador determina o tipo do sinal de voz decodificada (isto é, se é estacionária ou não-estacionária). O histórico dos parâmetros de voz é usado para classificar o sinal de voz decodificado (como estacionário ou não, e mais especificamente, como vocodificado ou não); o histórico que é usado pode ser derivado principalmente dos valores mais recentes LTP e dos parâmetros espectrais.

Os termos sinal de voz estacionário e sinal de voz vocodificado são praticamente sinônimos; uma sequência de voz vocodificada é usualmente um sinal relativamente estacionário, enquanto que uma sequência de voz não-vocodificada não o é. Nós usamos a terminologia sinal de voz estacionário e não-estacionário aqui, pois essa terminologia é mais precisa.

Um quadro pode ser classificado como vocodificado ou não-vocodificado (e também estacionário e não-estacionário) de acordo com a razão da potência de excitação adaptativa sob o total de excitação, como indicado no quadro para a voz correspondente ao quadro. (Um quadro contém os parâmetros de acordo com os quais ambas a excitação adaptável e total são construídas; depois disto, a potência total pode ser calculada).

Se a sequência de voz for estacionária, os métodos da técnica anterior dos

quais os parâmetros espectrais corrompidos são encobertos, como indicado acima, não são particularmente eficazes. Isto é devido aos parâmetros espectrais adjacentes estacionários estarem mudando vagarosamente, então os valores anteriores espectrais bons (os valores não corrompido ou espectral perdido) são usualmente bem estimados para os próximos coeficientes espectrais, e mais especificamente, são melhores do que os parâmetros espectrais do quadro anterior conduzidos em direção ao meio constante, o qual a técnica anterior poderia usar no lugar dos parâmetros espectrais ruins (para encobrí-los). A Fig. 2 ilustra, um sinal de voz estacionário (e mais particularmente um sinal de voz vocodificado), as características de LSFs, como um exemplo dos parâmetros espectrais; esta ilustra os coeficientes LSF [0 ... 4 kHz] dos quadros adjacentes de voz estacionária, o eixo Y sendo a freqüência e o eixo X sendo os quadros, mostrando que as LSFs realizam a troca relativamente lenta, de quadro para quadro, para a voz estacionária.

Durante os segmentos da voz estacionária, o encobrimento é executado de acordo com a invenção (para os quadros perdidos ou corrompidos) usando os seguintes algoritmos:

Para $i = 0$ a $N-1$ (elementos dentro de um quadro):

$$\begin{aligned} & \text{meio_adaptativo_LSF_vetor}(i) \\ &= \text{última_LSF_boa}(i)(0) + \text{última_LSF_boa}(i)(1) + \dots + \text{última_LSF_boa}(i)(k-1)/k; \\ & \text{LSF_q1}(i) \\ &= \alpha * \text{última_LSF_boa}(i)(0) + (1-\alpha) * \text{meio_adaptativo_LSF}(i); \\ & (2.1) \\ & \text{LSF_q2}(i) = \text{LSF_q1}(i). \end{aligned}$$

onde α pode ser aproximadamente 0,95, N é a ordem do filtro LP, e k é o comprimento de adaptação. $\text{LSF_q1}(i)$ é o vetor LSF quantizado do segundo sub-quadro e $\text{LSF_q2}(i)$ é o vetor LSF quantizado do quarto sub-quadro. Os vetores LSF do primeiro e terceiro sub-quadros estão interpolados a partir destes dois vetores. A quantidade $\text{última_LSF_boa}(i)(0)$ é igual ao valor da quantidade $\text{LSF_q2}(i-1)$ do quadro bom anterior. A quantidade $\text{última_LSF_boa}(i)(n)$ é um componente do vetor dos parâmetros LSF do quadro anteriormente bom $n+1^{\text{ésimo}}$ (isto é, o quadro bom que precede o quadro

ruim atual por $n + 1$ quadros). Finalmente, a quantidade `meio_adaptativo_LSF (i)` é o meio (média aritmética) dos vetores bons LSF anteriores (isto é, é um componente da quantidade do vetor, cada componente sendo um meio dos componentes correspondentes dos vetores bons LSF anteriores).

5 Tem sido demonstrado que o método do meio adaptativo da invenção melhora a qualidade subjetiva da voz sintetizada comparado ao método da técnica anterior. A demonstração usou simulações onde a voz é transmitida através do canal de comunicação de indução de erro. Toda vez que um quadro ruim foi detectado, o erro espectral foi calculado. O erro espectral foi obtido ao subtrair, do espectro original, o

10 espectro que foi usado para encobrir durante o quadro ruim. O erro absoluto é calculado através do valor absoluto do erro espectral. As Fig. 4 e 5 mostram os histogramas do erro de desvio absoluto de LSFs da técnica anterior e do método inventado, respectivamente. O encobrimento do erro ótimo possui um erro próximo de zero, isto é, quando o erro é próximo de zero, os parâmetros espectrais usados para encobrir estão bem próximos dos

15 parâmetros espectrais originais (corrompidos ou perdidos). Como pode ser visto dos histogramas das Figs. 4 e 5, o método do meio adaptativo da invenção (Fig. 5) encobre os erros melhor do que o método da técnica anterior (Fig. 4) durante as seqüências de voz estacionária.

Como mencionado acima, os coeficientes espectrais dos sinais não-estacionários (ou, menos precisamente, sinais não-vocodificados) flutuam entre os quadros adjacentes, como indicado na Fig. 3, a qual é um gráfico ilustrando as LSFs dos quadros adjacentes no caso da voz não-estacionária, o eixo Y sendo a freqüência e o eixo X sendo os quadros. Neste caso, o método de encobrimento ótimo não é o mesmo do caso do sinal de voz estacionária. Para a voz não-estacionária, a invenção proporciona o encobrimento

20 para os segmentos ruins de voz não-estacionária (corrompidos ou perdidos) de acordo com o algoritmo seguinte (o algoritmo não-estacionário):

Para $i = 0$ a $N-1$:

$$\begin{aligned} &\text{meio_parcialmente_adaptativo_LSF (i)} \\ &= \beta * \text{meio_LSF(i)} + (1-\beta) * \text{meio_adaptativo_LSF (i);} \\ &(2.3) \end{aligned}$$

$$\begin{aligned}
 & \text{LSF_q1 (i)} \\
 & = \alpha * \text{última_LSF_boa (i)} (0) + (1-\alpha) * \text{meio_parcialmente_adaptativo_LSF (i)}; \\
 & (2.2) \\
 & \text{LSF_q1 (i)} = \text{LSF_q2 (i)};
 \end{aligned}$$

5 onde N é a ordem do filtro LP, e α é, tipicamente, aproximadamente 0,90, onde LSF_q1 (i) e LSF_q2 (i) são dois conjuntos de vetores LSF para o quadro atual como na equação (2.1), onde o último LSF_q(i) é LSF_q2 (i) do quadro bom anterior, onde meio_parcialmente_adaptativo_LSF (i) é uma combinação do vetor LSF no meio adaptativo e o vetor LSF médio, e onde meio_adaptativo_LSF (i) é o meio dos últimos

10 vetores LSF bons K (o qual é atualizado quando o BFI não for estabelecido), e onde meio_LSF(i) é uma LSF média constante e é gerada durante o processo de projeto do codec sendo usado para a voz sintetizada; este é um LSF médio de alguma base de dado. O parâmetro β é tipicamente de aproximadamente 0,75, um valor usado para expressar a extensão pela qual a voz é estacionária como oposta a não-estacionária. (O

15 valor é algumas vezes calculado baseado na relação da energia de excitação da predição a longo prazo pela energia de excitação fixa do livro-código, ou mais precisamente, usando a fórmula:

$$\beta = \frac{1 + \text{Fator de voz}}{2}$$

20 em que:

$$\text{Fator de voz} = \frac{\text{energia}_{\text{tom}} - \text{energia}_{\text{inovação}}}{\text{energia}_{\text{tom}} + \text{energia}_{\text{inovação}}}$$

em que a energia_{tom} é a energia de excitação do tom e a energia_{inovação} é a energia da

25 inovação da excitação do código. Quando a maior parte da energia for uma excitação de predição a longo prazo, a voz sendo decodificada é principalmente estacionária. Quando a maior parte da energia for uma excitação fixa do livro-código, a voz é principalmente não-estacionária).

Para $\beta = 1,0$, a equação (2.3) reduz a equação (1.0), a qual é usada pela técnica

30 anterior. Para $\beta = 0,0$, a equação (2.3) reduz para a equação (2.1), a qual é usada pela presente invenção para os segmentos estacionários. Para as implementações sensíveis

á complexidade (nas aplicações onde é importante manter a complexidade a um nível razoável), β pode ser estabelecido para algum valor de compromisso, por exemplo, de 0,75, para ambos os segmentos estacionários e não-estacionários. O encobrimento do parâmetro espectral especificamente para quadros perdidos.

No caso do quadro perdido, apenas a informação dos últimos parâmetros espectrais está disponível. Os parâmetros espectrais substituídos são calculados de acordo com um critério baseado nos históricos do parâmetro de, por exemplo, os valores espectrais e LTP (predição a longo prazo – long-term prediction); os parâmetros LTP incluem o valor de ganho LTP e o valor de retardo. LTP representa a correlação do quadro atual para o quadro anterior. Por exemplo, o critério usado para calcular os parâmetros espectrais substituídos podem distinguir situações onde as últimas LSFs boas deveriam ser modificadas por um meio LSF adaptativo ou, como na técnica anterior, por um meio constante.

Encobrimento do parâmetro espectral alternativo para quadros corrompidos:

Quando o quadro de voz é corrompido (quando oposto ao perdido), o procedimento de encobrimento da invenção pode ser também otimizado. Neste caso, os parâmetros podem estar completamente ou parcialmente corretos quando recebidos no decodificador de voz. Por exemplo, a conexão baseada em pacote (como em uma conexão Internet ordinária TCP/IP), o método de encobrimento dos quadros corrompidos não é usualmente possível, pois com as conexões do tipo TCP/IP usualmente todos os quadros ruins são quadros perdidos, porém para outros tipos de conexões, tal como as conexões comutadas por circuito GSM ou EDGE, o método de encobrimento dos quadros corrompidos da invenção pode ser usado. Desse modo, para as conexões comutadas por pacote, o seguinte método alternativo não pode ser usado, porém para as conexões comutadas por circuito, esta pode ser usada, desde que em tais quadros ruins de conexões sejam pelo menos (e de fato usualmente) quadros corrompidos.

De acordo com as especificações GSM, um quadro ruim é detectado quando o indicador BFI é estabelecido seguindo uma verificação CRC ou outro mecanismo de detecção de erro usado no processo de decodificação do canal. Os mecanismos de

detecção de erro são usados para detectar os erros nos bits subjetivamente mais significativos, isto é, nos bits que possuem um efeito mais elevado na qualidade de voz sintetizada. Em alguns dos métodos da técnica anterior, estes bits mais significativos não são usados quando o quadro é indicado para ser um quadro ruim. Entretanto, um quadro

5 pode possuir apenas uns poucos erros de bit (mesmo que seja suficiente para estabelecer o indicador BFI), então todo o quadro poderia ser descartado, embora a maioria dos bits estivesse correto. Uma verificação CRC detecta simplesmente se ou não um quadro possui quadros enormes, mas não faz nenhuma estimativa do BER (taxa de erro em bit). A Fig.6 ilustra como os bits são classificados de acordo com a técnica anterior quando um quadro

10 ruim é detectado. Na Fig. 6, um único quadro é mostrado sendo comunicado, um bit de cada vez (da esquerda para direita), para um decodificador sobre o canal de comunicações com condições tais que alguns bits do quadro incluídos na verificação CRC sejam corrompidos, e então o BFI é estabelecido para um.

Como pode ser visto da Fig. 6, mesmo quando o quadro recebido as vezes

15 contém alguns bits corretos (o BER em um quadro usualmente sendo pequeno quando as condições de canal são relativamente boas), a técnica anterior não os usa. Em contraste, a presente invenção tenta estimar se os parâmetros recebidos são corrompidos e se não são, o método inventado os usa.

A Tabela 1 demonstra a idéia anterior do encobrimento do quadro

20 corrompido de acordo com a invenção no exemplo de um decodificador de banda larga (WB - wideband) de múltiplas-taxas adaptativas (AMR).

	C/I [dB]				
Modo 12.65 (AMW WB)	10	9	8	7	6
BER	3,72%	4,58%	5,56%	6,70%	7,98%
FER	0,30%	0,74%	1,62%	3,45%	7,16%
Índices de parâmetro espectral correto	84%	77%	68%	64%	60%
Espectro totalmente correto	47%	38%	32%	27%	24%

Tabela 1. Porcentagem de parâmetros espectrais corretos em um quadro de voz corrompido.

No caso do decodificador AMR WB, o modo 12,65 kbits/s é uma boa

25 escolha par usar quando a relação portadora do canal para interferência (C/I) está na faixa de aproximadamente 9 dB a 10 dB. A partir da Tabela 1, pode ser visto que no caso das

condições de canal GSM com uma C/I na faixa de 9 a 10 dB ao usar o esquema de modulação GSMK (Modulação por Desvio Mínimo Gaussiano), aproximadamente de 30-50% dos quadros ruins recebidos possuem um espectro totalmente correto. Também, aproximadamente 75-85% de todos os coeficientes do parâmetro espectral de quadro ruim estão corretos. Devido à natureza localizada do impacto espectral, como mencionado anteriormente, a informação do parâmetro do espectro pode ser usada nos quadros ruins. As condições do canal com uma C/I na faixa de 6-8 dB ou menos são tão pobres que o modo de 12,65 kbit/s não poderia ser usado.

A idéia básica da presente invenção no caso dos quadros corrompidos é que de acordo com o critério (descrito abaixo), os bits de canal do quadro corrompido são usados para decodificar o quadro corrompido. O critério para os coeficientes espectrais é baseado nos últimos valores dos parâmetros de voz do sinal sendo decodificado. Quando um quadro ruim é detectado, as LSFs recebidas ou outros parâmetros espectrais comunicados através do canal são usados se o critério for encontrado; em outras palavras, se as LSFs recebidas encontram o critério, elas são usadas na decodificação da mesma maneira que elas seriam se o quadro não fosse um quadro ruim. Por outro lado, isto é, se as LSFs do canal não encontram o critério, o espectro para o quadro ruim é calculado de acordo com o método de encobrimento descrito acima, usando equações (2.1) ou (2.2). O critério para aceitar os parâmetros espectrais pode ser implementado para usar, por exemplo, o cálculo da distância espectral, tal como o cálculo da chamada distância espectral de Itakura-Saito. (Ver, por exemplo, a página 329 do artigo Processamento de Tempo-Discreto dos Sinais de Voz de John R Deller Jr, John H. L. Hansen, e John G. Proakis, publicado pela editora IEEE, 2000).

O critério para aceitação dos parâmetros espectrais a partir do canal deve ser muito rigoroso no caso do sinal de voz estacionária. Como mostrado na Fig.3, os coeficientes espectrais são muitos estáveis durante a sequência estacionária (por definição) de modo que as LSFs corrompidas (ou outros parâmetros de voz) do sinal de voz estacionário pode ser usualmente detectado (desde que eles sejam distinguíveis das LSFs não corrompidas com base em que elas diferem dramaticamente das LSFs dos quadros adjacentes não-corrompidos). Por outro lado, para um sinal de voz não-estacionária, o

critério não necessita ser tão rígido, o espectro para o sinal de voz não-estacionária é permitido possuir uma variação maior. Para um sinal de voz não-estacionária, a exatidão dos parâmetros espectrais corretos não é rígida em relação aos artefatos audíveis, desde que para voz não-estacionária (isto é, mais ou menos voz não-vocodificada), os artefatos não audíveis são provavelmente indiferentes se os parâmetros de voz são ou não corretos. Em outras palavras, mesmo se os bits dos parâmetros espectrais forem corrompidos, eles podem ainda ser aceitos de acordo com o critério, uma vez que os parâmetros espectrais para voz não-estacionária com alguns bits corrompidos não irão usualmente gerar quaisquer artefatos audíveis. De acordo com a invenção, a qualidade subjetiva da voz sintetizada é para ser diminuída o menos possível no caso dos quadros corrompidos ao usar toda a informação disponível sobre as LSFs recebidas, e ao selecionando qual LSFs a usar de acordo com as características da voz sendo carregada.

Desse modo, embora a invenção inclua um método para encobrir os quadros corrompidos, também compreende como uma alternativa usar um critério no caso do quadro corrompido carregando voz não-estacionária, a qual, se encontrada, irá fazer com que o decodificador use o quadro corrompido como está; em outras palavras, embora o BFI seja estabelecido, o quadro será usado. O critério é na essência um limiar usado para distinguir entre um quadro corrompido que é útil e um que não é; o limiar é baseado em quantos parâmetros espectrais do quadro corrompido diferem dos parâmetros espectrais dos quadros bons recentemente recebidos.

O uso de possíveis parâmetros espectrais corrompidos é provavelmente mais sensível aos artefatos audíveis do que o uso de outros parâmetros corrompidos, tais como os valores de retardo de LTP corrompidos. Por esta razão, o critério usado para determinar se ou não usar possivelmente um parâmetro espectral corrompido deveria ser especialmente confiável. Em algumas configurações, é vantajoso usar como critério, uma distância espectral máxima (do parâmetro espectral correspondente ao quadro anterior, além do qual o parâmetro espectral suspeito não é para ser usado); em tal incorporação, o cálculo da distância Itakura-Saito bem conhecido poderia ser usado para quantificar a distância espectral para ser comparada com o limiar. Alternativamente, as estatísticas adaptativas ou fixas dos parâmetros espectrais poderiam ser usadas para determinar se ou

não usar os parâmetros espectrais possivelmente corrompidos. Outros parâmetros de voz também, tais como os parâmetros de ganho, podem ser usados para gerar o critério. (Se outros parâmetros de voz não são drasticamente diferentes do quadro atual, quando comparados com os valores no quadro bom mais recente, então os parâmetros são provavelmente bons para uso, ao fornecer os parâmetros espectrais recebidos que também encontram o critério. Em outras palavras, outros parâmetros, tais como os ganhos de LTP, podem ser usados como um componente adicional para estabelecer o critério próprio para determinar se ou não usar os parâmetros espectrais recebidos. O histórico dos outros parâmetros de voz pode ser usado para reconhecimento melhorado das características de voz. Por exemplo, o histórico pode ser usado para decidir se a seqüência de voz decodificada possui uma característica estacionária ou não-estacionária. Quando as propriedades da seqüência de voz decodificada são conhecidas, é mais fácil detectar os parâmetros espectrais possivelmente corretos do quadro corrompido e é mais fácil estimar que tipo de valores de parâmetro espectrais são esperados para serem carregados no quadro corrompido recebido).

De acordo com a invenção em uma incorporação preferida, e agora referindo-se a Fig. 8, o critério para determinar se ou não usar um parâmetro espectral para um quadro corrompido é baseado na noção da distância espectral, como mencionado acima. Mais especificamente, para determinar se o critério para aceitação dos coeficientes de LSF de um quadro corrompido for encontrado, o processador do receptor executa o algoritmo que verifica quantos coeficientes LSF têm movido ao longo do eixo da frequência comparado aos coeficientes LSF do último quadro bom, o qual é armazenado na memória LSF, junto com os coeficientes LSF de um número pré-determinado de quadros anteriores mais recentes.

O critério de acordo com a incorporação preferida envolve realizar uma ou mais de quadro comparações: uma comparação interquadro, uma comparação intraquadro, uma comparação de dois pontos, e uma comparação de ponto único.

Na primeira comparação, a comparação interquadro, as diferenças entre os elementos do vetor LSF nos quadros adjacentes do quadro corrompido são comparados às diferenças correspondentes dos quadros anteriores. As diferenças são determinadas como

se segue:

$$d_n(i) = |L_{n-1}(i) - L_n(i)|, \quad 1 \leq i \leq P-1,$$

onde P é o número de coeficientes espectrais para o quadro, $L_n(i)$ é o elemento LSF $i^{\text{ésimo}}$ do quadro corrompido, e $L_{n-1}(i)$ é o elemento LSF $i^{\text{ésimo}}$ do quadro antes do quadro corrompido. O elemento LSF, $L_n(i)$ do quadro corrompido é descartado se a diferença, $d_n(i)$, for tão alta comparada a $d_{n-1}(i)$, $d_{n-2}(i)$, ..., $d_{n-k}(i)$, onde k é o comprimento da memória LSF.

A segunda comparação, a comparação intraquadro, é uma comparação da diferença entre os elementos do vetor LSF adjacentes no mesmo quadro. A distância entre o elemento candidato LSF $i^{\text{ésimo}}$, o $L_n(i)$, e o elemento LSF $(i-1)^{\text{ésimo}}$, $L_{n-1}(i)$, do quadro $n^{\text{ésimo}}$ é determinado como se segue:

$$e_n(i) = L_n(i-1) - L_n(i), \quad 2 \leq i \leq P-1,$$

onde P é o número de coeficientes espectrais e $e_n(i)$ é a distância entre os elementos LSF. As distâncias são calculadas entre todos os elementos do vetor LSF do quadro. Um ou outro ou ambos os elementos LSF $L_n(i)$ e $L_n(i-1)$ será ou serão descartado(s) se a diferença $e_n(i)$, for tão grande ou tão pequena comparada a $e_{n-1}(i)$, $e_{n-2}(i)$, ..., $e_{n-k}(i)$.

A terceira comparação, a comparação de dois pontos, determina se uma intercessão tem ocorrido envolvendo o elemento candidato $L_n(i)$ de LSF, isto é, se um elemento $L_n(i-1)$ que é inferior para ao elemento candidato que possui um valor maior do que o elemento candidato $L_n(i)$ de LSF. A intercessão indica um ou mais valores LSF altamente corrompidos. Todos os elementos de intercessão são usualmente descartados.

A quarta comparação, a comparação de ponto único, compara o valor do elemento candidato do vetor LSF, $L_n(i)$ a um elemento LSF mínimo, $L_{\min}(i)$, e o elemento LSF máximo, $L_{\max}(i)$, ambos calculados a partir da memória, e descarta o elemento LSF candidato se este estiver fora da área delimitada pelos elementos LSF mínimo e máximo.

Se um elemento LSF de um quadro corrompido é descartado (baseado no critério acima ou de outra maneira), então um novo valor para o elemento LSF é calculado de acordo com o algoritmo usando a equação (2.2).

Referindo agora a Fig. 7, um fluxograma do método completo da invenção é mostrado, indicando as diferentes provisões para os quadros de voz estacionários e não-estacionários, e para os corrompidos quando oposto aos quadros de voz perdidos não-estacionários.

5

Discussão:

A invenção pode ser aplicada em um decodificador de voz ou em uma estação móvel ou em um elemento de rede móvel. Pode também ser aplicada a qualquer decodificador de voz usado em um sistema que possui um canal de transmissão errôneo.

Escopo da Invenção:

10

É para ser entendido que as incorporações descritas acima são apenas ilustrativas dos princípios do pedido da presente invenção. Em particular, deve ser entendido que embora a invenção tenha sido mostrada e descrita usando os pares de linha de espectro para uma ilustração concreta, a invenção também pode usar outros parâmetros equivalentes, tais como os pares de imitância espectral. Numerosas modificações e incorporações alternativas podem ser visualizadas pelo técnico no assunto sem se afastar do conceito inventivo e escopo da presente invenção, e as reivindicações apenas são pretendidas para cobrir tais modificações e incorporações.

15

REIVINDICAÇÕES

1. Método para encobrir os efeitos dos erros de quadro nos quadros a serem decodificados pelo decodificador para proporcionar uma voz sintetizada, os quadros sendo fornecidos pelo canal de comunicação para o decodificador, cada quadro
5 fornecendo os parâmetros usados pelo decodificador na sintetização de voz, **caracterizado** pelo fato de que compreende as seguintes etapas:

a) determinar se o quadro é um quadro ruim; e

b) proporcionar uma substituição dos parâmetros espectrais do quadro ruim baseada somente nos parâmetros espectrais para os quadros bons recentemente e
10 anteriormente recebidos e incluindo pelo menos uma média parcialmente adaptativa dos parâmetros espectrais de um número pré-determinado da maioria dos quadros bons recentemente e anteriormente recebidos.

2. Método de acordo com a reivindicação 1, **caracterizado** pelo fato de que compreende ainda a etapa de determinar se o quadro ruim carrega voz não-estacionária ou estacionária, e sendo que a etapa de proporcionar a substituição do quadro ruim é
15 realizada de uma maneira que depende se o quadro ruim carrega voz não-estacionária ou estacionária.

3. Método de acordo com a reivindicação 2, **caracterizado** pelo fato de que, no caso de um quadro ruim carregando voz estacionária, a etapa de proporcionar a substituição do quadro ruim é realizada usando a média dos parâmetros de um número
20 pré-determinado dos quadros bons recentemente recebidos.

4. Método de acordo com a reivindicação 3, **caracterizado** pelo fato de que, no caso de um quadro ruim carregando uma voz não-estacionária e no caso de um filtro de predição linear (LP) estar sendo usado, a etapa de proporcionar a substituição do
25 quadro ruim é realizada de acordo com o algoritmo:

Para $i = 0$ a $N-1$:

meio_adaptativo_LSF_vetor (i)

$= (\text{última_LSF_boa} (i) (0) + \text{última_LSF_boa} (i) (1) + \dots +$

$\text{última_LSF_boa} (i) (k-1))/k;$

LSF_q1(i)

$= \alpha * \text{última_LSF_boa} (i) (0) + (1-\alpha) * \text{meio_adaptativo_LSF} (i);$

(2.1)

$$\text{LSF_q2}(i) = \text{LSF_q1}(i).$$

onde α é um parâmetro pré-determinado, em que N é a ordem do filtro LP, sendo que k é o comprimento de adaptação, em que $\text{LSF_q1}(i)$ é o vetor LSF quantizado do segundo sub-quadro e $\text{LSF_q2}(i)$ é o vetor LSF quantizado do quarto sub-quadro, sendo que $\text{última_LSF_boa}(i)(0)$ é igual ao valor da quantidade $\text{LSF_q2}(i-1)$ do quadro bom anterior, onde $\text{última_LSF_boa}(i)(n)$ é um componente do vetor dos parâmetros LSF do quadro bom anterior $n + 1^{\text{ésimo}}$, e em que $\text{meio_adaptativo_LSF}(i)$ é a média dos vetores bons LSF anteriores.

5. Método de acordo com a reivindicação 2, **caracterizado** pelo fato de que, no caso de um quadro ruim carregando uma voz não estacionária, a etapa de proporcionar a substituição do quadro ruim é realizada usando, quando muito, uma parte pré-determinada da média dos parâmetros de um número pré-determinado dos quadros bons recebidos recentemente.

6. Método de acordo com a reivindicação 2, **caracterizado** pelo fato de que, no caso de um quadro ruim carregando uma voz não-estacionária e no caso de um filtro de predição linear (LP) estar sendo usado, a etapa de proporcionar a substituição do quadro ruim é realizada de acordo com o algoritmo:

Para $i = 0$ a $N-1$:

$$\text{meio_parcialmente_adaptativo_LSF}(i)$$

$$= \beta * \text{meio_LSF}(i) + (1-\beta) * \text{meio_adaptativo_LSF}(i);$$

(2.3)

$$\text{LSF_q1}(i)$$

$$= \alpha * \text{última_LSF_boa}(i)(0) + (1-\alpha) * \text{meio_parcialmente_adaptativo_LSF}$$

(i);

(2.2)

$$\text{LSF_q1}(i) = \text{LSF_q2}(i);$$

onde N é a ordem do filtro LP, em que α e β são, particularmente parâmetros pré-determinados, em que $\text{LSF_q1}(i)$ é o vetor LSF quantizado do segundo sub-quadro e $\text{LSF_q2}(i)$ é o vetor LSF quantizado do quarto sub-quadro, em que o $\text{último_LSF_q}(i)$ é o valor de $\text{LSF_q2}(i)$ do quadro bom anterior, em que $\text{meio_parcialmente_adaptativo_LSF}(i)$ é uma combinação do vetor LSF no meio

adaptativo e o vetor LSF médio, em que o meio_adaptativo_LSF (i) é a média dos últimos vetores LSF bons K, e em que meio_LSF (i) é uma LSF média constante.

7. Método de acordo com a reivindicação 1, **caracterizado** pelo fato de que compreende ainda a etapa de determinar se o quadro ruim encontra um critério pré-determinado, e se sim, usar o quadro ruim ao invés de substituir o quadro ruim.

8. Método de acordo com a reivindicação 7, **caracterizado** pelo fato de que o critério pré-determinado envolve preparar uma ou mais de quatro comparações: uma comparação interquadro, uma comparação intraquadro, uma comparação de dois pontos, e uma comparação de ponto único.

9. Método para encobrir os efeitos dos erros de quadro nos quadros a serem decodificados pelo decodificador para proporcionar voz sintetizada, os quadros sendo fornecidos pelo canal de comunicação para o decodificador, cada quadro fornecendo os parâmetros usados pelo decodificador na sintetização de voz, **caracterizado** pelo fato de que compreende as etapas de:

a) determinar se o quadro é um quadro ruim; e

b) proporcionar a substituição dos parâmetros do quadro ruim, a substituição na qual as últimas frequências de imitação espectral (ISF) são desviadas em direção a um meio parcialmente adaptativo fornecido por:

$$\text{ISF}_q(i) = \alpha * \text{última_ISF}_q(i) + (1 - \alpha) * \text{ISF}_{\text{meio}}(i), \quad \text{para } i = 0, 16,$$

em que:

$$\alpha = 0,9$$

ISF_q(i) é o componente i^{ésimo} do vetor ISF para o quadro atual,

última_ISF_q(i) é o componente i^{ésimo} do vetor ISF para o quadro anterior,

ISF_{meio}(i) é o componente i^{ésimo} do vetor que é uma combinação do meio

adaptativo e dos vetores ISF constantes do meio pré-determinado, e é calculado usando a fórmula:

$$\text{ISF}_{\text{meio}}(i) = \beta * \text{ISF}_{\text{meio_const}}(i) + (1 - \beta) * \text{ISF}_{\text{meio_adaptativo}}(i), \quad \text{para } i = 0, 16$$

em que:

$$\beta = 0,75,$$

em que: $\text{ISF}_{\text{meio_adaptativo}}(i) = \frac{1}{3} \sum_{i=0}^2 \text{última_ISF}_q(i)$ e é adaptado sempre que

BFI = 0, em que BFI é um indicador de quadro ruim, e sendo que $ISF_{meio_const}(i)$ é o componente $i^{ésimo}$ do vetor formado da média a longo prazo dos vetores ISF.

10. Aparelho para encobrir os efeitos dos erros de quadro nos quadros a serem decodificados pelo decodificador para proporcionar a voz sintetizada, os quadros sendo fornecidos por um canal de comunicação para o codificador, cada quadro fornecendo parâmetros usados pelo decodificador na sintetização da voz, **caracterizado** pelo fato de que compreende:

a) um dispositivo para determinar se o quadro é um quadro ruim; e

b) um dispositivo para proporcionar uma substituição dos parâmetros espectrais do quadro ruim baseado somente nos parâmetros espectrais para os quadros bons recentemente e anteriormente recebidos e incluindo pelo menos uma média parcialmente adaptativa dos parâmetros espectrais de um número pré-determinado da maioria dos quadros bons recentemente e anteriormente recebidos.

11. Aparelho de acordo com a reivindicação 10, **caracterizado** pelo fato de que compreende ainda um dispositivo para determinar se o quadro ruim carrega voz não-estacionária ou estacionária, e sendo que o dispositivo para proporcionar a substituição do quadro ruim realiza a substituição de uma maneira que depende se o quadro ruim carrega voz não-estacionária ou estacionária.

12. Aparelho de acordo com a reivindicação 11, **caracterizado** pelo fato de que, no caso do quadro ruim carregando voz estacionária, o dispositivo para proporcionar a substituição do quadro ruim usa a média dos parâmetros de um número pré-determinado dos quadros bons recentemente recebidos.

13. Aparelho de acordo com a reivindicação 12, **caracterizado** pelo fato de que, no caso do quadro ruim carregando uma voz não-estacionária e no caso de um filtro de predição linear (LP) estar sendo usado, o dispositivo para proporcionar a substituição do quadro ruim é operativo de acordo com o algoritmo:

Para $i = 0$ a $N-1$:

$meio_adaptativo_LSF_vetor(i)$

$= (última_LSF_boa(i)(0) + última_LSF_boa(i)(1) + \dots + última_LSF_boa(i)(k-1))/k;$

$LSF_q1(i)$

$$= \alpha * \text{última_LSF_boa} (i) (0) + (1-\alpha) * \text{meio_adaptativo_LSF} (i);$$

(2.1)

$$\text{LSF_q2} (i) = \text{LSF_q1} (i),$$

em que α é um parâmetro pré-determinado, em que N é a ordem do filtro LP, em que k é o comprimento de adaptação, sendo que LSF_q1 (i) é o vetor LSF quantizado do segundo sub-quadro e LSF_q2 (i) é o vetor LSF quantizado do quarto sub-quadro, em que última_LSF_boa (i) (0) é igual ao valor da quantidade LSF_q2 (i-1) do quadro bom anterior, em que última_LSF_boa (i) (n) é um componente do vetor dos parâmetros LSF do quadro n + 1^{ésimo} bom anterior, e em que meio_adaptativo_LSF (i) é a média dos vetores bons LSF anteriores.

14. Aparelho de acordo com a reivindicação 11, **caracterizado** pelo fato de que, no caso do quadro ruim carregando uma voz não-estacionária, o dispositivo para proporcionar a substituição do quadro ruim usa, quando muito, uma parte pré-determinada da média dos parâmetros de um número pré-determinado dos quadros bons recebidos recentemente.

15. Aparelho de acordo com a reivindicação 11, **caracterizado** pelo fato de que, no caso do quadro ruim carregando uma voz não-estacionária e no caso de um filtro de predição linear (LP) estar sendo usado, o dispositivo para proporcionar a substituição do quadro ruim é operativo de acordo com o algoritmo:

Para i = 0 a N-1:

$$\begin{aligned} & \text{meio_parcialmente_adaptativo_LSF} (i) \\ &= \beta * \text{meio_LSF}(i) + (1-\beta) * \text{meio_adaptativo_LSF} (i); \end{aligned}$$

(2.3)

$$\text{LSF_q1} (i)$$

$$= \alpha * \text{última_LSF_boa} (i) (0) + (1-\alpha) * \text{meio_parcialmente_adaptativo_LSF} (i);$$

(2.2)

$$\text{LSF_q1} (i) = \text{LSF_q2} (i);$$

em que N é a ordem do filtro LP, em que α e β são, particularmente, parâmetros pré-determinados, em que LSF_q1 (i) é o vetor LSF quantizado do segundo sub-quadro e LSF_q2 (i) é o vetor LSF quantizado do quadro sub-quadro, em que o último_LSF_q (i) é o valor de LSF_q2 (i) do quadro bom anterior, em que

meio_parcialmente_adaptativo_LSF (i) é uma combinação do vetor LSF no meio adaptativo e o vetor LSF meio, em que o meio_adaptativo_LSF (i) é a média dos últimos vetores LSF bons K, e em que meio_LSF (i) é uma LSF média constante.

16. Aparelho de acordo com a reivindicação 11, **caracterizado** pelo fato de que compreende ainda um dispositivo para determinar se o quadro ruim encontra um critério pré-determinado, e se sim, usar o quadro ruim ao invés de substituir o quadro ruim.

17. Aparelho de acordo com a reivindicação 16, **caracterizado** pelo fato de que o critério pré-determinado envolve preparar uma ou mais de quatro comparações: uma comparação interquadro, uma comparação intraquadro, uma comparação de dois pontos, e uma comparação de ponto único.

18. Aparelho para encobrir os efeitos dos erros dos quadros nos quadros a serem decodificados pelo decodificador para proporcionar voz sintetizada, os quadros sendo fornecidos pelo canal de comunicação para o decodificador, cada quadro fornecendo os parâmetros usados pelo decodificador na sintetização de voz, **caracterizado** pelo fato de que compreende:

a) dispositivo para determinar se o quadro é um quadro ruim; e

b) dispositivo para proporcionar a substituição dos parâmetros do quadro ruim, a substituição na qual as últimas frequências de imitância espectral (ISF) são desviadas em direção a um meio parcialmente adaptativo fornecido por:

$$\text{ISF}_q(i) = \alpha * \text{última_ISF}_q(i) + (1 - \alpha) * \text{ISF}_{\text{meio}}(i), \quad \text{para } i = 0, 16,$$

em que:

$$\alpha = 0,9$$

ISF_q(i) é o componente $i^{\text{ésimo}}$ do vetor ISF para o quadro atual,

última_ISF_q(i) é o componente $i^{\text{ésimo}}$ do vetor ISF para o quadro anterior,

ISF_{meio}(i) é o componente $i^{\text{ésimo}}$ do vetor que é uma combinação do meio adaptativo e dos vetores ISF constantes do meio pré-determinado, e é calculado usando a fórmula:

$$\text{ISF}_{\text{meio}}(i) = \beta * \text{ISF}_{\text{meio_const}}(i) + (1 - \beta) * \text{ISF}_{\text{meio_adaptativo}}(i), \quad \text{para } i = 0, 16,$$

em que $\beta = 0,75$, em que $\text{ISF}_{\text{meio_adaptativo}}(i) = \frac{1}{3} \sum_{i=0}^2 \text{última_ISF}_q(i)$ e é adaptado

sempre que $BFI = 0$, onde BFI é um indicador de quadro ruim, e sendo que ISF_{meio_const} (i) é o componente 1^{ésimo} de um vetor formado da média a longo prazo dos vetores ISF.

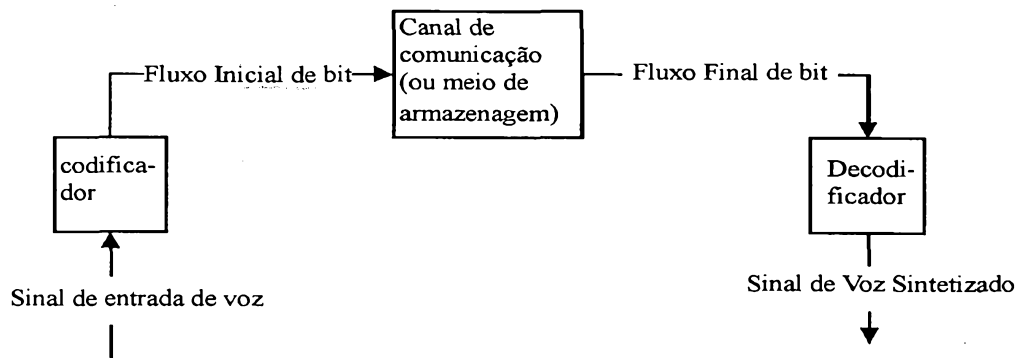


Fig. 1
(Técnica Anterior)

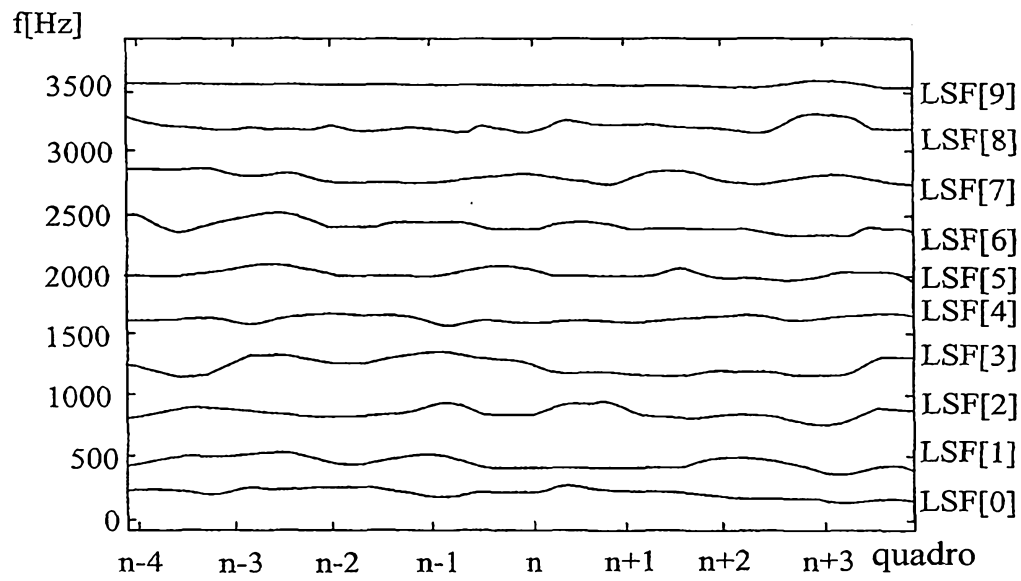


Fig. 2

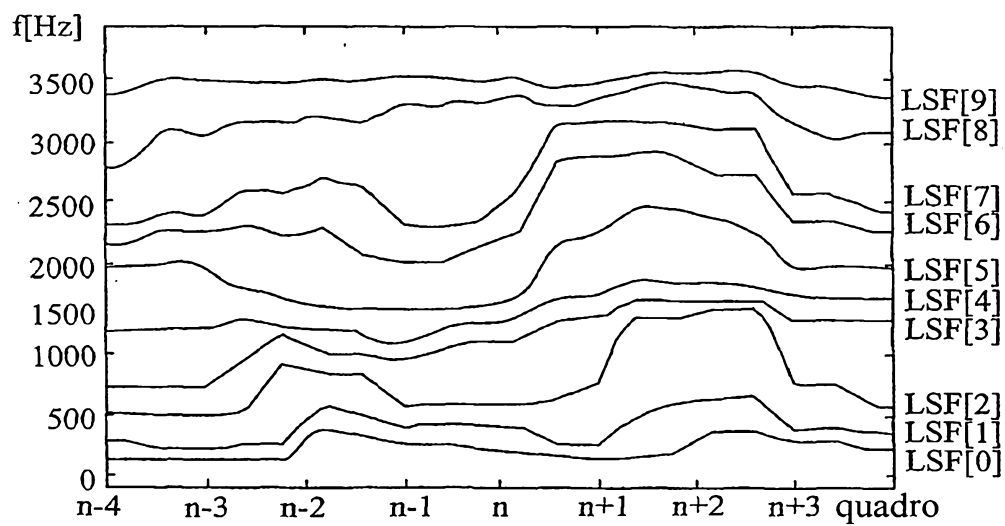


Fig. 3

3/6

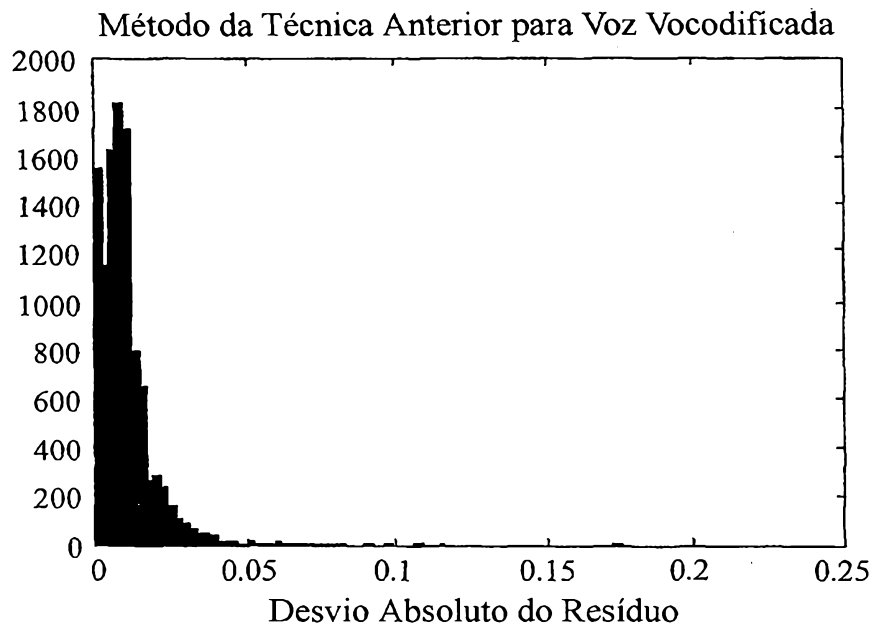


Fig. 4

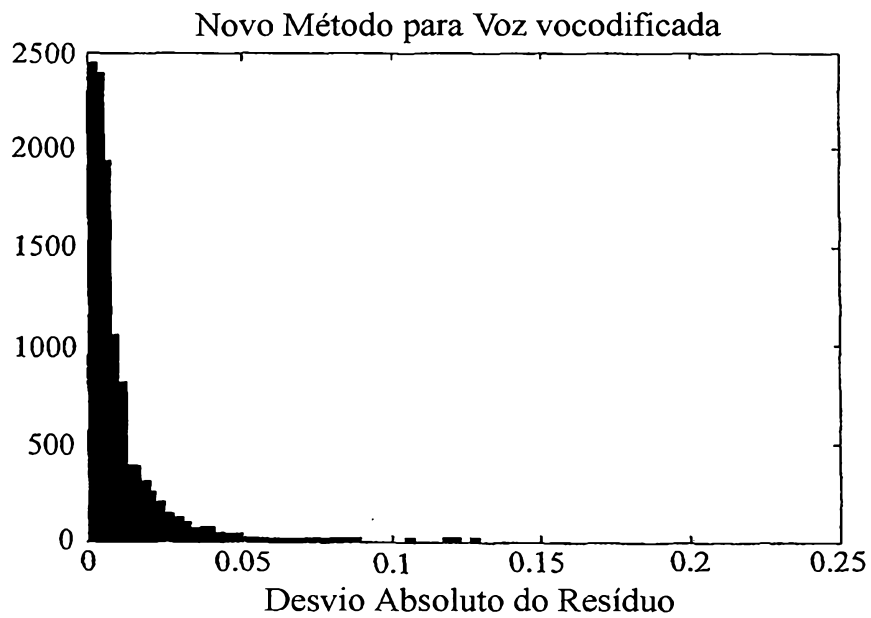


Fig. 5

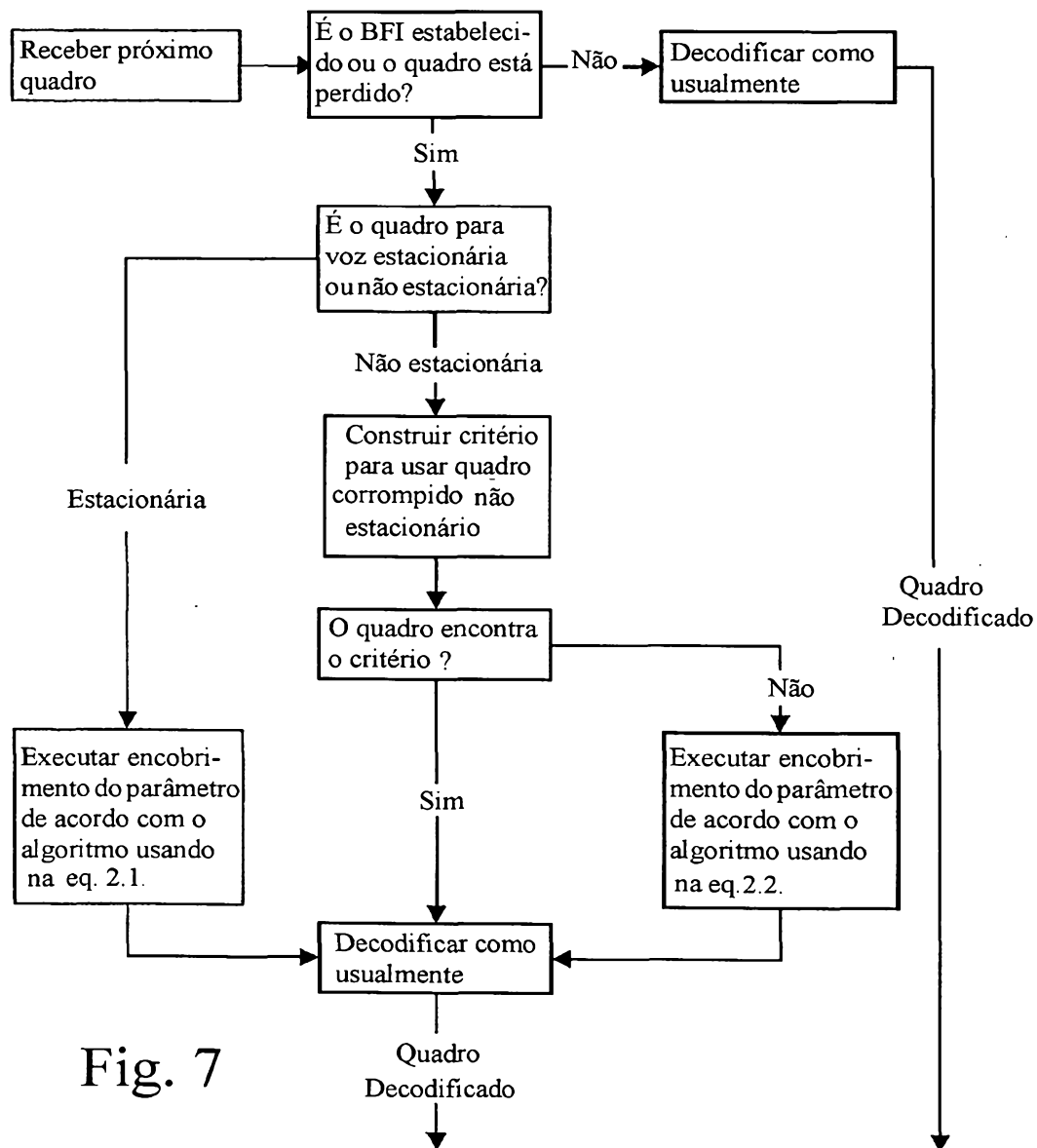


Fig. 7

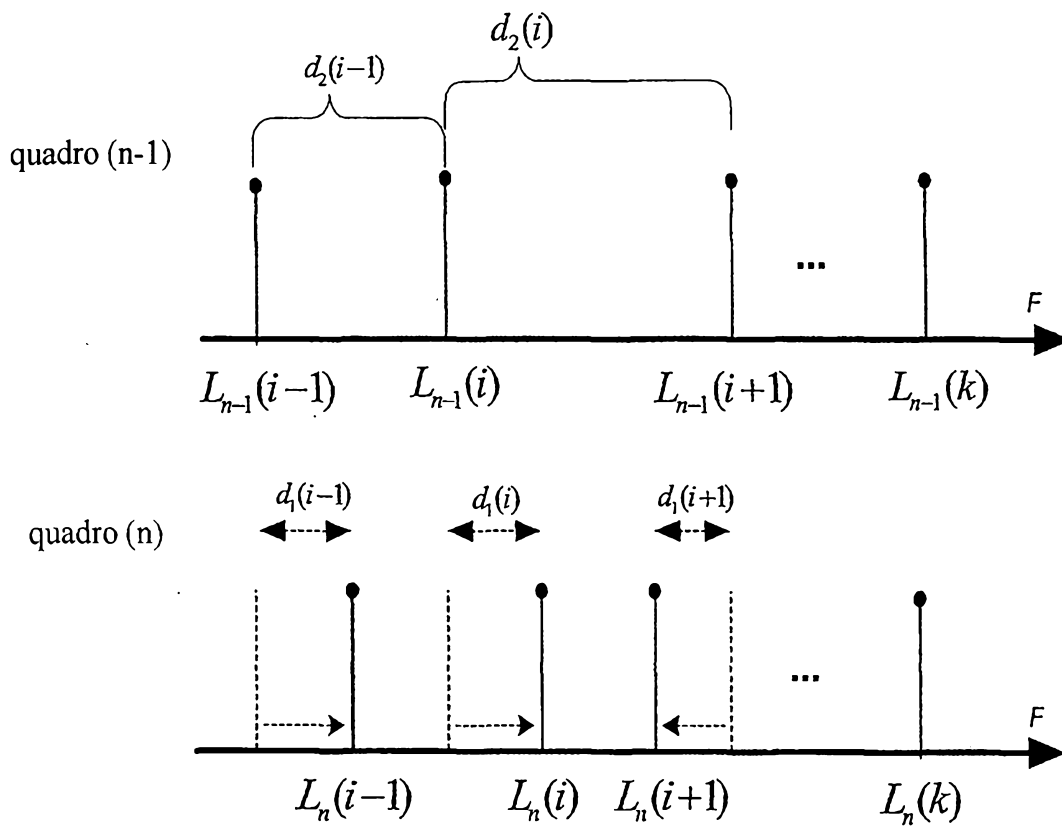


Fig. 8

RESUMO

P. 11. 4. 2. 7

**“MÉTODO E APARELHO PARA ENCOBRIR OS EFEITOS DOS
ERROS DE QUADRO NOS QUADROS A SEREM DECODIFICADOS PELO
DECODIFICADOR PARA PROPORCIONAR VOZ SINTETIZADA”.**

5 Os efeitos dos quadros ruins recebidos sobre o canal de comunicação pelo decodificador de voz são encobertos para substituir os valores dos parâmetros dos quadros ruins (um quadro ruim sendo um quadro corrompido ou um quadro perdido) com valores baseados em pelo menos um meio parcialmente adaptativo dos quadros bons recentemente recebidos, mas no caso do quadro corrompido (quando oposto ao quadro perdido), usando
10 o próprio quadro ruim se o quadro ruim encontrar um critério pré-determinado. O objetivo do encobrimento é encontrar os parâmetros mais adequados para o quadro ruim, de modo que a qualidade subjetiva de voz sintetizada seja a mais alta possível.