



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

(21) PI 0715978-1 A2



* B R P I 0 7 1 5 9 7 8 A 2 *

(22) Data de Depósito: 06/08/2007
(43) Data da Publicação: 06/08/2013
(RPI 2222)

(51) Int.Cl.:
G10L 21/04
G10L 19/14

(54) Título: QUADROS DE ALINHAMENTO TEMPORAL DE VOCODER DE BANDA LARGA

(30) Prioridade Unionista: 22/08/2006 US 11/508,396

(73) Titular(es): Qualcomm Incorporated

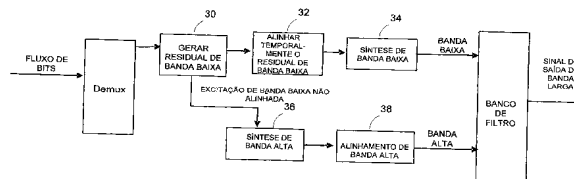
(72) Inventor(es): Rohit Kapoor, Serafin Spindola Diaz

(74) Procurador(es): Montaury Pimenta, Machado & Lioce

(86) Pedido Internacional: PCT US2007075284 de 06/08/2007

(87) Publicação Internacional: WO 2008/024615de 28/02/2008

(57) Resumo: "QUADROS DE ALINHAMENTO TEMPORAL DE VOCODER DE BANDA LARGA". Um método de comunicar fala compreendendo alinhamento temporal de um sinal de fala de banda baixa residual para uma versão expandida ou comprimida do sinal de fala de banda baixa residual, alinhamento temporal de um sinal de fala de banda alta para uma versão expandida ou comprimida do sinal de fala de banda alta, e fundindo os sinais de fala de banda baixa e banda alta com alinhamento temporal para fornecer um sinal de fala com alinhamento temporal inteiro. Na banda baixa, o sinal de fala de banda baixa residual é sintetizado após alinhamento temporal do sinal de banda baixa residual enquanto na banda alta, um sinal de banda alta não alinhado é sintetizado antes de alinhamento temporal do sinal de fala de banda alta. O método pode compreender ainda classificar segmentos de fala e codificar os segmentos de fala. A codificação dos segmentos de fala pode ser uma entre predição linear excitada por código, predição linear excitada por ruído ou codificação de quadro 1/8 (silêncio).



"QUADROS DE ALINHAMENTO TEMPORAL DE VOCODER DE BANDA LARGA"

FUNDAMENTOS

CAMPO

A presente invenção refere-se geralmente a
5 alinhamento temporal, isto é, expansão ou compressão,
quadros em um vocoder e, em particular, a métodos de
alinhamento temporal de quadros em um vocoder de banda
larga.

FUNDAMENTOS

10 Alinhamento temporal tem diversas aplicações em
redes comutadas por pacote onde pacotes de vocoder podem
chegar de forma assíncrona. Embora o alinhamento temporal
possa ser executado dentro ou fora do vocoder, a execução
no vocoder oferece diversas vantagens como melhor qualidade
15 de quadros alinhados e carga computacional reduzida.

SUMÁRIO

A invenção compreende um equipamento e método de
alinhamento temporal de quadros de fala por manipular um
sinal de fala. Em um aspecto, é revelado um método de
20 alinhamento temporal de quadros de Predição Linear excitada
por código (CELP) e Predição Linear excitada por ruído
(NELP) de um vocoder de banda larga de Vocoder de Quarta
geração (4GV). Mais especificamente, para quadros CELP, o
método mantém uma fase de fala por adição ou deleção de
25 períodos de pitch para expandir ou comprimir fala,
respectivamente. Com esse método, o sinal de banda mais
baixa pode sofrer alinhamento temporal no residual, isto é,
antes de síntese, enquanto o sinal de banda mais alta pode
sofrer alinhamento temporal após síntese no domínio de 8
30 kHz. O método revelado pode ser aplicado em qualquer
vocoder de banda larga que utiliza CELP e/ou NELP para a
banda baixa e/ou utiliza uma técnica de banda dividida para
codificar as bandas inferior e superior separadamente. Deve

ser observado que o nome de padrão para banda larga 4GV é EVRC-C.

Em vista do acima, as características descritas da invenção se referem, genericamente, a um ou mais sistemas, métodos e/ou equipamentos aperfeiçoados para comunicação de fala. Em uma modalidade, a invenção compreende um método de comunicar fala compreendendo alinhamento temporal de um sinal de fala de banda baixa residual para uma versão expandida ou comprimida do sinal de fala de banda baixa residual, alinhamento temporal de um sinal de fala de banda alta para uma versão expandida ou comprimida do sinal de fala de banda alta, e fundindo os sinais de fala de banda alta e banda baixa com alinhamento temporal para fornecer um sinal de fala com alinhamento temporal inteiro. Em um aspecto da invenção, o sinal de fala de banda baixa residual é sintetizado após alinhamento temporal do sinal de banda baixa residual enquanto na banda alta, a síntese é executada antes de alinhamento temporal do sinal de fala de banda alta. O método pode compreender ainda classificar os segmentos de fala e codificar os segmentos de fala. A codificação dos segmentos de fala pode ser um entre predição linear excitada por código, predição linear excitada por ruído ou codificação de quadro 1/8 (silêncio). A banda baixa pode representar a banda de frequência até aproximadamente 4 kHz e a banda alta pode representar a banda de aproximadamente 3,5 kHz até aproximadamente 7 kHz.

Em outra modalidade, é revelado um vocoder tendo pelo menos uma entrada e pelo menos uma saída, o vocoder compreendendo um codificador que compreende um filtro tendo pelo menos uma entrada eficientemente conectada à entrada do vocoder e pelo menos uma saída; e um decodificador compreendendo um sintetizador tendo pelo menos uma entrada

eficientemente conectada a pelo menos uma saída do codificador e pelo menos uma saída eficientemente conectada a pelo menos uma saída do vocoder. Nessa modalidade, o decodificador compreende uma memória, onde o decodificador
5 é adaptado para executar instruções de software armazenadas na memória compreendendo alinhamento temporal de um sinal de fala de banda baixa residual para uma versão expandida ou comprimida do sinal de fala de banda baixa residual, alinhamento temporal de um sinal de fala de banda alta para
10 uma versão expandida ou comprimida do sinal de fala de banda alta, e fundindo os sinais de fala de banda baixa e banda alta com alinhamento temporal para fornecer um sinal de fala com alinhamento temporal inteiro. O sintetizador pode compreender meio para sintetizar o sinal de fala de
15 banda baixa residual com alinhamento temporal, e meio para sintetizar o sinal de fala de banda alta antes de alinhamento temporal do mesmo. O codificador compreende uma memória e pode ser adaptado para executar instruções de software na memória compreendendo classificar segmentos de
20 fala como quadro 1/8 (silêncio), predição linear excitada por código ou predição linear excitada por ruído.

O escopo adicional de aplicabilidade da presente invenção tornar-se-á evidente a partir da seguinte descrição detalhada, reivindicações e desenhos. Entretanto,
25 deve ser entendido que a descrição detalhada e exemplos específicos, embora indique modalidades preferidas da invenção, são fornecidos somente como ilustração, uma vez que várias alterações e modificações compreendidas no espírito e escopo da invenção tornar-se-ão evidentes para
30 aqueles versados na técnica.

BREVE DESCRIÇÃO DOS DESENHOS

A presente invenção tornar-se-á entendida de forma mais completa a partir da descrição detalhada dada

abaixo, reivindicações apensas, e desenhos em anexo nos quais:

A figura 1 é um diagrama de blocos de um vocoder de Codificação preditiva linear (LPC);

5 A figura 2A é um sinal de fala contendo fala com voz;

A figura 2B é um sinal de fala contendo fala não-sonora;

10 A figura 2C é um sinal de fala contendo voz transiente;

A figura 3 é um diagrama de blocos ilustrando alinhamento temporal de banda baixa e banda alta;

A figura 4A representa a determinação de retardos de pitch através da interpolação;

15 A figura 4B representa a identificação de períodos de pitch;

A figura 5A representa um sinal de fala original na forma de períodos de pitch;

20 A figura 5B representa um sinal de fala expandido utilizando sobrepor/adicionar; e

A figura 5C representa um sinal de fala comprimido utilizando sobrepor/adicionar.

DESCRIÇÃO DETALHADA

25 A palavra "ilustrativo" é utilizada aqui para significar "servir como exemplo, ocorrência ou ilustração." Qualquer modalidade descrita aqui como "ilustrativa" não deve ser necessariamente interpretada como preferida ou vantajosa em relação a outras modalidades.

30 Alinhamento temporal tem diversas aplicações em redes comutadas por pacote onde pacotes de vocoder podem chegar de forma assíncrona. Embora alinhamento temporal possa ser executado dentro ou fora do vocoder, a execução do mesmo no vocoder oferece diversas vantagens como melhor

qualidade de quadros alinhados e carga computacional reduzida. As técnicas descritas aqui podem ser facilmente aplicadas a outros vocoders que utilizam técnicas similares como 4GV-banda larga, cujo nome de padrão é EVRC-C, para
5 fazer vocode de dados de voz.

DESCRIÇÃO DE FUNCIONALIDADE DE VOCODER

Vozes humanas compreendem dois componentes. Um componente compreende ondas fundamentais que são sensíveis a pitch (altura) e o outro é harmônica fixa que não é
10 sensível a pitch. O pitch percebido de um som é a resposta do ouvido à frequência, isto é, para fins mais práticos o pitch é a frequência. Os componentes de harmônica acrescentam características distintas a voz de uma pessoa. Mudam juntamente com as cordas vocais e com o formato
15 físico do trato vocal e são denominados formantes.

A voz humana pode ser representada por um sinal digital $s(n)$ (vide a figura 1). Considere que $s(n)$ é um sinal de fala digital obtido durante uma conversação típica incluindo diferentes sons vocais e períodos de
20 silêncio. O sinal de fala $s(n)$ pode ser dividido em quadros, como mostrado nas figuras 2A - 2C. Em um aspecto, $s(n)$ é digitalmente amostrado em 8 kHz. Em outros aspectos, $s(n)$ pode ser digitalmente amostrado em 16 kHz ou 32 kHz ou alguma outra frequência de amostragem.

25 Esquemas de codificação atuais comprimem um sinal de fala digitalizado em um sinal de taxa de bit baixa por remover todas as redundâncias naturais (isto é, elementos correlacionados) inerentes em fala. A fala apresenta, tipicamente, redundâncias de curta duração
30 resultando da ação mecânica dos lábios e língua, e redundâncias de longa duração resultando da vibração das cordas vocais. A Codificação preditiva linear (LPC) filtra o sinal de fala por remover as redundâncias que produzem

um sinal de fala residual. Modelada, então, o sinal residual resultante como ruído Gaussiano branco. Um valor amostrado de uma forma de onda de fala pode ser predito por ponderar uma soma de um número de amostras passadas, cada uma das
 5 quais é multiplicada por um coeficiente preditivo linear. Codificadores preditivos lineares, portanto, obtêm uma taxa de bit reduzida por transmitir coeficientes de filtro e ruído quantizado em vez de um sinal de fala de largura de banda total 10.

10 Um diagrama de blocos de uma modalidade de um vocoder LPC 70 é ilustrado na figura 1. A função do LPC é a de minimizar a soma das diferenças ao quadrado entre o sinal de fala original e o sinal de fala estimado em uma duração finita. Isso pode produzir um conjunto único de
 15 coeficientes preditor que são normalmente estimados em cada quadro 20. Um quadro 20 tem tipicamente 20 ms de comprimento. A função de transferência de um filtro digital de variação de tempo 75 pode ser dada por:

$$H(z) = \frac{G}{1 - \sum a_k z^{-k}}$$

20 onde os coeficientes de preditor podem ser representados por a_k e o ganho por G .

A soma é computada de $k = 1$ para $k = p$. Se um método de LPC-10 for utilizado, então $p = 10$. Isso significa que somente os 10 primeiros coeficientes são
 25 transmitidos para um sintetizador LPC 80. Os dois métodos mais comumente utilizados para computar os coeficientes são, porém não limitados a, o método de covariância e o método de autocorrelação.

Vocoders típicos produzem quadros 20 de 20 ms de
 30 duração, incluindo 160 amostras na taxa preferida de 8 kHz ou 320 amostras em taxa de 16 kHz. Uma versão comprimida com alinhamento temporal desse quadro 20 tem uma duração

menor do que 20 ms, enquanto uma versão expandida com alinhamento temporal tem uma duração maior do que 20 ms. Alinhamento temporal de dados de voz tem vantagens significativas ao enviar dados de voz através de redes comutadas por pacote, que introduzem instabilidade de retardo na transmissão de pacotes de voz. Em tais redes, alinhamento temporal pode ser utilizado para diminuir os efeitos dessa instabilidade de retardo e produzir um fluxo de voz que parece "síncrono".

As modalidades da invenção se referem a um equipamento e método para alinhamento temporal de quadros dentro do vocoder 70 por manipular o residual de fala. Em uma modalidade, o presente método e equipamento é utilizado em banda larga 4GV. As modalidades reveladas compreendem métodos e equipamentos ou sistemas para expandir/comprimir tipos diferentes de segmentos de fala de banda larga 4GV codificados utilizando codificação de Predição linear excitada por código (CELP) ou Predição linear excitada por ruído (NELP).

O termo "vocoder" 70 se refere, tipicamente, a dispositivos que comprimem fala com voz por extrair parâmetros com base em um modelo de geração de fala humana. Vocoders 70 incluem um codificador 204 e um decodificador 206. O codificador 204 analisa a fala que entra e extrai os parâmetros relevantes. Em uma modalidade, o codificador compreende o filtro 75. O decodificador 206 sintetiza a fala utilizando os parâmetros que recebe a partir do codificador 204 através de um canal de transmissão 208. Em uma modalidade, o decodificador compreende o sintetizador 80. O sinal de fala 10 é frequentemente dividido em quadros de dados e processado em bloco pelo vocoder 70.

Aqueles versados na técnica reconhecerão que fala humana pode ser classificada em muitos modos diferentes.

Três classificações convencionais de fala são sons de voz, não-sonora e fala transiente.

A figura 2A é um sinal de fala de voz $s(n)$ 402. A figura 2A mostra uma propriedade comum, mensurável de fala sonora conhecida como o período de pitch 100.

A figura 2B é um sinal de fala não-sonora $s(n)$ 404. Um sinal de fala não-sonora 404 se parece com ruído colorido.

A figura 2C representa um sinal de fala transiente $s(n)$ 406, isto é, fala que não é sonora ou não-sonora. O exemplo de fala transiente 406 mostrado na figura 2C poderia representar $s(n)$ transição entre fala não-sonora e fala sonora. Essas três classificações não são todas inclusive. Há muitas classificações diferentes de fala que podem ser empregadas de acordo com os métodos descritos aqui para obter resultados comparáveis.

VOCODER DE BANDA LARGA 4GV

O vocoder de quarta geração (4GV) provê características atraentes para uso através de redes sem fio como adicionalmente descrito no pedido de patente copendente número de série 11/123.467, depositado em 5 de maio de 2005, intitulado "Time warping frames inside the vocoder by modifying the residual", que é totalmente incorporada aqui a título de referência. Algumas dessas características incluem a capacidade de trade-off qualidade vs. taxa de bits, vocoding mais resiliente na face de taxa de erro de pacote (PER) aumentada, melhor ocultação de apagamentos, etc. Na presente invenção, o vocoder de banda larga 4GV é revelado que codifica fala utilizando uma técnica de banda dividida, isto é, as bandas inferior e superior são separadamente codificadas.

Em uma modalidade, um sinal de entrada representa uma fala de banda larga amostrada em 16 kHz. Um banco de

filtro de análise é fornecido gerando um sinal de banda estreita (banda baixa) amostrada em 8 kHz, e um sinal de banda alta amostrada em 7 kHz. Esse sinal de banda alta representa a banda a partir de aproximadamente 3,5 kHz até aproximadamente 7 kHz no sinal de entrada, enquanto o sinal de banda baixa representa a banda até aproximadamente 4 kHz, e o sinal de banda larga reconstruído final será limitado em largura de banda até aproximadamente 7 kHz. Deve ser observado que há uma sobreposição de aproximadamente 500 Hz entre as bandas baixa e elevada, permitindo uma transição mais gradual entre as bandas.

Em um aspecto, o sinal de banda estreita é codificado utilizando uma versão modificada do codificador de fala EVRC-B de banda estreita, que é um codificador CELP com um tamanho de quadro de 20 milissegundos. Vários sinais a partir do codificador de banda estreita são utilizados pela síntese e análise de banda alta; são: (1) o sinal de excitação (isto é, residual quantizado) a partir do codificador de banda estreita; (2) o primeiro coeficiente de reflexão quantizado (como um indicador da inclinação espectral do sinal de banda estreita); (3) o ganho de livro-código adaptável quantizado; e (4) o retardo de pitch quantizado.

O codificador de banda estreita EVRC-B modificado utilizado em banda larga 4GV codifica cada dado de voz de quadro em um de três tipos de quadros diferentes: Predição Linear excitada por código (CELP); Predição linear excitada por ruído (NELP); ou quadro de taxa de 1/8° de silêncio.

CELP é utilizado para codificar grande parte da fala, que inclui fala que é periódica bem como aquela com pouca periodicidade. Tipicamente, aproximadamente 75% dos quadros não silenciosos são codificados pelo codificador de banda estreita EVRC-B modificado utilizando CELP.

NELP é utilizado para codificar fala que é de caráter semelhante a ruído. O caráter semelhante a ruído desses segmentos de fala pode ser reconstruído por gerar sinais aleatórios no decodificador e aplicar ganhos apropriados aos mesmos.

Quadros de taxa de $1/8^\circ$ são utilizados para codificar ruído de segundo plano, isto é, períodos onde o usuário não está falando.

ALINHAMENTO TEMPORAL DE QUADROS DE BANDA LARGA 4GV

Uma vez que o vocoder de banda larga 4GV codifica bandas inferior e superior separadamente, a mesma filosofia é seguida em alinhamento temporal dos quadros. A banda inferior é com alinhamento temporal utilizando uma técnica similar como descrito no pedido de patente copendente acima mencionado intitulado "Time warping frames inside the vocoder by modifying the residual."

Com referência à figura 3, é mostrado um alinhamento de banda inferior 32 que é aplicado em um sinal residual 30. O principal motivo para fazer alinhamento temporal 32 no domínio residual é que isso permite que a síntese de LPC 34 seja aplicada no sinal residual com alinhamento temporal. Os coeficientes de LPC desempenham um papel importante em como sons de fala e a aplicação de síntese 34 após alinhamento 32 assegura que informações de LPC corretas são mantidas no sinal. Se alinhamento temporal for feito após o decodificador, por outro lado, a síntese de LPC já foi executada antes de alinhamento temporal. Desse modo, o procedimento de alinhamento pode alterar as informações de LPC do sinal, especialmente se a estimação de período de pitch não foi muito precisa.

ALINHAMENTO TEMPORAL DE SINAL RESIDUAL QUANDO SEGMENTO DE FALA É CELP

Para alinhar o residual, o decodificador utiliza

informações de retardo de pitch contidas no quadro codificado. O retardo de pitch é na realidade o retardo de pitch na extremidade do quadro. Deve ser observado aqui que mesmo em um quadro periódico, o retardo de pitch poderia estar levemente alterando. Os retardos de pitch em qualquer ponto no quadro podem ser estimados por interpolar entre o retardo de pitch na extremidade do último quadro e aquele na extremidade do quadro atual. Isso é mostrado na figura 4. Uma vez que os retardos de pitch em todos os pontos no quadro são conhecidos, o quadro pode ser dividido em períodos de pitch. Os limites de períodos de pitch são determinados utilizando os retardos de pitch em vários pontos no quadro.

A figura 4A mostra um exemplo de como dividir os quadros em seus períodos de pitch. Por exemplo, o número de amostra 70 tem retardo de pitch de aproximadamente 70 e o número de amostra 142 tem retardo de pitch de aproximadamente 72. Desse modo, períodos de pitch são de [1-70] e de [71-142]. Isso é ilustrado na figura 4B.

Após o quadro ter sido dividido em períodos de pitch, esses períodos de pitch podem ser então sobrepostos/adicionados para aumentar/diminuir o tamanho do residual. A técnica de sobrepor/adicionar é uma técnica conhecida e as figuras 5A-5C mostram como é utilizada para expandir/comprimir o residual.

Alternativamente, os períodos de pitch podem ser repetidos se o sinal de fala necessitar ser expandido. Por exemplo, na figura 5B, o período de pitch PP1 pode ser repetido (em vez de sobreposto-adicionado com PP2) para produzir um período de pitch extra.

Além disso, a sobreposição/adição e/ou repetição de períodos de pitch pode ser feita tantas vezes quanto é necessário para produzir a quantidade de

expansão/compressão necessária.

Com referência à figura 5A, o sinal de fala original compreendendo 4 períodos de pitch (PPs) é mostrado. A figura 5B mostra como esse sinal de fala pode ser expandido utilizando sobrepor/adicionar. Na figura 5B, períodos de pitch PP2 e PP1 são sobrepostos/adicionados de tal modo que a contribuição de PP2s continua diminuindo e aquela de PP1 está aumentando. A figura 5C ilustra como sobreposição/adição é utilizado para comprimir o residual.

Em casos quando o período de pitch está mudando, a técnica de sobrepor-adicionar pode exigir a fusão de dois períodos de pitch de comprimento desigual. Nesse caso, fusão melhor pode ser obtida por alinhar os picos dos dois períodos de pitch antes de sobrepor/adicionar os mesmos.

O residual expandido/comprimido é finalmente enviado através da síntese de LPC.

Após a banda inferior ser alinhada, a banda superior necessita ser alinhada utilizando o período de pitch a partir da banda inferior, isto é, para expansão, um período de pitch de amostras é adicionado, enquanto para compressão, um período de pitch é removido.

O procedimento para alinhar a banda superior é diferente da banda inferior. Com referência novamente à figura 3, a banda superior não é alinhada no domínio residual, porém em vez disso o alinhamento 38 é feito após síntese 36 das amostras de banda superior. O motivo para isso é que a banda superior é amostrada em 7 kHz, enquanto a banda inferior é amostrada em 8 kHz. Desse modo, o período de pitch da banda inferior (amostrada em 8 kHz) pode se tornar um número fracional de amostras quando a taxa de amostragem é 7 kHz, como na banda superior. Como exemplo, se o período de pitch for 25 na banda inferior, no domínio residual da banda superior, isso exigirá que $25 \cdot 7/8$

= 21,875 amostras sejam adicionados/removidos do residual de banda superior. Evidentemente, uma vez que um número fracional de amostras não pode ser gerado, a banda superior é alinhada 38 após ter sido reamostrada para 8 kHz, que é o caso após a síntese 36.

Após a banda inferior ser alinhada 32, a excitação de banda inferior não alinhada (consistindo em 160 amostras) é passada para o decodificador de banda superior. Utilizando essa excitação de banda inferior não alinhada, o decodificador de banda superior produz 140 amostras de banda superior em 7 kHz. Essas 140 amostras são então passadas através de um filtro de síntese 36 e reamostradas para 8 kHz, fornecendo 160 amostras de banda superior.

Essas 160 amostras em 8 kHz são então com alinhamento temporal 38 utilizando o período de pitch a partir da banda inferior e a técnica de sobrepor/adicionar utilizada para alinhar o segmento de fala CELP de banda inferior.

As bandas superior e inferior são finalmente adicionadas ou fundidas para fornecer o sinal alinhado inteiro.

ALINHAMENTO TEMPORAL DE SINAL RESIDUAL QUANDO SEGMENTO DE FALA É NELP

Para segmentos de fala NELP, o codificador codifica somente as informações de LPC bem como os ganhos de diferentes partes do segmento de fala para a banda inferior. Os ganhos podem ser codificados em "segmentos" de 16 amostras PCM cada. Desse modo, a banda inferior pode ser representada como 10 valores de ganho codificado (um de cada para 16 amostras de fala).

O decodificador gera o sinal residual de banda inferior por gerar valores aleatórios e então aplicar os

respectivos ganhos nos mesmos. Nesse caso, não há conceito de período de pitch e como tal, a expansão/compressão de banda inferior não tem de ser da granularidade de um período de pitch.

5 Para expandir/comprimir a banda inferior de um quadro codificado NELP, o decodificador pode gerar um número maior/menor de segmentos do que 10. A expansão/compressão de banda inferior nesse caso é por um múltiplo de 16 amostras, levando a $N = 16 \cdot n$ amostras, onde
10 n é o número de segmentos. No caso de expansão, os segmentos adicionados extra podem tomar os ganhos de alguma função dos primeiros 10 segmentos. Como exemplo, os segmentos extra podem tomar o ganho do 10º segmento.

 Alternativamente, o decodificador pode
15 expandir/comprimir a banda inferior de um quadro codificado NELP por aplicar os 10 ganhos decodificados em conjuntos de y (em vez de 16) amostras para gerar um residual de banda inferior expandido ($y > 16$) ou comprimido ($y < 16$).

 O residual expandido/comprimido é então enviado
20 através da síntese LPC para produzir o sinal alinhado de banda inferior.

 Após a banda inferior ser alinhada, a excitação de banda inferior não alinhada (compreendendo 160 amostras) é passada para o decodificador de banda superior.
25 Utilizando essa excitação de banda inferior não alinhada, o decodificador de banda superior produz 140 amostras de banda superior em 7 kHz. Essas 140 amostras são então passadas através de um filtro de síntese e reamostradas para 8 kHz, fornecendo 160 amostras de banda superior.

30 Essas 160 amostras em 8 kHz são então com alinhamento temporal em um modo similar à alinhamento de banda superior de segmentos de fala CELP, isto é, utilizando sobrepor/adicionar. Ao utilizar

sobrepor/adicionar para a banda superior de NELP, a quantidade a comprimir/expandir é igual à quantidade utilizada para a banda inferior. Em outras palavras, a "sobreposição" utilizada para o método de sobrepor/adicionar é considerada como sendo a quantidade de expansão/compressão na banda inferior. Como exemplo, se a banda inferior produziu 192 amostras após alinhamento, o período de sobreposição utilizado no método de sobreposição/adição é $192 - 160 = 32$ amostras.

As bandas superior e inferior são finalmente adicionadas para fornecer o segmento de fala NELP alinhado inteiro.

Aqueles versados na técnica entenderão que informações e sinais podem ser representados utilizando qualquer de uma variedade de técnicas e tecnologias diferentes. Por exemplo, dados, instruções, comandos, informações, sinais, bits, símbolos e chips que podem ser referenciados em toda a descrição acima podem ser representados por tensões, correntes, ondas eletromagnéticas, campos ou partículas magnéticas, partículas ou campos ópticos, ou qualquer combinação dos mesmos.

Aqueles versados na técnica reconhecerão adicionalmente que os vários blocos lógicos ilustrativos, módulos, circuitos, e etapas de algoritmo descritos com relação às modalidades reveladas aqui podem ser implementados como hardware eletrônico, software de computador ou combinação dos mesmos. Para ilustrar claramente essa capacidade de intercâmbio de hardware e software, vários componentes ilustrativos, blocos, módulos circuitos e etapas foram descritos acima genericamente em termos de sua funcionalidade. O fato de se essa funcionalidade é implementada como hardware ou software

depende da aplicação específica e limitações de desenho impostas no sistema geral. Técnicos especializados podem implementar a funcionalidade descrita em modos variáveis para cada aplicação específica, porém tais decisões de
5 implementação não devem ser interpretadas como causando afastando do escopo da presente invenção.

Os vários blocos lógicos ilustrativos, módulos e circuitos descritos com relação às modalidades reveladas aqui podem ser implementados ou executados com um
10 processador de propósito geral, um Processador de sinais digitais (DSP), um Circuito integrado de aplicação específica (ASIC), uma Disposição de porta programável em campo (FPGA) ou outro dispositivo lógico programável, porta discreta ou lógica de transistor, componentes de hardware
15 discretos, ou qualquer combinação dos mesmos projetada para executar as funções descritas aqui. Um processador de propósito geral pode ser um microprocessador, porém na alternativa, o processador pode ser qualquer processador convencional, controlador, microcontrolador ou máquina de
20 estado. Um processador também pode ser implementado como uma combinação de dispositivos de computação, por exemplo, uma combinação de um DSP e um microprocessador, uma pluralidade de microprocessadores, um ou mais microprocessadores em combinação com um núcleo de DSP, ou
25 qualquer outra configuração.

As etapas de um método ou algoritmo descrito com relação às modalidades reveladas aqui podem ser incorporadas diretamente em hardware, em um módulo de software executado por um processador, ou em uma combinação
30 dos dois. Um módulo de software pode residir em Memória de acesso aleatório (RAM), memória flash, Memória somente de leitura (ROM), ROM eletricamente programável (EPROM), ROM programável eletricamente apagável (EEPROM), registros,

disco rígido, um disco removível, um CD-ROM ou qualquer outra forma de meio de armazenagem conhecido na técnica. Um meio de armazenagem ilustrativo é acoplado ao processador de tal modo que o processador possa ler informações a partir de, e gravar informações para o meio de armazenagem. Na alternativa, o meio de armazenagem pode ser integral ao processador. O processador e o meio de armazenagem podem residir em um ASIC. O ASIC pode residir em um terminal de usuário. Na alternativa, o processador e o meio de armazenagem podem residir como componentes discretos em um terminal de usuário.

A descrição anterior das modalidades reveladas é fornecida para permitir que qualquer pessoa versada na técnica faça ou utilize a presente invenção. Várias modificações nessas modalidades serão prontamente evidentes para aqueles versados na técnica, e os princípios genéricos definidos aqui podem ser aplicados em outras modalidades sem se afastar do espírito ou escopo da invenção. Desse modo, a presente invenção não pretende ser limitada às modalidades mostradas aqui porém deve ser acordada o escopo mais amplo compatível com os princípios e aspectos novos aqui revelados.

REIVINDICAÇÕES

1. Método para comunicar fala, compreendendo:
 - alinhamento temporal de um sinal de fala de banda baixa residual para uma versão expandida ou comprimida do
5 sinal de fala de banda baixa residual;
 - alinhamento temporal de um sinal de fala de banda alta para uma versão expandida ou comprimida do sinal de fala de banda alta; e
 - fundir os sinais de fala de banda baixa e banda
10 alta com alinhamento temporal para fornecer um sinal de fala com alinhamento temporal inteiro.
2. Método, de acordo com a reivindicação 1, compreendendo ainda sintetizar o sinal de fala de banda baixa residual com alinhamento temporal.
- 15 3. Método, de acordo com a reivindicação 2, compreendendo ainda sintetizar o sinal de fala de banda alta antes de alinhamento temporal do mesmo.
4. Método, de acordo com a reivindicação 3, compreendendo ainda:
 - 20 classificar segmentos de fala; e
 - codificar os segmentos de fala.
5. Método, de acordo com a reivindicação 4, em que a codificação dos segmentos de fala compreende utilizar predição linear excitada por código, predição linear
25 excitada por ruído ou codificação de quadro 1/8.
6. Método, de acordo com a reivindicação 4, em que a codificação é codificação de predição linear excitada por código.
7. Método, de acordo com a reivindicação 4, em
30 que a codificação é codificação de predição linear excitada por ruído.
8. Método, de acordo com a reivindicação 7, em que a codificação compreende codificar informações de

Codificação preditiva linear como ganhos de partes diferentes de um quadro de fala.

5 9. Método, de acordo com a reivindicação 8, em que os ganhos são codificados para conjuntos de amostras de fala.

10 10. Método, de acordo com a reivindicação 9, compreendendo ainda gerar um sinal de banda baixa residual por gerar valores aleatórios e então aplicar os ganhos aos valores aleatórios.

15 11. Método, de acordo com a reivindicação 9, compreendendo ainda representar as informações de Codificação preditiva linear como 10 valores de ganho codificado para o sinal de fala de banda baixa residual, onde cada valor de ganho codificado representa 16 amostras de fala.

20 12. Método, de acordo com a reivindicação 7, compreendendo ainda produzir 140 amostras do sinal de fala de banda alta a partir de um sinal de excitação de banda baixa não alinhada.

25 13. Método, de acordo com a reivindicação 7, em que alinhamento temporal do sinal de fala de banda baixa compreende gerar um número mais elevado/mais baixo de amostras e aplicar alguma função dos ganhos decodificados das partes de um quadro de fala ao residual e então sintetizar o mesmo.

30 14. Método, de acordo com a reivindicação 13, em que a aplicação de alguma função dos ganhos decodificados de partes do quadro de fala ao residual compreende aplicar o ganho do último segmento de fala às amostras adicionais quando a banda inferior é expandida.

15. Método, de acordo com a reivindicação 7, em que alinhamento temporal do sinal de fala de banda alta compreende:

sobrepor/adicionar o mesmo número de amostras como foram comprimidas na banda inferior se o sinal de fala de banda alta for comprimido; e

5 sobrepor/adicionar o mesmo número de amostras como foram expandidas na banda inferior se o sinal de fala de banda alta for expandido.

16. Método, de acordo com a reivindicação 6, em que alinhamento temporal do sinal de fala de banda baixa residual compreende:

10 estimar pelo menos um período de pitch; e
adicionar ou subtrair pelo menos um dos períodos de pitch após receber o sinal de fala de banda baixa residual.

15 17. Método, de acordo com a reivindicação 16, em que alinhamento temporal do sinal de fala de banda alta compreende:

utilizar os períodos de pitch a partir do sinal de fala de banda baixa;

20 sobrepor/adicionar um ou mais períodos de pitch se o sinal de fala de banda alta for comprimido; e

sobrepor/adicionar ou repetir um ou mais períodos de pitch se o sinal de fala de banda alta for expandido.

25 18. Método, de acordo com a reivindicação 6, em que alinhamento temporal do sinal de fala de banda baixa residual compreende:

estimar retardo de pitch;

30 dividir um quadro de fala em períodos de pitch, onde limites dos períodos de pitch são determinados utilizando o retardo de pitch em vários pontos no quadro de fala;

sobrepor/adicionar os períodos de pitch se o sinal de fala de banda baixa residual for comprimido; e

sobrepor/adicionar ou repetir um ou mais períodos

de pitch se o sinal de fala de banda baixa residual for expandido.

19. Método, de acordo com a reivindicação 18, em que alinhamento temporal do sinal de fala de banda alta
5 compreende:

utilizar períodos de pitch a partir do sinal de fala de banda baixa;

sobrepor/adicionar os períodos de pitch se o sinal de fala de banda alta for comprimido; e

10 sobrepor/adicionar ou repetir um ou mais períodos de pitch se o sinal de fala de banda alta for expandido.

20. Método de acordo com a reivindicação 18, em que a estimação do retardo de pitch compreende interpolar entre um retardo de pitch de uma extremidade de um último
15 quadro e uma extremidade de um quadro atual.

21. Método, de acordo com a reivindicação 18, em que a sobreposição/adição ou repetição de um ou mais dos períodos de pitch compreende fundir os segmentos de fala.

22. Método, de acordo com a reivindicação 18, em
20 que a sobreposição/adição ou repetição de um ou mais dos períodos de pitch se o sinal de fala de banda baixa residual for expandido compreende adicionar um período de pitch adicional criado a partir de um primeiro segmento de pitch e um segundo segmento de período de pitch.

25 23. Método, de acordo com a reivindicação 21, compreendendo ainda selecionar segmentos de fala similares, em que os segmentos de fala similares são fundidos.

24. Método, de acordo com a reivindicação 21, compreendendo ainda correlacionar os segmentos de fala,
30 pelo que segmentos de fala similares são selecionados.

25. Método, de acordo com a reivindicação 22, em que a adição de um período de pitch adicional criado a partir de um primeiro segmento de pitch e um segundo

segmento de período de pitch compreende adicionar os primeiro e segundo segmentos de pitch de tal modo que a contribuição do primeiro segmento de período de pitch aumenta e a contribuição do segundo segmento de período de pitch diminui.

26. Método, de acordo com a reivindicação 1, em que a banda baixa representa a banda até e incluindo 4 kHz.

27. Método, de acordo com a reivindicação 1, em que a banda alta representa a banda de aproximadamente 3,5 kHz até aproximadamente 7 kHz.

28. Vocoder tendo pelo menos uma entrada e pelo menos uma saída, compreendendo:

um codificador compreendendo um filtro que tem pelo menos uma entrada eficientemente conectada à entrada do vocoder e pelo menos uma saída; e

um decodificador compreendendo um sintetizador tendo pelo menos uma entrada eficientemente conectada a pelo menos uma saída do codificador e pelo menos uma saída eficientemente conectada a pelo menos uma saída do vocoder.

29. Vocoder, de acordo com a reivindicação 28, em que o decodificador compreende:

uma memória, onde o decodificador é adaptado para executar instruções de software armazenadas na memória compreendendo:

alinhamento temporal de um sinal de fala de banda baixa residual para uma versão expandida ou comprimida do sinal de fala de banda baixa residual;

alinhamento temporal de um sinal de fala de banda alta para uma versão expandida ou comprimida do sinal de fala de banda alta; e

fundir os sinais de fala de banda baixa e banda alta com alinhamento temporal para fornecer um sinal de fala com alinhamento temporal inteiro.

30. Vocoder, de acordo com a reivindicação 29, em que o sintetizador compreende meio para sintetizar o sinal de fala de banda baixa residual com alinhamento temporal.

5 31. Vocoder, de acordo com a reivindicação 30, em que o sintetizador compreende ainda meio para sintetizar o sinal de fala de banda alta antes de alinhamento temporal do mesmo.

10 32. Vocoder, de acordo com a reivindicação 28, em que o codificador compreende uma memória e o codificador é adaptado para executar instruções de software armazenadas na memória compreendendo classificar segmentos de fala como quadro 1/8, predição linear excitada por código ou predição linear excitada por ruído.

15 33. Vocoder, de acordo com a reivindicação 31, em que o codificador compreende uma memória e o codificador é adaptado para executar instruções de software armazenadas na memória compreendendo codificar segmentos de fala utilizando codificação de predição linear excitada por código.

20 34. Vocoder, de acordo com a reivindicação 31, em que o codificador compreende uma memória e o codificador é adaptado para executar instruções de software armazenadas na memória compreendendo codificar segmentos de fala utilizando codificação de predição linear excitada por ruído.

25 35. Vocoder, de acordo com a reivindicação 34, em que a codificação dos segmentos de fala utilizando instrução de software de codificação de predição linear excitada por ruído compreende codificar informações de Codificação preditiva linear como ganhos de partes diferentes de um segmento de fala.

30 36. Vocoder, de acordo com a reivindicação 35, em que os ganhos são codificados para conjuntos de amostras de

fala.

37. Vocoder, de acordo com a reivindicação 36, em que a instrução de alinhamento temporal do sinal de fala de banda baixa residual compreende ainda gerar um sinal de
5 fala de banda baixa residual por gerar valores aleatórios e então aplicar os ganhos aos valores aleatórios.

38. Vocoder, de acordo com a reivindicação 36, em que a instrução de alinhamento temporal do sinal de fala de banda baixa residual compreende ainda representar a
10 informação de Codificação preditiva linear como 10 valores de ganho codificados para o sinal de fala de banda baixa residual, onde cada valor de ganho codificado representa 16 amostras de fala.

39. Vocoder, de acordo com a reivindicação 34,
15 compreendendo ainda produzir 140 amostras do sinal de fala de banda alta a partir de um sinal de excitação de banda baixa não alinhada.

40. Vocoder, de acordo com a reivindicação 34, em que a instrução de software de alinhamento temporal do
20 sinal de fala de banda baixa compreende gerar um número mais elevado/mais baixo de amostras e aplicar alguma função dos ganhos decodificados de partes de um quadro de fala ao residual e então sintetizar o mesmo.

41. Vocoder, de acordo com a reivindicação 40, em
25 que a aplicação de alguma função dos ganhos decodificados de partes do quadro de fala ao residual compreende aplicar o ganho do último segmento de fala às amostras adicionais quando a banda inferior é expandida.

42. Vocoder, de acordo com a reivindicação 33, em
30 que a instrução de software de alinhamento temporal do sinal de fala de banda alta compreende:

sobrepor/adicionar o mesmo número de amostras como foram comprimidas na banda inferior se o sinal de fala

de banda alta for comprimido; e

sobrepor/adicionar o mesmo número de amostras como foram expandidas na banda inferior se o sinal de fala de banda alta for expandido.

5 43. Vocoder, de acordo com a reivindicação 33, em que a instrução de software de alinhamento temporal do sinal de fala de banda baixa residual compreende:

estimar pelo menos um período de pitch; e

10 adicionar ou subtrair pelo menos um período de pitch após receber o sinal de fala de banda baixa residual.

44. Vocoder, de acordo com a reivindicação 43, em que a instrução de software de alinhamento temporal do sinal de fala de banda alta compreende:

15 utilizar o período de pitch a partir do sinal de fala de banda baixa;

sobrepor/adicionar um ou mais períodos de pitch se o sinal de fala de banda alta for comprimido; e

sobrepor/adicionar ou repetir um ou mais períodos de pitch se o sinal de fala de banda alta for expandido.

20 45. Vocoder, de acordo com a reivindicação 33, em que a instrução de software de alinhamento temporal do sinal de fala de banda baixa residual compreende:

estimar o retardo de pitch;

25 dividir um quadro de fala em períodos de pitch, onde limites dos períodos de pitch são determinados utilizando o retardo de pitch em vários pontos no quadro de fala;

sobrepor/adicionar os períodos de pitch se o sinal de fala residual for comprimido; e

30 sobrepor/adicionar ou repetir um ou mais períodos de pitch se o sinal de fala residual for expandido.

46. Vocoder, de acordo com a reivindicação 45, em que a instrução de software de alinhamento temporal do

sinal de fala de banda alta compreende:

utilizar os períodos de pitch a partir do sinal de fala de banda baixa;

5 sobrepor/adicionar os períodos de pitch se o sinal de fala de banda alta for comprimido; e

sobrepor/adicionar ou repetir um ou mais períodos de pitch se o sinal de fala de banda alta for expandido.

10 47. Vocoder, de acordo com a reivindicação 45, em que a instrução de sobreposição/adição dos períodos de pitch se o sinal de fala de banda baixa residual for comprimido compreende:

segmentar uma seqüência de amostra de entrada em blocos de amostras;

15 remover segmentos do sinal residual em intervalos de tempo regulares;

fundir os segmentos removidos; e

substituir os segmentos removidos com um segmento fundido.

20 48. Vocoder, de acordo com a reivindicação 45, em que a instrução de estimação do retardo de pitch compreende interpolar entre um retardo de pitch de uma extremidade de um último quadro e uma extremidade de um quadro atual.

25 49. Vocoder, de acordo com a reivindicação 45, em que a sobreposição/adição ou repetição de uma ou mais das instruções de períodos de pitch compreende fundir os segmentos de fala.

30 50. Vocoder, de acordo com a reivindicação 45, em que a sobreposição/adição ou repetição de uma ou mais das instruções de períodos de pitch se o sinal de fala de banda baixa residual for expandido compreende adicionar um período de pitch adicional criado a partir de um primeiro segmento de período de pitch e um segundo segmento de período de pitch.

51. Vocoder, de acordo com a reivindicação 47, em que a instrução de fusão dos segmentos removidos compreende aumentar a contribuição de um primeiro segmento de período de pitch e diminuir a contribuição de um segundo segmento de período de pitch.

52. Vocoder, de acordo com a reivindicação 49, compreendendo ainda selecionar segmentos de fala similares, em que os segmentos de fala similares são fundidos.

53. Vocoder, de acordo com a reivindicação 49, em que a instrução de alinhamento temporal do sinal de fala de banda baixa residual compreende ainda correlacionar os segmentos de fala, pelo que segmentos de fala similares são selecionados.

54. Vocoder, de acordo com a reivindicação 50, em que a instrução de adição de um período de pitch adicional criado a partir dos primeiro e segundo segmentos de período de pitch compreende adicionar os primeiro e segundo segmentos de período de pitch de tal modo que a contribuição do primeiro segmento de período de pitch aumenta e a contribuição do segundo segmento de período de pitch diminui.

55. Vocoder, de acordo com a reivindicação 29, em que a banda baixa representa a banda até e incluindo 4 kHz.

56. Vocoder, de acordo com a reivindicação 29, em que a banda alta representa a banda de aproximadamente 3,5 kHz até aproximadamente 7 kHz.

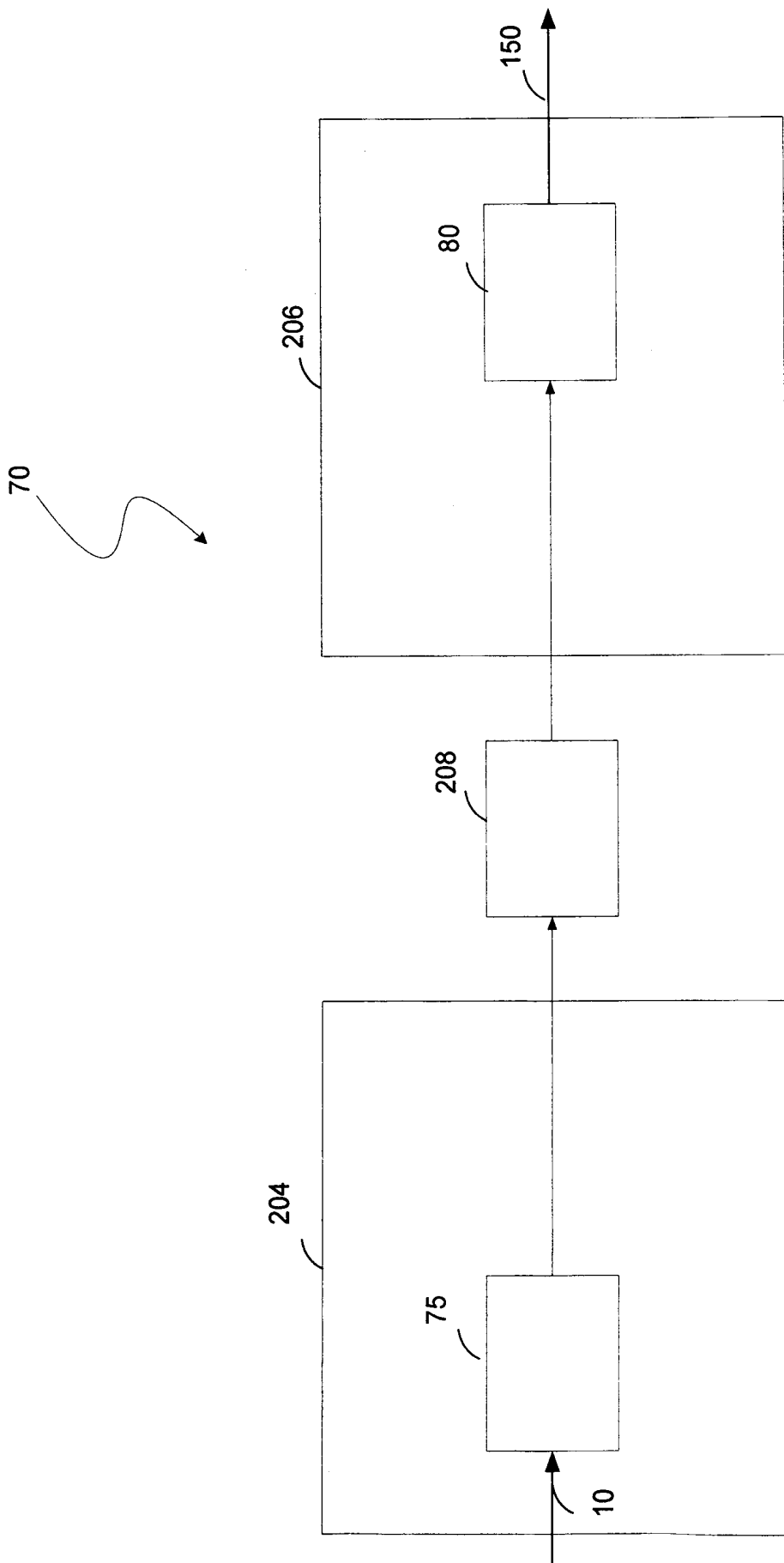


FIG.1

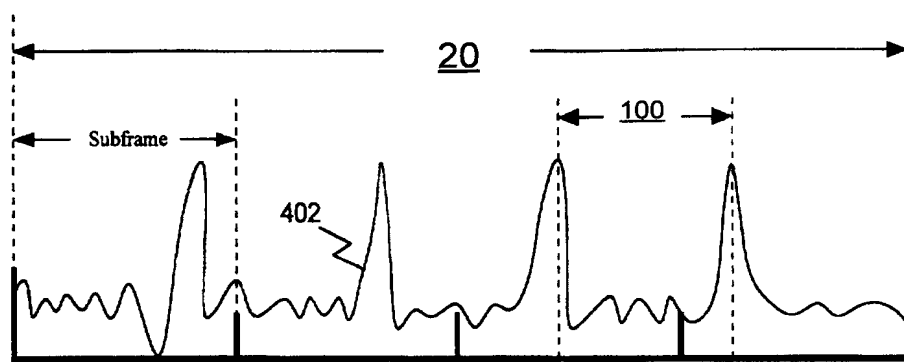


FIG. 2A

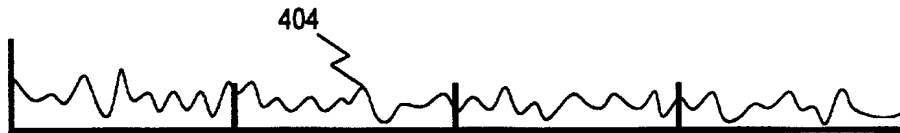


FIG. 2B

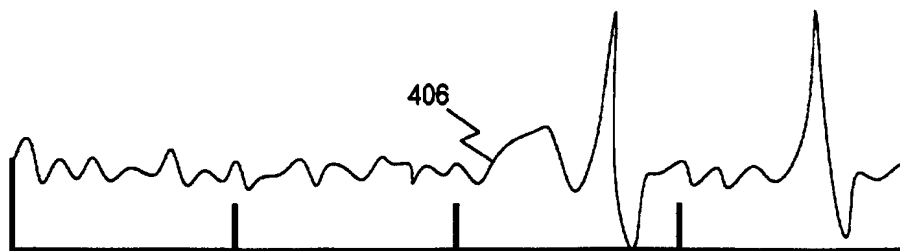


FIG. 2C

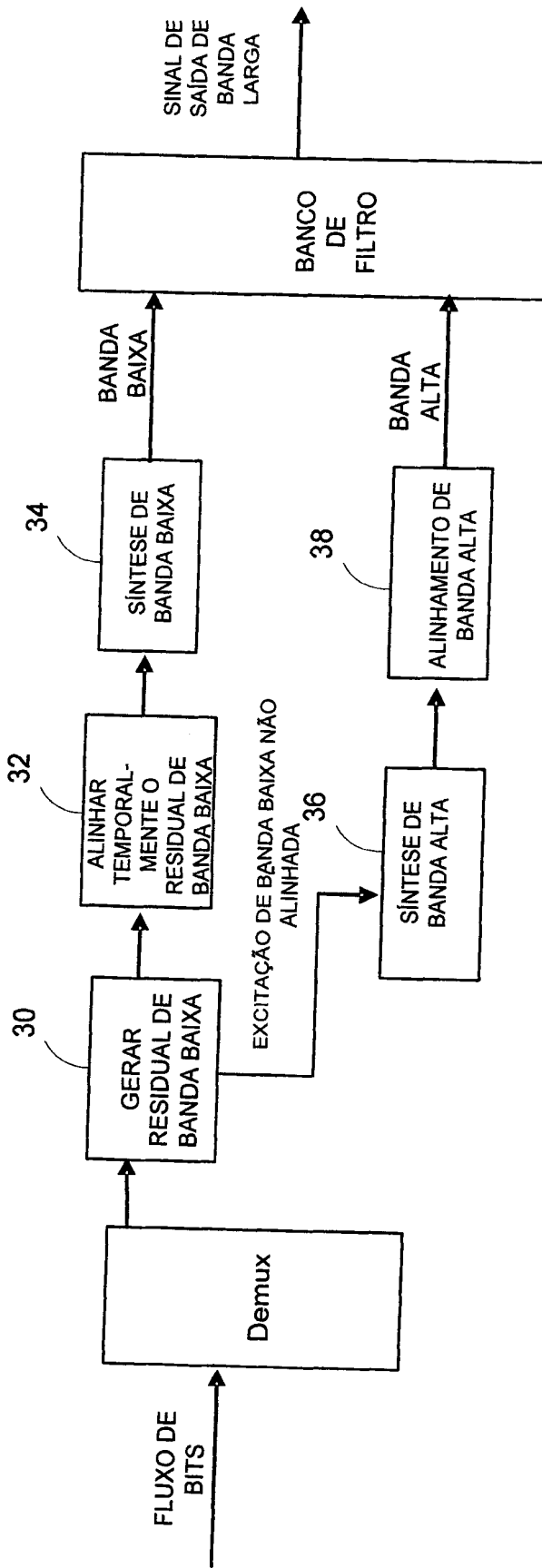


FIG. 3
ALINHAMENTO TEMPORAL DE BANDA BAIXA E BANDA ALTA

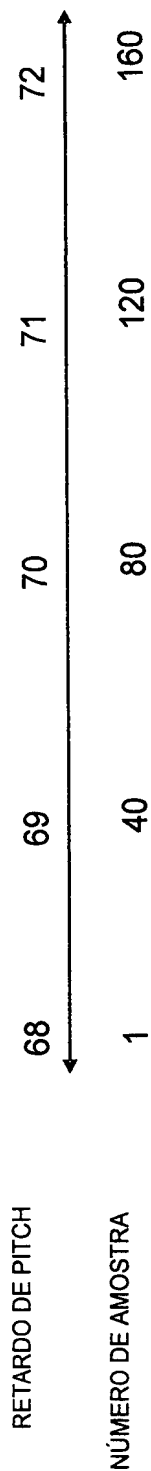


FIG. 4A

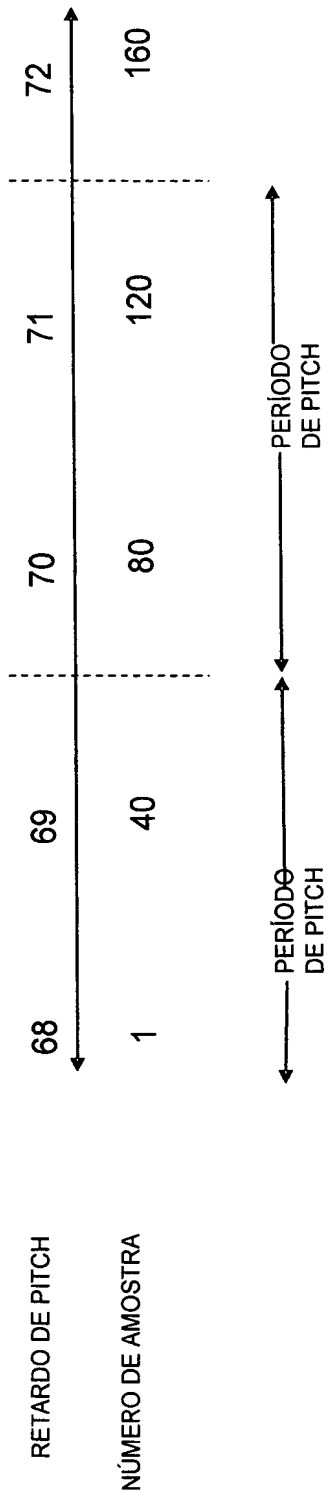


FIG. 4B



FIG. 5A

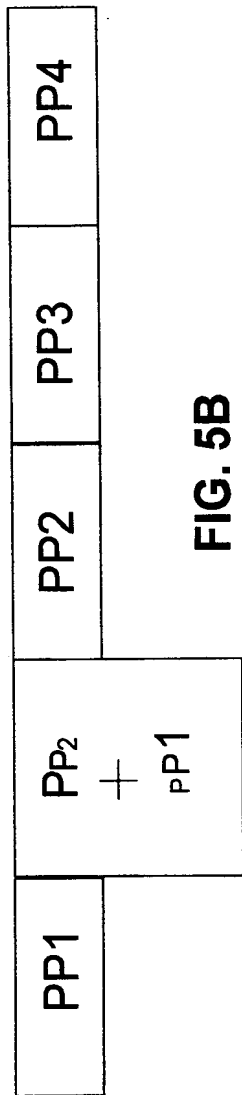


FIG. 5B

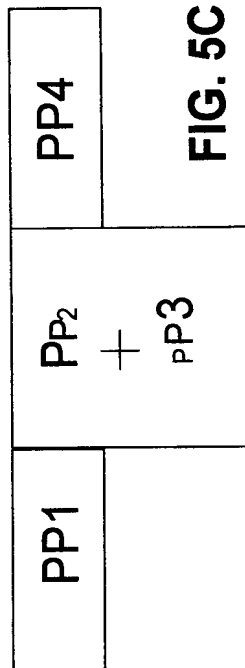


FIG. 5C

RESUMO

"QUADROS DE ALINHAMENTO TEMPORAL DE VOCODER DE BANDA LARGA"

Um método de comunicar fala compreendendo alinhamento temporal de um sinal de fala de banda baixa residual para uma versão expandida ou comprimida do sinal de fala de banda baixa residual, alinhamento temporal de um sinal de fala de banda alta para uma versão expandida ou comprimida do sinal de fala de banda alta, e fundindo os sinais de fala de banda baixa e banda alta com alinhamento temporal para fornecer um sinal de fala com alinhamento temporal inteiro. Na banda baixa, o sinal de fala de banda baixa residual é sintetizado após alinhamento temporal do sinal de banda baixa residual enquanto na banda alta, um sinal de banda alta não alinhada é sintetizado antes de alinhamento temporal do sinal de fala de banda alta. O método pode compreender ainda classificar segmentos de fala e codificar os segmentos de fala. A codificação dos segmentos de fala pode ser uma entre predição linear excitada por código, predição linear excitada por ruído ou codificação de quadro 1/8 (silêncio).